Peter Krope, Marta Popko

Uniwersytet im. Humbolda w Kilonii

Educational Measurement between Standardization and Individualization – A Learning Unit –

1. Introduction: About this Learning Unit

This learning unit is about a problem that confronts educational measurement in the Polish educational system. On the one hand standardization is aimed to compare test results. On the other hand it should be possible to refer test results to individual persons so that teachers can help them to improve their skills. While well-proven test procedures are available for comparison, no methods are offered for using the test results to support the individual learner. In school reality the teachers are left in the lurch. This learning unit shows how teachers can use standardized test results to obtain personalized statements on individual pupils.

The basis for this learning unit is an international research project. Members of the project are the Centre for Constructive Educational Science ("Zentrum für Konstruktive Erziehungswissenschaft", ZKE) in Kiel and the Examination Board "Okręgowa Komisja Egzaminacyjna" (OKE) in Cracow.

The learning unit is going to be improved by Polish teachers during a workshop at the diagnostics conference in Lublin in 2006. On the occasion of the workshop the practical exercises of this learning unit mainly include that the Polish teachers put the principles of the learning unit into own words and write down examples of their own school reality with reference to the learning unit examples. The teachers start to work on this learning unit at chapter 3.

If the elaboration is concluded successfully, it is planned to realize an empirical test next year.

307

2. Unexpected, but Inevitable: Theoretical Foundations

The following presents some theory centring on the relation between the particular and the general. This explanation shall help teachers to understand why they cannot rely on tests attending to individual persons. On the basis of this theory teachers shall be able to work independently in future.

2.1 Tests

Tests are scientific procedures to examine personality traits. By means of a test teachers can, for example, examine the summing skills of pupils. In the first place the procedures can be called scientific, if they have a theoretical foundation. Today two theories are mainly used. These are the classic test theory and the probabilistic test theory[1]. In the second place the scientificality of tests is founded on the condition that both the execution of the examination and the description of the test results fulfil methodological standards.

Further on the text centres on the description of test results although this aspect is not independent of other aspects.

There are various ways to describe test results. Firstly it is possible to use qualitative statements, this term being defined as a statement that declares facts to be equal or unequal (KROPE 2000, 51). For example, a teacher says: "Janek's solution of the multiplication problem is correct." The teacher goes on: "Piotrek's solution of the multiplication problem is incorrect." In this example the teacher has described two unequal facts, i.e., Janek's correct solution and Piotrek's incorrect solution.

Secondly test results can be described by using quantitative statements, this term being defined as a statement that represents the multiplicity of equal facts (KROPE 2000, 52). For example, a test consists of three arithmetic problems of which Janek has solved two. Here the teacher describes the multiplicity of equal facts as follows: "Janek's solutions of two arithmetic problems are correct."

If test experts use this second mode of description, they usually do not talk of "quantitative statements". Actually they tend to say that test results are described as "raw points". If, for example, one raw point is assigned to each correctly solved arithmetic problem the teacher may also describe Janek's result by the sentence: "Janek has achieved two points in the arithmetic test."

Test results are not always described as raw points. Sometimes the raw points are used as a basis for conversions by different methods to serve varying intentions. For example, such a conversion leads to the so-called Centile standards, described by the letter C. The scale of the C values virtually extends from -1 to 11 with an average of 5. If Janek's summing skills exactly correspond

---

[1] In the ZKE an alternative test theory concept is founded on the basis of the constructivist theory of educational measurement.

to the class average, his raw point value of 2 can be converted into a C value: $C_{Janek} = 5$. The C scale has 13 marks. If the C values from -1 to +1 are shortened to 1 and the C values from 9 to 11 are shortened to 9, a scale with a total of 9 marks results. It is known as Stanine scale.[2]

Test results are used to compare. According to the intention of a test, you can choose between two basic types. One basic type consists of the comparison of one person's test results to a second person's test result while both are members of the same group. For example, you can ask whether Janek has achieved more or less points than his classmates, or whether his result is exactly the class average. This type of comparison is called "norm-referenced".

The other basic type is founded on exactly described criteria as the learning objective of a school class. This basic type consists of a comparison between the result obtained by a person in a test and the learning objective this person shall achieve. For example, the question is whether Janek (regardless of the results of his classmates) has reached the arithmetic learning objective for his class level, or not. This type of comparison is called "criterion-referenced".

As mentioned above, test results are used to compare. However, test results do not provide any information about the question, why individual persons have passed a test or not. Example 1 explains this:

Example 1

A teacher tests his pupils' fractions skills. His class is experienced in testing. He uses a proven standardized arithmetic test with detailed instructions. The following shows one of the problems:

2: 4 =
a)   0.2
b)   0.4
c)   0.5
d)   0.8

The teacher has carefully practised fractions with his class for a long time. He thinks that the problem is very easy. The problem corresponds to the curriculum and marks out the learning objective that the class shall have reached at this point of time. In fact, the pupils chose 0.5 as the correct solution – with one exception. Ola does not know how to solve the problem. The teacher is mystified. He does not know on which level Ola has understood fractions and how he can help her. On the occasion of a visit to her home he meets Ola who is just handing out dinner to her three brothers and sisters. The pizza service has just delivered two pizzas. The teacher asks if she would know how to divide the pizzas in such a way that each of the four children gets an equally large piece of pizza. "Yes", is the answer of Ola and she divides the two pizzas into four halves.

---

[2] The name "Stanine" is compounded of the abbreviation for "standard" and the word "nine".

Example 1 shows that tests exclusively serve comparison. In the example above the problem makes clear that Ola does not come up to the expectation of the curriculum. She has furthermore obtained the worst result of her class. But the test does not show the reason. Searching for the cause, tests leave the teachers in the lurch. Regarding to this example, we phrase the following principle:

Principle 1

Test results do not give any information about the reason for the measured skills of individual persons.

2.2 Particular and General

Example 1 shows that a test always draws the teacher's attention to something else than the individual person. The norm-reference-measurement asks which average level all pupils of a class have achieved. The criterion-reference-measurement deals with the question whether the learning objectives have been reached that are valid for the grade of the class as a whole. The starting point of the test is something general, whereas, in order to give personal assistance, the teacher starts from an individual person. In Example 1 the teacher wants to know something about Ola. The teacher asks why Ola has not solved the problems and has not shown the expected skill.

In the scientific literature the problem resulting from Example 1 is described as the "problem of the general and the particular". What do the two terms "the general" and "the particular" mean? The ancient philosopher ARISTOTLE explains them i. a. as follows: The general is what belongs to various individual items. Consequently the particular is what does not belong to various individual items (ARISTOTELES Metaphysik 1038 b, 11/12). For example, there are two objects, a car on a parking space and a banana on a breakfast plate. The car and the banana may have the common characteristic that they are yellow. The common characteristic "yellow" is the general by means of which the diversity of the two objects can be reduced to a certain unity. This unity is characterised by the statement: "The car and the banana are yellow." The general ("yellow") is the linking factor. Via the general the diversity of objects is reduced to a unity.

The particular as well as the general can be described linguistically. In order to describe a particular proper names can be used in the first place. If the teacher calls Bogumila Biedronka by name in class, a certain pupil is meant who exists only once. In the second place the individual can be described by a definite description (Kennzeichnung). The expression "definite description" stands for a so-called showing gesture (Zeigehandlung). The teacher points at the pupil by adding expressions as "this one here" or "that one there". The general, however, is described by concepts (Begriffe). "Yellow" is such a concept. To describe the general we use concepts that we may also call "common names" to distinguish them from proper names. Principle 2 refers to the linguistic description of the particular and the general.

310

Principle 2

The particular is linguistically described via proper names or definite descriptions, whereas the general is linguistically described via concepts.

Principle 2 also refers to the differences between the objectives of achievement tests and the objectives of teachers who ask for the reason for the measured skills of individual persons. Tests are about the general. In Example 1 the teacher marks the work on a problem "right" or "wrong". "Right" or "wrong" are concepts that may be valid for the act of not only one individual person but also several persons. Also the quantitative statement explaining the diversity of right (or wrong) solutions describes the general. In contrast to this the teacher wants to learn something especially about Ola. During the handing out of dinner he observes an act that does not occur again in this context. To observe the particular in Ola's act helps him to find a way to help her.

This short linguistic explanation shows that teachers are in the centre of a conflict between the regularity of general relationships and the impossibility to foresee individual acts. There, tests always aim at the general. Like all empirical sciences, tests do not aim at the particular. Example 2 shows the consequences arising, if you insist on describing the particular in educational measurement[3].

Example 2

During a physical education test two experienced teachers evaluate the final examination of their pupils. After a handstand at the parallel bars the pupil Leszek H. has ended the optional exercises with a sequence of movement unexpected by the teachers and unseen until now. Since the dismount was furthermore painful to the pupil, the teachers think that this might have been a mistake and offer Leszek H. a second attempt. If the pupil, however, insists that he intended to end his parallel bar exercises exactly this way and thus wanted to show a unique skill, and if the two teachers accept their pupil's statement, it will be difficult to describe the facts: How may they describe the dismount from the parallel bars while their terminology with concepts like squat vault, straddle and flank vault is inadequate? Finally they describe the dismount as follows: "This dismount from the parallel bars performed by Leszek H. on the occasion of his final examination on March 10, 1994." This means: The two teachers use a definite description. It belongs to just one single person or one single event, but it depends on a situation. It is impossible to make the mode of parallel bars dismount clear to anyone who has not seen Leszek H.'s physical education examination. If you wanted to describe the particular in educational measurement, you would get into similar serious difficulties as Jonathan SWIFT's[1] scholars who, after abolishing language, had to carry everything in big sacks on their backs about which they wanted to talk.

---

[3] Example according to KROPE et al. (2000, 74).

Still today we must consider the problem of the general and the particular as unsolved in principle. As a consequence of this dilemma we can observe empirical scientists phrasing general statements on the one side whereas practitioners are expecting individual statements on the other side. Both sides are facing each other being completely at a loss. The PISA study (cf. BAUMERT et al. 2000) and the consequential discussion are an example for this misunderstanding. In German educational science this unsolved problem has lead to the formation of two opposite positions that confront each other in a rather irreconcilable way. The advocates of the one position think that the main concern of education must be to support the individual. To prove this opinion they refer to the classics ROUSSEAU, PESTALOZZI and HUMBOLDT. This is opposed by the request to give priority to the promotion of capability and willingness to perform. In modern society sworn to achievement principle also school should measure capability and document the measured result.

Two conceptions taken up and advanced in the Centre for Constructive Educational Science at the University of Kiel intend to find a way out of the dilemma shown. The present learning unit is based on them. The first conception is the constructive concept formation focussing on the relationship of everyday language and scientific language (cf. KROPE 1986, KROPE and WOLZE 2005). The second conception discusses the question in which way it may be possible to logically conclude the general from the particular. To this purpose a reference is made to the "Eliminative Induction" method according to v. WRIGHT (1957).

3. Expected: Practical Aids

The preceding has explained the theoretical background of this learning unit. The following will describe the four steps that teachers must follow when the test results are in their hands and they, nevertheless, need information on individual acts.

3.1 Phrase Hypotheses!

Imagine the following situation: Your class had to make an examination. The examination results are at hand. A pupil asks you to explain his result. He says: "I do not understand my result. Could you help me?" However, you do not have any information except the test data.

What may you do? First you write down your suppositions. Wild conjectures are welcome! Be inspired by the social environment of the child! Certainly you know from experience that the young sometimes follow rare paths to find solutions. Often more than one explanation may be true. Collect as many suppositions as possible. The suppositions shall extremely differ from each other. Example 3 shows such a procedure.

312

Example 3

Ms Zielonka teaches arithmetic in pre-school education. A learning objective for the numbers from 1 to 10 is as follows: "Condensing simultaneous reception and counting skills in order to create counting strategies." The teacher wants to know whether the children already have a good command of this ability. She uses a standardized test.

One problem shows a picture of fishes.
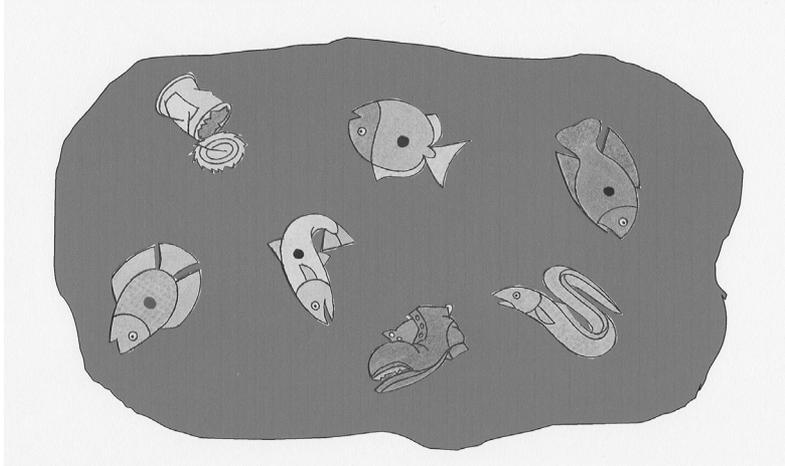
The children are to count how many fishes are in the pond. Jassir[2] answers: "Seven fishes!".

Jassir's parents come to Ms Zielonka. They ask her to explain the test result of their son. Ms Zielonka says: "In the entire test, Jassir has achieved only one point. The class average is five points. I use a Stanine scale. Jassir's result is at the lower end of the scale". "You think only of numbers, aren't you!" interjects Jassir's father. Ms Zielonka shows the picture of the fishes and replies: "For example, Jassir could not solve this problem correctly." The father retorts: "You just make notes of who has found the correct and the incorrect solution. But what does this wrong solution especially mean for my son? How can I help him to improve in counting?" Jassir's father is disappointed. To formulate her answer the teacher has used sentences that are meaningful only within test theory. But Jassir's father cannot see any reference of the comments to his son's counting difficulties.

Ms Zielonka promises to think about it. First she makes some suppositions on her own that must differ very much from each other:

1. Jassir does not know what fishes are.
2. Jassir cannot count yet.
3. This is typical of Jassir.
4. Jassir suffers from visual defects.
5. At home Jassir suffers emotional distress.

For a review you should convert the suppositions into hypotheses. We can talk of a hypothesis, if a statement complies with the following four conditions:

1. A hypothesis is a supposition that goes beyond the individual case. A popular example is: All swans are white.

2. A hypothesis has – at least logically – the shape of a conditional clause. It is written as "if – then" clause. For example, the hypothesis for the swans will be: "If swan, then white".

3. A hypothesis must contain at least two semantically filled concepts. In our example these are "swan" and "white".

4. The conditional clause must be falsifiable. It must be possible to contradict the conditional clause. For the supposition on the swans this condition is easily fulfilled. It is only necessary to find one swan that is black to contradict the hypothesis.

The fourth example shows how the teacher has converted her suppositions into hypotheses.

Example 4

Ms Zielonka writes down four hypotheses about Jassir:

1. If Jassir does not dominate the language, then he does not solve the problem.
2. If Jassir cannot count, then he does not determine the number of things.
3. If Jassir cannot see sufficiently, then he does not solve his problems correctly.
4. If Jassir suffers emotional distress, then he does not work on problems in school.

In Example 4 the four suppositions are converted into hypotheses. The hypotheses refer to one single person and are called "individual case hypotheses". They comply with all conditions of a hypothesis. The condition of general validity is also fulfilled. This is true because no particulars regarding faculty of speech, knowledge of numbers, vision, emotional distress at home and processing of problems are defined. It was not possible to convert the third supposition. The third condition of a hypothesis could not be fulfilled.

The first step to be followed to obtain test results valid for an individual person is summarized in Principle 3.

Principle 3

As a first step phrase as many hypotheses as possible with reference to the test result of your pupil!

314

3.2 Observe! Observe! Observe!

As a second step keep a watch!

It is true that it is your everyday task as a teacher to observe your pupils. You have always been observing. You cannot do without observation. It supplies you with important information about your class. But in the following we will differentiate. On the one hand there is everyday observation as you know it from your school work. On the other hand there is a scientific method called "systematic observation" to distinguish it from everyday observation.

If we talk of systematic observation, we think of gaining experience via all perceptive potentials. Systematic observation is more purposeful than everyday observation. It stands out for the use of instruments that help us to extend the limits of our perceptive abilities.

In contrast to everyday observation carried out more or less arbitrarily, systematic observation requires an observation plan. The plan prescribes,

- what you have to observe,
- what is insignificant for observation,
- how to record the observed.

Example 5 describes a systematic observation.

Example 5

In the preceding, Ms Zielonka has written down several hypotheses about her pupil. On this basis she observes Jassir systematically. By means of a language test she first discovers that Jassir perfectly dominates the language. Then she tests his counting skills. Little Jassir already knows how to count to 7. The ophthalmologist diagnoses a perfect eyesight. When Ms Zielonka makes a visit to Jassir's family, she finally describes an optimal familiar situation by means of a standardized education questionnaire.

After having falsified her four hypotheses she chooses once again the picture showing the fishes. She plays the Count Number Game with Jassir[3]. Jassir is Count Number who loves to count the things of his environment. Standing on the castle wall the Count looks into the pond and counts the fishes. The teacher asks: "Count Jassir Number, how many fishes are in your pond?" "Seven!", answers the child. "Please count again!", asks Ms Zielonka. "Put your finger on each fish that you count!". She observes:

- Jassir says "1" and puts his finger on the fishes with the blue and the red dot,
- Jassir says "2" and puts his finger on the fishes with the green, the blue and the red dot,
- Jassir says "3" and puts his finger on the fish without a dot,
- Jassir says "4" and puts his finger on the fishes with the black and the blue dot,
- Jassir says "5" and puts his finger on the fishes with the green dot and the one without a dot,
- Jassir says "6" and puts his finger on the fishes with the blue dot,
- Jassir says "7" and puts his finger on all fishes one after the other.

The teacher shows the boy a paper with seven empty boxes. She asks Jassir to count aloud. Every time he says a number he shall enter it into an empty box. He shall start at the left taking care not to skip any box. Then a fish hook is located under each box. Now Jassir shall take one fish after the other out of the pond and put it to a hook. Again he shall start at the left without skipping any hook. The numbers to which no fish is belonging are crossed out. Finally Jassir shall read one number after the other that are not crossed out starting from the left side.
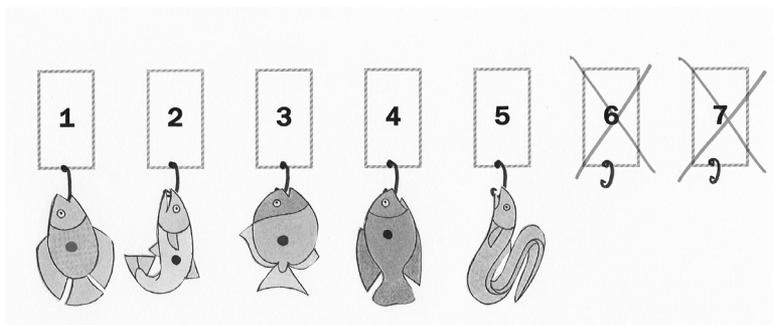
The teacher and Jassir are repeating the Count Number Game with nuggets from the castle well and with roast chicken from the castle kitchen.

The second step on the way to obtain test results valid for an individual person is summarized in Principle 4.

Principle 4
In a second step systematically observe your pupil!

3.3 From Everyday Language to Terminology!

In a third step record your observations!

The language of your record must comply with two conditions. First it must be comprehensible in school. For this purpose it must be founded on everyday life. Second the language must be well-defined and unambiguous. For this purpose you need terms.

These conditions are fulfilled by constructive language design. This method includes the following principles:
• Start language design with an everyday situation without communicative problems.
• Point to examples and counter-examples.
• Introduce explicit rules for the use of the term.

316

The sixth example demonstrates this method.

Example 6

Ms Zielonka ends the Count Number Game. She discusses the result with Jassir.

Teacher: That's it.

Jassir: (laughs)

Teacher: I think that you now know how to count.

Jassir: No jumble any more.

Teacher: That really was a jumble in the beginning.

Jassir: The same as with grandma's hens.

Teacher: ?

Jassir: In grandma's henhouse they are all running around.

Teacher: There the hens are in a jumble. In the beginning, however, it was **you** who made the jumble. That weren't the fishes, either.

Jassir: But no more now.

Teacher: What do we want to call what you have learnt?

Jassir: No counting jumble!"

Teacher: ?

Jassir: (Looks at the picture of Example 5). "I can count with fish hooks!"

Teacher: But Count Number has not only counted with fish hooks. At the well he has used fishing-nets. For each nugget one. And in the castle kitchen! Do you remember? There Count Number has sneaked two roast chicken by means of long spits through the chimney.
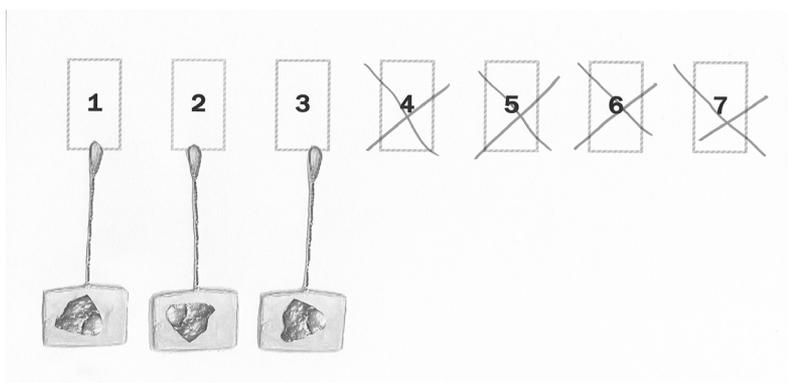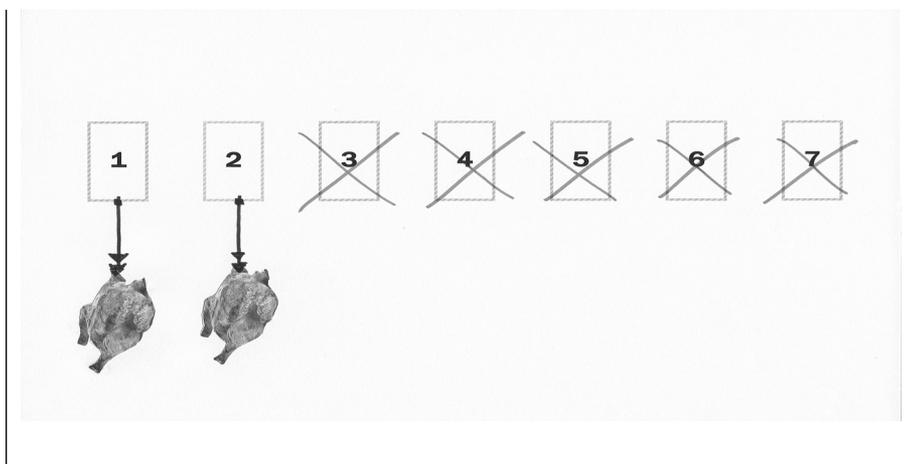

Jassir: ?

Teacher:

First Count Number has used fish hooks. Then fishing-nets. Then spits.

The teacher suggests to call the counting aids "counting signs" (Zählzeichen). Once again she repeats the rules for counting signs that Jassir now knows to use

---

317

The word "counting sign" is a term from arithmetic. The teacher has used a known everyday phrase ("no counting jumble") to define the phrase "to correctly use a counting sign" that has not been familiar to Jassir and his parents. Here, "jumble" is an observation that the teacher has not made before. Thus this is an individual experience that the teacher has observed for just one single pupil until now. In contrast to this, the term "counting sign" is characterized as a general meaning. It describes the acts of all pupils that are successfully introduced into arithmetic.

The third step on the way to obtain test results valid for an individual person is summarized in Principle 5.

Principle 5
Record the result of your observation in a language bridging the gap between everyday life and science!

3.4 Test!

The aim of the method presented in this learning unit consists in making hypotheses that extremely differ from each other and in verifying and rejecting these hypotheses until one hypothesis remains. By recording the observation using terminology, you can then empirically verify this remaining hypothesis. If the hypothesis is considered to be approved, it is called the "most probable hypothesis".

The topic of this learning unit consists in the possibility to refer the test result to the capability of an individual person. According to what we have

318

learnt about the method above, the statement on the personal significance of a test result can be characterized as a hypothesis. This hypothesis can again be tested empirically. And this can be done using the test that has been the starting point of the procedure.

Example 7 shows this process.

Example 7

Ms Zielonka has tested Jassir a second time. For this purpose she has chosen a parallel form of the test she already used in the pre-school class before. Jassir has improved his result considerably. He has achieved five points and is thus in the class average. From this result the teacher concludes that the pupil has achieved the learning objective of the class.

Ms Zielonka tells her colleagues of Jassir's progress in arithmetic. The bad result in the first test she puts down to the fact that in the beginning the boy did not know how to use counting signs. Having doubts about that her colleagues object that she should also have checked other possible causes. These could have been insufficient command of language, as his family has only lived in this country for a short time, insufficient vision, family problems or in general a disability to count.

"I have checked all this", Ms Zielonka replies. Finally the two colleagues agree: "We cannot think of other causes, either. You seem to have found the most probable explanation!".

The fourth step on the way to obtain test results valid for an individual person is summarized in Principle 6.

Principle 6

Repeat the test to verify the hypothesis referring to the personal significance of a test result!

---

4. References

1. ARISTOTELES: Metaphysik. 2 volumes. Translated by H. BONITZ, new edition, with an introduction and comment by H. SEIDL, Greek text edited by W. CHRIST. Greek-German. Hamburg 1980.
2. BAUMERT, J., et al. (Eds.): PISA 2000. Opladen 2001.

319

3. KROPE, P.: Der Beitrag der Prüfungsforschung zur erziehungswissenschaftlichen Theoriebildung. Dargestellt am Problem des Allgemeinen und des Einzelnen. In: Hochschulausbildung, 4/1986, 29 – 41.

4. KROPE, P., with the collaboration of O. ALBERS, B. FRIEDRICH, A. KOHRS, P. PETERSEN, M. VESPER: In acht Schritten zum Modernen Berichtszeugnis. Münster etc. 2000.

5. KROPE, P.: Konstruktive Pädagogische Diagnostik. 2nd ed. Münster etc. 2000.

6. KROPE, P., W. WOLZE: Konstruktive Begriffsbildung. Vom lebensweltlichen Wissen zum wissenschaftlichen Paradigma der Physik. Münster etc. 2005.

7. SWIFT, J.: Gullivers Reisen. Complete edition. Translated from English to German by K. H. HANSEN. München 1990.

8. von WRIGHT, G. H.: The Logical Problem of Induction. 2nd ed.Oxford 1957.

(Footnotes)
1 In "Gullivers Reisen" (New translation to German: SWIFT 1990).
2 The name "Jassir" characterizes a pupil whose family has migrated into the country.
3 Count Number: "Graf Zahl" (English "Count von Count") is a puppet from children's TV program "Sesamstraße".