PSYCHOLOGIA WYCHOWAWCZA
NR 15/2019, 99–110
ISSN 0033-2860
DOI: 10.5604/01.3001.0013.2967
Data wpływu: 09.2018
Data przyjęcia: 02.2019

KULTUROWE I SPOŁECZNE KONTEKSTY WYCHOWANIA

GEORG LIND
Universität Konstanz

THE DILEMMA OF PSYCHOLOGY: OBJECTIVE WITHOUT AN OBJECT?

Abstract: Anyone who seeks the service of psychology (which translates to "science of the mind") faces a persisting dilemma. One has to choose between two psychologies: "Subjective", also called "qualitative", psychologists believe that the focus on studying the internal structure of the human mind will provide important insights needed in therapy and education. Yet the human mind, they argue, can be studied only with *subjective* methods like clinical interview, not with standardized tests. In contrast, "objective" or "quantitative", psychologists argue that if psychology wants to be recognized as a science, it must enlist only *objective* methods of measurement. Yet this excludes, they argue, the study of *internal* psychological factors of the human mind. While the subjective approach is based on psychological assumptions regarding the nature of the target measurement object, the objective approach is based on purely statistical theories. Must we really have to abandon psychological objects like intellectual and moral capacities if we want our measurement to be objective? In this paper I show that both approaches are based on questionable theories about the relationship between visible behavior on the one side and psychological objects on the other. I also show that we can measure psychological traits objectively and validly if we use an experimental approach. *Experimental Questionnaires* can be used in all fields of psychology in which testable theories about the nature of its object have been developed. We have successfully used this new approach, for example, for the construction and validation of the *Moral Competence Test* (MCT).

Keywords: psychological measurement, standardized tests, theory, objectivity, validity.

THE PERSISTING DILEMMA OF PSYCHOLOGY

Millions of dollars are spent every year on tests of character, academic abilities, vocational skills, mental disorders and so on, in the hope that their findings will help to improve therapy, education and the politics of mental health and education (Gregory, 2018, p. 22). But anyone who seeks the service of psychology (which translates to "science of the mind") faces a persisting dilemma: One has to choose between two opposite approaches to the observation and measurement of psychological traits. "Subjective" (also called "qualitative") psychologists believe that the focus on studying the

Correspondence address: georg.lind@uni-konstanz.de, ORCID: https://orcid.org/0000-0002-8411-6363.

internal structure of the human mind will indeed provide badly needed insights on the human condition. The human mind, they insist, can be studied only by using *subjective* methods like clinical interview. In contrast, "objective" ("quantitative") psychologists argue that if psychology wants to be recognized as a science, it must use only *objective* methods of measurement.

For many years eminent scholars have argued that this dilemma has prevented psychology from developing into a real science (Travers, 1951; Loevinger, 1957; Miller, 1962) and from playing a more constructive role in evaluating and improving education (Schoenfeld, 1999; Ravitch, 2013).

For centuries, psychology was part of philosophy and, therefore, the domain of subjective psychology. Philosopher-psychologists focused on the nature of the human mind and used subjective methods in studying it. Their methods were mostly *ideographic* (acknowledging the individuality of the person) and *holistic* (taking the whole structure of the individual personality into account).

These subjective psychologists were challenged in the 19th century by objective psychologists who were at home in physics, biology and medicine. They argued that psychological research must be *nomothetic* (searching for general laws) and *objective,* studying people's behaviors instead of the structure of their mind: "The behaviorist recognizes no such things as mental traits, dispositions or tendencies," postulated Watson (1970/1924, p. 98), whose *behaviorism* is still very influential. Objective psychologists still believe that psychological measurement should focus on behavior instead of on psychological traits: "A test is a standardized procedure for sampling behavior and describing it with categories or scores" (Gregory, 2018, p. 23). Their object is only somehow "related" to psychology: "We define psychological assessment as the gathering and integration of psychology-related data" (Cohen, Swerdlik, 2018, p. 2). So, ironically, objective psychological observation lacks an object.

This antagonism of the two psychologies has caused a deep "crisis of psychology" (Bühler, 1927). The philosopher Wittgenstein (1953) noted: "In psychology there are experimental methods and conceptual confusion. The existence of experimental methods makes us think we have the means of solving the problems which trouble us; though problem and methods pass one another by." Eminent psychologists agree. Graumann (1960) wrote: "Theoretical frameworks and methodological convictions are still too divergent, if not partially incommensurable" (p. 146, my transl.). Similarly Block (1977) asserts that "perhaps 90% of the studies are methodologically inadequate, without conceptual implication, and even foolish" (p. 39). The educational researcher Travers (1951) noted "that the rather meager advances made in many areas of psychological measurement during the last 20 years are mainly a consequence of the fact that these areas are staffed mainly by technicians interested in producing useful instruments and not by scientists interested in expanding knowledge" (p. 137). The statistician and psychologist Kempf (1981) wrote: "What usually is called psychological test theory is actually a statistical theory of item selection in order to produce a test with some desirable features" (p. 3, my transl.). Ten years later, the former president of the *American Psychological Association* Alan Schoenfeld (1999), an accomplished educational researcher and mathematician, complained that still "virtually none of the current assessments in wide use are grounded in well-developed theories of competence" (p. 12). He called for a moratorium on standardized testing until this basic issue of standardized testing has been solved.

In psychology, it seems, we have to choose between Scylla and Charybdis, that is, between a subjective approach object which is not accessible through objective methods of measurement, on the one hand, and an objective method without a psychological object, on the other.

Measurement theories are the link between reality and our images of reality. If these theories are not true then measurement does not provide us with valid data for reviewing our images of the world and for developing effective methods of education and therapy. Hence it is essential that measurement theories are testable – and that such tests are actually made.

THE "SUBJECTIVE" PSYCHOLOGISTS' DILEMMAS

Subjective psychologists base their measurement on the assumption that our behavior is determined mainly by unconscious affects and cognitions. In other words, they believe that only through the study of the unconscious level of our mind we can really understand human behavior and make education, therapy, and politics more effective. They also believe that unconscious affects and cognitions cannot be assessed directly but only indirectly, namely through interpreting people's visible performances in certain situations or their answers to the psychologist's questions. Interpretation means that measurement requires the researcher to make some assumptions about the relationship between people's visible behavior and their unconsciousness. They derive these assumptions from elaborate theories of the human mind that is, for example, from psychoanalytic theory, cognitive-developmental theory, or hermeneutics, to name a few.

Take for example Lawrence Kohlberg's research on the nature and development of moral judgment competence. He defined this "as the capacity to make decisions and judgments which are moral (i.e. based on *internal* principles) and to act in accordance with such judgments" (Kohlberg 1964, p. 425; emphasis added). "Morality […] defined as 'having moral principles' includes […] inner conformity to them in difficult situations rather than outer conformity in routine situations" (Kohlberg 1958, p. 3). Kohlberg's definition agrees well with most moral philosophy: (1) It defines the affective aspect of morality in terms of the individual's *inner* moral principles or orientations, (2) it defines the cognitive aspect of morality as the *structure* of his or her moral judgments, and (3) it defines both as *aspects of visible action* or *behavior*. Subjective moral psychologists argue that objective approaches to observing morality, like behaviorism, must fail because they either do not define their concept of morality at all, or, if they do, their definition is faulty: "The trouble with such [studies] is that they describe the person externally in terms of his impact on and relation to his culture and to the observer's values. They will not tell us how the individual thinks, what values he *actually* holds" (Kohlberg 1958, p. 82; emphasis added).

In order to find out how the individual thinks about moral issues, Kohlberg developed a clinical interview method, the *Moral Judgment Interview* (MJI). In this, the interviewer confronts participants with several dilemma stories in which the protagonists have to make a presumably difficult decision: Whatever they decide, they violate a moral principle. The participants should say whether they agree or disagree with the protagonist's decision and why. The interviewer follows up their answers to get a rich picture of their reasons, and also probes into counter-arguments: Which reasons

could justify the opposite opinion? The answers of the interviewees are recorded, transcribed, and then categorized by a trained scorer into one of the six "cognitive-developmental stages" which Kohlberg (1984) had defined.

This method is based on two postulates, namely (a) that people's moral cognitions are organized as a structural whole and (b) that they develop in a pre-determined invariant sequence. He considered the scoring to be valid only if it agrees with these two postulates. Since the interview data did not agree well enough with these two postulates he and his students revised the scoring system several times in order make it better fit with the data (Kohlberg 1984). But through this the *Moral Judgment Interview* has become immune against falsification, which makes us doubt whether its underlying postulates are true and the interview really measures what it was originally supposed to measure (Lind, 1989).

THE DILEMMA OF "OBJECTIVE" PSYCHOLOGY

"Objective" psychologists believe that if psychological theories bias measurement the best way to avoid such bias is to avoid theories. Measurement, they demand, should be based only on visible acts or behavior, which, they believe, requires no theory (Watson, 1923). If we want to know from them what "intelligence" means, they merely say, "intelligence is what an intelligence test measures" (Bailey, 1994, p. 57). However, they cannot avoid theoretical assumptions. Instead of psychological theory they base their measurement on a statistical theory. This theory determine which items and which scoring methods are regarded as valid. In other words, their statistical models define their object of measurement. Statistical models, they seem to believe, are more objective than psychological models.

The famous *Studies in the Nature of Character* by Hartshorne and May (1928) are a good example of objective behaviorists' approach to psychological measurement. Funded by a church organization, they wanted to test experimentally the hypothesis that character exists and that it is fostered through religious instruction. They confronted participants with situations in which they were tempted to cheat and observed how they reacted. They recorded the agreement or disagreement of these reactions with their standard of honesty. They explicitly discarded any psychological and philosophical interpretations of their subjects' behavior, because "no progress [of psychological science] can be made, however, unless the overt act be observed and, if possible, measured without any reference, for the moment, to its motives and its rightness or wrongness" (p. 11).

Obviously, the authors believed that we can read the character strength of a person directly from his or her reactions to temptations to deceive, like reading a temperature scale: the current temperature is simply the reading on the display plus/minus some error of reading or of malfunctioning of the scale. Similarly they believe that we can reduce measurement error just by reading those reactions several times and calculating the average score.

However, objective psychologists is also based on a theory, too, not on psychological theory but on a statistical theory such as "Classical Test Theory" or the *Theory* of *Mental Tests* (Gulliksen, 1950), and its variants like "Item Response *Theory*," IRT. Notably their theory is not about psychological objects but about statistical constructs, for example, about "latent variables," "latent classes," or statistical

"factors." Through this theory, they create their own object of measurement, which may best be described as a "*homo statisticus*." Although the textbooks on CTT and ITR are usually voluminous (e.g., Cohen, Swerdlik, 2012, has 612 pages), this homo statisticus is described by a very simple statistical formula: Y = T + e. This formula means that the reading of the scale ("Y") is simply the addition of the subjects' "true" behavior ("T") and some random error ("e"). The formulas of more sophisticated statistical test theories are more complex but are essentially based on similar statistical assumptions (Wilson, 2005).

The shortcomings of this test theory become obvious when we translate the statistical formulas into plain language:

- *Observation is simple*. Objective psychologists believe that we can directly read the participants' behavior without any psychological interpretation. As we have seen above, they believe that, for example, the participants' behavior in an honesty experiment enables us to directly read his or her character. By definition this behavior is not affected by any other factor like the type of temptation, the participants understanding of the test, or by their moral competence.
- *Error is random*. Any aberration of this statistical model from the real data is believed to be caused only by a *random* error of "reading the scale," meaning that no systematic factor of the participant or of the testing circumstances affects our reading of people's behavior.
- *Repeated observation of identical behavior is possible*. Because objective psychologists belief that any error is purely random it averages to zero. Therefore, they assert, measurement error can be simply reduced to any smallness simply by repeating the reading as often as needed (the so called "law of large numbers"). But this forces us to believe that people respond to replications of the questions or task always in the same way, and that they are willing to do so. But not even objective psychologists seem to believe this.
- *Similarity of behavior can be determined purely statistically*. Because objective psychologists avoid psychological concepts, they use statistical means for defining the "similarity" of tasks and questions. They define that two behaviors are similar if the participants show them together. So, for example, if people answer two different questions in the same way, they are considered similar, or, if they solve task A and also task B in a math test, these two tasks are considered similar. If the items do not show statistical similarity they are excluded from the test even though they may be considered as highly valid by experts on the subject matter. Note that if all test items which threaten the reliability of the test are excluded from the test as "dissimilar," the measurement model becomes immune against refutation through data. This immunization violates a basic standard of good science, namely refutability (Popper, 1967). It also calls the objectivity of objective psychology into question and creates an illusionary reality.
- *Error and reliability are an attribute of the measurement instrument*. If they were an attribute of the measurement, they would not change from one application to another. But they do. Item selection does NOT lead to a stable estimate of a test's reliability but it varies from one test sample to another and from one test administration to another. For example, even though PISA tests are carefully trimmed on the basis of many prior studies and replacement of "unreliable" tasks, the final tests still deviate substantially from the statistical model on which their construction was based (Wuttke, 2006; Jablonka, 2007).

Because objective psychologists avoid testing their hypothesis of random error, they overlook any systematic bias and ambiguity of their measurements (Wuttke, 2007; Jablonka, 2007). They overlook, as Scott (1968) showed, that scores in the middle range of an attitudes scale can have three very different meanings: (1) they can mean, as researchers mostly assume, that the respondents have a medium attitude toward the declared object of the scale (like "conservatism"). But these scores can also mean (2) that they do not have such an attitude at all, but rate the items in regard to other criteria. Or these scores could mean (3) that the respondents have a differentiated attitude which involves more than the one attitude.

The deliberate blindness of objective psychology for structural aspects of human behavior explains Hartshorne & May's (1928) failure to produce evidence for the existence of character. Only after completion of their study did Hartshorne and his colleagues admit that excluding internal traits from their observations was a mistake. "The essence of the act is its pretense. Hence [character] can be described and understood only in terms of the human elements in the situation. It is not the act that constitutes the deception, nor the particular intention of the actor, but the relation of this act to his intentions and to the intentions of his associates" (p. 377). They also admitted the blindness of their measurement model for the competence aspect of character: "A trait such as honesty or dishonesty is an achievement like ability in arithmetic, depending of course on native capacities of various kinds" (Hartshorne, May, 1928, p. 379).

In spite of their abstinence in regard to psychological theories, objective "psychologists" claim that their statistical models can be used to evaluate psychological theories, therapeutic methods, educational policies and competencies of people. They underpin their claim with a naming trick: (a) They give their statistical constructs psychological names like intelligence, character, or conservatism, and (b) they equate pattern of correlations across groups of people with an individual mind's "structure". But like families names, these names do not actually establish a real relationship between statistics and psychology. Or would Mrs. Miller allow an unrelated Mr. Miller to share her bedroom, just because he bears the same family name?

SOLVING THE DILEMMA OF PSYCHOLOGY: TESTING OUR MEASUREMENT THEORIES

How can we overcome the dilemma of psychology? Measurement is always based on theoretical assumptions. It provides reliable information if these assumptions are true. However the assumptions of both approaches to psychological are questionable. They have hardly ever been tested or are testable at all. Subjective psychologists base their methods of observation on psychological theories and objective psychologists base them on statistical theories. The dilemma of both approaches is that in both cases the theoretical assumptions are non-testable postulates instead of testable hypotheses. Objective psychologists fail be really objective. Instead of rigorously examining their postulates, like the postulate of pure random error and of onedimensionality, they select tasks and questions to fit these postulates. Subjective psychologists fail to examine the psychological assumptions on which they base their observation and interviewing methods. As Jean Piaget (1965) admits: "The point, then, that we have to settle is whether the things that children say to us constitute, as compared to the real conduct,

a conscious realization […], reflection […] or psittacism […]. We do not claim to have solved the problem completely. Only direct observation can settle it" (p. 115).

To make sure that our measurement data reflect reality and are not only an artifact of our theoretical assumptions, we must spell out these assumptions and submit them to rigorous experimental testing. How this can be done has been shown by the Dutch psychologist Franciscus Donders (1969/1868) a hundred and fifty years ago. Donders was probably the first who showed that we can examine the truth of psychological hypotheses about *internal* processes of the mind in the same way as we examine hypotheses about the impact of an external factor on human behavior. He hypothesized that humans do not just react to stimuli like an automat, but that they also *think* when it is needed. To test this hypothesis he designed a simple experiment for which he constructed an ingenious time recorder for measuring very short reaction times. When he gave his participants clearly distinct stimuli, they reacted as quickly as an automat. But when he gave them similar stimuli they caused a "dilemma" for them, so that their reactions took much longer. Obviously under the second condition they had to *think* before reacting.

Our behavior is often determined at once by more than one factor. Therefore, tests based on the assumption of onedimensionality are not suited for psychological measurement. These tests, as we have seen above, regard everything as "error" which does not fit their onedimensional test theory. If they speak of multi-dimensionality they refer to pooled data across many individuals instead of the structure of an individual's mind. This problem has also been solved long time ago, namely by Egon Brunswik (1955). He has shown how we can disentangle multiple factors of behavior with the "diacritical method." This method means that the experimenter confronts the participants with a purposefully designed *pattern* of tasks, questions, or situations. This pattern is designed as an individual *multi-variate experiment*. The design-factors of this dispositional experiment are chosen to directly correspond with the dispositions that are believed to be involved in the participant's behavior. If the test is designed in such a way that the factors are independent of each other, we can directly read from the pattern of the participant's reactions how much each dispositional factor determines his or her behavior (Lind, 1982).

On the basis of the ideas of Donders and Brunswik I have developed a new methodology for measuring psychological objects, *Experimental Questionnaire*, EQ (Lind, 1982). EQs make is based on psychological theory, instead of statistical theory, and makes it possible to measure traits of the human mind truly objectively. In other words, it allows us to test our measurement hypotheses and thus safeguard it against subjective or collective biases.

I have used this methodology to construct the first objective *Moral Competence Test*, MCT (formerly called *Moral Judgment Test)* (Lind, 1978; 2016). Informed by the moral theories of Piaget (1965), Kohlberg (1963; 1984) and Habermas (1990) we define moral competence *as the ability to solve problems and conflicts on the basis of moral principles through thinking and discussion* (Lind, 2016). A core element of this competence is the ability to judge arguments pro and contra a decision in regard to their moral quality instead of in regard only to their opinion agreement. This seems to be a very difficult task for many people, as Keasey (1974) showed in a series of experiments. So we decided to use this task for measuring moral competence. Hence we designed the MCT to let us measure people's ability to *rate arguments in a controversy in regard to their moral quality instead of to their opinion-agreement or to other non-moral aspects*. We

regard this ability as the core of moral competence. If people have not developed such a moral sense, they cannot solve problems and conflicts through moral thinking and moral discussion but must use violence, deceit or submission to others.

Because the MCT has not been immunized against falsification through empirical item selection, as is the case with classical tests, it can be used to rigorously test the four measurement hypotheses which informed its construction without saving circularity. These are the four hypotheses on which the MCT is based:

- *Competence nature of morality*: In contrast to many other psychologists Piaget (1965) and Kohlberg (1958; 1964) have hypothesized that moral behavior is not only affective in nature but also cognitive, that is, it is not only determined by people's moral orientations (values, attitudes, principle, and so on) but also by their moral competence. Experiments using the MCT clearly support this competence hypothesis. While participants can be instructed to fake their moral orientations upward (Emler et al., 1983), the same kind of instruction fails to make participants fake their competence-scores upward (Lind, 2002). Experiments also showed that the ability to estimate other people's moral competence is positively correlated with their own moral competence (Wasel cited in Lind, 2002).
- *Moral competence is a unique skill*. The ability to solve moral dilemmas is not just a linguistic skill but is a unique competence. This has been shown by the research team of Kristin Prehn (2013) of the Charité Hospital in Berlin. Moral competence as measured with the MCT correlates highly with neural activities in the right dorso-lateral prefrontal cortex (DLPC) when subjects' brain activities are studied in a brain scanner while they are confronted with moral problems: the lower their C-score is the longer their right DLPC is busy. This phenomenon does not show when the subject is confronted with linguistic problems.
- *Hierarchical preference order*: Kohlberg (1958, 1984) has hypothesized that the six types of moral orientations – which he has identified on the basis of theoretical analysis and interview studies – form a universal hierarchy of moral adequacy. This hypothesis lets us predict that people will prefer them according to this hierarchy. Note that this is a very risky hypothesis. Since a person's preference in regard to six types of orientation can be ordered in 720 (= 6!) different ways, the risk of a coincidental confirmation of the hypothesis is very small (p = 1/720 = 0.0014) compared to the very small risk of accidently confirming a statistical hypothesis ($p < 0.05$). The risk of accidental confirmation becomes extremely tiny if, for example, we test this hypothesis with ten people ($p = 0.0014^{10}$). In contrast, the likelihood of confirming a statistical hypothesis accidentally increases when the number of tested people increases. In spite of the extremely low likelihood of confirming this hypothesis accidently, empirical studies in various cultures almost unanimously confirm it (Rest, 1969; Lind, 1986, 2002).
- *Simplex structure of moral orientations*: Kohlberg (1958) has hypothesized that the correlations between the six types of moral orientation show a *simplex structure*, which means that neighboring orientations correlate more highly with each other than with more distant orientations. All MCT studies support this hypothesis with only very few exceptions (Lind, 1978, 2002).
- *Affective-cognitive parallelism*: Piaget (1976) has hypothesized that affective and cognitive aspects of behavior are *parallel*. This lets us predict, for example, that the higher people's moral competence is, the more they will prefer higher types of moral orientations, and the more they will reject low, inadequate types. With the

MCT it became possible to test this hypothesis because only this test allows us to measure both aspects distinctly and simultaneously. So far all studies very clearly support Piaget's notion (Lind, 2002, 2013).

Experimentally designed tests accomplish what subjective psychologists always wanted but objective psychologists could not deliver: They make it possible to measure internal properties of humans' mind objectively that is transparently and through direct observation of their behavior. Moreover, experimentally designed tests allow us to examine the truth of the assumptions underlying them. Thus psychology is no longer hampered by a dilemma. Experimentally designed tests like the *Moral Competence Test* (MCT) enable us to measure psychological objects like moral competence and moral orientations validly and objectively. We can rely on them when we use them for evaluating methods of education, therapy and policy-making.

When psychologists chose to base their measurement on psychological theory instead of statistical theory they will find it easy to apply the concept of experimentally designed tests to virtually all fields of psychological measurement. A simple application would be to apply it to conventional attitude tests: If we would measure the variances of the individual responses due to the measured factor, we could easily answer Scott's (1968) question to which degree an individual's response pattern manifests the assumed attitude, and manifests something else. If we had grounded hypotheses about other factors which could possibly influence responses to attitude scales, we could add them to our test design as we did with the MCT. Similarly we could think of applications of this concept to tests of intelligence and of academic achievements.

REFERENCES

Allerup, P. (2007). Identification of group differences using PISA scales – considering effects of inhomogeneous items. In: S.T. Hopmann, G. Brinek, M. Retzl (Eds.), *PISA zufolge PISA* (pp. 175–202). Berlin: LIT-Verlag.

Amrein, A., Berliner, D. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10* (18). Download: http://epaa.asu.edu/ojs/article/view/297.

Ausch, R. (2016). *Methodological problems with the academic sources of popular psychology*. Lanham: Lexington Books.

Bailey, K. (1994). *Method of social research*. New York: The Free Press.

Block, J. (1967). Advancing the psychology of personality. In: D. Magnusson, N. Endler (Eds.), *Personality at the crossroads* (pp. 37–63). Hillsdale, NJ: Lawrence Erlbaum.

Bühler, K. (1927). *Die Krise der Psychologie* [The crisis of psychology]. Jena: Gustav Fischer.

Cohen, R.J., Swerdlik, M.E. (2018). *Psychological testing and assessment. An introduction to tests and measurement.* 9th edition. New York: McGraw Hill.

Colby, A., Kohlberg, L., Colby, A., Speicher, B., Hewer, A., Candee, O., Gibbs, J., Power, C. (1987). *The measurement of moral judgment: Standard issue scoring manual*, Vol. I & II. New York: Cambridge University Press.

Donders, F.C. (1969/1868). On the speed of mental processes. *Acta Psychologica, 30, Attention and Performance II* (pp. 412–431). Download: http://www2.psychology.uiowa.edu/faculty/mordkoff/InfoProc/pdfs/Donders 01868.pdf.

Emler, N., Renwick, S., Malone, B. (1983). The relationship between moral reasoning and political orientation. *Journal of Personality and Social Psychology, 45*, 1073–1080.

Graumann, C.F. (1960). Eigenschaften als Problem der Persönlichkeitsforschung. In: P. Lersch, H. Thomae (Eds.), *Hand-*

buch der Psychologie, Bd. 4, Persönlichkeitsforschung und Persönlichkeitstheorie (pp. 87–154). Göttingen: Hogrefe.

Gregory, R.J. (2018), *Psychological testing: History, principles, and applications*. Pearson Publisher.

Habermas, J. (1990). *Moral consciousness and communicative action*. Cambridge: MIT Press.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, *316*, 998–1002.

Hartshorne, H., May, M.A. (1928). *Studies in the nature of character. Vol. I: Studies in deceit, Book one and two*. New York: Macmillan.

Jablonka, E. (2007). Mathematical literacy: Die Verflüchtigung eines ambitionierten Testkonstrukts in bedeutungslosen PISA-Punkten. In: T. Jahnke, W. Meyerhöfer (Eds.), *Pisa & Co. Kritik eines Programms* (pp. 247–280). Hildesheim: Franzbecker.

Keasey, Ch.B. (1974). The influence of opinion-agreement and qualitative supportive reasoning in the evaluation of moral judgments. *Journal of Personality and Social Psychology, 30*, 477–482.

Keller, M. (1990). Zur Entwicklung moralischer Reflexion: Eine Kritik und Rekonzeptualisierung der Stufen des präkonventionellen moralischen Urteils in der Theorie von Kohlberg. In: M. Knopf & W. Schneider (Eds.), *Entwicklung. Allgemeine Verläufe – Individuelle Unterschiede* (pp. 19–44). Göttingen: Verlag für Psychologie.

Kempf, W. (1981). *Testtheorie.* University of Konstanz, Dept. of Economy and Statistics, mimeographed paper.

Kohlberg, L. (1958). *The development of modes of moral thinking and choice in the years 10 to 16*. University of Chicago. Unpublished doctoral dissertation.

Kohlberg, L. (1964). Development of moral character and moral ideology. In: M.L. Hoffman, L.W. Hoffman (Eds.), *Review of child development research*, Vol. I (pp. 381–431). New York: Russel Sage Foundation.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. In: T. Lickona (Ed.), *Moral development and behavior: Theory, research and social issues* (pp. 31–53). New York: Holt, Rinehart & Winston.

Kohlberg, L. (1979). *The meaning and measurement of moral development. The Heinz Werner Lecture Series,* vol. 13. Worcester, MA: Clark University Press.

Kohlberg, L. (1984). *The psychology of moral development. Vol. 2: Essays on moral development*. San Francisco: Harper & Row.

Koretz, D. (2017). The *testing charade. Pretending to make schools better*. Chicago: University of Chicago Press.

Kurtines, W.M., Greif, E.B. (1974). The development of moral thought: Review and evaluation of Kohlberg's approach. *Psychological Bulletin*, *81*, 453–470.

Lind, G. (1978). Wie misst man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts. In: G. Portele (Ed.), *Sozialisation und Moral* (pp. 171–201). Weinheim: Beltz.

Lind, G. (1982). Experimental Questionnaires: A new approach to personality research. In: A. Kossakowski, K. Obuchowski (Eds.), *Progress in psychology of personality* (pp. 132–144). Amsterdam, NL: North-Holland.

Lind, G. (1986). Cultural differences in moral judgment competence? A study of West and East European university students. *Behavior Science Research*, *20*, 208–225.

Lind, G. (1989). Essay Review: "The measurement of moral judgment" by Anne Colby, Lawrence Kohlberg et al. *Human Development*, *32*, 388–397.

Lind, G. (2002). *Ist Moral lehrbar? Ergebnisse der modernen moralpsychologischen Forschung.* [Can morality be taught? Research findings from modern moral psychology]. 2nd edition. Berlin: Logos-Verlag.

Lind, G. (2005). *The cross-cultural validity of the Moral Judgment Test:* Findings from 28 cross-cultural studies*. Presentation at the conference of the American Psychological Association, 18–21, 8. 2005, Washington, D.C.

Lind, G. (2013). Thirty years of the Moral Judgment Test:* Support for the dual-

aspect theory of moral development. In: Claudio Simon Hutz, Luciana Karine de Souza (Eds.), *Estudos e pesquisas em psicologia do desenvolvimento e da personalidade: uma homenagem a Angela Biaggio* (pp. 143–170). Sao Paulo, Brasilien: Casa do Psicólogo.

Lind, G. (2016). *How to teach morality. Promoting deliberation and discussion. Reducing violence and deceit*. Berlin: Logos.

Lind, G. (2017a). From Donder's dilemma to objective internal assessment: How experimenttal developmental psychology can contribute to moral education. *Psychologia Rozwojowa, 22*(3), 15–24.

Lind, G. (2017b). *Wie kann man moralische Orientierung im Verhalten erkennen?* Beitrag zur Moralforschertagung 26.–28. 1. 2017 in Leipzig. Universität Leipzig. Download: http://www.uni-konstanz.de/ag-moral/moral/demokratie/2017_01_26%20Moralforscher%20Leipzig/Lind-2017_Brunswiks_Diakritische_Methode_d.pdf.

Lind, G., Nowak, E. (2015). Kohlberg's unnoticed dilemma – The external assessment of internal moral competence? In: B. Zizek, D. Garz, E. Nowak (Eds.), *Kohlberg revisited* (pp. 139–154). Rotterdam: Sense Publisher.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3,* 635–694.

Meehl, P.E. (1958). When to use your head instead of the formula? In: H. Feigl, M. Scriven, G. Maxwell (Eds.), *Minnesota studies in the philosophy of science* (pp. 498–506). Minneapolis, MI University of Minnesota Press.

Piaget, J. (1965). The *moral judgment of the child*. New York: The Free Press. (Originally published 1932).

Piaget, J. (1976). The affective unconscious and the cognitive unconscious. In: B. Inhelder, H.H. Chipman (Eds.), *Piaget and his school* (pp. 63–71). New York: Springer.

Popper, K.R. (1968). *Conjectures and Refutations*. New York: Harper & Row.

Prehn, K. (2013). Moral judgment competence: A re-evaluation of the Dual-Aspect Theory based on recent neuroscientific research. In: E. Nowak, D. Schrader, B. Zizek (Eds.), *Educating competencies for democracy* (pp. 9–22). Frankfurt: Peter Lang.

Ravitch, D. (2013). *The death and life of the great American school system: How testing and choice are undermining education*. New York: Basic Book.

Rest, J.R. (1969). Level of moral development as a determinant of preference and comprehension of moral judgments made by others. *Journal of Personality, 37*(1), 220–228.

Rest, J.R. (1979). *Development in judging moral issues.* Minneapolis, MI: University of Minnesota Press.

Rosnow, R.L., Rosenthal, R. (1997). *People studying people. Artifacts and effects on behavioral research*. New York: Freeman & Co.

Ryle, G (1949). *The concept of mind*, Chicago: University of Chicago Press.

Scott, W.A. (1968). Attitude measurement. In: G. Lindsey, E. Aronson (Eds.), *Handbook of social psychology* (pp. 204–273). Vol. II. Reading, MA: Addison-Wesley.

Schoenfeld, A.H. (1999). Looking toward the 21th century: Challenges of educational theory and practice. *Educational Researcher, 28*, 4–14.

Sjoberg, S. (2007). PISA and 'Real Life Challenge': Mission impossible? In: S.T. Hopfmann et al. (Eds.), PISA zufolge PISA (pp. 203–224). Berlin: LIT-Verlag.

Sjoberg, S. (2017). PISA testing – A global educational race? *Europhysics News.* https://www.europhysicsnews.org/articles/epn/pdf/2017/04/epn2017484p17.pdf

Travers, R.M. (1951). Rational hypotheses in the construction of tests. *Educational and Psychological Measurement, 11*, 128–137.

Watson, J.B. (1970). *Behaviorism*. New York: Norton (orginally published 1924).

Wilson, M. (2005). *Constructing measures. An item response modeling approach*. Mawah, NJ: Erlbaum Associates Publishers.

Wittgenstein, L. (1953). *Philosophical investigations*. Translated by G.E.M. Ans-

combe. New York: The Macmillan Company.
Wuttke, J. (2005). Fehler, Verzerrungen, Unsicherheiten in der PISA-Auswertung. In: T. Jahnke, W. Meyerhöfer (Eds.), *Pisa & Co. Kritik eines Programms* (pp. 101–154). Hildesheim: Franzbecker.
Wuttke, J. (2007). Uncertainties and bias in PISA. In: S.T. Hopmann, G. Brinek, M. Retzl (Eds.), *PISA zufolge PISA* (pp. 241–264). Berlin: LIT-Verlag.

Notes
* The MJT has been renamed as *Moral Competence Test*, MCT.

DYLEMAT PSYCHOLOGII: OBIEKTYWIZM VS. SUBIEKTYWIZM

Streszczenie: Każda osoba zainteresowana psychologią (która oznacza „naukę o umyśle") mierzy się ze stałym dylematem. Jesteśmy zmuszeni do wyboru pomiędzy dwiema „psychologiami": „subiektywni" czy też "jakościowi" psychologowie utrzymują, że skupienie się na badaniu wewnętrznej struktury umysłu człowieka dostarczy kluczowego zrozumienia i wiedzy potrzebnych w terapii i edukacji. Zarazem twierdzą oni, że umysł człowieka może być badany jedynie za pomocą metod *subiektywnych,* takich jak wywiad kliniczny, a nie za pomocą wystandaryzowanych testów. Dla kontrastu, „obiektywni" bądź „ilościowi" psychologowie twierdzą, że skoro psychologia dąży do uznania jej za naukę, powinna ona używać jedynie *obiektywnych* metod pomiaru. To z kolei wyklucza według nich badanie wewnętrznych czynników psychologicznych w umyśle ludzkim. Podejście subiektywne bazuje na psychologicznych założeniach dotyczących natury docelowego mierzonego konstruktu, podejście obiektywne z kolei opiera się całkowicie na teoriach statystycznych. Czy rzeczywiście jesteśmy zmuszeni do porzucenia konstruktów psychologicznych, takich jak zdolności intelektualne lub moralne, jeśli chcemy, aby nasze narzędzia pomiaru były obiektywne? W artykule ukazuję, że oba podejścia bazują na wątpliwych teoriach dotyczących relacji pomiędzy obserwowanym zachowaniem z jednej strony a konstruktami psychologicznymi z drugiej strony. Pokazuję także, że możliwe jest obiektywne i wiarygodne zbadanie cech psychologicznych, jeśli zastosujemy podejście eksperymentalne. *Kwestionariusze eksperymentalne* mogą być używane w każdej dziedzinie psychologii, w której rozwinięte zostały testowalne teorie o naturze danego konstruktu. Użyliśmy tego nowego podejścia z sukcesami, np. podczas konstrukcji i walidacji Testu Moralnej Kompetencji.

Słowa kluczowe: pomiar psychologiczny, wystandaryzowane testy, teoria, obiektywizm, wiarygodność.