X Konferencja z cyklu Diagnostyka edukacyjna
20-22 IX 2004

Peter KROPE
Wyższa Szkoła Pedagogiczna w Kilonii

FROM LIFE — WORLDLY EXPERIENCE TO SCIENTIFIC SKILLS: MEASURING EDUCATIONAL OUTCOMES ON A LEARNING PATHWAY

1. THE PROBLEM

This contribution proceeds from a fundamental methodical differentiation. That is the differentiation between norm-reference-tests and criterion-reference-tests.

The purpose of a norm-reference-test is to make a statement about the relative position of a person within a group of people. This orientation on norm-groups is up to now seen as almost the only possible base for measurements in the social science. In that sense GUILFORD writes: "If modern psychology and education have taught anything about measurement, they have amply demonstrated the fact that there are few, if any, absolute measures of human behaviour. The search for absolute measures has given way to emphasis upon the concept of individual differences. The mean of the population has become the reference point, and out of the differences between individuals has come the basis for scale units" (GUILFORD 1965, p. 510/511).

In contrast to this the purpose of a criterion-reference-test is to make a statement about to what extent a person fulfils an exact defined criterion. In an achievement-test the method leads to compare the actually produced achievement of a person with the demanded achievement. This method takes up the suggestions of FLANAGAN (1951), EBEL (1962) and GLASER (1963), who mentioned the term "criterion-referenced measurement".

The criterion-referenced measurement has two functions in the pedagogic. In the pedagogical diagnostic it is the starting point for the probabilistic test-theory (ROST 1995). In the didactics it is the base for learning pathways.

In this contribution the diagnostic and didactic possibilities of the criterion-reference approach are being described against the background of an international, interdisciplinary empirical study (cf. KROPE & WOLZE 2003).

Starting point of the study was the observation that with the imparting and acquisition of science in school and university very often simply the ability to reproduce knowledge and not the ability to apply this knowledge is placed in the foreground. In the study this restriction of learning-results for scientific lessons has been examined. The object was the development of terms in the area of elementary mechanics.

In the study the differentiation between common language and scientific language has found special attention. It has been assumed that the difference between common language and scientific language presents one of the relevant barriers for teaching and

learning science. In the study an instruction unit was developed. Its strategy was to initiate a learning process by partly defining expressions of the common language.

The strategy based upon the development of a learning pathway leading from everyday life to science. The study started with the assumption that students who are being questioned about a physical phenomenon would generally choose their answer from one of the three stages of a learning pathway. Before the start of the physics instruction they would give common language answers — stage 1. After the start of the instruction they would show physical knowledge — stage 2. After a successful completion of the instruction they could apply the knowledge — stage 3. An example for the three stages of this learning pathway is shown in table 1.

Table 1: Answer to the question why a ball rebounds after it has been thrown vertical into a wall. The common language answer at stage 1 derives from a student of a secondary school in Gdansk (cf. Krope & Wolze 2003). The answers at stage 2 and 3 represent the partly defined expressions of the common language answer. Yet with the answer at stage 2 a knowledge has been reproduced that has not been applied to the specific situation.

Stages of a learning pathway		Answers
1	Common language answer to a question about a physical fact	That is a physical law. After rebounding of the wall the ball at first does not know where to go. But because it wants its old shape back it pushes itself from the wall and therefore rebounds. The thrower knows the law and uses it to be able to catch the ball again.
2	Ability to reproduce physical knowledge	The term "exertion of strength" describes a relation.
3	Ability to apply physical knowledge	The wall exerts strengths towards the ball.

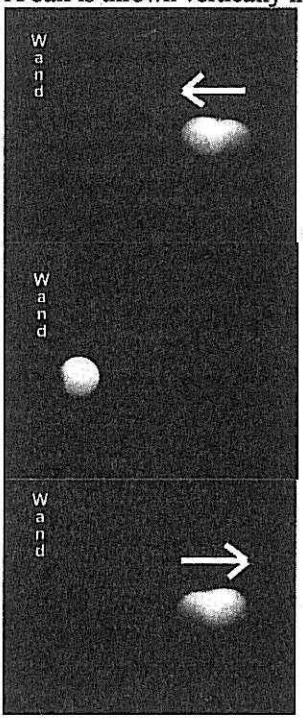
In the international interdisciplinary study an instruction was developed that is based on three stages of a learning pathway. The hypothesis reads that the students usually take up these three stages one after another — from stage 1 through stage 2 to stage 3. In the following the testing of this hypothesis is represented.

2. FROM EVERYDAY LIFE TO SCIENCE

To analyse the success of the instruction a test with a total of 13 items of the multiple choice type was constructed. The items 1 to 9 measure the ability of application while the items 10 to 13 measure transfer. Every item consists of the question or respectively of the problem which are being illustrated by photos or drawings. In the standard version every item contains four alternatives of which one is regarded as the right answer. Every item contains two alternatives formulated in physical terminology and two alternatives formulated in common language (compare figure 1). The simultaneous use of common language and scientific language was determined by the method: Instead of the probabilistic test-theory the classical test-theory (cf. GULLIKSEN 1950) has been used.

Item 1

A ball is thrown vertically into a wall. It rebounds. Why does it rebound?



The wall pushes the ball away.
The wall exerts strengths towards the ball.
The ball exerts strengths towards the wall.
The ball pushes itself off the wall.

Figure 1: The first item of the test constructed for the international interdisciplinary study.

The hypothesis was tested in two steps. First step 1.

Transformation from every day thinking to physical thinking was expected of the students. Therefore in the first step it was analysed if after the instruction the number of common language solutions had actually decreased. The search had been carried out with test-groups and control-groups in a pre-test-post-test experiment. Accordingly the hypothesis could be formulated (compare table 2).

Table 2.: The hypothesis about the number of common linguistic solutions

Group	Pre-test expected are:	Intervention	Post-test expected are:
Test-group	Common language solutions	Physics instruction	Less common linguistic and more terminological correct solutions than in the pre-test
Control-group	Common language solutions	No instruction	Just as many common linguistic solutions as in the pre-test

To test the hypothesis students from Gorzów Wielkopolski in Poland had been chosen as representative for the whole study.

The results are shown in table 3. The number of test-persons is 33 in the test-group and 32 in the control-group. Differences in table 3 to these figures are explained due to missing data.

With the t-test for independent samples (two-sided question) the differences between pre-test and post-test results have been tested for significance. As expected the difference in the items 1 to 9 in the test-group is significant but not in the control-group. In the items 10 to 13 no significant differences have been observed in the test- and control-group. This result as well turned out according to the expectations. These items measure a transfer which in the experiment was neither prepared nor occurred.

Table 3: Number of life-worldly solutions. For item 10 to 13 only the answers of the respective second part of the questions have been evaluated because only they are relevant to the difference between common language and terminological formulated answers. Explanations: $\overline{X}$ — arithmetic mean, s — standard deviation, n — number of test-persons. The double star (**) indicates a highly significant difference (1 % error probability) between pre-test and post-test results.

Group	Items	Pre-test			Post-test		
		$\overline{X}$	s	n	$\overline{X}$	s	N
Test-group	1 — 9	4,94	1,13	32	2,81**	1,40	32
	10b —13b	1,19	0,82	32	1,12	0,99	33
Control group	1 —9	5,16	1,35	32	4,78	1,10	32
	10b — 13b	1,34	0,97	32	1,37	0,85	30

The test-persons in Gorzów Wielkopolski used in the post-test after the instruction in the item-group 1 — 9 as expected significantly less common language solutions and more terminologically formulated solutions as in the pre-test. It is assumed that this result is valid for all groups within the study. From this finding it is concluded that the

instruction really resulted in a transformation from common linguistic to scientific thinking.

3. FROM KNOWLEDGE TO APPLICATION

In step 2 it is tested if the instruction really taught the ability to apply physical terms — or if in reality the alleged ability of application is only concerned with reproductive abilities.

The common language expression "to apply a term" has been defined as "to predicate a term upon a state of affairs (Sachverhalt)"; "terminological knowledge" has been defined as the ability to differentiate between common language expressions and physical terms. For the following the term "application" is described with the help of a classification of educational goals, developed by BLOOM and collaborators (BLOOM et al. 1971). In the taxonomy the cognitive domain is divided into six major classes. They are knowledge, comprehension, application, analysis, synthesis and evaluation (defined as the making of judgements of methods etc). BLOOM and his collaborators assume that the goals of a class are based on the goals of the preceding class. Corresponding to this assumption persons whose achievements are to be classified on step 3 (application), show achievements on step 2 (comprehension) and step 1 (knowledge). Thereby "knowledge" is among other things described as "terminological knowledge".

Under the assumption of a hierarchical order learners achieve the ability to apply physical terms via the knowledge of physical terms. As the results of the first step of the hypothesis-testing have shown the starting point of the learners is in general the use of common language expressions instead of physical terms. Thus it can be assumed that learners in the learning unit of the study take three steps of a learning pathway. Step 1 is marked by common language solutions, step 2 by the mastery of the physical terminology and step 3 by the ability to apply the physical terms.

An experiment that has been carried out at the university of Kiel proved the correctness of this assumption. For the experiment the film "Term-building. From the living-world to physics", developed for the study, was used. The film demonstrates the learning pathway from life-worldly to scientific paradigm of the physics by showing how terms are being developed, learned and applied.

To analyse the learning-results a two-piece test was developed. Part 1 contains four items to test knowledge (see figure 2). Part 2 contains four items of the test constructed for the whole international interdisciplinary study to test application (see figure 1).

Item 2
The term "to exert strength"
describes
a relation
an intention
an excess
an extension

Figure 2: Item number two to test the knowledge of terms in the area of elementary mechanics. The correct solution is a.

Test-persons were the students of a seminar on educational measurement. They were divided by random into 4 groups:

1. Group 1 did not receive instructions.
2. Group 2 received instructions about terms (corresponding to part one of the film "term-building").
3. Group 3 received instructions about how to apply terms (corresponding to part two of the film).
4. Group 4 received instructions throughout the whole film (part one and two).

The size of the group was between 11 and 13 students.

The hypothesis reads that the test-persons take three steps of a hierarchically set up learning pathway and that the learning results can be differentiated as follows:

1. Group 1 shows both less knowledge in regard to the term-building as well as less abilities to its application.
2. Group 2 shows more knowledge in the term-building and less abilities to its application.
3. Group 3 shows less knowledge in regard to the term-building as well as less abilities to its application.
4. Group 4 shows both more knowledge in the term-building as well as more abilities to its application.

The expressions "more" and "less" in the preceding enumeration relate to the differences of the learning results which exist between the groups. Group 1 and group 4 are considered to be reference-groups.

What are the findings? First of all the differences within the learning-groups. The results calculated with the t-test (two-sided) for independent samples are shown in table 4. The differences between pre-test and post-test in the learning-groups 2 and 4 are highly significant with an error probability of 1 %. The differences in the learning-group 3 are in the item-groups 5 to 8 significant for 5 %.

Table 4: Learning-increases (specified as arithmetic means) within the learning-groups 1 to 4, items 1-4 and items 5-8. The shadings refer to significant differences. Explanations: ** — highly significant differences (1 % error probability), * — significant difference (5 % error probability).

Group	1		2		3		4	
Items	1 - 4	5 - 8	1 - 4	5 - 8	1 - 4	5 - 8	1 - 4	5 - 8
Pre —test	2,45	1,82	2,00	1,58	2,31	1,54	2,33	1,42
Post — test	2,55	1,73	** 3,75	** 2,92	3,08	* 2,69	** 3,83	** 3,42

Now about the learning-increases between the learning-groups. To eliminate interpersonal variations interpersonal differences form the basis for the computation. The significances are represented in table 5. The towards zero moving learning-increases of group 1 differ highly significantly from the ones of all the other groups. In group 2

there is in the items 1-4 a significant difference to group 3 and in the items 5-8 a significant difference to learning-group 4. Group 3 shows in the items 5-8 a significantly less learning-increase than group 4.

Table 5: Learning-increases between the learning-groups 1 to 4, items 1-4 and 5-8. Explanations: n.s. — no significant difference, ** — highly significant difference (1% error probability), * — significant difference (5 % error probability).

	Group 2		Group 3		Group 4	
	Items 1-4	Items 5-8	Items 1-4	Items 5-8	Items 1-4	Items 5-8
Group 1	**	**	**	**	**	**
Group 2			*	n.s.	n.s.	*
Group 3	*	n.s.			n.s.	*

For the moment unexpected is the significant learning-increase in the learning-groups 2 and 3 with the item-group 5-8 (compare table 4). Nevertheless the result in learning-group 2 does not necessarily speak against the hypothesis. The difference to the reference-group 4 is significant. The weak difference can be explained by the fact that the film of the instruction does not strictly separate between the introduction of the terminology in part 1 and its application in part 2 as it would have been desirable for the experiment and that the test in both item-groups for the purpose of the experiments is too easy and a ceiling-effect took place. Also the result of item-group 5-8 in learning-group 3 turns out according to the hypothesis when it is compared with the reference-group. It differs significantly from the corresponding result of the learning-group 4 (compare table 5).

4. RESULTS

On the basis of the two reported experiments the assumption is supported that in the study cognitive achievements are being described which belong to three steps of a learning pathway. At the first step, before the beginning of the instruction, the learners use common language expressions to describe physical state of affairs. At the second step, after the beginning of the instruction, the learners show physical knowledge which they can reproduce. At step three, after a successful completion of the instruction, the learners are capable to apply the physical knowledge.

The cognitive achievements are hierarchically arranged on the three steps of the learning pathway. Namely persons who have achieved step three are also in command of step two. But in reverse a person with only physical knowledge — step two — is not at the same time able to apply this knowledge — step three. In view of the limited research methods these relations do not have absolute validity. Rather they are only valid with a certain probability. According to this a person who takes up a higher step on the learning pathway will answer a physical question with greater probability physically correct than a person who takes up a lower step.

What is the gain of the experiments for the pedagogic diagnostic? In principle a rank-row as described through the learning pathways stays maintained for every agreeable random sample of persons who belong to the population for whom the

learning pathway is valid. Learning pathways are a diagnostic instrument with results that do not depend upon random samples.

And what is the gain for the didactics? The differentiation between pre-scientific and scientific language raises a general problem for the teaching and learning. The problem is that learners usually can not overcome the outlined gap. Students who are introduced to a science are generally more likely to be capable to reproduce the scientific knowledge by for example repeating formulas. Skills that go beyond reproducing knowledge can be less often be proven. To analyse this problem and its solutions a learning pathway leading from life-worldly to scientific paradigm of physics has been developed together with a successful instruction technique.

REFERENCES

BLOOM, B. S., et al. (Eds.), (1971), *Taxonomy of Educational Objectives*. Handbook I: Cognitive Domain. New York, sixteenth printing.

EBEL, R. L., (1962), *Content Standard Test Scores* In: Educational and Psychological Measurement., pp.15 — 25.

FLANAGAN, J. C., (1951), *Units, Scores, and Norms* In: E. T. LINDQUIST (Ed.): Educational Measurement. Washington, pp. 695 — 763.

GLASER, R., (1963), *Instructional Technology and the Measurement of Learning Outcomes* In: American Psychologist., pp. 519 — 521.

GUILFORD, J. P., (1965), *Fundamental Statistics in Psychology and Education*. 4. edn. New York etc..

GULLIKSEN, H., (1950), *Theory of Mental Tests*. New York.

KROPE, P., WOLZE, W., (2003), *Ordinary Language and Scientific Language*. Test Validity from a Constructivistic Point of View. In: Wyższa Szkoła Humanistyczno-Ekonomiczna w Łodzi et al. (Ed.): Trafność pomiaru jako podstawa obiektywizacji egzaminów szkolnych. Lodz, pp. 51 — 60.

ROST J., (1995), *Lehrbuch Testtheorie Testkonstruktion*. Bern etc.