

PETER KROPE, NORBERT BEHRENS, MARTIN GORLIKOWSKI, PAUL LORENZ

EDUCATIONAL MEASUREMENT WITHOUT CATEGORY MISTAKE

1. The Unbearable Category Mistake

The answer to the question what is to be seen on an empty plate is: “Nothing!”. After an apple has been added it is correct to reply: “An apple!”. The response that there are now two objects on the plate, that is to say the apple and the nothing, might be amazing. But it is incorrect. The answer is wrong, because two circumstances have been connected that logically belong to different categories.

Exactly this mistake is common in educational measurement. In questionnaires we are asked questions like “Are you contented?” and are requested to mark a digit as an answer. We are used to read sentences like: “The average contentedness of the test persons was 2.4”.

The mistake does not only appear in the daily routine of diagnostics, but also in the theory. Because of its clearer representation, the probabilistic test theory is chosen here. The general starting-point of the probabilistic test theory reads as follows: $P(a_{vi}) = f(V_1, V_2, \dots)$. This formula means that the probability P of a particular reaction a of the person v to the item i is a function of the latent variables V_1, V_2, \dots (ROST and SPADA 1978, 63/64). In the formula something manifest is mentioned indirectly, that is to say the reaction of a person. In the context of multiple-choice items, it is the behaviour of choice in ticking boxes. In the formula something latent is directly mentioned which is essentially identical with dispositions. A simplified explanation suggests that the probability of the (manifest) behaviour is explained by (latent) dispositions (ROST 1995). In theory, it is assumed that the probability of manifest facts may be explained entirely by identifying a finite number of latent variables. This assumption is based on models, which axiomatically describe the connection between the manifest and the latent parameters (SCHEUCH and ZEHNPENNIG 1974, 134).

Confusion arises, if circumstances (like behaviour and behaviour disposition, manifest facts and latent characteristics, body and mind) are assembled in

the same category, though they belong to different types. RYLE (1949) calls this wrong classification a category mistake. According to BENNETT and HACKER (2003) the category mistake casts doubt on theories, methods and results in psychology and neuroscience.

The category mistake has been one of the reasons for developing a constructivist theory of educational measurement (KROPE 2000). The theory has been developed further, an important part of which is the constructivist principle of generic acts ("*Handlungsschemata*", KROPE and PETERSEN 2003). It is to be expected that this principle will contribute to an elimination of the category mistake.

The following serves to illustrate the use of the principle of generic acts in the educational measurement and to determine the abilities that test persons must have as a precondition for the application of the constructivist approach.

2. The Definition of Contentedness

From 1999 to 2002 the research team of the Centre for Constructive Educational Science conducted the first study of contentedness. The aim was to measure the contentedness of young trainees (KROPE et al. 2002).

At the beginning of an empirical study the basic term has to be defined. Starting with terms defined by LORENZEN (1987, 263) and KAMLAH and LORENZEN (1973, 39-40) the term "discontentedness" has being defined as

$$\text{discontentedness } (S, y, \sigma) \stackrel{\Leftarrow}{=} \text{desire } (S, y, \sigma) \wedge \neg V (S, y, \sigma).$$

In this formula $\stackrel{\Leftarrow}{=}$ is the sign for a definition, \wedge the sign for „and“ and \neg the sign for a negation. The definition reads: The subject S is discontented in a situation σ with the object y exactly when it firstly desires the object y in a situation σ and secondly does not dispose of the object y (means) in this situation σ or rather this object (circumstances) is not fulfilled. In step 2 the term "contented" is identified with the term "not discontented".

The following example is given. A child comes home after a long day at school. It informs his dad that it is hungry and wants to have something to eat (desire) because it did not get anything to eat all morning (availability). The father serves lunch (change in the object of the availability) and the child is contented.

3. Connecting Scientific and Ordinary Language

Due to the terminological agreement pre-scientific and scientific language differ in two ways. Firstly a difference in the contents is to be noticed. Secondly there is a difference in the forms. The differences represent a problem in questionnaire surveys. The different languages require translation rules which ensure an understanding between science and everyday life. The rules have to take care

that the questioning scientists who make use of the terminology and the answering persons who use a pre-scientific language does not misunderstand each other.

Several methodical details served to bridge the gap between scientific and ordinary language in the study. Firstly. The definition of “contentedness” is based on the term “desire” introduced by LORENZEN “as the endeavour that we support by our acts” (LORENZEN 1987, 263). These acts are observable to a large extent. Due to this definition scientists are able to decide when a desire is fulfilled and when not, because the fulfilment of the desire has to find expression in acts.

Secondly. Because everybody who expresses a wish desires something, the term “desire” has been translated by the everyday expression “wish”.

Thirdly. Persons questioned about their contentedness answer that they are “more or less” contented. To interpret these statements traditionally, assumptions are being made which can not be accepted. The traditional assumptions lead to an error that is described as “category mistake”. To avoid the mistake the first study of contentedness asks for acts and not for behaviour on the empirical level and for generic acts and not for behaviour dispositions or attitudes on the rational level (KROPE and PETERSEN 2003). A “more or less” is understood as a frequency of acts. That is why the questions are being introduced with “how often” in the first study of contentedness (see table 1).

Table 1: The introduction to the questions on contentedness in the first study of contentedness (KROPE et al. 2002, 48).

How often did you express last week the wish that ...	State here the number of days! ↓
...	0 1 2 3 4 5 6 7

4. Support as a Methodical Problem

In 2003 the research-group of the Centre for Constructive Educational Science conducted the second study of contentedness (FELST et al. 2004). In the second inquiry the concept of the first study was adopted.

As it turned out during the preparation of the survey, numerous persons were only insufficiently or not at all able to read or answer the questions of a questionnaire. They are dependent on help. In this second study those methodical problems had come up which were described already in the first study of contentedness.

In the first study several of the young people were handicapped so that they could not answer the questionnaires without help. Support was given by coaches who were familiar with the problems of the young trainees. Their task was strictly limited to technical aid. Among other things it consisted of reading out texts, translating questions into sign language, to dispel understanding problems and to put down answers. The coaches were very well aware of the danger of the inter-

vention. They tried to restrict their influence because they were interested in the objectivity of the study.

Despite all precautions the results of the questioned handicapped persons showed a special feature. The reliability of their data was significantly higher than in the entire study. Because the reliability of the questionnaire was calculated according to the split half method it informed about the internal consistency. In other words the answers of the young handicapped people turned out to be more homogenous than in the whole study.

How could the homogeneity be explained? The method of support raises the question of the authorship of the answers. Do the answers come from the questioned young people? Or are the answers formed by the helping persons? That the helpers run the risk to see their own achievements as the achievements of those whom they want to help becomes clear from the research on facilitated communication. Through facilitated communication, according to the advocates of the method, humans with communication disorders like autism are supposed to be able to communicate with their environment. User of the method assert that via the support, such as holding the writing arm, meaningful texts at a high level have been produced. However, in controlled double-blind studies under laboratory conditions as well as in field studies it has been proven that these texts did not derive from the handicapped persons but from the helping persons (JACOBSEN et al. 1995).

5. The Talking Questionnaire

To analyse the problem and its solution, the “talking questionnaire” project was started. Starting point was an analysis of ordinary language according to the living-world (“*Lebenswelt*”) concept of HUSSERL (1992). Different factors can determine the way of pre-scientific statements of a person. The determining factors can be the gender, the age and the place of residence. Male pupils can make other pre-scientific statements than female pupils, young people make statements differing from those of older people, the people in the north of a city again are different from the people in the south of the same city. Different people can belong to different living-worlds which can link up with different habits of the ordinary language. The differences can be so large that the people from different living-worlds can be unable to understand each other. Therefore it was impossible to easily adopt the concept from the first study in the second study. Rather a communicative code had additionally to be developed for the illiterate persons.

To question the illiterate persons the computer program “the talking questionnaire” was developed. It contains 10 questions of the questionnaire which was developed for the second study. The questions are presented by a voice stored in

a computer program¹. Because of the multiple-choice type the answers can be given with the help of three differently coloured keys (see table 2). These keys also serve to control the entire course of the oral questioning. After an explicit agreement about how to deal with the talking questionnaire the questioned persons independently answer the instructions and questions which are entirely verbally presented using the keys. Not a single word or sign has to be read or written. The results of the questioning are stored in a data file.

Table 2: The three keys to operate the computer program called “talking questionnaire”



After the program had been developed successfully ten test persons (not belonging to the group of illiterate persons) were selected by random. These persons had to answer the ten questions of the talking questionnaire and additionally the same questions in a written presentation. The group was divided in such a way that five persons started with answering the oral version and the other five persons with the written version. The question was: Do the oral answers and the written answers differ? The cases of coinciding answers are given in table 3.

Table 3: Cases of coinciding answers

Testperson	1	2	3	4	5	6	7	8	9	10
Cases of coinciding answers	10	10	3	10	10	7	9	8	8	10

¹ The program was developed by Daniel Mattsson under the supervision of Prof. Jörg Petersen. The text was spoken by Wiebke Skala.

It was to be expected that all ten answers in the oral and the written version coincided for all ten test persons. The χ^2 -test showed no significant difference. In general the test persons answered the oral questions and the written questions in the same way.

6. The Ability to Distinguish

Topic of this contribution is the attempt to avoid the category mistake in educational measurement. The two reported empirical studies demonstrate a three-step-solution. Step 1 is the development of a definition of basic terminology. Step 2 includes the task to bridge the gap between scientific terms and everyday language expressions. Step 3 experimentally furnishes proof of a successful communication between scientists and test persons.

Via the “talking questionnaire” program not only the possibilities but also the limits of educational measurement can be demonstrated. The written questionnaire on contentedness is based on the ability to distinguish, which is ensured by the ability to read and to write. In the second study the written questionnaire had to be adapted to the illiterate persons’ ability to distinguish. The questionnaire of the first study was modified with regard to the living-world of the test persons for the second study. An explicit agreement was made, formal (with ε for “is” and \rightarrow as a transition-arrow):

$$\begin{aligned} x \varepsilon \text{ YES} &\rightarrow \text{yellow key} \\ x \varepsilon \text{ NO} &\rightarrow \text{blue key} \end{aligned}$$

In words: For all questions to be answered with YES or approximately with YES, the yellow key is to be pressed and for all questions to be answered with NO or approximately with NO, the blue key is to be pressed. Thus through an inter-subjective, explicit agreement a correspondence (YES \rightarrow yellow key and NO \rightarrow blue key) is created to guarantee reliability and understanding.

With reference to this introductory situation the general limits of educational measurement can be described. The unavoidable beginning is the ability of the test persons to make linguistic and pragmatic differentiations².

References

1. BENNETT, M. R., HACKER, P. M. S.: *Philosophical Foundations of Neuroscience*, Oxford 2003.
2. FELST, M., KROPE, P., LATUS, K., PETERSEN, J. P., SKALA, W., STENDER, D., WEIS, T.: *Wie zufrieden sind Jugendliche mit der Beratung? Abschlußbericht einer Evaluationsstudie auf methodisch-konstruktiver Grundlage*, [in:] *Arbeitsgruppe konstruktive erziehungswissenschaft am institut für pädagogik* (ed.): *Monographien zur konstruktiven Erziehungswissenschaft*, vol. 5, Kiel 2004.

² For details of this epistemological principle see MITTELSTRAB (1996).

3. HUSSERL, E.: *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*, [in:] *Gesammelte Schriften*, E. STRÖKER (ed.), vol. 8, Hamburg 1992.
4. JACOBSON, J. W., MULICK, J. A., SCHWARTZ, A. A.: *A History of Facilitated Communication: Science, Pseudoscience, and Antiscience: Science Working Group on Facilitated Communication*, [in:] "American Psychologist" 1995, vol. 50, No 9, 750-765.
5. KAMLAH, W., LORENZEN, P.: *Logische Propädeutik. Vorschule des vernünftigen Redens*, 2nd ed. Mannheim etc., 1973.
6. KROPE, P.: *Konstruktive Pädagogische Diagnostik*, 2nd ed. Münster etc., 2000.
7. KROPE, P., unter Mitwirkung von FRIEDRICH, B., GREFE, S., KLEMENZ, D., LORENZ, P., PETERSEN, J. P., THIEBACH, J., WOLZE, W.: *Die Kieler Zufriedenheitsstudie. Evaluation und Intervention auf konstruktiver Grundlage*, Münster etc., 2002.
8. KROPE, P., PETERSEN, J. P.: *Generic Acts: The Constructivistic Contribution to the Solution of the Body-And-Mind-Problem*, [in:] *Relationship between Theory and Method in Educational Research*, B. KOŽUH, A. KOZŁOWSKA, A. I. PALERMO (eds.), Buenos Aires/Częstochowa 2003, 111–122.
9. LORENZEN, P.: *Lehrbuch der konstruktiven Wissenschaftstheorie*, Mannheim etc., 1987.
10. MITTELSTRAß, J.: *Unterscheidung*, [in:] *Enzyklopädie Philosophie und Wissenschaftstheorie*, J. MITTELSTRAß (ed.), vol. 4, Stuttgart/Weimar 1996, 429–431.
11. ROST, J.: *Lehrbuch Testtheorie Testkonstruktion*, Bern etc., 1995.
12. ROST, J., SPADA, H.: *Probabilistische Testtheorie*, [in:] *Handbuch der Pädagogischen Diagnostik*, K. J. KLAUER (ed.), vol. 1, Düsseldorf 1978, 59 - 97.
13. RYLE, G.: *The Concept of Mind*, New York 1949.
14. SCHEUCH, E. K., ZEHNPFENNIG, H.: *Skalierungsverfahren in der Sozialforschung*, [in:] *Handbuch der empirischen Sozialforschung. Taschenbuchausgabe*, R. KÖNIG (ed.), vol. 3a, 3rd ed. Stuttgart 1974, 97-203.