Deflating Validity

or: The use and abuse of "validity" in survey research

G. A. Reynolds

February 10, 2014

Abstract

In philosophy and logic, validity and truth are closely related. Truth is a property of sentences (propositions); validity is a property of inferences. In recent decades, "deflationary" (or "minimalist") accounts of truth have become increasingly popular among philosophers. Broadly speaking, these accounts deny that truth is a substantial property, and instead treat the term "truth" as a kind of expressive device; it adds nothing significant to the expressions in which is appears, but it makes the language significantly more powerful. It allows us to say things we otherwise could not say, or could only say in cumbersome ways. For example, with a locution like "... is true", we can endorse claims by naming them (e.g. "Fermat's last theorem is true"); without such a locution, we would have to explicitly repeat the theorem as a claim (e.g. "There is no integer z greater than 2 such that ..."). And some things we can say with "... is true" would be practically impossible to express without it, such as "everything the policeman said is true" (since it would not be possible to repeat everything he said) or "the theorems of group theory are true" (since there are (I assume) infinitely many such theorems).

A third aspect: sentences contain referring components. To the truth of a sentence corresponds the "referentiality" of its components. "Snow is white" is true; it is true because "Snow" refers to the famous cold stuff, and "white" refers to the famous color. We need (but generally speaking do not have) a technical term to refer to the property of such refering relations that corresponds to the property of truth of sentences. It is a category mistake to say "'Snow' is true", but we would like to say "'Snow' is x" in order to bring attention to this truth-like referential condition. In Survey Research (and social science in general), the term "validity" is often recruited to serve this need in measurement vocabulary. The inadvisability of this becomes obvious when you move from measurement to description: "2.3 meters is a valid measurement of the length of x" is a common way to talk, but "'Snow' is valid" sounds decidedly off-key.

This paper has two goals. The theoretical goal is to do with validity what deflationists have done with truth. The more practical goal is to examine the use and role of the concept (term) validity in Survey Research.

The first part of the paper thus explores the plausibility of a deflationary or minimalist concept of validity. Not just logical (inferential) validity, but validity as used by the social sciences, as a property of referential relations.

The second part of the paper examines the notion of validity as used in Survey Research. Suffice it to say that vocabulary of validity in the social sciences, especially psychology and education

research, is very, *very* confused. Generally speaking, the term is used to refer, not to inferences and their properties, but to referential relations. Classic definitions of validity in the social sciences usually say something like "measures what it purports to measure", which is to say, measurement expressions (e.g. "2.3 meters") *refer* to entities (properties, relations) in the world. But it is also used to refer to inferences and a variety of other concepts.

The connection between the first and second parts is that the social sciences usually treat validity as a substantial property. Theories of validity often take on a metaphysical hue; they attempt to say what validity *is*, as if it were some kind of entity or substance – validity stuff – that referring terms "have", possibly in greater or lesser degrees. On the deflationary view, this is a mistake that inevitably leads to unresolvable problems.

Contents

List of Figures

1 Deflationism

Remark 1 *Deflationism seems to depend essentially on some form of expressivism. Or maybe they amount to the same thing?*

1.1 Validity, Reliability, Error

Remark 2 *What is the point of worrying about validity? Is it something in the world that we are trying to discover? Then we're trying to find "the right description of the world" (Putnam). Or is it a concept, so that validity talk is about conceptual analysis and definition?*

Or: we try to find the right description, and validity talk is part of how we decide that we have found it.

Remark 3 *Why do psychometricians and the like worry so about validity?*

Hypothesis: when they say "validity", what they're really interested in is scientific legitimacy. Effectively, to say that a test (etc.) is valid is to say that it is in fact scientific. Thats the practical import of the concept of validity for them.

Unpack this. Expose the assumptions and implications.

Remark 4 *The problem with validity (quantifiability) is circularity. If the task is to show that some property is quantitative, we have to do this without relying on quantitative vocabulary. So for example, if we want to show that temperature is quantitative, we cannot use the concept of a unit of temperature to do so, because that presupposes just the outcome we are supposed to demonstrate. This is similar to the problem we face in seeking to account for representational vocabulary in non-representational terms.*

quantifiability v. validity? distinct problems, but the latter depends on the former?

key concepts:

- validity treated as a special kind of property - of what?
- constructs
- (latent) variables
- indicators

"validity" as code for:

- legitimacy
- vindication
- credibility
- proof (good premises + valid inference)

Remark 5 *On the idea that validity something (a property, etc.) that we look for in scientific theories in order to distinguish good ones from bad: see Putnam on fact/value distinction. We use value judgments - simplicity, parsimony, etc. - in every aspect of science (thought), esp. in weeding out bad theories. For there is no external or objective criterion of acceptability for theories to which we can appeal, nor is there any such citerion that does not involve value judgments.*

Remark 6 *So along with the fact/value distinction, and the analytic/synthetic distinction, the internal/external distinction also collapses? Or do we just exclude the notion of external? No; we need to retain the idea of an external world that is independent of us and to which some of our judgments are answerable. We don't get to just make stuff up and call it true (correct) for at least some of our claims. There is no external absolute authority that can decide for us which theories are true, or rather which we should endorse, but that does not mean there is no external world that is authoritative for some of our sayings. But isn't that trying to have it both ways? How can our theories answer to the world if we cannot appeal to the world or some other external authority to sort them out? See Brandom.*

Related issue: what counts as evidence? How do we decide? What are we doing when we decide that something counts as strong (weak) evidence in support of a theory? What are the criteria of adequacy for an account of evidence?

1.2 RCT and Self-validation

See Cartwright on RCT as self-validating. This seems to mean that RCTs are valid by construction.

This nicely parallels industrial QA notions of guaranteeing quality by designing a production process that prevents defects.

What's the logic here? Is self-validation really possible? How can a process validate itself - isn't the very idea inherently circular? Or rather, don't we land in a regress? After all, if the idea is to specify a process that yields validity, how do we know that that process is itself valid?

1.3 Deflation

How can we get out of this mess? One way is to deflate the notion of validity, just deny that it is a substantive property. When we claim that a result is valid etc. what we are really saying is that we endorse it, approve of it, etc. It's an expressive device. Compare the semantic deflationist's idea that calling something true amounts to endorsing or approving of it.

So if we discard the notion of validity (since it does no real work), don't we find ourselves lacking something essential? Well, we just need a vocabulary that allows us to say explicitly the sorts of things we find it useful to be able to express with respect to a study or qx technique. For example: credibility, utility, legitimacy, vindication, justification, etc.

Remark 7 *The notion of validity seems to be connected to the problem of deciding which theories we should endorse. What are the criteria of adequacy for any notion (or theory) of validity? Or: what are the requirements that should be met by any purported explanation of validity? Both particular cases and the general idea. Tarski gives us something like this for logical validity; what about "validity" as the term is used by psychometricians, test theorists, etc.?*

Contrast: claims of validity for a case, v. explanation of what validity is.

The objection will no doubt be that we need some kind of standard, which is just to say that we want to measure this something (validity, credibility, whatever). Implicit in all this is the notion that

there is some "objective" fact of the matter to which our study/technique/etc. is ansswerable. A study is valid iff - what? If it meets some definite "objective" criteria. Methodological criteria, conditions of validity, etc. In the psychometrics and testing tradition this appeal to external authority is expressed as something along the lines of "measures what it purports to measure". Which is only meaningful insofar as a) there is actually something there to measure, and b) it is in fact susceptibel to measurement.

And usually this is expressed in statistical terms. But that dog won't hunt either - you cannot get to validity via statistics. All you can do is measure central tendencies and variance - not enough to establish validity, which is a substantive notion. (analysis elsewhere).

To say that sth is valid is just to say that it is admirable (Peirce?), or perhaps that it is virtuous, that it has the virtues we prize.

2 Previous Work

Remark 8 *Fact/value; inference;*

2.1 Landmark articles

Cronback & Meehl, Messick, Lissitz, Michell, etc.

Special issues:

Social Indicators Research Volume 45, Issue 1-3, November 1998: Special Issue on Validity Theory and the Methods Used in Validation: Perspectives from the Social and Behavioral Sciences
Schwarz, "Is Psychology Based on a Methodological Error?"

2.2 Validity

2.3 Measurement

2.4 Assessment and Evaluation

Brinkmann, "Psychology's Facts and Values"
Cicourel, "Interviews, Surveys, and the Problem of Ecological Validity"
Hood, "Validity in Psychological Testing and Scientific Realism"
McDonald, "Measuring Latent Quantities"
Messick, "Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning"
Putnam, "The Collapse of the Fact/Value Dichotomy"

3 Kinds of Validity

Validity of what? Question, answer, both, or something else.

Remark 9 *The critical idea is that representational validity ("measures what it purports to measure") is itself validate by causality. What matters is that the measurand (attribute to be measured) should cause change in the Instrument of Measurement.*

Example: thermometer. The readings of a (good) thermometer are "valid" measurements of temperature just because temperature (or heat or?) causes them. NOT because the recording measurements represent. In other words, representation comes after causality.

This seems to imply that validity is a property of the Instrument of Measurement. Which in turn argues that we should come up with a different term. In any case, it is essential to distinguish between the causal relationship between measurand and Instrument of Measurement on the one hand, and the behavior of the instrument and the measurement scale on the other, and on the third hand, the reading of the measurement and its recording as text in the proverbial lab notebook.

Classic definitions: validity means it (test, etc.) measures what it purports to measure. This is not a very clear definition; it fails to make the essential distinction between referentiality and accuracy, etc. Do they mean to say that e.g. an inaccurate thermometer is valid, so long as it does measure temp? What this means is so long as its (inaccurate) readings are caused by temperature? (But if they are inaccurate there must be other causes at work.)

A better definition: a valid measurement is one that is caused by (and only by) the measured property, regardless of accuracy.

But the standard definition also may be read as implying that "measuring what it purports to measure" means that it measure the true value. But since all meassurements contain error, what can this mean? Seems "measure true value" must mean "is cause by the measured property".

Remark 10 *This characterization of measurement in terms of representation is not to be confused with the Representational Theory of Measurement, which is but one of several competeing theoretical models of measurement. A criterion of adequacy for any such theory is that it must be able to account for the representational aspect of measurement.*

The term is used in so many ways that it effectively means not much more than "good, valuable". In the literature one can find valid: questions, answers, data, etc. etc. These are different things; what property common to them all does "validity" name? Is there such a common property? If "good" is the best we can do, then the concept of validity is largely trivial if not vacuous.

Need to distinguish between validity and accuracy. An inaccurate thermometer may not measure temperature accurately, but it does measure (i.e. respond to) temperature. So bad results are not necessarily invalid, only inaccurate.

The classic concept is that validity is about what gets measured, or whether the measurement indeed measures what it purports to measure. Underlying this notion seems to be the idea that the measurement outcome (the number on a scale) stands in a particular relation to the measurand. In other words, this is a representationalist notion: validity is about goodness of representation, where representation is viewed as a relation between signifier and signifiee.

Notice that such a concept of validity can only apply to attributes that are in fact quantitatively structured (measurable).

Thus "question validity" would describe the relation between the question and what it purports to measure. But since questions do not measure anything, the real idea seems to be that question validity is about the (representing) relation between the question and what it purports to represent.

Pragmatic alternative: good measurement useful, in that it allows us to make good predictions.

Example: simple math test of one question: what is the sum of 1 and 1? Possible correct answers: 2; square root of 4; difference between the 2nd and 3rd prime numbers; etc. Not to mention: II (Roman numberals); deux (French), dos (Spanish), etc.

Moral of the story: q-a pairs do not measure.

Re: use of tests as a matter of validity (i.e. many uses, each of which must be validated, so many validities, etc.). Compare use of temperature. The financial authorities could set monetory policy based on average temperature, if they wanted to. That would be an invalid (bad, incorrect, unjustifiable) use of temperature measurement, but the use of temperature measurements has nothing whatsoever to do with the "validity" temperature measurement.

The fact that different uses of social measurement do have something to do with validity should be taken as a sign that such "measurements" are not measurements at all, at least not in the way temperature measurements are.

Why? Because temperature *causes* change in the instrument of measurement. What makes it valid is reference to temperature and only temperature, which happens because temperature causes the reading. Compare use of a barometer as a proxy for thermometer. Why is this invalid? Because changes in the reading may be caused by factors other than temperature, and changes in temperature do not *necessarily* cause changes in the reading. So barometric measurements do not refer to temperature – "refer" meaning something like "be directly caused by". Directly, because temperature does (or at least can) cause a change in barometric pressure, thus causing a change in the reading of a barometer, but the causal link is indirect. It causes a change in pressure, which causes a change in temperature.

Compare this with any social measurement, e.g. how old are you, or what is 2+2 or the like. There is no direct causality involved.

That's in addition to the problem of whether putative latent variables are in fact measurables (have quantitative structure, etc.) Even if we could find a convincing latent psychological variable and somehow demonstrate that it is quantitative, we would still have no warrant for treating questions and answers involving it as measurements.

The task of survey research is to investigate normative commitments and entitlements, rather than causes and effects. Deontic rather than natural necessity. The means of doing so is controlled, explicit, deontic scorekeeping.

References:

Messick, "Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning"
Hood, "Validity in Psychological Testing and Scientific Realism"

3.1 Success Conditions

Remark 11 *The ideal is that we should say that some instrument "works", that is, that it has worked in the past and will work in the future.*

"Validity" is used by the orthodoxy to convey this ideal notion - if a survey instrument or procedure is "valid" then we can use it in future research.

Validity is variously described as a property of instruments, or of inferences involving "interpretation or use" of a test, etc. etc.

A better approach is to talk of success conditions rather than validity. In the case of the ART, we would like to claim that it works, but what does this mean? It works IF it is properly used, and if various other things are done - e.g. if quex items preceding the ART item are "right", if the project as a whole is "correctly" presented to respondents, etc.

In short, what we should claim is not "ART works" tout court, but that we have identified the conditions under which it is likely to work - the success conditions. Compare the notion of truth conditions as giving the meaning of a proposition. Instead of "Is P true?" we ask "What are the truth conditions of P?" - what conditions make it true. Similarly, instead of "Is the technique valid?" we ask "What are the success conditions of the technique?" - what conditions make it valid.

3.2 Representational Validity

survey research seems to derive its notions of validity from psychometrics, test theory, etc.

"The problem of validity is that of whether a test really measures what it purports to measure." (Kelly quoted in Saint-Mont).

This seems to be pretty clearly a representationalist idea.

Note that it is also rather vague, since it does not address accuracy. If all that is require is that a test measure (i.e. refer to) what it purports to measure, then it is an all-or-nothing affair, just like logical validity of inference. There would be no question of degrees of validity, and validity itself would not be measurable on a scale.

3.3 Epistemic Validity

I.e. truth-value of propositions.

3.4 Logical Validity

The logical concept of validity is relatively straightforward: it is a property of inferences only.

Remark 12 *NB: distinction between rules of inference (schemata, generalizations), and particular inferences. To keep this clear, we might use "formal inference" for the former and "vernacular inference" for the latter?*

Prawitz, "Inference and Knowledge"
Prawitz, "The epistemic significance of valid inference"

3.5 Material Validity

A concept of *material inference* was introduced by Wilfrid Sellars in x.

3.6 Statistical Validity

Statistical concepts of validity are pervasive in the survey research literature.…

3.7 Discursive Validity

"Validation" of a question = exploration of proprieties of discursive practices governing use of the question.

This is a substantially different approach from the orthodox approach. Usually validation is taken to involve question semantics. The meaning of the question is often antecedently assumed, often based on some variety of Gricean intentionality; the literature often discusses this in terms of the respondents understanding of the "intended meaning of the question."

Pragmatic "validation" seeks instead to discover the discursive *norms* governing the performance of the question. Since the question is construed in practical terms, as a discursive turn, its meaning cannot be antecedently presumed. The task is not to determine whether the respondent understands the intended sense of the question, but to discover the norms implicit in the way respondents treat the question. Questions establish conditions of correct response; but those conditions may differ for every person, since they depend on collateral commitments and entitlements. The goal is to explore this; only after we have gotten a grip on how respondents treat a question can we move on to the question of how this corresponds to the researcher's understanding (i.e. the intended sense).

In other words, "intended meaning" does not establish a standard against which respondent performance is to be judged. That would beg the question, which is just what the question means (i.e. how is it normatively used).

3.8 Example: Age

Suppose we want to validate the question "How old are you?"

Most of us can respond appropriately to this question by citing a number representing age in years. But the correctness of this response is a matter of discursive practice rather than representational accuracy. It counts as knowledge not because the number cited represents age, but because we have been trained to respond to the question appropriately. We do not examine the number, see that it represents our age, then therefore cite it; nor do we examine our age (a latent variable?) and see that it is represented by a number which we then cite. We have been trained from an early age to cite a number in response to questions of this sort, and to increment that number by one on every birthday. This is a matter or normative discursive practice. If we do not respond to the question by stating our names or something else unrelated to age, that is because doing so would violate the discursive norm, not because hidden cognitive processes prevent us from doing so for purely semantic reasons.

That is not to say that our adherence to such practical norms is a matter of mere response to a stimulus. To borrow a bit of imagery favored by Brandom, a parrot can be trained to reliably respond to our age question by producing a number noise, say "three". But it will never age; in the absence of retraining, every year it will announce the same age. That's because its response is non-conceptual. It does not *understand* the noises it responds to nor those it produces. It presumably has no grasp of the inferential articulation of the concepts of age, three, you, etc. Furthermore it has no notion of correctness; it cannot decide that "three" is incorrect and therefore decline to respond. And although one might concede that its reliable responsive disposition is a kind of discursive practice (since it involves turn-taking or sequencing), it does not qualify as *linguistic* discursive practice.

What's the significance of this example? First, it elucidates the distinction between non-conceptual and conceptual (that is, rational) responsiveness in terms of grasp of inferential structure.

Remark 13 *But what we need is clarification of how our conceptual life is grounded in normative practice; do parrot examples do this?*

4 Causality and Normativity

Norms make the correct answer "necessary" (i.e. obligatory), but not compulsory.

So questions do not *cause* responses. They establish the criteria by which the answer may be judged as correct or not, but they do not causally *compel* the correct response.

Furthermore, neither questioning nor answering is purely individualistic. They involve discursive practices that are essentially social. So the Respondent's answer to a question cannot be construed as something cut off from social relations, nor as referencing a purely private state, believe, disposition, etc. Since meaning is socially constituted, statements always involve more than the personal; they essentially involve, for example, positioning, role, status, etc.

5 Measurement and Error

5.1 Measurement

survey research (at least as far as interviewing methodology is concerned) seems to depend almost entirely on notions of measurement drawn from psychology in general and *psychometrics* in particular. But quantitative psychology, and especially psychometrics, has its critics; one authority on measurement has gone so far as to claim that psychometrics is a form of "pathological" science, arguing that "[Q]uantitative psychology manifests methodological thought disorder"[1], and that "[P]sychometricians are not only uncritical of an issue basic to their discipline, but that, in addition, they have constructed a conception of quantification that disguises this."[2] The gist of his argument:

[1] Michell, "Quantitative science and the definition of measurement in psychology".
[2] Michell, "Normal Science, Pathological Science and Psychometrics", p. 639.

> Consider any attribute that psychometricians currently believe they are able to measure (such as any of the veraious intellectual abilities, personality traits or social attitutes that the textbooks mention), and ask the question, *Is that attribute quantitative?* The hypothesis that such an attribute is quantitative underwrites the claim to be able to measure it. However, *there has never been any serious attempt within psychometrics to test such hypotheses.*[3]

Remark 14 *TODO: concise description of our approach to measurement in the context of this research. Summary: the answers provided by Respondents in response to Interviewer questions are not to be taken as proper measurements of anything. Nor are the data that the Interviewer records and the analyst eventual manipulates to be considered quantified representations of measurable (quantifiable) properties. Instead, answers are discursive turns or moves whose contents are implicit propositions endorsed by the Respondent, and the recorded data are representations of those endorsements. Measurement only enters the picture when the statistician describes the collection of recorded endorsed propositions. In other words, the statistical description of the collected data serves to characterize what people assert (the propositions they endorse; their speech practices) rather than "levels" of hidden properties.*

Objection 1 *How does this apply specifically to our question about status? Is it not correct to interpret a Respondent's true correct answer to our question as a "measurement" of status? On the orthodox view, the answer corresponds to the Respondent's actual status; isn't this properly considered measurement?*

Response to Objection 1.

A Respondent who truthfully claims e.g. "I am a Legal Permanent Resident" thereby signals a *commitment* to the propositional content of the statement (namely, that R is a Legal Permanent Resident). Any such commitment may be challenged, in which case the Respondent's *entitlement* to the claim comes to the fore. The Respondent might *justify* the claim by citing any number of *reasons*, including personal knowledge, etc., or may exhibit non-linguistic evidence (such as an official LPR card).

The critical point is that the Respondent's answer is not *caused* by anything; in particular it is not caused by some occult "variable". What matters to the meaningfulness of the discursive exchange is not a matter of natural causes, but of propositions, commitments, entitlements, authority, and the other aspects of discursive practice involved in "the game of giving and asking for reasons".

Objection 2 *If that is the case, if individual answers are not in fact measurements of anything, then what becomes of the statistical analysis of the collected data? Is it not measurement?*

Response to Objection 2.

[3] Michell, "Normal Science, Pathological Science and Psychometrics", 648; *emph. added.*

The collected data is just that, a set of data, so like any set it can be described statistically. The critical question is what is the *meaning* of that statistical description. The orthodox approach claims that it measures something – a latent variable or the like. Implicitly at least, it claims that there is a determinate relation between the statistical description and the levels (or whatever) in the individual Respondents. The claim here is that this claim is unwarranted.

The claim here is that the business of the Survey Interviewing is just deontic scorekeeping.

References:

Michell, "Normal Science, Pathological Science and Psychometrics"

Sherry, "Thermoscopes, thermometers, and the foundations of measurement"

See British Journal of Psychology, Aug 1997 vol 88 issue 3: Michell, "Quantitative science and the definition of measurement in psychology" and six commentaries.

5.2 Variables

References:

Toomela, "Variables in Psychology"

Schwarz, "Is Psychology Based on a Methodological Error?"

Stam, "The Fault is Not in Ourselves, but in Our Methods"

5.3 Error

"Problems" in a Survey Interview should not automatically be treated as indications of *error*, whether "Respondent error", "Interviewer error" or some other class of error. Discrepancies between Respondent understanding and researcher understanding should often be treated as evidence of normal variance in how different members treat the question. "Problems" in question-answer interactions may be traceable to mistakes, misunderstandings, etc.

5.4 The Metrological Corkscrew

The fundamental problem confronting any purported measurement is the metric equivalent of the hermeneutical circle.

For a measure of an attribute to be "valid", the attribute must be measureable in the first place. The problem is that the only way to determine whether or not an attribute is quantifiably measurable is to measure it. Measurability cannot be antecedently established.

Sherry's account of the history of the development of temperature measurement[4] clearly illustrates the problem. This problem also lies at the heart of Michell's attack on psychometrics[5].

[4] Sherry, "Thermoscopes, thermometers, and the foundations of measurement".

[5] Michell, "Normal Science, Pathological Science and Psychometrics".

5.5 The "Referential Gap"

By *referential gap* I mean the conceptual chasm between the meaning of statistical description and the meanings of the individual data so described.

Consider a canonical example of measurement: using a thermometer to measure temperature of a liquid. An individual (specific occasion of) measurement yields a number on a scale that corresponds to the "true" temperature of the liquid.

Such an account is not particularly controversial[6], but there is much more than meets the eye here. First, this account assumes but does not demonstrate that "true temperature" is meaningful; second, it takes "valid" (or, good, or correct, etc.) measurement as essentially involving accuracy of representation or correspondence.

Let's assume for the sake of argument that things really do have a true temperature, and that our thermometer accurately represents that temperature. In this case, if we replicate our trial n times and then describe the resulting dataset statistically, we will obtain a statistical description not only of the data but of the temperature measured. That's because each datum in the statistical dataset has a determinate meaning, which can be expressed in terms of representation.

Remark 15 *This is of course an oversimplification; the notion of reference or denotation with respect to measurement is not clear-cut. The idea here is that a particular measurment yields a number (on a scale) that* ***refers*** *to temperature; but does it* ***denote*** *"the" temperature? Since perfect measurement is not possible (all measurements "contain" error), it cannot be the case that the measurement denotes the true temperature. So although the measurement (the number) in some sense refers to the temperature of the sample, it does not denote the true temperature. This sort of reference is fuzzy.*

Things are quite different for social scientific data, however.

A specific question performance, by contrast, yields an answer, which is interpretable in terms of discursive norms rather than correspondence (accuracy of representation).

5.6 Person

5.7 Reference

The person. Orthodox view is that answers measure some property or attribute of the respondent. One problem with this is that it assumes a individualist autonomy. But discursive practice is essentially social, and the meanings of our sayings cannot be private.

Subjectivity is social, relational; even first-person claims implicate essentially social stuff, spill over the boundaries of the individual speaker.

[6] For most of us, anyway; the theory of measurement is quite subtle, and there are several different approaches, some of which may not agree with the emphasis I put on representation here. Nonetheless, here we keep it simple for the sake of clarity. The point of this toy example is to clarify the issues involved in order to show how measurement concepts in survey research are problematic and to make an alternative pragmatist concept intelligible.

> Strongly constitutive sociocultural perspectives in psychology have become more developed and influential in recent decades, particularly during the past 15 years. These approaches include constructionist, discursive, relational, dialogical, and neo-Vygotskian theories. They diverge from one another in some respects, but are alike in that they all consider psychological processes, such as mind and self, to emerge out of social, cultural, and historical contexts.[7]

Bucholtz and Hall, "Identity and interaction: a sociocultural linguistic approach"
Kirschner, "Sociocultural Subjectivities Progress, Prospects, Problems"
Andreouli, "Identity, Positioning and Self-Other Relations"

6 Reliability

Intuitively, reliability is about replication. Something is reliable if we can count on it to "work" the next time (or every time) we use it. It's a kind of prediction.

Remark 16 *Brandom: reliabilism as a key strategy of representationalism. Importance of the gerrymandering objection.*

6.0.1 Statistical Reliability

7 Assessment

The goal is usually stated as "comparable data", but what we really want is *commensurable* data; that is, data that can be meaningfully statistically analyzed.

7.1 Validation

Give a concept of validity, what must we do in order to validate our research?

Standard survey research uses several "methods" of validation. The best method involves checking responses against external data, usually administrative data. Note that even here the assumption is that the external data is itself valid and reliable, which is not necessarily the case. Administrative data, after all, is just as vulnerable to error as any other data, and so is the process of obtaining and using such data for verification purposes. So even the best method of survey research validation is already fairly weak.

Another common method involves some form of triangulation. If the verification method corresponds to correspondence theories of truth, triangulation methods correspond to coherence theories. The idea is simply that the data obtained in an interview should be internally consistent. Contradictory responses are taken as indicators of invalidity. And so forth.

[7] Kirschner, "Sociocultural Subjectivities Progress, Prospects, Problems".

Regardless of what methods are used, we have no means of truly establishing validity and reliability. The problem is that these concepts implicitly rely on theoretical presuppositions that do not hold for the human sciences, though they work just fine for the "hard" sciences.

A "valid" question really just means we have "good" reasons to believe that we have provided (necessary) and sufficient *resources* for the respondent to use in producing a correct and cooperative response. That is, with the resources we have provided we have good reason to believe that Respondents *should*:

- grasp the intended significance of the question; and
- cooperatively disclose a correct response

This is really the best we can hope for. Real "verification" of answers is not truly possible even for simple non-threatening questions; for the sort of question we are working with, there is no means of even approximating verification, since no external source of information is available.

7.2 Establishing Reliability

8 Quality Assurance: Failure, Defect, Cause

Instead of "error" and its sources we talk of failure and its causes.

8.1 Protocol

8.2 Instrument

8.3 Interviewer

8.4 Respondent

8.5 Environment

8.6 Other Causes of Breakdown

9 Validation

Here the critical question is how we can determine whether or not Respondents answer the status question truthfully; that is, whether they pick the correct box. The problem is that there is no way even in principle to verify answers based on externally obtained data.

9.1 Triangulation

A standard method of validating answers is by collectiing collateral information that can be used to check the internal consistency of answers. For example, we might ask whether the Respondent is currently enrolled in school, and compare the response to the claimed immigration status group.

There are two ways to obtain triangulation data: we can ask such questions either before or after we ask the immigration status question.

The problem is that triangulation amounts to a form of disclosure analysis. If we can use it to infer status, so can a hostile analyst. On the other hand, we might use it to determine whether the Respondent is responding truthfully, rather than to try to detect status.

We do not want to use the response to the GAM question as a means of inferring status in conjunction with other items in the questionnaire. On the contrary, for ethical reasons the instrument must be designed so that such analysis is not possible.

For extra protection we neither asked for nor collected any identifying information.

10 Quality

11 Design

Appendices

A Validity in Logic

B Bibliography

[1] Peter Abell. “Narrative Explanation: An Alternative to Variable-Centered Explanation?” In: *Annual Review of Sociology* 30 (Jan. 2004), pp. 287–310. URL: http://www.jstor.org/stable/29737695 (visited on 12/08/2013).

[2] E. Andreouli. “Identity, Positioning and Self-Other Relations”. In: *Papers on Social Representations* 19.1 (2010), pp. 14.1–14.13 (cit. on p. 15).

[3] Mikhail Mikhaĭlovich Bakhtin. *Speech genres and other late essays*. Trans. by Vern W. McGee. University of Texas Press, 1986. ISBN: 9780292775602.

[4] Mikhail Mikhaĭlovich Bakhtin. “The Problem of Speech Genres”. In: *Speech Genres and Other Late Essays*. University of Texas Press, 1986, pp. 60–102. ISBN: 9780292775602.

[5] Randy Elliot Bennett. “Consequences That Cannot Be Avoided: A Response to Paul Newton”. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 30–32. DOI: 10.1080/15366367.2012.686865.

[6] Jack Bilmes. “Ethnomethodology, Culture, and Implicature: Toward an Empirical Pragmatics”. In: *Pragmatics* 3.4 (Feb. 2010). ISSN: 1018-2101. URL: http://elanguage.net/journals/index.php/pragmatics/article/viewArticle/181.

[7] Paul Black. “EPMA Professionals—Servants or Masters?” In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 33–37. DOI: 10.1080/15366367.2012.677342.

[8] S. F. Blinkhorn. “Past imperfect, future conditional: Fifty years of test theory”. In: *British Journal of Mathematical and Statistical Psychology* 50.2 (1997), pp. 175–185. DOI: 10.1111/j.2044-8317.1997.tb01139.x.

[9] Herbert Blumer. “Sociological Analysis and the ”Variable””. In: *American Sociological Review* 21.6 (Dec. 1956), pp. 683–690. DOI: 10.2307/2088418.

[10] Denny Borsboom. “The attack of the psychometricians”. In: *Psychometrika* 71.3 (Sept. 2006), pp. 425–440. DOI: 10.1007/s11336-006-1447-6.

[11] Denny Borsboom. “Whose Consensus Is It Anyway? Scientific Versus Legalistic Conceptions of Validity”. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 38–41. DOI: 10.1080/15366367.2012.681971.

[12] Denny Borsboom and Gideon J. Mellenbergh. “Why Psychometrics is Not Pathological A Comment on Michell”. In: *Theory & Psychology* 14.1 (Feb. 2004), pp. 105–120. DOI: 10.1177/0959354304040200.

[13] Denny Borsboom, Gideon J. Mellenbergh, and Jaap van Heerden. “The theoretical status of latent variables”. In: *Psychological Review* 110.2 (2003), pp. 203–219. DOI: 10.1037/0033-295X.110.2.203.

[14] Tom Bramley. ““Measurement” and “Construct” Need to Be Clarified First. Commentary on Newton, P. E. “Clarifying the Consensus Definition of Validity””. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 42–45. DOI: 10.1080/15366367.2012.677344.

[15] Henry Braun. “Conceptions of Validity: The Private and the Public”. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 46–49. DOI: 10.1080/15366367.2012.679159.

[16] Svend Brinkmann. “Psychology’s Facts and Values: A Perennial Entanglement”. In: *Philosophical Psychology* 18.6 (2005), pp. 749–765. DOI: 10.1080/09515080500355244 (cit. on p. 6).

[17] Svend Brinkmann. “The ethical subject: Accountability, authorship, and practical reason”. In: *SATS* 11.1 (Sept. 2010), pp. 75–89. URL: http://www.degruyter.com.proxy.uchicago.edu/view/j/sats.2010.11.issue-1/sats.2010.007/sats.2010.007.xml (visited on 04/21/2012).

[18] Svend Brinkmann. “The Mind as Skills and Dispositions: On Normativity and Mediation”. In: *Integrative Psychological and Behavioral Science* (Aug. 2011). ISSN: 1932-4502, 1936-3567. DOI: 10.1007/s12124-011-9183-6. URL: http://www.springerlink.com.proxy.uchicago.edu/content/g2222336gjn7x61k/ (visited on 01/02/2012).

[19] Svend Brinkmann. “Towards an Expansive Hybrid Psychology: Integrating Theories of the Mediated Mind”. In: *Integrative Psychological and Behavioral Science* 45.1 (Oct. 2010), pp. 1–20. ISSN: 1932-4502, 1936-3567. DOI: 10.1007/s12124-010-9146-3. URL: http://www.springerlink.com.proxy.uchicago.edu/content/ejq3033k4l08g430/ (visited on 02/04/2012).

[20] Mary Bucholtz and Kira Hall. “Identity and interaction: a sociocultural linguistic approach”. In: *Discourse Studies* 7.4-5 (Oct. 2005), pp. 585–614. ISSN: 1461-4456, 1461-7080. DOI: 10.1177/1461445605054407. URL: http://dis.sagepub.com.proxy.uchicago.edu/content/7/4-5/585.short (visited on 02/12/2012) (cit. on p. 15).

[21] Edward G. Carmines and Richard A. Zeller. *Reliability and Validity Assessment*. en. SAGE, Nov. 1979. ISBN: 9780803913714.

[22] Robyn Carston. “Linguistic Communication and the Semantics/Pragmatics Distinction”. In: *Synthese* 165.3 (Dec. 2008). ArticleType: research-article / Full publication date: Dec., 2008 / Copyright © 2008 Springer, pp. 321–345. ISSN: 0039-7857. URL: http://www.jstor.org/stable/40271119.

[23] Miguel Castillo-Díaz and José-Luis Padilla. “How Cognitive Interviewing can Provide Validity Evidence of the Response Processes to Scale Items”. In: *Soc Indic Res* (), pp. 1–13. DOI: 10.1007/s11205-012-0184-8.

[24] A M Chang, J P Chau, and E Holroyd. “Translation of questionnaires and issues of equivalence”. In: *J Adv Nurs* 29.2 (Feb. 1999). PMID: 10197930, pp. 316–322. URL: http://www.ncbi.nlm.nih.gov/pubmed/10197930 (visited on 08/28/2011).

[25] Hasok Chang. *Inventing temperature: measurement and scientific progress*. Oxford studies in philosophy of science. Oxford ; New York: Oxford University Press, 2004. ISBN: 0195171276.

[26] Hasok Chang. “Measurement, Justification, and Scientific Progress”. In: *Inventing temperature: measurement and scientific progress*. Oxford studies in philosophy of science. Oxford ; New York: Oxford University Press, 2004, pp. 220–234. ISBN: 0195171276.

[27] Hasok Chang. “Spirit, Air, and Quicksilver”. In: *Inventing temperature: measurement and scientific progress*. Oxford studies in philosophy of science. Oxford ; New York: Oxford University Press, 2004, pp. 57–102. ISBN: 0195171276.

[28] Jeasik Cho and Allen Trent. “Validity in qualitative research revisited”. In: *Qualitative Research* 6.3 (2006), pp. 319–340. DOI: 10.1177/1468794106065006.

[29] Aaron V. Cicourel. “Interviews, Surveys, and the Problem of Ecological Validity”. In: *The American Sociologist* 17.1 (Feb. 1982), pp. 11–20. URL: http://www.jstor.org/stable/27702491 (visited on 03/11/2012) (cit. on p. 6).

[30] Andy Clark. *Being there : putting brain, body, and world together again*. Cambridge, Mass.: MIT Press, 1997.

[31] Andy Clark and David Chalmers. “The Extended Mind”. In: *Analysis* 58.1 (Jan. 1998), pp. 7–19. URL: http://www.jstor.org/stable/3328150 (visited on 01/24/2012).

[32] Andy Clark and Josefa Toribio. “Doing without Representing?” In: *Synthese* 101.3 (Dec. 1994), pp. 401–431. DOI: 10.1007/BF01063896.

[33] Herbert H. Clark. *Using language*. Cambridge University Press, May 1996. ISBN: 9780521567459.

[34] David A. Cook and Thomas J. Beckman. “Current Concepts in Validity and Reliability for Psychometric Instruments: Theory and Application”. In: *The American Journal of Medicine* 119.2 (Feb. 2006), 166.e7–166.e16. DOI: 10.1016/j.amjmed.2005.10.036.

[35] Deborah J. Coon. “Standardizing the Subject: Experimental Psychologists, Introspection, and the Quest for a Technoscientific Ideal”. In: *Technology and Culture* 34.4 (Oct. 1993), p. 757. DOI: 10.2307/3106414.

[36] Jeff Coulter. “Discourse and Mind”. In: *Human Studies* 22.2/4 (Oct. 1999), pp. 163–181. URL: http://www.jstor.org/stable/20011238.

[37] Jeff Coulter. “What Is ”Discursive Psychology”?” In: *Human Studies* 27.3 (Jan. 2004), pp. 335–340. URL: http://www.jstor.org/stable/20010378.

[38] Angélique O. J. Cramer. “Why the Item “23 +1” Is Not in a Depression Questionnaire: Validity From a Network Perspective”. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 50–54. DOI: 10.1080/15366367.2012.681973.

[39] Lee J. Cronbach and Paul E. Meehl. “Construct validity in psychological tests”. In: *Psychol Bull* 52.4 (July 1955). PMID: 13245896, pp. 281–302. URL: http://www.ncbi.nlm.nih.gov/pubmed/13245896 (visited on 03/17/2012).

[40] Simon Cullen. “Survey-Driven Romanticism”. en. In: *Review of Philosophy and Psychology* 1.2 (June 2010), pp. 275–296. DOI: 10.1007/s13164-009-0016-1.

[41] Bronwyn Davies and Rom Harré. “Positioning: The Discursive Production of Selves”. en. In: *Journal for the Theory of Social Behaviour* 20.1 (Mar. 1990), pp. 43–63. DOI: 10.1111/j.1468-5914.1990.tb00174.x.

[42] Robyn M. Dawes. “Psychological measurement”. In: *Psychological Review* 101.2 (1994), pp. 278–281. DOI: 10.1037/0033-295X.101.2.278.

[43] Daniel C. Dennett. “Styles of Mental Representation”. In: *Proceedings of the Aristotelian Society*. New Series 83 (Jan. 1982), pp. 213–226. ISSN: 00667374. URL: http://www.jstor.org.proxy.uchicago.edu/stable/4545000 (visited on 02/19/2011).

[44] Jacques Derrida. “Signature Event Context”. In: *Limited Inc*. Evanston, Ill.: Northwestern Univ. Press, 1988.

[45] Robert F DeVellis. *Scale development : theory and applications*. English. Thousand Oaks, Calif.: Sage Publications, 2003. ISBN: 0761926046 9780761926047 0761926054 9780761926054.

[46] Santo Di Nuovo. “Variables and Quantity: What Else?” In: *Integrative Psychological and Behavioral Science* 43.1 (Nov. 2008), pp. 84–88. DOI: 10.1007/s12124-008-9081-8.

[47] Eli Dresner. “Language and the Measure of Mind”. In: *Mind & Language* 25.4 (Sept. 2010), pp. 418–439. DOI: 10.1111/j.1468-0017.2010.01396.x.

[48] Hubert L. Dreyfus. “How Representational Cognitivism Failed and is being replaced by Body/World Coupling”. In: *After Cognitivism*. Springer Netherlands, 2009, pp. 39–73. URL: http://www.springerlink.com.proxy.uchicago.edu/content/q4kt33m51hh41456/abstract/ (visited on 10/16/2012).

[49] Hubert L. Dreyfus. “Intelligence without representation – Merleau-Ponty’s critique of mental representation The relevance of phenomenology to scientific explanation”. In: *Phenomenology and the Cognitive Sciences* 1.4 (Dec. 2002), pp. 367–383. DOI: 10.1023/A:1021351606209. URL: http://dx.doi.org/10.1023/A:1021351606209.

[50] Judith A. Droitcour, Eric M. Larsen, and Fritz J. Scheuren. “The Three-Card Method: Estimating Sensitive Survey Items—With Permanent Anonymity of Response”. In: *Proceedings of the Annual Meeting of the American Statistical Association*. Aug. 2001. URL: http://www.amstat.org/sections/srms/proceedings/y2001/Proceed/00582.pdf.

[51] Judith A. Droitcour and Eric M. Larson. “An Innovative Technique for Asking Sensitive Questions: the Three-Card Method”. In: *Bulletin de Méthodologie Sociologique* 75.1 (July 2002), pp. 5–23. DOI: 10.1177/075910630207500103. URL: http://bms.sagepub.com/content/75/1/5.abstract (visited on 10/25/2011).

[52] Susan E. Embretson. “Construct Validity: A Universal Validity System or Just Another Test Evaluation Procedure?” In: *Educational Researcher* 36.8 (Nov. 2007), pp. 449–455. DOI: 10.3102/0013189X07311600.

[53] George Engelhard and Nadia Behizadeh. “Epistemic Iterations and Consensus Definitions of Validity”. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 55–58. DOI: 10.1080/15366367.2012.681974.

[54] Christopher Essex and William E. Smythe. “Between Numbers and Notions A Critique of Psychological Measurement”. In: *Theory & Psychology* 9.6 (Dec. 1999), pp. 739–767. DOI: 10.1177/0959354399096002.

[55] Ronald Fischer and Taciano L. Milfont. “Standardization in psychological research.” In: *International Journal of Psychological Research* 3.1 (July 2010), pp. 88–96. URL: http://mvint.usbmed.edu.co:8002/ojs/index.php/web/article/view/463 (visited on 12/09/2013).

[56] J A Flaherty et al. “Developing instruments for cross-cultural psychiatric research”. In: *J. Nerv. Ment. Dis.* 176.5 (May 1988). PMID: 3367140, pp. 257–263. URL: http://www.ncbi.nlm.nih.gov/pubmed/3367140 (visited on 10/23/2011).

[57] J. A. Fodor. “Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology”. In: *Behavioral and Brain Sciences* 3.01 (1980), pp. 63–73. DOI: 10.1017/S0140525X00001771.

[58] Melissa Freeman. “Validity in Dialogic Encounters With Hermeneutic Truths”. In: *Qualitative Inquiry* 17.6 (July 2011), pp. 543–551. DOI: 10.1177/1077800411409887.

[59] Barbara Fultner. “Inferentialism and Communicative Action: Robust Conceptions of Intersubjectivity”. In: *Philosophical Studies* 108.1 (Mar. 2002), pp. 121–131. DOI: 10.1023/A:1015768316114.

[60] U.S. General Accounting Office (GAO). *Estimating the Undocumented Population: A “Grouped Answers” Approach to Surveying Foreign-Bord Respondents*. Tech. rep. GAO-06-775. Washington D.C.: U.S. General Accounting Office, Sept. 2006. URL: http://www.gao.gov/new.items/d06775.pdf.

[61] Charles Goodwin. “The Interactive Construction of a Sentence in Natural Conversation”. In: *Everyday Language: Studies in Ethnomethodology*. Irvington Publishers, 1979, pp. 97–121. ISBN: 9780470266700.

[62] Joanna S. Gorin. “Reconsidering Issues in Validity Theory”. In: *Educational Researcher* 36.8 (Nov. 2007), pp. 456–462. DOI: 10.2307/4621100. URL: http://www.jstor.org/stable/4621100 (visited on 01/27/2013).

[63] H. P. Grice. “Meaning”. In: *The Philosophical Review* 66.3 (July 1957). ArticleType: research-article / Full publication date: Jul., 1957 / Copyright © 1957 Duke University Press, pp. 377–388. ISSN: 00318108. URL: http://www.jstor.org/stable/2182440.

[64] H. P. Grice. “Utterer’s Meaning and Intention”. In: *The Philosophical Review* 78.2 (Apr. 1969). ArticleType: research-article / Full publication date: Apr., 1969 / Copyright © 1969 Duke University Press, pp. 147–177. ISSN: 00318108. URL: http://www.jstor.org/stable/2184179.

[65] H. P. Grice. “Utterer’s Meaning, Sentence-Meaning, and Word-Meaning”. In: *Foundations of Language* 4.3 (Aug. 1968). ArticleType: research-article / Full publication date: Aug., 1968 / Copyright © 1968 Springer, pp. 225–242. ISSN: 0015900X. URL: http://www.jstor.org/stable/25000329.

[66] H. Paul Grice. *Studies in the Way of Words*. Harvard University Press, 1989.

[67] Paul Grice. “Logic and Conversation”. In: *Studies in the Way of Words*. Cambridge Mass ;;London: Harvard University Press, 1989.

[68] Robert M Groves. “Measurement Error Across the Disciplines”. In: *Measurement Errors in Surveys*. John Wiley & Sons, Inc., 1991.

[69] Brian D. Haig. “From Construct Validity to Theory Validation”. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 59–62. DOI: 10.1080/15366367.2012.681975.

[70] J. Haiman. “Ritualization and the Development of Languages”. In: *Perspectives on Grammaticalization*. Amsterdam: John Benjamins, 1994.

[71] John Hale. “Uncertainty about the Rest of the Sentence”. In: *Cognitive Science* 30.4 (2006), pp. 643–672. URL: http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ747089 (visited on 03/19/2012).

[72] Rom Harré. “Discursive Psychology and the Boundaries of Sense”. In: *Organization Studies* 25.8 (Oct. 2004), pp. 1435–1453. DOI: 10.1177/0170840604046351. URL: http://oss.sagepub.com/content/25/8/1435.abstract (visited on 02/06/2012).

[73] Rom Harré. “The Discursive Creation of Human Psychology”. In: *Symbolic Interaction* 15.4 (Nov. 1992). ArticleType: research-article / Full publication date: Winter 1992 /, pp. 515–527. ISSN: 0195-6086. URL: http://www.jstor.org/stable/10.1525/si.1992.15.4.515 (visited on 01/19/2012).

[74] Rom Harré et al. “Recent Advances in Positioning Theory”. en. In: *Theory & Psychology* 19.1 (Feb. 2009), pp. 5–31. ISSN: 0959-3543, 1461-7447. DOI: 10.1177/0959354308101417. URL: http://tap.sagepub.com.proxy.uchicago.edu/content/19/1/5 (visited on 01/23/2013).

[75] John Heritage. *Garfinkel and ethnomethodology*. Cambridge [Cambridgeshire]; New York, N.Y.: Polity Press, 1984. ISBN: 0745600603 9780745600604 0745600611 9780745600611.

[76] C. S. Herrman. “Fundamentals of Methodology Part I: Definitions and First Principles”. In: *SSRN eLibrary* (Apr. 2009). URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1373976 (visited on 01/21/2012).

[77] S. Brian Hood. “In Defense of an Instrument-Based Approach to Validity”. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 63–65. ISSN: 1536-6367. DOI: 10.1080/15366367.2012.681976.

[78] S. Brian Hood. “Validity in Psychological Testing and Scientific Realism”. In: *Theory & Psychology* 19.4 (Aug. 2009), pp. 451–473. DOI: 10.1177/0959354309336320 (cit. on pp. 6, 8).

[79] Lisa Tsoi Hoshmand. “Can Lessons of History and Logical Analysis Ensure Progress in Psychological Science?” In: *Theory & Psychology* 13.1 (Feb. 2003), pp. 39–44. DOI: 10.1177/0959354303131003.

[80] David Houle et al. “Measurement and Meaning in Biology”. In: *The Quarterly Review of Biology* 86.1 (Mar. 2011), pp. 3–34. DOI: 10.1086/658408.

[81] Dorothy Howie and Michael Peters. “Positioning Theory: Vygotsky, Wittgenstein and Social Constructionist Psychology”. en. In: *Journal for the Theory of Social Behaviour* 26.1 (Mar. 1996), pp. 51–64. ISSN: 1468-5914. DOI: 10.1111/j.1468-5914.1996.tb00285.x. URL: http://onlinelibrary.wiley.com.proxy.uchicago.edu/doi/10.1111/j.1468-5914.1996.tb00285.x/abstract (visited on 01/20/2012).

[82] Frank Jackson. “Mental Causation”. In: *Mind* 105.419 (July 1996), pp. 377–413. URL: http://www.jstor.org/stable/2254828 (visited on 12/08/2013).

[83] Michael Kane. “All Validity Is Construct Validity. Or Is It?” In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 66–70. ISSN: 1536-6367. DOI: 10.1080/15366367.2012.681977.

[84] Antti Kauppinen. “The Rise and Fall of Experimental Philosophy”. In: *Philosophical Explorations* 10.2 (2007), pp. 95–118. DOI: 10.1080/13869790701305871.

[85] Gary King et al. “Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research”. In: *American Political Science Review* 98 (2004), pp. 191–207. DOI: 10.1007/978-3-531-91826-6_16.

[86] Suzanne R. Kirschner. “Sociocultural Subjectivities Progress, Prospects, Problems”. en. In: *Theory & Psychology* 20.6 (Dec. 2010), pp. 765–780. ISSN: 0959-3543, 1461-7447. DOI: 10.1177/0959354310375745. URL: http://tap.sagepub.com.proxy.uchicago.edu/content/20/6/765 (visited on 01/23/2013) (cit. on p. 15).

[87] Paul Kline. “Commentary on Michell, Quantitative Science and the definition of measurement in psychology”. In: *British Journal of Psychology* 88.3 (1997), pp. 358–387. DOI: 10.1111/j.2044-8295.1997.tb02642.x.

[88] Joshua Knobe. “Experimental Philosophy and Philosophical Significance”. In: *Philosophical Explorations* 10.2 (2007), pp. 119–121. DOI: 10.1080/13869790701305905.

[89] Joshua Knobe. *Experiments in Philosophy*. http://opinionator.blogs.nytimes.com/2010/09/07/experimental-philosophy/. Sept. 7, 2010. URL: http://opinionator.blogs.nytimes.com/2010/09/07/experimental-philosophy/ (visited on 12/24/2012).

[90] Joshua Knobe et al. “Experimental Philosophy”. In: *Annual Review of Psychology* 63.1 (2012), pp. 81–99. DOI: 10.1146/annurev-psych-120710-100350.

[91] Tetsuya Kono. “The ‘extended mind’ approach for a new paradigm of psychology”. In: *Integr Psychol Behav Sci* 44.4 (Dec. 2010). PMID: 20499217, pp. 329–339. DOI: 10.1007/s12124-010-9128-5.

[92] Frederick Kroon. “Theory-dependence, warranted reference, and the epistemic dimensions of realism”. In: *Euro Jnl Phil Sci* 1.2 (May 2011), pp. 173–191. DOI: 10.1007/s13194-010-0004-4.

[93] Donald Laming. “A critique of a measurement–theoretic critique: Commentary on Michell, Quantitative science and the definition of measurement in psychology”. In: *British Journal of Psychology* 88.3 (1997), pp. 389–391. DOI: 10.1111/j.2044-8295.1997.tb02643.x.

[94] Suzanne Lane. “Consequences of Assessment and Accountability Systems Are Integral to the Argument-Based Approach to Validity”. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 71–74. DOI: 10.1080/15366367.2012.679160. (Visited on 01/24/2013).

[95] Jeanette A Lawrence and Agnes E Dodds. “Three approaches to the disjunction between psychological measurement and psychological persons: methodological and ethical considerations”. In: *Integr Psychol Behav Sci* 44.4 (Dec. 2010). PMID: 20490955, pp. 299–309. DOI: 10.1007/s12124-010-9129-4.

[96] David Lewis. “Psychophysical and theoretical identifications”. In: *Australasian Journal of Philosophy* 50.3 (1972), pp. 249–258. DOI: 10.1080/00048407212341301.

[97] Robert W. Lissitz and Tiago Caliço. “Validity is an Action Verb: Commentary on: “Clarifying the Consensus Definition of Validity””. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 75–79. DOI: 10.1080/15366367.2012.677346.

[98] Robert W. Lissitz and Karen Samuelsen. “A Suggested Change in Terminology and Emphasis regarding Validity and Education”. In: *Educational Researcher* 36.8 (Nov. 2007), pp. 437–448. DOI: 10.3102/0013189X07311286.

[99] Robert W. Lissitz and Karen Samuelsen. “Further Clarification regarding Validity and Education”. In: *Educational Researcher* 36.8 (Nov. 2007), pp. 482–484. DOI: 10.3102/0013189X07311612.

[100] Nathalie Lovasz and Kathleen L. Slaney. “What makes a hypothetical construct “hypothetical”? Tracing the origins and uses of the ‘hypothetical construct’ concept in psychological science”. In: *New Ideas in Psychology* 31.1 (Apr. 2013), pp. 22–31. DOI: 10.1016/j.newideapsych.2011.02.005.

[101] A. D. Lovie. “Commentary on Michell, Quantitative science and the definition of measurement in psychology”. In: *British Journal of Psychology* 88.3 (1997), pp. 393–394. DOI: 10.1111/j.2044-8295.1997.tb02644.x.

[102] R. Duncan Luce. “Quantification and symmetry: Commentary on Michell, Quantitative science and the definition of measurement in psychology”. In: *British Journal of Psychology* 88.3 (1997), pp. 395–398. DOI: 10.1111/j.2044-8295.1997.tb02645.x.

[103] R. Duncan Luce and Patrick Suppes. “Representational Measurement Theory”. In: *Stevens' Handbook of Experimental Psychology*. John Wiley & Sons, Inc., 2002. URL: http://onlinelibrary.wiley.com.proxy.uchicago.edu/doi/10.1002/0471214426.pas0401/abstract (visited on 01/27/2013).

[104] William Lycan. “Representational Theories of Consciousness”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2008. 2008. URL: http://plato.stanford.edu/archives/fall2008/entries/consciousness-representational/.

[105] Michael Lynch. “Method: Measurement - ordinary and scientific measurement as ethnomethodological phenomena”. In: *Ethnomethodology and the Human Sciences*. Cambridge Univ. Press, 1991, pp. 77–108.

[106] Norman Malcolm. “Wittgenstein: The relation of language to instinctive behaviour”. In: *Philosophical Investigations* 5.1 (Jan. 1982), pp. 3–22. DOI: 10.1111/j.1467-9205.1982.tb00531.x.

[107] Michael D. Maraun. “Validity and Measurement”. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 80–83. ISSN: 1536-6367. DOI: 10.1080/15366367.2012.682523.

[108] Michael D. Maraun and Stephanie M. Gabriel. “Illegitimate concept equating in the partial fusion of construct validation theory and latent variable modeling”. In: *New Ideas in Psychology* 31.1 (Apr. 2013), pp. 32–42. DOI: 10.1016/j.newideapsych.2011.02.006.

[109] S.W. Margulis. “Measurement Error and Reliability”. In: *Encyclopedia of Animal Behavior*. Ed. by Editors-in-Chief: Michael D. Breed and Janice Moore. Oxford: Academic Press, 2010, pp. 424–428. URL: http://www.sciencedirect.com/science/article/pii/B978008045337800231X.

[110] Keith A. Markus. “Constructs and Attributes in Test Validity: Reflections on Newton’s Account”. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 84–87. DOI: 10.1080/15366367.2012.677348.

[111] Keith A. Markus. “Science, Measurement, and Validity: Is Completion of Samuel Messick’s Synthesis Possible?” In: *Social Indicators Research* 45.1/3 (Nov. 1998), pp. 5–34. DOI: 10.1023/A:1006960823277.

[112] Keith A. Markus and Denny Borsboom. “Reflective measurement models, behavior domains, and common causes”. In: *New Ideas in Psychology* 31.1 (Apr. 2013), pp. 54–64. DOI: 10.1016/j.newideapsych.2011.02.008.

[113] Keith A. Markus and Denny Borsboom. “The cat came back: Evaluating arguments against psychological measurement”. In: *Theory & Psychology* 22.4 (Aug. 2012), pp. 452–466. DOI: 10.1177/0959354310381155.

[114] Jack Martin. “Positivism, Quantification and the Phenomena of Psychology”. In: *Theory & Psychology* 13.1 (Feb. 2003), pp. 33–38. DOI: 10.1177/0959354303013001760.

[115] Per Martin-Löf. “Truth of a proposition, evidence of a judgement, validity of a proof”. In: *Synthese* 73.3 (Dec. 1987), pp. 407–420. DOI: 10.1007/BF00484985.

[116] Rosa L Matzkin. “Nonparametric Survey Response Errors”. In: *International Economic Review* 48.4 (Nov. 2007), pp. 1411–1427. DOI: 10.1111/j.1468-2354.2007.00468.x.

[117] Roderick P. McDonald. “Measuring Latent Quantities”. In: *Psychometrika* 76.4 (Oct. 2011), pp. 511–536. DOI: 10.1007/s11336-011-9223-7 (cit. on p. 6).

[118] Samuel Messick. “Test Validity: A Matter of Consequence”. In: *Social Indicators Research* 45.1-3 (Nov. 1998), pp. 35–44. DOI: 10.1023/A:1006964925094.

[119] Samuel Messick. “Validity of psychological assessment: Validation of inferences from persons’ responses and performances as scientific inquiry into score meaning”. In: *American Psychologist* 50.9 (1995), pp. 741–749. DOI: 10.1037/0003-066X.50.9.741 (cit. on pp. 6, 8).

[120] Joel Michell. “Constructs, inferences, and mental measurement”. In: *New Ideas in Psychology* 31.1 (Apr. 2013), pp. 13–21. DOI: 10.1016/j.newideapsych.2011.02.004.

[121] Joel Michell. “Item Response Models, Pathological Science and the Shape of Error Reply to Borsboom and Mellenbergh”. In: *Theory & Psychology* 14.1 (Feb. 2004), pp. 121–129. DOI: 10.1177/0959354304040201.

[122] Joel Michell. “Normal Science, Pathological Science and Psychometrics”. In: *Theory & Psychology* 10.5 (Oct. 2000), pp. 639–667. DOI: 10.1177/0959354300105004 (cit. on pp. 11–13).

[123] Joel Michell. “Quantitative science and the definition of measurement in psychology”. In: *British Journal of Psychology* 88.3 (1997), pp. 355–383. DOI: 10.1111/j.2044-8295.1997.tb02641.x (cit. on pp. 11, 13).

[124] Joel Michell. ““The constantly recurring argument”: Inferring quantity from order”. In: *Theory & Psychology* 22.3 (June 2012), pp. 255–271. DOI: 10.1177/0959354311434656.

[125] Joel Michell. "The psychometricians' fallacy: Too clever by half?" In: *British Journal of Mathematical and Statistical Psychology* 62.1 (2009), pp. 41–55. DOI: 10.1348/000711007X243582.

[126] Joel Michell. "The Quantitative Imperative Positivism, Naive Realism and the Place of Qualitative Methods in Psychology". In: *Theory & Psychology* 13.1 (Feb. 2003), pp. 5–31. DOI: 10.1177/0959354303013001758.

[127] Robert J. Mislevy. "The Case for Informal Argument". In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 93–96. DOI: 10.1080/15366367.2012.682525. (Visited on 01/24/2013).

[128] Robert J. Mislevy. "Validity by Design". In: *Educational Researcher* 36.8 (Nov. 2007), pp. 463–469. DOI: 10.3102/0013189X07311660.

[129] Michael Morgan. "Measurement in psychology: Commentary on Michell's Quantitative Science and the definition of measurement in psychology". In: *British Journal of Psychology* 88.3 (1997), pp. 399–400. DOI: 10.1111/j.2044-8295.1997.tb02646.x.

[130] Pamela A. Moss. "Reconstructing Validity". In: *Educational Researcher* 36.8 (Nov. 2007), pp. 470–476. DOI: 10.3102/0013189X07311608.

[131] Pamela A. Moss. "Recovering a Dialectical View of Rationality". In: *Social Indicators Research* 45.1/3 (Nov. 1998), pp. 55–67. DOI: 10.1023/A:1006925226003.

[132] Kevin R. Murphy. "Validity for What? The Peril of Overclarifying". In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 97–99. DOI: 10.1080/15366367.2012.677362.

[133] Thomas Nadelhoffer and Eddy Nahmias. "The Past and Future of Experimental Philosophy". In: *Philosophical Explorations* 10.2 (2007), pp. 123–149. ISSN: 1386-9795. DOI: 10.1080/13869790701305921.

[134] Paul E. Newton. "Clarifying the Consensus Definition of Validity". In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 1–29. DOI: 10.1080/15366367.2012.669666.

[135] Paul E. Newton. "Questioning the Consensus Definition of Validity". In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 110–122. DOI: 10.1080/15366367.2012.688456.

[136] Hugh J. Parry and Helen M. Crossley. "Validity of Responses to Survey Questions". In: *The Public Opinion Quarterly* 14.1 (Apr. 1950), pp. 61–80. DOI: 10.1086/266150.

[137] Elazar J. Pedhazur and Liora Pedhazur Schmelkin. *Measurement, design, and analysis: an integrated approach*. Psychology Press, 1991. ISBN: 9780805810639.

[138] Charles Sanders Peirce. "How to Make Our Ideas Clear". In: *Popular Science Monthly* 12 (1878), pp. 286–302.

[139] David Pitt. “Mental Representation”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2012. 2012. URL: http://plato.stanford.edu/archives/win2012/entries/mental-representation/.

[140] Alastair Pollitt. “Validity Cannot Be Created, It Can Only Be Lost”. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 100–103. DOI: 10.1080/15366367.2012.686868. (Visited on 01/24/2013).

[141] Deborah Poole. “Routine Testing Practices and the Linguistic Construction of Knowledge”. In: *Cognition and Instruction* 12.2 (Jan. 1994), pp. 125–150. DOI: 10.1207/s1532690xci1202_3.

[142] Jonathan Potter. “Post-Cognitive Psychology”. en. In: *Theory & Psychology* 10.1 (Feb. 2000), pp. 31–37. DOI: 10.1177/0959354300010001596.

[143] Dag Prawitz. “Inference and Knowledge”. In: *Logica Yearbook 2008*. London: College Publications, 2009, pp. 183–200 (cit. on p. 9).

[144] Dag Prawitz. “The epistemic significance of valid inference”. In: *Synthese* (Mar. 2011), pp. 1–12. DOI: 10.1007/s11229-011-9907-7 (cit. on p. 9).

[145] Dag Prawitz. “Truth as an Epistemic Notion”. In: *Topoi* (), pp. 1–8. DOI: 10.1007/s11245-011-9107-6.

[146] Dag Prawitz. *Validity of Inference*. Abstract. Feb. 2008.

[147] George Psathas. *Everyday language: studies in ethnomethodology*. Irvington Publishers : distributed by Halsted Press, 1979. ISBN: 9780470266700.

[148] Hilary Putnam. *Representation and reality*. Cambridge, Mass.: MIT Press, 1991. ISBN: 0262660741 9780262660747.

[149] Hilary Putnam. “The Collapse of the Fact/Value Dichotomy”. In: *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge, MA: Harvard University Press, 2002 (cit. on p. 6).

[150] Hilary Putnam. *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge, MA: Harvard University Press, 2002. ISBN: 0674009053.

[151] Hilary Putnam. “The Meaning of ‘Meaning’”. In: *Minnesota Studies in the Philosophy of Science* 7 (1975), pp. 131–193.

[152] Hilary Putnam. “The Three Enlightenments”. In: *After Cognitivism: A Reassessment of Cognitive Science and Philosophy*. Springer, 2009.

[153] W. V. Quine. “Two Dogmas of Empiricism”. In: *The Philosophical Review* 60.1 (Jan. 1951), pp. 20–43. URL: http://www.jstor.org/stable/2181906 (visited on 11/04/2009).

[154] Willard Van Orman Quine. *From a Logical Point of View: Nine Logico-Philosophical Essays*. 2d ed., rev. Cambridge, Mass: Harvard University Press, 1980.

[155] W.V.O. Quine. *Word and object*. English. Cambridge: The MIT Press, 1960.

[156] Panu Raatikainen. “Causation, Exclusion, and the Special Sciences”. In: *Erkenn* 73.3 (Nov. 2010), pp. 349–363. DOI: 10.1007/s10670-010-9236-0.

[157] François Recanati, Gerhard Preyer, and Georg Peter. “Literalism and Contextualism: Some Varieties”. In: *Contextualism in Philosophy: Knowledge, Meaning and Truth*. Clarendon Press, 2005, pp. 171–196.

[158] Mark D. Reckase. “The Interaction of Values and Validity Assessment: Does a Test’s Level of Validity Depend on a Researcher’s Values?” In: *Social Indicators Research* 45.1-3 (Nov. 1998), pp. 45–54. DOI: 10.1023/A:1006973109164.

[159] A. Kimball Romney, Susan C. Weller, and William H. Batchelder. “Culture as Consensus: A Theory of Culture and Informant Accuracy”. In: *American Anthropologist*. New Series 88.2 (June 1986), pp. 313–338. DOI: 10.1525/aa.1986.88.2.02a00020.

[160] Philip J. Rosenbaum and Jaan Valsiner. “The un-making of a method: From rating scales to the study of psychological processes”. In: *Theory & Psychology* 21.1 (Feb. 2011), pp. 47–65. DOI: 10.1177/0959354309352913.

[161] Nora Ruck and Thomas Slunecko. “A Stale Challenge to the Philosophy of Science”. In: *Integr. psych. behav.* 44.2 (June 2010), pp. 168–175. DOI: 10.1007/s12124-010-9121-z.

[162] Gilbert Ryle. *The concept of mind*. University of Chicago Press ed. Chicago: University of Chicago Pres, 1984. ISBN: 0226732959.

[163] Uwe Saint-Mont. “What measurement is all about”. In: *Theory & Psychology* 22.4 (Aug. 2012), pp. 467–485. DOI: 10.1177/0959354311429997.

[164] Annette C Scherpenzeel and Willem E Saris. “The Validity and Reliability of Survey Questions A Meta-Analysis of MTMM Studies”. In: *Sociological Methods & Research* 25.3 (Feb. 1997), pp. 341–383. DOI: 10.1177/0049124197025003004.

[165] Verena D. Schmittmann et al. “Deconstructing the construct: A network perspective on psychological phenomena”. In: *New Ideas in Psychology* 31.1 (Apr. 2013), pp. 43–53. DOI: 10.1016/j.newideapsych.2011.02.007.

[166] Peter Schroeder-Heister. “Validity Concepts in Proof-theoretic Semantics”. In: *Synthese* 148.3 (Feb. 2006), pp. 525–571. DOI: 10.1007/s11229-004-6296-1.

[167] Michael Schwarz. “Is Psychology Based on a Methodological Error?” In: *Integrative Psychological and Behavioral Science* 43.3 (Sept. 2009), pp. 185–213. DOI: 10.1007/s12124-009-9089-8 (cit. on pp. 6, 13).

[168] Wilfrid Sellars. *Empiricism and the Philosophy of Mind*. Cambridge, Mass.: Harvard University Press, 1997.

[169] Wilfrid Sellars. “Language as Thought and as Communication”. In: *Philosophy and Phenomenological Research* 29.4 (June 1969), pp. 506–527. DOI: 10.2307/2105537.

[170] David Sherry. “Thermoscopes, thermometers, and the foundations of measurement”. In: *Studies In History and Philosophy of Science Part A* 42.4 (Dec. 2011), pp. 509–524. DOI: 10.1016/j.shpsa.2011.07.001 (cit. on p. 13).

[171] Klaas Sijtsma. “Correcting Fallacies in Validity, Reliability, and Classification”. In: *International Journal of Testing* 9.3 (2009), pp. 167–194. DOI: 10.1080/15305050903106883.

[172] Klaas Sijtsma. “Future of Psychometrics: Ask What Psychometrics Can Do for Psychology”. In: *Psychometrika* 77.1 (Jan. 2012), pp. 4–20. DOI: 10.1007/s11336-011-9242-4.

[173] Klaas Sijtsma. “Psychological measurement between physics and statistics”. In: *Theory & Psychology* 22.6 (Dec. 2012), pp. 786–809. DOI: 10.1177/0959354312454353.

[174] Stephen G. Sireci. “On Validity Theory and Test Validation”. In: *Educational Researcher* 36.8 (Nov. 2007), pp. 477–481. DOI: 10.3102/0013189X07311609.

[175] Stephen G. Sireci. “The Construct of Content Validity”. In: *Social Indicators Research* 45.1-3 (Nov. 1998), pp. 83–117. DOI: 10.1023/A:1006985528729.

[176] Tom W. Smith. “Refining the Total Survey Error Perspective”. In: *International Journal of Public Opinion Research* 23.4 (2011), pp. 464–484. DOI: 10.1093/ijpor/edq052.

[177] Ernest Sosa. “Intuitions and meaning divergence”. In: *Philosophical Psychology* 23.4 (2010), pp. 419–426. DOI: 10.1080/09515089.2010.505859.

[178] Henderikus J. Stam. “The Fault is Not in Ourselves, but in Our Methods: Comment on Schwarz”. In: *Integrative Psychological and Behavioral Science* 44.4 (Dec. 2010), pp. 281–287. DOI: 10.1007/s12124-010-9144-5 (cit. on p. 13).

[179] S. S. Stevens. “On the Theory of Scales of Measurement”. In: *Science* 103.2684 (June 1946). PMID: 17750512, pp. 677–680. DOI: 10.1126/science.103.2684.677.

[180] R. Steyer. “Classical (Psychometric) Test Theory”. In: *International Encyclopedia of the Social & Behavioral Sciences*. Ed. by Editors-in-Chief: Neil J. Smelser and Paul B. Baltes. Oxford: Pergamon, 2001, pp. 1955–1962. DOI: 10.1016/B0-08-043076-7/00721-X. (Visited on 02/18/2013).

[181] Milton E. Strauss and Gregory T. Smith. “Construct Validity: Advances in Theory and Methodology”. In: *Annual Review of Clinical Psychology* 5.1 (2009). PMID: 19086835, pp. 1–25. DOI: 10.1146/annurev.clinpsy.032408.153639.

[182] B. G Sundholm. “Inference versus Consequence”. In: *LOGICA Yearbook 1997, 26 - 35 (1998)*. Prague: Filosofia Publishers, Czech Academy of Science, 1998, pp. 26–35. URL: https://openaccess.leidenuniv.nl/handle/1887/10419 (visited on 01/16/2012).

[183] Shirley Anne Tate. “Foucault, Bakhtin, Ethnomethodology: Accounting for Hybridity in Talk-in-Interaction”. In: *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 8.2 (May 2007). ISSN: 1438-5627. URL: http://www.qualitative-research.net/index.php/fqs/article/view/247/545.

[184] L. L. Thurstone. “A Law of Comparative Judgment”. In: *Psychological Review* 34.4 (1927), pp. 273–286. DOI: 10.1037/h0070288.

[185] Michael Tomasello. *Origins of human communication*. 1st. MIT Press pbk. ed. Cambridge Mass. ;;London: MIT Press, 2010. ISBN: 9780262515207.

[186] Aaro Toomela. “Variables in Psychology: A Critique of Quantitative Psychology”. In: *Integrative Psychological and Behavioral Science* 42.3 (June 2008), pp. 245–265. DOI: 10.1007/s12124-008-9059-6 (cit. on p. 13).

[187] Günter Trendler. “Measurement Theory, Psychology and the Revolution That Cannot Happen”. In: *Theory & Psychology* 19.5 (Oct. 2009), pp. 579–599. DOI: 10.1177/0959354309341926.

[188] Anand J. Vaidya. “Philosophical methodology: The current debate”. In: *Philosophical Psychology* 23.3 (2010), pp. 391–417. ISSN: 0951-5089. DOI: 10.1080/09515089.2010.490940.

[189] Guy C. Van Orden, Christopher T. Kello, and John G. Holden. “Situated Behavior and the Place of Measurement in Psychological Theory”. In: *Ecological Psychology* 22.1 (2010), pp. 24–43. DOI: 10.1080/10407410903493145.

[190] Stéphane Vautier et al. “The ambiguous utility of psychometrics for the interpretative foundation of socially relevant avatars”. In: *Theory & Psychology* 22.6 (Dec. 2012), pp. 810–822. DOI: 10.1177/0959354312450093.

[191] Meike Watzlawik. ““Der Weisheits letzter Schluss””? Wisdom’s Last Conclusion?” In: *Integr. psych. behav.* 43.3 (Sept. 2009), pp. 214–220. DOI: 10.1007/s12124-009-9094-y.

[192] Michael A. Westerman. “Conversation analysis and interpretive quantitative research on psychotherapy process and problematic interpersonal behavior”. In: *Theory & Psychology* 21.2 (Apr. 2011), pp. 155–178. DOI: 10.1177/0959354310394719.

[193] Michael A. Westerman and Stephen C. Yanchar. “Changing the terms of the debate: Quantitative methods in explicitly interpretive research”. In: *Theory & Psychology* 21.2 (Apr. 2011), pp. 139–154. DOI: 10.1177/0959354310393565.

[194] Susan E. Whitely. “Construct validity: Construct representation versus nomothetic span”. In: *Psychological Bulletin* 93.1 (1983), pp. 179–197. DOI: 10.1037/0033-2909.93.1.179.

[195] Peter Winch. *The Idea of a Social Science and its Relation to Philosophy*. English. 2nd. London; New York: Routledge, 1958.

[196] Ludwig Wittgenstein. *Philosophical Investigations*. Ed. by P.M.S. Hacker and Joachim Schulte. Trans. by G. E. M. Anscombe, P. M. S. Hacker, and Joachim Schulte. Rev. 4th ed. Chichester, West Sussex, U.K. ; Malden, MA: Wiley-Blackwell, 2009.

[197] R. Yusoff and R. Mohd Janor. “Generation of an Interval Metric Scale to Measure Attitude”. In: *SAGE Open* 4.1 (Jan. 2014). ISSN: 2158-2440. DOI: 10.1177/2158244013516768. URL: http://sgo.sagepub.com/content/4/1/2158244013516768.full-text.pdf+html (visited on 01/16/2014).

[198] Annemarie Zand Scholten. “The Consensus Definition Redefined From a Representational Perspective”. In: *Measurement: Interdisciplinary Research & Perspective* 10.1-2 (2012), pp. 104–109. DOI: 10.1080/15366367.2012.681978.

[199] Donald W. Zimmerman. “Commentary on ‘Science, Measurement, and Validity: Is Completion of Samuel Messick’s Synthesis Possible?’ by Keith A. Markus”. In: *Social Indicators Research* 45.1-3 (Nov. 1998), pp. 69–72. DOI: 10.1023/A:1006977210073.

[200] Donald W. Zimmerman. “How Should Classical Test Theory Have Defined Validity?” In: *Social Indicators Research* 45.1-3 (Nov. 1998), pp. 233–251. DOI: 10.1023/A:1006949915525.

[201] Bruno D. Zumbo. “Opening Remarks to the Special Issue on Validity Theory and the Methods Used in Validation: Perspectives from the Social and Behavioral Sciences”. In: *Social Indicators Research* 45.1/3 (Nov. 1998), pp. 1–3. DOI: 10.1023/A:1006956722368.