

GEORG LIND

Universität Konstanz
e-mail: Georg.lind@uni-konstanz.de

From Donders' Dilemma to Objective Internal Assessment: How Experimental Developmental Psychology Can Contribute to Moral Education¹

Abstract. For a century, developmental psychology, and psychology as a whole, has been caught in an objectivity-validity dilemma which divides the scientific community: when measuring psychological traits, researchers either strive for *objectivity* at the expense of validity in regard to the object of the measurement, or they strive for *validity* by using subjective information from participants and subjective methods of rating this information, thus lacking objectivity.

While the main problem of physical measurement is precision, the main problem of psychology is validity: does an instrument really measure what it should? On the one hand, psychologists like Sigmund Freud, Jean Piaget and Lawrence Kohlberg believed that valid inferences on psychological functioning can only be made through subjective methods like clinical interviews – at the expense of objectivity. On the other hand, scientists working in the tradition of Louis Leon Thurston and Karl Pearson wanted to turn psychology into an objective science by assessing psychological traits through behavioral data even if that meant deleting important psychological traits like moral orientations and moral competence from the scientific research agenda.

In this article I will show that this methodological dilemma can be (and has been) resolved on the basis of the ideas of Franciscus Cornelis Donders (1868), who, a long time ago, showed how we can study internal mental processes in an objective way, that is, without relying on the subjective experience of the participant and without subjective scoring. Interestingly, Donders used behavioral dilemmas in order to measure thinking.

It took us a long time to understand the significance of Donders' invention and to apply it to psychological measurement. Our Moral Competence Test (MCT) is probably the first method of psychological measurement based on it. The MCT has helped us to test many hypotheses on moral orientations and moral competence objectively, which hitherto could be studied only with subjective methods. It has also made possible to design and evaluate new, effective methods of moral education.

Key words: moral-democratic competence, moral development, moral psychology, education, democracy.

THE DWINDLING IMPACT OF DEVELOPMENTAL PSYCHOLOGY ON MORAL EDUCATION

Great developmental psychologists like Jean Piaget and Lawrence Kohlberg, to name a few, have had a strong impact on moral education. But their impact is dwindling mainly because

most, if not all, supportive data are subjective. This means that they are based on subjective accounts of the participants about their reasoning, and on subjective scoring of these accounts (Piaget 1965; Lind 1989). In spite of much supportive evidence for the fruitfulness of their approaches, mainstream psychologists remain skeptical. "For a large part of the early

20th century it was generally held that moral phenomena cannot be studied scientifically," wrote Bill Kurtines and Jacob Gewirtz in their handbook article on moral development (Kurtines, Gewirtz 1995, p. 3; see also Emler 1996). We can hardly develop better methods and programs of fostering moral competence if we cannot empirically demonstrate their efficacy with adequate measurement. To prove that certain methods (like direct teaching of values, values clarification, dilemma discussion, just community) can effectively foster moral competence requires that we can assess this trait objectively and validly.

THE METODOLOGICAL DILEMMA

Modern theories of moral development like that of Piaget (1965) and Kohlberg (1984) identified *structural* aspects of morality – a person's moral consistency – to be the best predictor of human behavior (Blasi 1980; Kohlberg, Candee 1984; Sprinthall, Sprinthall, Oja 1997; Lind 2016). However, these findings have been at the center of many controversies. Today we understand that these controversies are caused by a deep methodological dilemma. In regard to measuring structural aspects of human behavior, the two main methodologies of measuring psychological traits available – the classical-test-theory-based psychometrics and clinical interview – both have severe shortcomings.

A structure is a particular set of relationships between dispositions which are manifested in a person's pattern of behaviors. In developmental psychology such relationships are the essential focus of study. Human development does not merely mean a change of attitudes and increase of competencies but, foremost, a re-structuring of these dispositions and their relationships. However, in classic psychometrics behavioral structures and relationships are largely ignored or considered merely as a source of measurement error. By relegating structural aspects of behavior like moral consistency to the error term, these tests prevent the assessment of structure.

The classical test theory (and also its modern variants like the item-response theory) is

based on two dubious assumptions: first, that *every* response to a test-item (Y) is caused by one, and only one, trait or disposition (X , e.g., a certain moral orientation like the preference for principled moral reasoning), and second, that it is camouflaged only by a random measurement error (e) (see Gulliksen 1950). The famous equation of the classical test theory ($Y = X + e$) leaves no room for the simultaneous impact of several dispositions on an individual's behavior, nor does it take into account the fact that the relationships (structure) between these dispositions can widely vary between individuals. Modern variants of the test theory are more complex but are essentially based on the same assumption.

If psychometrists study structure they do not mean the relationships between dispositions operating *within* an individual but their correlations *between* the members of an accidental sample of individuals. Obviously, such inter-individual correlations tell us nothing about an individual's personality. These shortcomings have been known for long but were of little consequence for psychological research because no better objective methods were available (cf., e.g., Loevinger 1957; Pittel, Mendelsohn 1966; Mischel 1973; Broughton 1978; Lind 1978; 1982; Kohlberg 1984; Lohman, Ippel 1993; Messick 1995; Mischel, Shoda 1995). In sum, the measurement model of the classical and modern test theory may be adequate for engineering (Thurston 1928) and eugenics (Pearson 1913), where these methods come from, but certainly not for developmental psychology.

In contrast, the qualitative methods of measurement such as clinical interviews (Piaget 1965; Kohlberg 1984) seek to remedy this shortcoming of main-stream psychology by trying to assess the structure. Thus, at first sight, they seem better suited to assess structural dispositions like moral competence (Broughton 1978). However, they are not based on behavior but on the participants' accounts of their internal motives and competencies and on subjective, intuitive interviewing and scoring methods (Piaget 1965; Colby et al. 1987). These methods are prone to biases through unconscious distortions on both sides, on the side of the interviewee and that of the interviewer, and even through

explicit scoring instructions (“upper stage inclusion rule,” “guess scores,” “aligning scores with age development”). The subjective scoring of Kohlberg’s Moral Judgment Interview (MJ) was rightly criticized for inflating consistency and immunizing the theory against refutation (see, e.g., Kurtines, Greif 1978; Arbuthot 1979; Nicolayev & Phillips 1979; Rest 1979; Cortese 1984; Lind 1989; Anderson 1991; Flanagan 1991; Keller 1992; Wark, Krebs 1996; Haidt 2001; Beck et al. 2002).

DONDERS’ DILEMMA HELPS TO SOLVE THE DILEMMA

Donders (1986) showed that objective measurement of internal structure is possible in principle. With his experimental study of mental processes involved in reacting to behavioral dilemmas, Donders provided objective evidence that the human mind is not – as many of his contemporaries believed – a hard-wired apparatus but that mental structures are involved when subjects have to cope with behavioral dilemmas. As he had predicted, the subjects’ reaction times were much longer when two stimuli, which the subjects had to memorize, conflicted with each other and the subjects had to think about the right reaction, than when they did not conflict. In other words, when people are confronted with a dilemma they do not rely on inborn “reflex-arcs” but engage in (tacit) thinking (even when they are not aware of it).

But how can we measure a certain disposition if peoples’ responses to test-items are determined by more than one disposition, as is usually the case? Here another ground-breaking experimental-psychological method can be used which was invented by the Hungarian-born American psychologist Egon Brunswik (1955), namely, the “diacritical method.” This method allows us to disentangle experimentally the impact of various dispositions that may determine an individual’s behavior. This method is also objective and does not involve any subjective scoring of data, nor does it rely on introspection or self-report. He showed how we can apply concepts like hypothesis-testing and

multi-variate experimental design to the study of psychological dispositions of individuals. For some time his ideas triggered a considerable amount of research under the heading of “N = 1 experiments,” “cognitive algebra” (Anderson 1991) and “response-scaling” (Torgerson 1958). Also Kelly’s (1955) “personal construct theory” evolved out of these ideas.

However, today the ingenious ideas of Donders and Brunswik have disappeared out of most psychology text-books. Paul Meehl’s famous admonition “when to use your head instead of a formula” (Meehl 1958) has not made a lasting impression on psychological research and practice. In most areas of psychology, including moral development and education research, the classical-test-theory-based measurement has won over all rivals, not only because it has been lobbied by a strong test-industry but because no serious alternatives have been developed from the experimental-psychological ideas of Donders, Brunswik, Torgerson, Anderson, Kelly and others, which were both objective and valid.

THE EMERGENCE OF EXPERIMENTAL-PSYCHOLOGICAL MEASUREMENT

Beginning in the 1970s, I have been thinking about possibilities to use the ideas of experimental psychology in order to design an objective method for assessing internal, structural dispositions of moral behavior. I suggested a new objective methodology for psychological measurement which I called Experimental Questionnaire (EQ, Lind 1982). “Experimental” means (a) that a bold theory and its predictions are submitted to rigorous tests of their empirical truth, that is, that our assumptions about the nature of the disposition under consideration are examined, and (b) that the empirical data are produced through confronting participants with certain stimuli (test items) and observing their responses (usually choosing an answer from a list or marking a number on a response scale).

Tests based on classical and modern test theories agree only partly with this ideal of

experimental measurement. Since classical test construction is rarely based on an elaborate psychological theory but on simple statistical models, no real theory-testing can be done with them. Items are rarely selected on the basis of an elaborated psychological theory but on the basis of their “face validity,” and their goodness of fit into the statistical test model of one-dimensionality. Often the measurement object is not defined in advance, so that their theoretical validity cannot be determined at all. Because of this and because no structural information is used, test scores are mostly ambiguous. Scott (1968) showed that a person’s score in the middle range of, for example, a conservatism scale can have three completely different meanings: (1) that this person has a mean preference for conservatism, (2) has no attitude toward conservatism at all, or (3) has a differentiated opinion on the items of this attitude scale. In each case, the test-taker can get the same score. Because of this ambiguity of classical tests, their scores can create the illusion of a mere attitude change when in fact cognitive-structural functions are transformed (Lind 2010).

In contrast, EQs require as a basis an elaborated psychological theory and clear operational definition of the measurement object. Accordingly, item-construction and test-design are based on a psychological theory and a clear operational definition of the construct to be measured (e.g., moral competence, see below). Because EQs are strictly theory-based, their underlying assumptions can be empirically tested, and refuted, unlike classical tests, which have criteria for reliability or precision of measurement but no built-in criteria for validity. In contrast, EQs provide rigorous criteria for empirical validity which do not rely on the unknown validity of other measurements.

THE MORAL COMPETENCE TEST (MCT)

The first test which was constructed and validated according to the concept of Experimental Questionnaires is the Moral Competence Test (MCT), which can be used for assessing the

two main determinants of moral behavior, namely moral orientations and moral competence. Invoking Kohlberg’s famous definition of moral judgment competence 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), as well as Habermas’ (1990) definition of moral decision-making as decisions which are based on free discourse rather than on violence and deceit, we defined moral competence as *the ability to solve problems and conflicts on the basis of moral principles through thinking and discussion rather than through violence, deceit and power* (cf. Lind 2008; 2016). For the purpose of measurement, we specified this definition as *the ability to rate arguments of others about someone’s decision in regard to their moral quality rather than other criteria like opinion-agreement*.

The Moral Competence Test has two features which sets it apart both from subjective methods of assessment in the moral domain like those of Jean Piaget (1965) and Lawrence Kohlberg (1958; 1964; 1984), but also from test-theory-based psychometrics used by Hugh Hartshorne and Mark May (1928), James Rest (1979), Jonathan Haidt (2001) and many others:

- The MCT is *objective* and transparent, that is, no subjective accounts of internal processes and no subjective interpretations of data are involved.
- At the same time, the MCT allows us to study more than one internal disposition underlying moral behavior and their interrelationship simultaneously. The number of dispositions that can be studied with experimental-psychological measurement is limited only by the patience of the participants and the ingenuity of the test constructor (Lind 1978; 1982). Thereby, the test-experiment can very be short since all design-factors are “crossed.” Even though the Moral Competence Test allows us to assess the impact of three dispositions (A, B, C) and their four interactions (A x B, A x C, B x C, A x B x C), as well as six types of moral orientations, it contains only 24 items (Lind 1978; 2016).

The MCT also fulfills the basic criteria for valid measurement of moral competence (see also Lind 2013; 2016):

- The MCT allows us to measure objectively the structural disposition of moral competence.
- The measurement of moral competence is fully aligned with its definition as the ability to behave according to one's own *internal* moral principles: the MCT uses participants' *own* moral orientations as standards for scoring their moral competence instead of external standards as virtually all classical tests do (Lind, Nowak 2015).
- Finally, the MCT makes it possible to measure the affective (orientation) and cognitive (structural) aspect of moral behavior simultaneously. This is especially noteworthy because these two aspects can be (and need to be) clearly distinguished but cannot be separated as components. Other tests either measure only the moral affect (or similar concepts like orientations, attitudes, preferences etc., (see, e.g., Hartshorne, May 1928, Rest 1979, Haidt 2001)) or a mixture of moral affect and moral cognition like Kohlberg's "cognitive-developmental stage" score (1984).

For reasons of space, I can discuss here only briefly the similarities and differences between the MCT and Kohlberg's approach to the measurement of structure (for an extensive discussion see Lind 2016, Lind, Nowak 2015).

Like Kohlberg's (1964) concept of *moral judgment competence*, our concept of *moral competence* refers to the internal structure that determines an individual's moral judgment behavior. Unlike Kohlberg, we do not equate structure with the stage of the subject's conscious moral reasoning. Instead of using the participants' overt moral reasoning to indirectly infer their structure (stage) of moral behavior, the MCT lets us assess this structure directly through their response pattern.

Like Kohlberg's Moral Judgment Interview (MJI), the MCT is designed to measure the "intelligible" ordering of the manifest items" (Kohlberg 1984, p. 408). We agree with Kohlberg, who noted that "the responses of subjects

to the dilemmas and their subsequent responses to clinical probing are taken to reflect, exhibit, or manifest the structure" (Kohlberg 1984, p. 407) – except that with the MCT, the subjects respond to written arguments instead of producing their own arguments.

We agree with Kohlberg's postulate that "structure [...] will not emerge through pure factor-analytic responses classified by content" (pp. 401–402) because such "factors" are not based on individual data but on *aggregated* data across individuals, and they are constructed using certain statistical methods, which contain hidden subjective assumptions (Lind 2008; 2016).

In contrast to Kohlberg's MJI, which attempts to assess structure through clinical interview, the MCT does not require the use of subjective "guess scores," the "upper-stage inclusion rule," or similar scoring rules (Colby et al. 1987). Rather, it relies solely on participants' evaluative reactions to arguments on scales from "–4" to "+4": "How much do you agree or reject the following arguments?"

Like Kohlberg in his early phase of research (Kohlberg 1958) but in contrast to the classical test theory to which Kohlberg's scoring team later turned (Colby et al. 1987), the MCT uses as a basis for scoring the *whole pattern* of an individual's reactions. Specifically, we regard the "within-subject variance (some of which represents systematic performance variation)," which the test-theory "relegates [...] to the error term" (Lohman, Ippel 1993, p. 46), as the long-sought manifestation of moral competence (Lind 1978; 2016).

Structure cannot be inferred from individual responses but only from the whole *pattern* of responses to an experimentally designed *pattern* of test items. In contrast, Kohlberg and his colleagues require that "each item in the manual [is] to clearly reflect the structure of the stage to which it is keyed" (Kohlberg 1984, p. 403; Colby et al. 1987). This requirement does not agree with the experimental-psychological assessment theory or with Kohlberg's (1958) own original ideas regarding structural measurement. Rather, it is adapted from the classical test-theory.

The crux of the experimentally designed MCT is that it can show which of *three* hypothesized dispositions (moral orientation, opinion-agreement, and dilemma-contexts) and their combinations determine an individual’s pattern of judgment behavior. Inspired by Brunswik’s diacritical method, I have built these three dispositions into the design of the MCT as three fully-crossed, orthogonal factors (6 x 2 x 2) resulting in 24 test-items. Because of this design, we can see the influence of these three dispositions on the participants’ behavior from the participants’ response *pattern*.

Figure 1 depicts, as an example, the pattern of evaluative responses to moral arguments by two fictitious persons, *A* and *B*. For the sake of simplicity, the figure shows their responses only to one dilemma context. They respond to 6 arguments in favor and 6 arguments opposing the decision of the protagonist of a story. Each argument represents a different moral quality (namely, the six types of moral orientation defined by Kohlberg 1984). The algorithm for transforming this visual pattern into a numeric index (C-score) is explained in Lind (1978; 2008; 2016).

Figure 1 shows how important structural scoring is. For example, the very positive ra-

ting (“+4”) of arguments which express a moral orientation of Type 6 can only be adequately understood when considering the whole pattern of a person’s responses. For example, both participants rate Type 6 pro-arguments very high. Only structural scoring reveals that they differ strongly in regard to their moral competence. Person *A* obviously rated this argument high because it supported her opinion, whereas person *B* rated it high because of its moral quality. She evaluated all arguments in regard to their *moral quality*, not in regard to their opinion-agreement. If Kohlberg had been aware of the possibilities of experimental-psychological assessment of structure when he set up his research, he would have probably used it. In his lectures *Meaning and Measurement...* he stated that “[structure] is a construct rather than an inference, and is warranted only on the grounds of ‘intelligible’ ordering of the manifest items. One might say that the hypothetical structure is the principle of organization of the responses” (Kohlberg 1984, p. 408). This describes well the ideas on which the MCT has been built. Consequently, he endorsed the MCT, with which “a scoring algorithm can be arrived at for assigning pure stage structure score for an individual” (Kohlberg 2010, p. xvii).

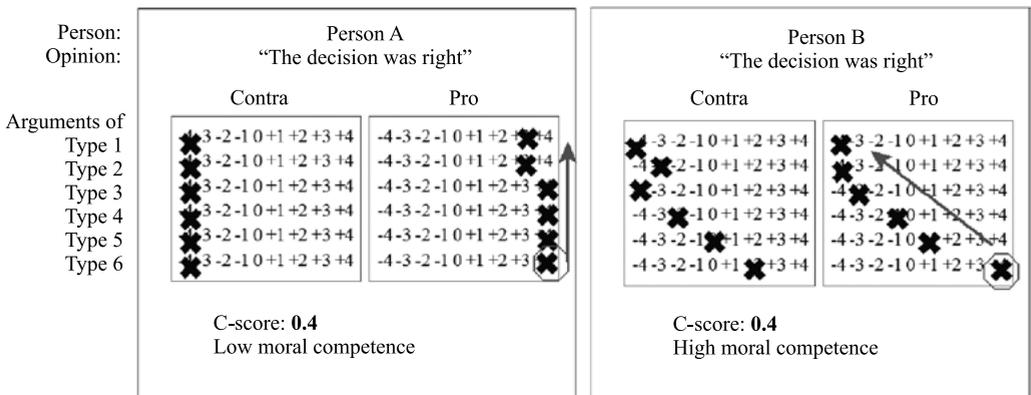


Figure 1. This figure depicts the pattern of evaluative responses of two fictitious participants to moral arguments of different moral quality. It shows that an isolated scoring of the responses would be ambiguous. For example, the true meaning of the positive rating of a type-6 supporting arguments (“+4”) can be determined only when considering the whole pattern of responses: Person A, with low moral competence, obviously rated arguments only in regard to their opinion-agreement, whereas Person B, with high moral competence, evaluated the arguments in regard to their moral quality. The C-score ranges from 0 to 100 (Lind 2016).

The MCT is also empirically valid. In all known studies, MCT-data agree almost perfectly with the four predictions derived from moral psychological research (Lind, 2016): (1) the preferences for each of the six types of moral orientation (as defined by Kohlberg, 1984) are ordered in the predicted way, (2) the correlations between preferences for “neighboring” types are higher than for more distant types, (3) the higher a person’s moral competence (structure), the more clearly he or she prefers high type moral arguments and rejects low type moral reasoning, and (4) moral competence cannot be simulated upward – otherwise the MCT could not be called a competence test (Lind 2016).

CONCLUSION

A hundred and fifty years after Donders’ ground-breaking study on the mental processing of cognitive dilemmas, his experimental approach (and Brunswik’s diacritical method) inspired a new methodology of objective assessment in developmental psychology: Experimental Questionnaires (EQ). This ends the researchers’

dilemma of having to choose between objective but invalid methods on the one hand, and valid but subjective methods on the other. We can now study *internal*, structural dispositions of human behavior *objectively*. The Moral Competence Test is one of the first applications of the EQ methodology, but this can be used in various fields of developmental psychology and beyond. With the MCT we can now objectively measure people’s moral competence without relying on their subjective accounts and on subjective scoring.

The new, experimental measurement methodology has made possible a breakthrough in applied developmental moral psychology and education. Because the MCT is not only objective and valid but requires little time for test-taking and scoring, it has also helped to increase enormously the efficacy of educational methods and educational systems for fostering moral competence (see, e.g., Lerkiatbundit et al. 2006; Schillinger 2006; Mouratidou et al. 2007; Comunian, Gielen 2006; Slováčková, Slovacek 2007; Yang, Wu 2008; Nowak, Schrader, Zizek 2013; Lupu 2013; Feitosa et al. 2013; Hemmerling 2014; Lind 2016).

NOTE

¹ This article is based on my invited presentation to the 25th anniversary meeting of the Polish Association of Developmental Psychology in Krakow, June 16, 2016.

Contact: Georg.lind@uni-konstanz.de. <https://www.uni-konstanz.de/ag-moral/>

REFERENCES

- Anderson, N.H. (1991), Moral-Social Development. In: N. H. Anderson (ed.), *Information Integration Theory*. (Vol. III: *Developmental*, pp. 137–187). Hillsdale, NJ: L. Erlbaum.
- Arbuthnot, J. (1979), Error in Self-Assessment of Moral Judgment Stages. *The Journal of Social Psychology*, 107, 289–290.
- Beck, K., Dransfeld, A., Minnameier, G., Wuttke, E. (2002), Autonomy in Heterogeneity? Development of Moral Judgment Behavior During Business Education. In: K. Beck (ed.), *Teaching-learning processes in vocational education*, pp. 87–119. Frankfurt: Peter Lang.
- Blasi, A. (1980), Bridging Moral Cognition and Moral Action: A Critical Review of the Literature. *Psychological Bulletin*, 88, 1–45.
- Brunswik, E. (1955), Representative Design and Probabilistic Theory in a Functional Psychology. *Psychological Review*, 62, 193–217.
- Broughton, J. M. (1978), The Cognitive-Developmental Approach to Morality: A Reply to Kurtines and Greif. *Journal of Moral Education*, 7, 81–86.

- Colby, A., Kohlberg, L., Abrahami, A., Gibbs, J., Higgins, A., Kauffman, K., Lieberman, M., Nisan, M., Reimer, J., Schrader, D., Snarey, J., Tappan, M. (1987). *The Measurement of Moral Judgment*. (Vol. I). New York: Cambridge University Press.
- Comunian, A. L., Gielen, U. P. (2006), Promotion of Moral Judgment Maturity Through Stimulation of Social Role-Taking and Social Reflection: An Italian Intervention Study. *Journal of Moral Education*, 35(1), 51–69.
- Cortese, A. J. (1984), Standard Issue Scoring of Moral Reasoning: A Critique. *Merrill Palmer Quarterly*, 30, 227–246.
- Donders, F. C. (1986), On the Speed of Mental Processes. *Acta Psychologica*, 30, *Attention and Performance II*, 412–431. (Original work published 1868). Retrieved from <http://www2.psychology.uiowa.edu/faculty/mordkoff/InfoProc/pdfs/Donders%201868.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.
- Emler, N. (1996), How can we Decide whether Moral Education Works? *Journal of Moral Education*, 25, 117–126.
- Feitosa, H., Rego, S., Bataglia, P.U., Sancho, K.F.C., Regor, G., Nunes, R. (2013). Moral Judgment Competence of Medical Students: A Transcultural Study. *Advances in Health Science Education*, 18(5), 1067–1085. <https://doi.org/10.1007/s10459-013-9449-5>
- Flanagan, O. F. (1991), *Varieties of Moral Personality. Ethics and Psychological Realism*. Cambridge, MA: Harvard University Press.
- Gulliksen, H. (1950), *Theory of Mental Tests*. New York: Wiley.
- Habermas, J. (1990), *Moral Consciousness and Communicative Action*. Cambridge, MA: MIT Press.
- Haidt, J. (2001), The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review*, 108(4), 814–834.
- Hartshorne, H., May, M. A. (1928), *Studies in the Nature of Character*. (Vol. I: *Studies in deceit*, Book one and two). New York: Macmillan.
- Hemmerling, K. (2014), *Morality behind Bars – An Intervention Study on Fostering Moral Competence of Prisoners as a new Approach to Social Rehabilitation*. Frankfurt/New York: Peter Lang.
- Keller, M. (1992), Soziales Verstehen und moralisches Urteilen im Kindesalter. In: F. Oser & W. Althof (hrsg.), *Moralische Selbstbestimmung: Modelle der Entwicklung und Erziehung im Wertebereich. Ein Lehrbuch*, pp. 196–198, Stuttgart: Klett-Cotta.
- Kelly, G. A. (1955), *The Psychology of Personal Constructs*. New York: Norton.
- Kohlberg, L. (1958), *The Development of Modes of Moral Thinking and Choice in the Years 10 to 16* (Doctoral dissertation). University of Chicago. (Published with Garland Press)
- 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. (1984), *Essays on Moral Development*. (Vol. II: *The Psychology of Moral Development*). San Francisco, CA: Harper & Row.
- Kohlberg, L. (2010), Foreword. In: G. Lind et al. (eds.), *Moral Judgment and Social Education*, pp. xv–xvii. New Brunswick, NJ: Transaction Publ. Inc. (Original work published 1985).
- Kohlberg, L., Candee, D. (1984), The Relationship of Moral Judgment to Moral Action. In: L. Kohlberg (ed.), *Essays on moral development*. (Vol. II: *The Psychology of Moral Development*, pp. 498–582). San Francisco, CA: Harper & Row.
- Kurtines, W.M., Gewirtz, J.L. (1995), An Introduction and Overview. In: W.L. Kurtines, J.L. Gewirtz (eds.), *Moral development: An introduction*, pp. 1–16. Needham, MA: Allyn & Bacon.
- Kurtines, W.M., Greif (1978), The Development of Moral Thought: Review and Evaluation of Kohlberg's Approach. *Psychological Bulletin*, 81, 453–470.
- Lerkiatbundit, S., Utaipan, P., Laohawiriyanon, C., Teo, A. (2006), Randomized Controlled Study of the Impact of the Konstanz Method of Dilemma Discussion on Moral Judgment. *Journal of Allied Health*, 35(2), 101–108.
- Lind, G. (1978), Wie misst man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts [How can we measure moral judgment? Problems and possibilities of measuring a complex construct]. In: G. Portele (hrsg.), *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. (1989), Measuring Moral Judgment: A Review of the Measurement of Moral Judgment by Anne Colby and Lawrence Kohlberg. *Human Development*, 32, 388–397.
- Lind, G. (2008), The Meaning and Measurement of Moral Judgment Competence Revisited: A Dual-Aspect Model. In: D. Fasko, W. Willis (eds.), *Contemporary philosophical and psychological perspectives on moral development and education*, pp. 185–220. Cresskill, NJ: Hampton Press.
- Lind, G. (2010), Attitude Change or Cognitive-Moral Development? How to Conceive of Socialization at the University. In: G. Lind, H.A. Hartmann, R. Wakenhut (eds.), *Moral judgment and social education*, pp. 173–220. Rutgers, NJ: Transaction Books.
- Lind, G. (2013), Thirty Years of the Moral Judgment Test* – Support for the Dual-Aspect Theory of Moral Development. In: C.S. Hutz, L.K. de Souza (eds.), *Estudos e pesquisas em psicologia do desenvolvimento e da personalidade: Uma homenagem a Angela Biaggio*, pp. 143–170. Sao Paulo, Brazil: Casa do Psicólogo. [* now renamed into Moral Competence Test]
- Lind, G. (2016), *How to Teach Morality. Promoting Deliberation and Discussion. Reducing Violence and Deceit*. Berlin: Logos.
- 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.
- Loevinger, J. (1957), Objective Tests as Instruments of Psychological Theory. *Psychological Reports*, 3, 635–694.
- Lohmann, D.F. & Ippel, M.J. (1993), Cognitive Diagnosis: From Statistically Based Assessment Toward Theory-Based Assessment. In: N. Frederiksen, R. J. Mislevy, I. I. Bejar (eds.), *Test Theory for a New Generation of Tests*, pp. 41–71. Hillsdale, NJ: Lawrence Erlbaum Ass., Publ.
- Lupu, I. (2013), Moral Competence and Dogmatic Religiosity. In: E. Nowak, D. Schrader, B. Zizek (eds.), *Educating Competencies for Democracy*, pp. 379–390. Frankfurt am Main: Peter Lang Verlag.
- 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.
- Messick, S. (1995), Validity of Psychological Assessment. Validation of Inferences from Persons' Responses and Performances as scientific Inquiry into Score Meaning. *American Psychologist*, 50, 741–749.
- Mischel, W. (1973), Toward a Cognitive Social Learning Reconceptualization of personality. *Psychological Review*, 80, 252–283.
- Mischel, W., Shoda, Y. (1995), A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure. *Psychological Review*, 102, 246–268.
- Mouratidou, K., Chatzopoulos, D., Karamavrou, S. (2007), Moral Development in Sport Context: Utopia or Reality? *Hellenic Journal of Psychology*, 4, 163–184.
- Nicolayev, J., Philips, D.C. (1979), On Assessing Kohlberg's Stage Theory of Moral Development. In: D.B., Cochrane; C.M., Hann, A. Kazepides (eds.). *The Domain of Moral Education*, pp. 231–250. New York.
- Nowak, E., Schrader, D., Zizek, B. (eds.) (2013), *Educating Competencies for Democracy*. Frankfurt am Main: Peter Lang Verlag.
- Pearson, K. (1913), *On the Correlation of Fertility with Social Value: A Cooperative Study*. London: Dulau & Co.
- Piaget, J. (1965), *The Moral Judgment of the Child*. New York: The Free Press. (Original work published 1932)
- Pittel, S. M., Mendelsohn, G. A. (1966), Measurement of Moral Values: A Review and Critique. *Psychological Bulletin*, 66, 22–35.
- Rest, J. R. (1979), *Development in Judging Moral Issues*. Minneapolis, MI: University of Minnesota Press.
- Schillinger, M. (2006), *Learning Environments and Moral Development: How University Education Fosters Moral Judgment Competence in Brazil and Two German-Speaking Countries*. Aachen: Shaker-Verlag.
- Scott, W. A. (1968), Attitude Measurement. In: G. Lindzey, E. Aronson (eds.), *Handbook of Social Psychology*, pp. 205–270. Reading MA: Addison-Wesley.
- Slovácková, B., Slovacek, L. (2007), Moral Judgment Competence and Moral Attitudes of Medical Students. *Nursing Ethics*, 14(3), 320–328.
- Sprinthall, R. C., Sprinthall, N. A., Oja, S. N. (1997), *Educational Psychology: A Developmental Approach* (7th ed.). New York: McGraw Hill.

- Thurstone, L. L. (1928), Attitudes can be Measured. *Journal of Sociology*, 33, 529–554.
- Torgerson, W. S. (1958), *Theory and Methods of Scaling*. New York: Wiley.
- Wark, G. R., Krebs, D. L. (1996), Gender and Dilemma Differences in Real-Life Moral Judgments. *Developmental Psychology*, 32(2), 220–230.
- Wiggins, J. S. (1973), *Personality and Prediction. Principles of Personality Assessment*. Reading MA: Addison-Wesley.
- Yang, S., Wu, H. (2008), The Features of Moral Judgment Competence Among Chinese Adolescents. *Asia Pacific Review*, 9(3), 296–307.