

The Art of Experimental Moral Psychology¹

Thoughts in Progress

Georg Lind²
University of Konstanz, Germany

This essay has been written for you if you plan to do empirical or experimental psychological research in . If you have asked me to be one of your readers, my essay will also give you an idea by which I will evaluate your final work. Please correct my summary as far as it is based on a false interpretation of your intentions. If my points seem not clear enough to you, please ask me to clarify them, or read some of my recent methodological publications:

<http://www.uni-konstanz.de/ag-moral/b-publik.htm>

As I understand it, your dissertation is to be an empirical, if not experimental study, which is to clarify an important question in a field which can be described broadly as psychology of moral behavior and moral education.

Definition & measurement

To be sure that we talk about the same object of research, let me say how it is defined. *Moral competence* (also called *moral-democratic competence* or *moral judgment and discourse competence*) is defined as *the ability to solve dilemmas and conflicts on the basis of shared moral ideals through deliberation and discussion rather than through violence, deceit, fraud, or abuse of power*.

Moreover, moral competence is an *aspect* or attribute of moral behavior, not a component or a different kind of behavior. It is the *cognitive* aspect, while moral orientations (or ideals, or values, or principles) represent the affective aspect of moral behavior. Because of this special feature, we call our theory *dual-aspect theory of moral behavior and development*.

The operational meaning of these two aspects is given through the *Moral Judgment Test* (MJT) and its scoring algorithm. The two aspects differ in regard to an important feature. While a moral orientation can be simulated (or faked) in any direction if the participants want to do so, moral competence can be simulated only in one direction, namely downward.

Currently, the MJT is the only instrument for measuring moral competence and moral orientations in an un-confounded manner (for more details and FAQs concerning the MJT, see Lind, 2008a; <http://www.uni-konstanz.de/ag-moral/mut/mjt-engl.htm>). Kohlberg's Moral

¹ The last revision of this essay was Nov. 29, 2013

² Contact: Prof. Dr. Georg Lind, Schottenstr. 65, 78462 Konstanz, Germany. E-mail: georg.lind@uni-konstanz.de.

Judgment Interview (MJI) has been designed to measure ‘moral stages,’ which are an attempt to provide a joint measure of moral competence and moral orientation (Lind, 1989).

Basically, there are two kinds of studies which involve the concept of moral competence:

First, studies which are interested in the question whether moral competence is influenced by cultural constraints and whether it can be fostered through means like education, and how it can be done effectively and sustainably. More specifically, one can ask whether methods like the Konstanz Method of Dilemma Discussion (KMDD) improve substantially moral competence?

A second kind of study is interested in the impact of moral competence on certain human behavior (like helping, cheating, political participation, learning, decision-making, prejudice, etc.), and social phenomena (like rate of criminality, welfare, learning climate, peace, etc.).

At the end of this paper I present a collection of research questions which seem rather important but have not been addressed yet by empirical and experimental studies. If you take up one of these questions, please let me know.

I propose the following guidelines for such studies:

- Be focused. Your study should focus on testing only *one main* hypothesis or prediction. You may also examine some auxiliary hypotheses which are closely related to your main hypothesis. But you should always be aware of the fact that each additional research question is not just an addition to your work but means always a multiplication of work efforts, especially if you do it carefully. Besides you readers and the marketing department of your publisher will appreciate a clear focus.
- Be in charge. You must take full responsibility for your research question and research findings. This means that you should take authorship for your main predictions and your examination of its validity. So there should be statements starting with "I believe that ...," or "my research has found that ...", and "Statement X is true / false." This means also that you should be able to give examples for the application of the main theory on which you base your predictions, and for the application of your recommendations based on your findings. And it means that you should write clearly, use your terminology consistently and present a coherent argument.
- Honor others. Quotes are not intellectual crutches but means of honoring important sources of your own insights. Quotes are no substitution for a lack of own reasoning. However, you should use quotes when you take over an important definition from another source. But put it in quotation marks only once, not every time you use it. If you paraphrase other authors because you find their reasoning especially convincing, or if you rely on their research findings, you should honor your sources through references in parentheses, e.g., (see Lind, 2008) and in the list of references, or even through footnotes, in which you explain how much you are indebted to those sources.

- Be fair. Show that you really know and understand the theory which you put to test. Through your writing the reader should exactly learn what this theory says. Do not devalue this theory unintentionally by paraphrasing its core concepts and statements in a loose or even sloppy manner. Quote always the most recent writings about this theory, and tell the reader if the theory has changed over time because of new insights, especially if the publications lie far apart in time.
- Be precise. Since you have chosen my Dual-Aspect Theory of moral behavior and development as a major object of empirical examination and, if confirmed, as basis for practical recommendations, you must always use its core terms as they are defined in (recent!) literature, and not use synonyms or circumscriptions without labeling them as such. The core concepts of this theory are "moral orientation," "moral competence," "affective-cognitive parallelism," morally related "behavior" or "decision," and "aspects" as contrasted with "components."
- Lost in translation. Social sciences are still in a pre-scientific stage. Social scientists rarely define their core concepts clearly, but use them in a variety of ways, and sometimes even inconsistently and contradictory. On the hand this means you are free, or even forced, to develop your own definitions (within the boundaries explained in the previous paragraph). On the other hand it means you must do translation work when you quote and discuss the work of other authors. This means, for example, that you need to study careful the usage of the term "competence" in the literature about ethical/moral behavior and development, because it often does not agree with our definition of competence, but means a moral orientation.
- Perspective taking. Every-day language and scientific language are mostly so different that they also need translation. Especially the term "moral" has many meanings in every-day talking, which, in turn, is very different from the use of this term in dual aspect-theory. Use theoretical terms cautiously in every-day contexts. Else your audience may misunderstand you. Also the term "dilemma discussion" should be rarely used in praxis contexts. In the KMDD we do not present "dilemmas" but stories which seem to contain a dilemma or many. A dilemma lies in the eyes of the beholder!
- Mistrust statistical significance (use them only for conventional reasons). Statistical significance signifies first and mostly how big the sample was which the author could afford to study. If possible, convert F-statistics, t-statistic and alike into a measure of effect size (preferably the coefficient of correlation 'r') or, even better, into a measure of absolute effect size (differences between means or percentages). Coefficients of correlation are better than test-statistics yet they are also problematic, because they depend too much on the variance of the independent variable which can vary greatly from one study to another but is hardly ever reported or controlled for (Lind, 2012).
- Mistrust so-called correlation studies. Correlation studies, that is, studies in which many variables are measured and subjected to correlation analysis, are very difficult to interpret because of a number of problems. Therefore, correlation studies can be interpreted in many different ways and can support even mutually exclusive theories. Correlations do

not mean causation ...

- because a correlation cannot distinguish between the direction of causation. ‘A can cause B’ or ‘B can cause A’. In both cases, there can be the same correlation;
 - because correlations can be spurious. In young age, shoe size and English test scores can correlate positively ... because both correlate with age;
 - because correlations can be masked. A zero correlation can be the result of two strong but opposing trends, as in the case of curvo-linear correlations. For example, the Yerkes-Dodson law states that level of excitement and achievement are positive correlated ... until a certain point of excitement. If people become over-excited or anxious, growing excitement results in decreasing achievement; hence the overall correlation between level of excitement and achievement is zero.
 - because correlations depend on the variance of the independent variable(s), which is hardly reported and controlled for.
- Be experimental. You are in a much better position if you can manipulate your main independent variable (e.g., through an intervention) and observe how it affects your core dependent variables. Of course, even then you cannot be absolutely sure whether your interpretation stands the test of your critiques: (a) your intervention is usually a compound of elements and you may believe the wrong component to be effective or overlook that no single element is effective but their composition), and (b) the effects of your intervention depend on characteristics of the environment or of the participants. But with an experimental design you are in a MUCH better position to wheat out competing explanations for your finding than with a correlational design.
 - Use follow-up. If you hope to make recommendations about lasting effects you can either tell your reader that you assume that the effects are lasting, or you can show it by doing a follow-up measurement or observation. Follow-up observations should be done at least a couple weeks after the post-test, but should not be done too much later because interventions can hardly be expected to have "eternal" effects. You can hope for that your intervention can start a perfectly self-sustaining process. But it is more realistic to assume that the effects of your treatment fade out if the treatments are not repeated or if they are not supported by institutionalized practices.
 - Protect privacy / be polite. If you do pretest-posttest comparison and follow-up then you must find a way to protect the privacy of your participants and you must give them a short but good explanation why you want them to fill out the same test(s) twice or more often, otherwise they may not want to give you a sincere answer or refuse to respond at all. Also the privacy of the participants' institution and its staff may need to be protected.
 - Compare findings. You can learn from the effects of interventions only if you compare the changes in the experimental group with changes in other groups. For example, there can be a strong effect even though there was no change found in an experimental group (e.g., prison inmates) but at the same time a control group (of prison inmates) without intervention showed strong regression. Hence the intervention had an effect. It stopped

regression. Matched control-groups are helpful but not the only way to compare. In the case of moral competence research there are many data available that you can also compare with them. The more comparisons you can make the better you can discount competing interpretations for your findings, and the more you can trust in the interpretation which survives the experimental examination of truth (for appropriate studies to compare your data with, see

<http://www.uni-konstanz.de/ag-moral/b-publik.htm>.

- Randomization is not fool-proof. Randomization is often requested in experimental studies, but it is debated because it does not rule out false interpretations. Full randomization is never possible because some very important variables cannot be randomized at all. You can never randomize over time-samples; past and future are out of reach for science. You can never randomize over all participants in a study; competent instructors cannot be replaced by incompetent ones; voluntary participants cannot be replaced by mandated participants. Often you cannot randomize control groups for practical reasons, because they are either given or they are too expensive (in a well-documented study on the effects of a reading program, the control schools had to be paid 32,000 US dollars. By the way, there was hardly any effect found). Moreover, studies with randomized samples have not shown any superiority of studies with matched samples. Matching, however, should be done carefully, based on important variables, not on conventional variable.
- Do not invent the wheel over again. Fortunately, you have chosen the dual-aspect theory of moral behavior and development. In this research field some findings are so well established that you can trust them and built on them without the burden to prove them in your own study again. I claim that this field of research is a rare example of “progressive problem shift” (Imre Lakatos) in psychological research. We have really made some scientific progress! Of course, there are still a lot of questions which are not yet answered or are still debated. But we do not need to start from the scratch as is often the case when you enter a field in the social sciences. Applied to your study: (a) You can trust that moral competence exists and has a strong impact on many areas of our moral life. (b) You can trust that with the KMDD one can effectively foster moral competence – however, only if the user of this method is well trained and certified.
- Your challenge. What, in my view, needs to be answered are questions like these: (a) has moral competence (as measured with the MJT) an impact on behavior and decision-making in the private sphere, in the public life, and in business context (like corruption, whistle-blowing, theft of company property, but also quality of work, satisfaction of employees, illness rates, and so on)? [Of course, it is up to you on which behaviors you want to focus] (b) Can the KMDD be successfully implemented in a particular context? Are there restrictions of its application? What costs and conditions are involved? What benefits should be expected? [I believe that the returns are high]
- Style of writing / design of report. The first and the last chapter (introduction, conclusion) of your research report (i.e., book) should be written in clear, simple, non-technical language, speaking to a general audience (that is, all people belong to this group besides

the very few experts in the field on which you do the research). If you think you should make the reader acquainted with some core technical concepts you must explain them with familiar examples (not with other unfamiliar concepts). The chapter 'theory' should exhibit your own theory; this does not need to be a completely new theory, but a theory for which you take over ownership as you take over ownership over a car which you have neither designed nor built. It is your car, and it is your theory. It means that you explore the world within the confines of this theory and try out where you can get with it and how far it carries you without a severe breakdown. The theory should be of a unique make, but it can get add-ons from other theories as well. The chapters 'method' and 'findings' should contain all technical information that is needed by researchers who want to analyze and replicate your study. With one exception: in order to prevent abuse, I do not allow re-printing the MJT in publications, rather you must refer the reader to the available sources (my web-site, and some older publications).

- Tables are necessary, but big tables with many numbers should be placed in the appendix; only simplified versions should go into the findings-chapter. Graphs are a must to present core findings. However, they should not be overburdened. Tables and figures should speak for themselves; avoid abbreviations if they are not in general use. The text of the findings-chapter should not replicate the information that you give in tables and figures, but rather they should (a) help the reader to read and understand the tables and figures if they cannot be made fully self-speaking, and (b) link the numerical and graphical data to the theory and the hypotheses which you are testing.
- Be a composer. To make the report powerful and convincing, you should compose it like a piece of music; you should repeat (and vary) your main research question in each chapter like the composer Johannes Bach has used a theme in his fugues. Then your readers will sing your song with you and remember it long after they have read your book. (I wished someone had told me this at the beginning of my career.)
- Last will. Read as much about your research topic as is possible but do not agree with everything you read; if you agree with more than 10 percent of what you are reading you may not have fully understood it. Criticize only those theories and authors which deserve it ("deserve" in the best sense of this word!); really bad pieces of research should not be criticized but ignored. If you have a better idea than you found in literature, say it! To say something simple is not simple at all. It took me 40 years of research to come up with these suggestions here.

Best wishes!

References

- Lind, G. (1989). Essay Review: 'The measurement of moral judgment' by Anne Colby, Lawrence Kohlberg. *Human Development* 32, pp. 388 - 397. [PDF](#)
- Lind, G. (2008a). 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. [PDF](#)

Lind, G. (2008b). Teaching students to speak up and to listen to others: Cultivating moral democratic competencies. In: D.E. Lund & P.R. Carr, eds., *Doing democracy and social justice in education: Political literacy for all students*, pp. 319 - 335. New York: Peter Lang Publishing. [PDF](#)

Lind, G. (2012). Effektstärken: Statistische versus praktische und theoretische Bedeutsamkeit. Erweiterter Vortrag. [*Effect sizes: statistical, practical, and theoretical significance of empirical studies.*] . [PDF](#)

E. Nowack, D. Schrader & B. Zizek, eds. (2013). *Educating competencies for democracy*. New York: Peter Lang Verlag ([table of contents](#)).

Unsystematic collection of research ideas

(work in progress)