McClimans, L. 2010a. A Theoretical Framework for Patient-reported Outcome Measures. *Theoretical Medicine and Bioethics* 31: 225–40.

McClimans, L. 2010b. Towards Self-determination in Quality of Life Research: a Dialogic Approach. *Medicine, Health Care, and Philosophy* 13: 67–76.

Mendlowitz, M. 1979. Some Theories of Hypertension: Fact and Fancy. *Hypertension* 1: 435–41.

Oort, F. J., Visser, M. R. M., and Sprangers, M. A. G. 2009. Formal Definitions of Measurement Bias and Explanation Bias Clarify Measurement and Conceptual Perspectives on Response Shift. *Journal of Clinical Epidemiology* 62:*11*: 1126–37.

Powers, B. J., Olsen, M. K., Smith, V. A., Woolson, R. F., Bosworth, H. B, and Oddone, E. Z. 2011. Measuring Blood Pressure for Decision Making and Quality Reporting: Where and How Many Measures. *Annals of Internal Medicine* 154:*12*: 781–8.

Salzberger, T. 2013. Attempting Measurement of Psychological Attributes. *Frontiers in Psychology* 4:*75*: 2–3.

Schwartz, C. E. and Sprangers, M. A. 1999. Methodological Approaches for Assessing Response Shift in Longitudinal Health-Related Quality-of-Life Research. *Social Science & Medicine* 48:*11*: 1531–48.

Streiner, D. L. and Norman, G. R. 2008. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford, New York: Oxford University Press.

Tal, E. 2011. How Accurate Is the Standard Second? *Philosophy of Science* 78:*5*: 1082–96.

Van Fraassen, B. C. 2010. *Scientific Representation: Paradoxes of Perspective*. Oxford: Clarendon Press.

Wasserman, D., Wachbroit, R., and Bickenbach, J. 2005. *Quality of Life and Human Difference: Genetic Testing, Health Care, and Disability*. Cambridge: Cambridge University Press.

11 Measuring Intelligence Effectively: Psychometrics from a Philosophy of Technology Perspective

Andreas Kaminski

The history of psychometrics has one constant: the enthusiastic development of means for measuring various kinds of mental performance (to which I will refer simply as measurement of mind) is accompanied by a shadow-like critique that is no less forceful. In the mid-19th century, when Gustav Theodor Fechner—following on from Johann Friedrich Herbart and Ernst Heinrich Weber—presented the idea of psychophysics with great verve, he often met with vehement rejection alongside widespread approval. Critical objections have often been raised on principle, such as that in the domain of the psyche only intensive quantities exist which cannot be measured (see Kant 1956 [1787]: B207–218). Another point of critique has been that there is no adequate psychological measure for the psyche. When research on intelligence began around 1900: it very quickly became an area of international research which was rapidly applied in schools, the military, corporations, and psychiatry. To this day, intelligence research has been vehemently criticized, partly because of its alleged dependency on language and culture. Likewise, the measurement of competence, which developed in the 1970s, has also been accompanied by its Siamese twin, critique of the measurement of competence. Most of these critiques originate from the philosophy of science. In the following, I would like to show that this approach is not sufficient. It needs to be supplemented by arguments from the philosophy of technology.

In order to explain why I believe this to be the case, I begin by presenting two theses that serve to illuminate the connection between psychology and the philosophy of technology.

The first thesis is: Psychology has practical and not just epistemological implications. The American philosopher Richard Rorty summed up the debate around psychologism by noting that "the rise of empirical psychology had raised the question 'What do we need to know about knowledge which psychology cannot tell us?'" (Rorty 1979: 165). While Rorty's question manages pointedly to expose the heart of the debate around psychologism, the way he formulates it simultaneously reveals an obvious blind spot. Psychology and psychometrics have been understood as challenges to the theory of cognition and the philosophy of science. But is this primarily what

they are? I have my doubts, and I would like to justify them using the case of psychological examination techniques. The debate about psychometrics has revolved particularly around the epistemological question of whether it is *possible* to measure mental performance. Beyond this epistemological question, however, lies a further issue, which I elucidate in the following.

Psychometrics is effective inasmuch as it influences the ways in which people relate to themselves (their self-relation), and it does so regardless of the answer to the question of its feasibility or ability to produce reliable knowledge. Psychometrics may be an impossible science epistemologically, but it is nonetheless an effective technology because it changes a person's self-relation. This is because the benchmark used in psychometrics is not external. If I were to measure the length of an object by placing a measuring rod next to it, my act of measuring does not change the object. If I measure the intelligence of a person, however, this benchmark does not remain external to that person. Why?

The difference is not that, in the one case, the influence of the measuring tool on the object remains negligible, whereas in the other case—the measurement of intelligence—the measuring tool is of practical relevance. In the second case, it is not a matter of *influence*, but rather a matter of the *change in the subject's self-relation*, which may result in the individual forming a concept of his or her self from the point of view of the benchmark. This in turn may lead them to adjust their behavior in order to conform to the benchmark. Thus, the problem with any psychological examination technique lies in the fact that—to use the analogy of the measurement of length—the subject (the person) begins to "stretch" maybe as soon as the psychometric benchmark is applied, because it knows that it is being measured.

The case of psychometrics is thus highly suitable as an example to point out what seems to me to be a characteristic shared by psychology and by the debates surrounding it: it is concerned not only with implications in the fields of scientific theory or the theory of cognition but also with social and technical implications.

This is the second thesis: The success of psychology is due not to its theoretical principles but rather to its potential for technical application. The rise of psychology is not correlated with advances made in basic knowledge or fundamental understanding but with the fact that it opened up a number of fields of application for itself after 1900. The conviction that psychology offers solutions to "practical problems"—the "tasks of culture"—is based on the notion that it is capable of providing answers to certain questions in a technical form, as it were, specifically in the form of measurements. However, the models used in psychology (such as cognitive load theory) represent the psyche in a manner analogous to technology. Where psychologists—far away from psychoanalysis—have been engaged in practical work, it is certainly not their underlying theories that have been most relevant. The lasting contribution of William Stern, for example, has not been his theory of intelligence or his coining of the concept of psychotechnics or even his theory of personality;

what has attracted attention and gained widespread acceptance is his notion of an intelligence quotient, or "IQ." To put the point somewhat starkly: in contexts where it has had practical relevance, psychology in the 20th century has always also been "psychotechnics".

My contribution is divided into three parts. In the first part, I will discuss three key arguments in the critique of psychometrics. My purpose is to show that the critique of psychometrics has an underlying blind spot. The different positions I discuss share a common perspective which is so self-evident to them that they fail to recognize a major missing factor. The question they all pose is whether it is *possible* to measure the mind. In doing so, however, they fail to take account of the *effectiveness* of psychometrics, that is, the fact that psychometrics changes not only social relations but also self-relations.

In the second part, I will attempt to show what the reasons are for the effectiveness of psychometrics by comparing psychometric and classical forms of measurement. My thesis is that there is one important difference in psychometrics: self-relations are influenced by psychometric measurement techniques, because subjects behave in relation to the scale which is used to measure them.

In the third and final part, and using the example of the Lynn-Flynn effect, I will try to define the consequences of this behavior of subjects in respect to scale. This part will demonstrate the extent to which the philosophy of subjectivity is relevant to psychometrics.

The Critique of Psychometrics—and its Blind Spot

It is possible to identify three distinct positions within the critical epistemological debates surrounding psychometrics. They offer valuable analyses of the problems entailed in acquiring knowledge by measuring mental performance and they are also helpful when it comes to evaluating the legitimate and illegitimate claims made by psychometrics. However, these three critical positions share a common perspective that serves to render a key issue invisible. In the following I seek to make this blind spot visible by looking at each position in turn.

The first critical position assumes that measuring mental performance is *possible in principle*, even if in practice the methods for doing so require improvement. Denny Borsboom, for example, assumes that psychometrics is a valid approach in principle but also points out that its weakness lies in the lack of appropriate theories to address the object being measured. Borsboom sees considerable effort being expended upon the development of methods and technologies of measurement, yet without any corresponding development of a substantial theory of the measured object. Without such a theory, says Borsboom, the validity of the measurements cannot be verified (see Borsboom 2005).

The second critical position argues that it has not yet been established *whether* the measurement of mental properties and performance is *possible*.

This might or might not be the case—it has not yet been determined. Joel Michell has developed this position in a series of publications. Michell frames the question of the feasibility of psychometrics by noting that it needs to accomplish two tasks, a scientific one and a methodical and technical one. The scientific task consists in showing that the prerequisite for measurement—namely the quantitative nature of the object—is satisfied. In order to establish criteria that determine whether or not an object is quantitative, Michell turns to the work of mathematician Otto Hölder (1901) who defined the "axioms of quantity" (see Hölder 1901). Michell assumes that the proof of an object having a quantitative nature can only be provided empirically, and he develops methods for doing this. His criticism of psychometrics is that it has never taken on this first scientific task. Instead of demonstrating that its object is quantitative, psychometrics has simply acted as if it were so. It has largely devoted its efforts to the development and refinement of methods, none of which would be of any use if applied to an a priori inadmissible object. In particular, Michell critically notes that psychometrics does not recognize that the question of whether or not psychometrics itself is possible remains unresolved (see Michell 1997; 2000; 2005). In other words, the feasibility of psychometrics has yet to be determined.

The third position is based on the assumption that measuring mental properties or performance is (probably) *impossible*. Even if a proof of the quantitative nature of the mind were given—as required by Michell—psychometrics would be an impossible science because the object, the psyche, does not satisfy other pre-requisites for measurement. A critique of psychometrics along these lines has been put forward by Trendler (see Trendler 2009). In order to be measurable, an object must be capable of being clearly located within a simplified causal nexus: apart from a few causal relations, all others would have to be excluded. In addition, it would have to be possible to manipulate rather precisely the remaining causal relations. Within the natural sciences devices and technical environments have been created which satisfy both conditions. The same has not been true regarding the psyche. Neither causal reduction nor causal manipulation is given to a sufficient extent, and there is no prospect of any change in this situation. For this reason, psychometrics is not feasible as a science.[1]

What do these three critical positions have in common? The first assumes that while it is theoretically possible to gain information about the mind by means of measurement, methodological improvements are required to do so. The second deems it to be undetermined whether measurement of the mind is possible in principle, while the third denies this possibility on practical grounds. These three strands of criticism are focused upon the possibility or impossibility of psychometrics. What is being addressed is the relationship between measurement and mind. All three positions examine whether this relationship is suited to the task. Is the object suitable for measurement (positions 2 and 3)? Is the measurement theory suited to producing an adequate measurement of the object (position 1)?

These approaches seem so self-evident that it is indeed difficult to identify what they do not take into account. As valuable as the critiques of the claims of psychometrics are, they disregard an important point, namely, that psychometrics is *effective*, no matter whether—epistemologically speaking—it is feasible or not. By focusing on the issue of whether or not psychometrics is *feasible* we have lost sight of the fact that it is *effective* (even if its claims are epistemologically inadmissible). The effectiveness of psychometrics can be observed in the way it changes social self-relations. People appreciate themselves differently when they see themselves reflected in the mirror of intelligence tests or measurements of competence. They regulate their behavior in different ways, aligning themselves with the scales of the measurement being conducted.

In order better to understand this effectiveness, we need to proceed to the next issue: what distinguishes psychometrics from classical methods of scientific measurement?

An Object that Behaves According to the Scale Applied to It

Let us turn to the question of how psychometrics differs from classical measurement practices. By way of example, we can compare the measurement of length or temperature with that of intelligence. The most familiar attempts at establishing the differences can be traced to the differentiation of intensive and extensive properties, as described by Kant in his *Critique of Pure Reason* (see Kant 1956 [1787]: B207–218). In order to contest the measurability of mental properties, let us examine the attempts at differentiation which (more or less clearly) followed Kant's line of argument.

Mental properties are not additive

This approach dates from Norman Campbell's theory of fundamental measurement. Campbell assumed that criteria can be stated for objects that are measurable (see Campbell 1920: 267–94). Among these criteria is the requirement that the objects' properties must be transitive, asymmetrical and, above all, additive. Objects' properties are additive if empirical operations of addition can be stated for them. Only then can corresponding mathematical operations be conducted with these properties. Concerning weights and lengths, this means that weights can be added to the scale pan or lengths can be increased by placing another object next to the original one: these are empirical operations of addition. As far as mental properties are concerned, no such operation can be readily identified. The problem with this criterion is that Campbell mentions fundamental as well as derived measurements. Density, for example, does not allow for an immediate empirical operation of addition. However, it is measurable by derivation, through the use of other properties which themselves are additive. In principle, such procedures could be found for mental properties, as psychophysics has attempted to do.

By doing this however, the proposal would lead back to the debate about whether mental properties are quantitative or not. This is why the criterion is inadequate: it merely leads back to the unanswered question asked by Michell (see Michell 1997).

There is no mental measure

Could the fact that there is no mental measure (or that no such measure can be set as a stable standard) constitute a distinctive feature? If this were the case, measurement of the mind would be different from that of nature. While neither has any pre-ordained measures, nature provides us with measures which, due to their far greater stability and commensurability, are much more suited as scales of measurement. As a way out of this, Fechner devised a very clever notion of using those differences in perception which are just noticeable as a scale. Quite apart from the controversy surrounding psychophysics, though, a similar approach seems impossible for other mental phenomena (other than sensation).

However, the history of measurement practices shows that other classical measurement methods posed similar problems in their early stage of development. In his detailed account of how temperature came to be a scientific concept, Hasok Chang has shown how difficult it was to find "fixed points" in temperature measurement and to define them in a way that would allow them to be applied in practice (see Chang 2004). When does water boil? When does it freeze? Those seemingly natural "fixed points" seem to the naked eye to be more like "ranges" rather than points and are therefore quite unsuitable to fix the calibration of a thermometer, as Chang has shown. Similarly, intelligence does not display an immediate, quasi-natural reference point, either. In any case, the mental measure of 1 IQ has a different status than 1 m. But here, too, statistical procedures have made it possible to find "fixed points," namely, by generating average points on a distribution curve (see Terman 1916).

Failure and alternative

These obvious attempts at differentiating between classical and mental measurement therefore fail. Indeed, the attempts at differentiation suggest that one might conceptualize subjectivity as something that is romantically indeterminable, as it were. The following proposed means of differentiating between classical and mental measurement does not proceed on the assumption of such a romantically inscrutable subject.

In psychometrics—unlike in classical forms of measurement—the object exists in active relation to the benchmark applied. This is not the case with the measurement of so-called external nature. No body that has a ruler laid against it adapts its size to the scale—i.e. by becoming or seeming to become larger. No temperature that is measured adapts its heat to the scale

of the thermometer applied to it. People who have psychometric measurement techniques applied to them, however, will generally adapt themselves to the benchmark, that is to say, to the test criterion. People who are subjected to an intelligence test will by and large seek to behave in such a way that the test result certifies high intelligence. In general, when someone is subjected to a competence test, they will attempt to appear at their most competent. Comparing this to the measurement of length, this would be the equivalent of the object being measured stretching itself to appear as long as possible.

A general difference between psychometric testing techniques and classical measurement techniques thus becomes plain: in psychometrics the criterion is not external to the subject. Instead, the subject relates to the criterion which is applied to it. The classical relation is that between measurement and object. This is the point on which psychometrics's classical critical positions focus, as seen above. They ask whether this relation is possible and whether it has been adequately conceptualized. However, this is where it becomes apparent that the relation between measurement and object occurs in the object—in psychometrics, the subject—itself; it is reflected by it and affects its behavior. Within the context of psychometrics, the subject behaves relative to this relationship, and therein lies the effectiveness of psychometrics, which exists regardless of its epistemological possibility.

The effectiveness of psychometrics is by no means restricted to the actual test event. In order to understand this, it is important to realize what a benchmark is when seen in the context of psychometrics. It is more than merely a scale; it is also a dimension of subjectivity, as a benchmark indicates what subjects are capable of being (or becoming). Intelligence tests not only determine a person's general ability to perform certain mental tasks; in order to do this, they also need to state what subjects may be (i.e. "intelligent" or, to take other standards, "weak-/strong-willed," "extroverted," "able to deal with conflict"). What subjects can be, what defines their subjectivity, what their subjectivity consists of, changes with the appearance of intelligence tests. At an elementary level, the self-conception of those who ask themselves "who am I?" and "who can I be?" changes when their intelligence is measured.

This comes to the fore with intelligence tests in particular, given that even today such tests are associated with scholarly and professional achievement. Intelligence tests are regarded as predictive tools—they purport to predict the future of a person in an important area of their personality. Measurement procedures also change people's self-conception in other areas of cognition. Thus personality and competence testing define a grid of potential subjectivity. They construct a "space of possibility" that shows how subjects can be—and they assign an individual a certain location within this space.

Alongside the issue of self-conception, psychometrics acts on another level, namely, that of *work done on the self*. The benchmark not only creates an understanding of what and how subjects can be but also serves as a prompt for self-control and self-formation. The scale of the benchmark is a scale of

values. It allows for both downward and, especially, upward orientation. It gives feedback on where a person is and whether progress is being made. Practice, self-control and self-formation are aspects of the effectiveness of psychometric processes.

For the reasons just outlined, then, it is not sufficient merely to ask whether measurement is possible. Rather, the issue that requires additional attention is that it is effective. Regardless of whether or not the mind is quantitatively constituted and regardless of whether or not it is accessible to measurement from an epistemological perspective: since the object being measured exists in active relation to the means of measurement, the act of measurement can be described as effective because of the behavior it induces. In other words, the benchmark is not external to the subject but can rather be acquired by the subject as a powerful social benchmark. The consequence of this is that our society and our self-reference (because an ambitious notion of *self* can only be conceived of in the context of social relations with others) changes with psychometrics—and with the introduction of new psychometric methods of testing. These change people's self-conceptions and the way they work on (act upon) themselves. To name two of the breaks within this psychometric history of subjectivity, humans became intelligent around 1900 and became competent around 1970!

Traces of Psychometrics: the Lynn-Flynn Effect as a Change in Ways of Thinking

In the first part of this chapter I showed that, in asking first and foremost whether psychometrics is possible, scientific theory loses sight of the reality and effectiveness of the measurement practices of the mind. In the second part I established this effectiveness by showing that the object—the subject—behaves relative to the benchmark applied to it. With psychological measurements, the self-conception and the work a subject does on herself change. These two observations may create the impression of two perspectives that co-exist side by side without having much to do with each other: the epistemological perspective on the one side and the perspective inspired by the philosophy of subjectivity on the other. I do not believe this to be the case, however. The discussion in the second part of the chapter regarding the reflexivity prompted by a benchmark has consequences for epistemological issues. In this third part I would like to offer, all too briefly, a few indications of what these consequences are.

The so-called Lynn-Flynn effect belongs to the history of the measurement of intelligence. It was first observed by psychologist Richard Lynn and was made more widely known by political scientist James R. Flynn in the 1980s. This effect consists in a significant rise in the level of intelligence recorded over the course of a few decades. The background is as follows: intelligence tests have to be standardized (that is, re-calibrated) repeatedly. This is done by selecting a representative sample of test subjects (the standardization sample)

in order to establish new measurement norms. The standardization is done using the sample average, which is defined as being 100 points. When this process of standardization is repeated, changes in the level of intelligence of the standardization samples become invisible. This was the starting point for Flynn's idea. He calculated the scores for later test samples using the earliest standardization, which went back to the year 1932 in his data set. The result was that all the later samples showed a continuous, relatively linear, and equally observable increase in the level of intelligence for all groups within the normal distribution range when they were calculated uniformly using the standardization from 1932. Flynn examined test samples from 1932 to 1978 in this way and observed an increase in the average by about 13.8 points (see Flynn 1984).

Various explanations have been offered for this effect, from changes in nutrition through a change in environmental influences (media) to improvements in medical care. Some researchers have commented that intellectual and abstract ways of thinking have become more a part of everyday life—from schools to work environments—than was the case 100 years ago. Without wishing to ally myself with any single monocausal explanation, any one of these inferred explanations would provide a close match for the arguments presented in the second part of this chapter. Intelligence tests apply a different benchmark for measurement than that of mere memorization assignments: subjects adjust their behavior relative to an intelligence benchmark—which is simultaneously a value-based one and one which holds out the prospect of scholarly, professional and financial success. Work on the self and self-control relative to this benchmark, lessons at school and exercises at home have changed. Given that intelligence is a (not merely natural) disposition and that every disposition must be developed by work, it may not be so surprising after all that subjects' ways of thinking have changed over time; after all, intelligence tests have had and still have a good reputation. Broadly speaking, the subjects of psychometric testing are not just thinking about different *content*, they are thinking in a different *way* than they did 200 years ago.

Note

1 For objections to this argument see Markus and Borsboom (2012).

References

Borsboom, D. 2005. *Measuring the Mind. Conceptual Issues in Contemporary Psychometrics*. Cambridge, New York: Cambridge University Press.

Campbell, N. R. 1920. *Physics. The Elements*. University of Michigan: University Press.

Chang, H. 2004. *Inventing Temperature. Measurement and Scientific Progress*. Oxford, New York: Oxford University Press.

Flynn, J. R. 1984. The Mean IQ of Americans: Massive Gains 1932 to 1978. *Psychological Bulletin* 95:*1*: 29–51.
Hölder O. 1901. Die Axiome der Quantität und die Lehre vom Mass. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Classe* 53:*1*: 1–64.
Kant, I. 1956 [1787]. *Kritik der reinen Vernunft*. Hamburg: Meiner.
Markus, K. A. and Borsboom, D. 2012. The Cat Came Back: Evaluating Arguments Against Psychological Measurement. *Theory & Psychology* 22:*4*: 452–66.
Michell, J. 2005. *Measurement in Psychology. A Critical History of a Methodological Concept*. Cambridge: Cambridge University Press.
Michell, J. 2000. Normal Science, Pathological Science and Psychometrics. *Theory & Psychology* 10:*5*: 639–67.
Michell, J. 1997. Quantitative Science and the Definition of Measurement in Psychology. *British Journal of Psychology* 88:*3*: 355–83.
Rorty, R. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
Terman, L. M. 1916. *The Measurement of Intelligence. An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet-Simon Intelligence Scale*. Cambridge: The Riverside Press.
Trendler, G. 2009. Measurement Theory, Psychology and the Revolution That Cannot Happen, *Theory & Psychology* 19:*5*: 579–99.

12 The Klein Sexual Orientation Grid and the Measurement of Human Sexuality

Donna J. Drucker

As sex research developed into a scientific field in the late 19th century, researchers have struggled with how something as mysterious, ephemeral or to use the word that many other scholars have—fluid—could be ordered for scientific study. Sex researchers from the 1920s onward, including Robert Latou Dickinson, Katharine Bement Davis, and Gilbert V. Hamilton, experimented with quantitative measurements for their work to complement the qualitative (see Dickinson and Beam 1932 and 1934; Hamilton 1929; Davis 1929). Examining quantitative forms of sexual measurement illuminates not only what such measures have been for, but also whom they have been for—how effective (or not) they are in developing scientific thought, and also how useful they are for public and individual understanding of sexuality. This essay focuses on one of the best known quantitative measurements, the Klein Sexual Orientation Grid (aka KSOG, or Klein Grid)—its theoretical and historical development, its use in scientific study, its transference to the Internet, and its use in personal discovery, and what it does and does not show about human sexuality. Overall, the Klein Grid is a tool best designed for researchers and professional psychologists working with clients, although it, along with other tools, remains popular with and relevant to individuals seeking instruments for personal discovery.

Fritz Klein, the creator of the grid that bears his name, was born in Vienna in 1932 but moved to New York as a small child. He returned to Europe as an adult to attain his doctorate in medicine at the University of Bern, Switzerland, and later completed a master's of business administration degree at Columbia University in New York City. He divided his adult life as a board-certified psychiatrist between New York and San Diego, California. As part of his involvement in the sexual revolution of the 1970s, he founded "The Bisexual Forum" in those two cities, where bisexuals met to discuss common problems and to create community. He authored *The Bisexual Option* in 1978 (reprinted 1993). He was also the founder and first editor of the academic *Journal of Bisexuality*, and remained involved in research and education concerning bisexuality until his death in 2006.[1]

While the scientific discourse about bisexuality began with Sigmund Freud and Freud's term "psychological bisexuality," Klein's work initiated a new