Using A Multiple Evidence Model (MUEMO)
for Testing the Effectiveness of Educational Interventions

Hermann Astleitner, Josef Kriegseisen, and Franz Riffert
University of Salzburg

Paper presented at:
European Conference of Educational Research (ECER)
of the EERA (European Educational Research Association),
Vienna, 28.-30.9.2009

Using A Multiple Evidence Model (MUEMO)

for Testing the Effectiveness of Educational Interventions

1. A Multiple Evidence Model (MUEMO) for Educational Interventions

Investing large amounts of resources within educational systems has stimulated the development of scientific approaches dealing with the validation and evaluation of educational interventions (e.g., Campbell & Russo, 1999). Such approaches are empirical resp. evidence-based in nature and although subject of sophisticated methodological concepts, have not had yet significant impact on educational practice and related research. There are two main reasons for this situation. First, a lacking emphasis on interventions reduces the probability of the implementation of methodological concepts which deal with criteria for effectiveness: Many prominent international programs for improving educational practice are orientated on describing persons or situations, but not on systematically influencing conditions and testing intervention effects (see, for example, the PISA(Programme for International Student Assessment)-activities). Second, many empirical educational effectiveness studies have traditionally tested post-treatment-effects, but have not discussed given results based on other aspects of intervention effectiveness (e.g., Prenzel & Allolio-Näcke, 2006).

As far as there are research-guided educational interventions programs, then they are - in most cases - not based on comprehensive models of different forms of effectiveness (e.g., Astleitner, in print).

Such models could help to achieve a more reliable and valid picture of the effectiveness of educational interventions in question (see, for example, related issues in the field of psychological interventions; Hager, Patry, & Brezing, 2000). In addition, such models could stimulate the development of standards for the design of interventions (see, for example, the concept of "design research"; Kelly, Lesh, & Baek, 2008). It is the major goal of this paper to present such a recently developed model of evidence-based educational effectiveness and to show how it can be used within educational intervention programs.

Include Figure 1 about here!

Within Figure 1, a "multiple evidence model (MUEMO)" for evaluating the effectiveness of educational interventions is depicted. This model is related to some ideas from Hager and Hasselhorn (2000, p. 81) and from Middleton, Gorard, Taylor, and Bannan-Ritland (2008, p. 32), but it deals only with "effectiveness"; it does not focus on ethical norms, traditional scientific criteria (e.g., internal or external validity and design sensitivity), or efficiency standards for educational interventions, but assumes them to be fulfilled (e.g., Campbell & Russo, 1999; Lipsey, 1990). According to this model, "effectiveness" of educational interventions is given, when goal competencies are persistently increased by interventions (e.g., instructions, trainings, or counselling) in comparison to a non-intervention situation. Such non-intervention situations can be found in contexts given before the intervention ("Pre-Condition") or in contexts without interventions ("No-Conditions"). Goal competencies before interventions are measured as pretests; goal competencies in contexts without interventions are given in control groups. Comparisons with pretests and control groups represent a standard procedure in traditional experimental design and related validity concepts (e.g., Lissitz & Samuelsen, 2007).

In order to be able to evaluate the effectiveness of educational interventions comprehensively, effectiveness has to be related to "time", to "contexts", and to "different-intervention-functioning".

"Effectiveness in time". This aspect points to the necessity that intervention effects should be tested during intervention (A), at the end of the intervention (B), but also after the intervention on the long run (C). A three-times measurement allows not only for testing effectiveness at the end of an intervention, but also for finding some data about the process or development of effectiveness and its duration over time. To observe the development of effectiveness represents an idea that is based on methodological concepts of traditional single case, process-orientated, and also longitudinal research (e.g., Julius, Schlosser, & Goetze, 2000; Menard, 2008).

"Effectiveness in contexts". This aspect aims at testing whether intervention effects show improvements in different contexts. Effects should be tested within similar (relevant) contexts (D), within nested contexts (E), and within daily life situations (F). Similar contexts concern situations in which target competencies as established by intervention (e.g., reasoning skills) represent an important source of improving problem solving (e.g., finding solutions for complex biological problems). Nested contexts are more distant to goal competencies. In these contexts, goal competencies are important, but only in combination with other skills and competences at different organizational levels (e.g., for improving the overall performance of a school program, students should improve their reasoning skills, but also teachers should improve their daily instruction and so on). Daily life situations represent a type of context that lies clearly outside educational situations (e.g., applying reasoning skills to reduce the price for a product during selling negotiations). The aspect of being effective in different more or

less similar contexts brought into focus traditionally, for example, by “action research”; concepts associated with it are “ecological validity”, “design research”, “technological theories”, “pragmatic research” or approaches trying to bridge the “theory-practice-gap” (e.g., Patry et al., 2005; Stachowiak, 1995).

“Different-Intervention-Functions”. In addition to main effects of educational interventions side-effects (G) and interaction effects with other variables (H) should be tested too. Knowledge about these types of effects helps to calibrate and adapt interventions. Side-effects concern significant but unintended outcomes of educational interventions. The idea to consider side-effects originally comes from “utility research” and “evaluational research” and approaches which try to connect medical or ethical standards to the field of education (e.g., Bortz & Döring, 2006). Interaction effects are given when the educational intervention only show effects when certain situations and/or personality characteristics are given (e.g., an instructional program increases the achievements of students with high prior knowledge, but decreases the achievements of students with low prior knowledge). The focus on interactions is traditionally anchored with concepts like “Aptitude-Treatment-Interaction” or with complex methods for data analysis (e.g., Caspi & Bell, 2004; Corno et al., 2002; Jaccard, Turrisi, & Wan, 1990).

Within the MUEMO, it is assumed that all these tests (from A to H) should be undertaken in a hierarchically organized sequence: After A to C have been realized, then D to F should be done; finally G and H should be undertaken. A to C is dealing with effectiveness. If effectiveness of an intervention is given, then this effectiveness can be tested under different circumstances (D to F). If an intervention is effective and has shown effectiveness under different circumstances, it can be calibrated (G to H). Of course, these tests can be done within one study, more or less, at the same time.

In addition, all tests should be based on pre- and no-condition comparisons. Overall, the MUEMO combines and integrates different lines of educational effectiveness concepts, mainly not to stimulate methodological concept development, but to improve the quality of tests related to educational interventions. Within this paper, we tried to connect MUEMO to an educational intervention based on a certain instructional method, i.e., “Learning cycles”.

2. An Example of an Intervention Study: Learning Cycles and Science Instruction

Kriegseisen and Riffert (in preparation; see also Hascher, Hagenauer, Kriegseisen & Riffert, 2009) have done an educational intervention study in which “learning cycles” were implemented for improving learning in Science instruction.

The “learning cycles” approach represent an instructional method, originally developed by Whitehead (1928/1967) and Dewey (1938/1997), which is based on sequence of different phases of learning (i.e., romance/exploration, precision/concept introduction, generalization/application) in combination with specific instructional methods (e.g., dealing with life problems, asking questions, developing explanations, testing hypotheses, ... etc.) and related behaviour of students and teachers. It has been developed by Karplus (see Karplus & Thier 1967) by combining it with Piaget’s Genetic Structuralism and modified by Lawson, also including an observation instrument; for an overview on this instructional method, see Riffert (2005, 2008).

The educational intervention study was based on a quasi-experimental pre-posttest-design with a control group. The experimental group consisted of 25 8th-grade-students (13 girls and 12 boys) from one class of a primary school (Hauptschule). The control group consisted of 24 8th-grade-students (11 girls and 13 boys) from another class. Within the experimental group of this intervention study, Learning Cycles were used within Chemistry courses (3 Learning Cycles [about water, air, acids and bases] from November until January) and Physics courses (also 3 Learning Cycles [about electrical devices, light and colours, and radioactivity] from the end of February until the end of June). In the Control group, no Learning Cycles have been implemented; all other elements of the courses were identical. Instruction based on Learning Cycles overall took about 32 hours within the experimental group.

For measuring the dependent variable, i.e., the cognitive development according to Piaget’s stage theory, Science Reasoning Tasks (SRT) (Wylam & Shayer, 1980) were used before (2 tasks=2 pretests within an interval of 14 days), during (1 task), and after the end of intervention (2 tasks=2 posttests with an interval of 7 days). Before starting the intervention, Intelligence of all students was measured using the Culture Fair Intelligence Test-Scale 2 (CFT 20) from Weiß (1987). Reliability and validity of the SRT and CFT have been proved successfully repeatedly in the past, so they have not been tested for the given study (see Riffert, 2008; Weiß, 1987).

Before starting the intervention and immediately after the intervention parts of a measurement tool for students’s interest conceptions concerning physics and chemistry developed by Hoffmann, Häußler, and Lehrke (1998) has been used. Only one part of the collected data, i.e., the subscale of Subject Interest (with 20 items and good reliability, see Table 3, Part A), has been analysed within this paper. In addition, at the end of intervention, grades were collected. Within this study, grades as indicators of overall achievement in the subject concerning Chemistry, Physics, Mathematics, and German Language were compared.

3. Testing the Effectiveness of a Learning Cycles Intervention using MUEMO

Data from the study from Kriegseisen and Riffert (in preparation) can be re-analysed in order to find correspondence with MUEMO. This re-analysis aims at illustrating how MUEMO can be used for the designing and the evaluating effectiveness studies.

3.1. Testing Effectiveness in Time and in Comparison to a Control Group

First, we tested - according to MUEMO - the effectiveness of the learning cycles intervention in respect to time and in comparison to a non-intervention control group. Such tests are more or less standard in educational effectiveness studies (see, for example, Zohar & Peled, 2008).

To determine the effects of the learning cycles intervention a repeated ANOVA (with SPSS 16.0 for Mac) was performed with time as the within-subjects factor and treatment (experimental vs. control) as between-subjects factor (Bortz, 1993, p. 313). N differs in comparison to the original number of subjects because of missing data (overall n=34 instead of 52). The ANOVA revealed a main effect of time, $F(4, 32)=6.59$, $p<0.001$, partial $ETA^2=0.17$, indicating that children changed their Science reasoning skills over the course of instructional sessions, and a main effect of treatment, $F(1, 32)=8.95$, $p<0.001$, partial $ETA^2=0.22$, indicating that students in the Experimental group outperformed students in the Control group. The interaction between time and treatment was not significant, $F(4, 32)<1$. However, to examine the differences between the Experimental and Control groups prior to the intervention we compared their mean scores in the first and second pretest (see Table 1). *T-tests* for independent groups showed that the differences between the two groups were nonsignificant, $t(45)=-1.92$, $p>0.05$ in the first pretest, but significant, $t(41)=-4.02$, $p<0.001$, in the second pretest. This result indicates that the differences between Experimental and Control group are due to significant differences prior to interventions. Therefore, no conclusions can be made about the effectiveness of the intervention, except that the intervention did not reduce Science reasoning skills in comparison to a Control group. Based on MUEMO, further comparisons of *mean* scores between time and groups could be made in order to gain a comprehensive picture of the intervention effects. However, as intervention effects could not be attributed to the intervention itself, no further comparisons are performed.

Include Table 1 about here!

For the purpose of our paper, this result is not important because we used the learning cycles study in order to show how the MUEMO could be applied. Our major goal was not to develop an effective instructional intervention based on learning cycles, but to illustrate by an example how MUEMO could stimulate educational effectiveness testing.

3.2. Testing Effectiveness in Contexts

According to MUEMO, effects of educational interventions have also to be tested in contexts that are relevant for the intended goal areas. These contexts can be highly similar to the given situation of the intervention, they can have some nested relationship, or they can represent real life situations.

In our study, we have tested the effects of learning cycles on Science reasoning tasks. These skills are used in Science classes dealing with Physics and Chemistry. At the end of the school year, we have collected grades in Physics and Chemistry what represent similar contexts. We also have measured grades in Mathematics and German language what should measure effectiveness in nested contexts. Science reasoning skills concern subjects that are closely related to science, like Physics and Chemistry. However, these skills also represent some general thinking, hypothesis testing or argumentation skills that are also relevant in other subjects in which thinking and language skills are important, like Mathematics and Language education.

Mann-Whitney-U-Tests were performed to compare the grades within similar and nested contexts between Experimental and Control group (Bortz, 1993, p. 141). Table 2 shows *middle ranks*, *Z*-, and *p*-values for all four grades indicating that they are no significant differences between Experimental and Control group (all $p>0.05$).

Include Table 2 about here!

Using MUEMO, it can be stated that the learning cycle intervention has not shown any effectiveness in goal relevant contexts when comparing an intervention (condition) with a non-intervention situation (no condition). There have been no comparisons with a pre-condition as suggested by the MUEMO, because of missing data. Data about effects in daily life situations were not collected. Such data could be found in solving daily problems in which reasoning skills are important, like, for example, proofing statements, testing products, or confronting assumptions about situations with real experiences.

3.3. Different-Intervention-Functioning

In order to test side-effects according to MUEMO, the effects of learning cycles on subject interest have been tested. Subject interest represents a non-cognitive variable that has repeatedly been found to be relevant for Science learning (e.g., Kessels & Hannover, 2004). *T-Tests for independent samples* have been used to identify significant effects of the experimental condition on subject interest (see Table 3, part A). Results show that students in the Experimental group did not differ significantly from students in the Control group in both pre- and posttests of subject interest (Pretest: $t(43)$=1.82, p=0.076; Posttest: $t(41)$=0.49, p=0.63). Because of these insignificant results, no further analyses based on statistical tests have been made. However, an inspection of *mean* scores indicates that subjects' interest has been stable in the Experimental group (from pre- to posttest: 64.00 vs. 65.32), but has been reduced in the Control group (58.46 vs. 67.05) (a larger *mean* score indicates lower interest).

Include Table 3 about here!

For identifying different-intervention-functioning according to the MUEMO, Aptitude-Treatment-Interaction-tests have been performed (see Table 4). An *ANCOVA* (Bortz, 1993, p. 333) with interactions of the treatment-variable and Intelligence as co-variable has been used calculated to search for indications of different intervention effects on the first posttest (see Table 3, part B). The first posttest on Science Reasoning was used, because it was expected that for this measurement effectiveness should be increased.

Results show a significant Group x Intelligence-Interaction ($F(1, 41)=4.01$, $p<0.05$, partial $ETA^2=0.09$). This interaction indicates that the intervention of learning cycles had different effects based on the Intelligence of students. Students did not differ in Intelligence between Experimental and Control group ($t(46)=-0.46$, $p=0.65$). However, correlations between Intelligence and Science Reasoning are different in Experimental group ($r=0.56$, $p=0.005$) in comparison to the Control group ($r=0.21$, $p=0.369$). Using a *test for correlation differences* based on *Fishers-Z-Transformation* (Bortz, 1993, p. 203) shows that both correlations are different ($z=-1.31$, $p<0.10$). These results indicate - as expected - that students with high Intelligence gained more cognitive skills in the Experimental group in comparison to students with low Intelligence. Within the Control group, the Intelligence of students did not lead to different gains in learning. Therefore, there are some indications that the learning cycles intervention was helpful for learning, but - in tendency - especially when students had high Intelligence.

4. Discussions

Within this paper, a multiple evidence model for evaluating the effectiveness of educational interventions has been presented and applied to an intervention study dealing with the learning cycle approach in Science education. This model integrates different lines of educational effectiveness concepts that are focussing on time, contexts, and person-situation-interactions. It represents holistic standards for testing interventions effects, which can be used to improve the quality of interventions in practical educational settings.

A holistic approach for educational effectiveness testing and exploring. This way of effectiveness testing is holistic in a sense that it covers many possible proximal and distal effects of educational interventions. It therefore has a function for hypothesis testing (about expected effectiveness), but also represents stimulation for exploring hypothesis (about so far unknown or unexpected effectiveness that was discovered by using MUEMO). Therefore, MUEMO should not only have positive impact on applied, but also on basic research, because it can deliver effect patterns that must be explained by innovative theoretical concepts. Methods of how to build theories out of different effectiveness patterns can be found, for example, in Miles and Huberman (1984, p. 215).

A testing model for educational development. Especially, within the field of applied research in education, the research-driven development of products (courses, trainings, instructional methods, etc.) represents an important task. MUEMO can be implemented on different stages of such a development process in order to get data about the quality of a product. Such data can be used continuously throughout a whole process and contribute significantly to quality assurance. MUEMO can be used in combination with approaches that deal with the application of scientific research results to educational practice of about how knowledge from Educational Science can be transformed into educational practice, especially, in the field of data-based interventions. For example, Rolff (2008, p. 151) has postulated a model of an "action-circle" with theories, applied theories, and data-based interventions as a core of evidence-based educational development and quality assurance.

Implementing a decision mechanism. Within a next step in the development of MUEMO, decision mechanisms have to be established that help to make modifications of educational interventions based on different patterns of effectiveness. For example, it is an open question how designers of educational interventions should modify their interventions when the interventions have positive effects in time, but negative ones in contexts, or vice versa. Such decision mechanisms have to consider a) strong vs. weak effects, b) sizes of effects found in past, c) the importance resp. urgency of a problem that is related to goal competences, d) the possibility of improving the effectiveness by combining interventions, and other factors.

References

Astleitner, H. (in print). Methodische Rahmenbedingungen zur Entdeckung der Wirksamkeit von pädagogischen Interventionen [Methodological conditions for identifying the effectiveness of educational interventions]. In T. Hascher & B. Schmitz (Hrsg.), *Handbuch Pädagogische Interventionsforschung*. Weinheim: Juventa.

Bortz, J. (1993). *Statistik für Sozialwissenschaftler* [Statistics for social researcher]. Berlin: Springer.

Bortz, J. & Döring, N. (2006). Forschungsmethoden und Evaluation for Human- und Sozialwissenschaftler [Research methods and evaluation for human- and social researcher] (4. Aufl.). Berlin: Springer.

Campbell, D. T. & Russo, J. M. (1999). *Social experimentation*. Thousand Oaks, CA: Sage.

Caspi, O. & Bell, I. R. (2004). One size does not fit all: Aptitude x Treatment Interaction (ATI) as a conceptual framework for complementary and alternative medicine outcome research. The Journal of Alternative and Complementary Medicine, 10, 580-586.

Corno, L. et al. (2002). *Remaking the concept of Aptitude: Extending the legacy of Richard E. Snow*. Mahwah, NJ: Erlbaum.

Dewey, J. (1938/1997). *Experience and education*. New York: Macmillan.

Hager, W. & Hasselhorn, M. (2000). Psychologische Interventionsmaßnahmen: Was sollen sie bewirken können? [Psychological interventions: What should they be able to achieve?] In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Handbuch Evaluation psychologischer Interventionsmaßnahmen* (S. 41-85). Bern: Huber.

Hager, W., Patry, J.-L., & Brezing, H. (Hrsg.). (2000). *Handbuch Evaluation psychologischer Interventionsmaßnahmen* [Handbook on the evaluation of psychological interventions]. Bern: Huber.

Hascher, T., Hagenauer, G., Kriegseisen, J. & Riffert, F. (2009). Lernzirkel im Physik- und Chemieunterricht in der Sekundarstufe 1. Submitted.

Hoffmann, L., Häußler, P., & Lehrke, M. (1998). *Die IPN-Interessensstudie Physik* [The IPN-Interest-Study in Physics]. Kiel: Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.

Julius, H., R. W. Schlosser, & Goetze, H. (2000). Kontrollierte Einzelfallstudien [Controlled single case studies]. Göttingen: Hogrefe.

Karplus, R. & Thier, H. D. (1967). *A new look at elementary school science*. Chicago: Rand McNally.

Kelly, A. E., Lesh, R. A., & Beak, J. Y. (Eds.). (2008). *Handbook of design research methods in education. Innovations in science, technology, engineering, and mathematics learning and teaching*. New York, NY: Routledge.

Kessels, U. & Hannover, B. (2004). Entwicklung schulischer Interessen als Identitätsregulation [The development of school-based interest as identity regulation]. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule* (S. 398-412). Münster: Waxmann.

Kriegseisen, J. & Riffert, F. (in preparation). *Learning cycles im naturwissenschaftlichen Unterricht* [Learning cycles in Science instruction]. Department of Educational Research, University of Salzburg.

Lipsey, M. R. (1990). *Design sensitivity. Statistical power for experimental research.* Newbury Park, CA: Sage.

Lissitz, R. W. & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher, 36,* 437-448.

Menard, S. (Ed.).(2008). *Handbook of longitudinal research: Design, measurement, and analysis.* Amsterdam: Elsevier.

Middleton, J., Gorard, S., Taylor, C. & Bannan-Ritland, B. (2008). The „compleat“ design experiment. From soup to nuts. In A. E. Kelly, R. A. Lesh, & J. Y. Beak (Eds.), *Handbook of design research methods in education. Innovations in science, technology, engineering, and mathematics learning and teaching* (pp. 21-46). New York, NY: Routledge.

Miles, M. B. & Huberman, A. M. (1984). *Qualitative data analysis. A sourcebook of new methods*. Beverly Hills, CA: Sage.

Patry, J.-L. et al. (2005). Feedback-Schlaufen: Von der Rückmeldung in der Unterrichtssituation bis zu Forschungsauftrag - eine Herausforderung für die Lehrerbildung [Feedback-loops: From feedback to instructional situations up to a research order - a challenge for teacher education]. In C. Giordano & J.-L. Patry (Hrsg.), *Theorie und Praxis - Brüche und Brücken* (S. 9-27). Münster: LIT.

Prenzel, M. & Allolio-Näcke, L. (Hrsg.). (2006). *Untersuchungen zur Bildungsqualität von Schule* [Studies on the educational quality of schools]. Münster: Waxmann.

Riffert, F. (2005). Whitehead’s cyclic theory of learning and contemporary empirical educational research. In F. Riffert (Ed.): *Alfred North Whitehead on learning and educaton: theory and application* (pp. 89-119). Newcastle: Cambridge Scholar Press.

Riffert, F. (2008). Lernzyklen und ihre Beobachtung im Unterricht - Zwischenbericht zur Entwicklung eines Beobachtungsinstrumentes [Learning cycles and their observation during instruction - Preliminary report on the development of an observation instrument]. *Salzburger Beiträge zur Erziehungswissenschaft, 12/1*, 67-99.

Rolff, H.-G. (2008). Konsequenzen aus Schulleistungsstudien und ihre Umsetzung auf Schulebene [Consequences from school achievement studies and their application in schools]. In Landesinstitut für Schule und Medien Berlin-Brandenburg, et al. (Hrsg.), *Bildungsmonitoring, Vergleichsstudien und Innovation. Von evidenzbasierter Steuerung zur Praxis* (S. 147-159). Berlin: BWV.

Stachowiak, H. (Hg.). (1995). *Pragmatik. Handbuch pragmatischen Denkens* [Pragmatics. Handbook on pragmatic thinking]. Band V. Hamburg: Meiner.

Weiß, R.H. (1987). *Grundintelligenztest Skala 2. CFT 20*. Göttingen: Hogrefe.

Whitehead, A. N. (1928/1967). The *aims of education and other essays*. New York: Free Press.

Wylam, H. & Shayer, M. (1980). CMCS Science Reasoning Tasks: General guide. Windsor: NFER Publishing.

Zohar, A. & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low- and high-achieving students. Learning and Instruction, 18, 337-353.

Figure 1. A multiple evidence model (MUEMO) for educational interventions (related to Hager & Hasselhorn, 2000, p. 81 and Middleton, Gorard, Taylor, & Bannan-Ritland, 2008, p. 32).

INTERVENTION VS. NON-INTERVENTION
(PRE- AND NO-CONDITION)

EFFECTIVENESS IN TIME

A. Increasing goal competencies during intervention

B. Increasing goal competencies at the end of intervention

C. Increasing goal competencies after the intervention on the long run

EFFECTIVENESS IN CONTEXTS

D. Improving in similar contexts

E. Improving in nested contexts

F. Improving in daily life situations

DIFFERENT-INTERVENTION-FUNCTIONING

G. Side-effects

H. Interaction-effects

Table 1

Mean Scores (and SD) of Students´ Science Reasoning Tasks (Wylam & Shayer, 1980) Results

Group	1. Pretest	2. Pretest	During Intervention	1. Posttest	2. Posttest
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Experimental (*n*=20)	4.85 (1.18)	5.65 (0.49)	5.90 (0.55)	5.40 (1.19)	5.65 (0.75)
Control (*n*=14)	4.64 (1.08)	5.00 (0.78)	5.29 (0.61)	4.79 (0.58)	4.93 (0.48)

Table 2

Middle Ranks, Z-, and p-Values of Context Effectiveness Variables: Grades

Group	*Similar contexts*		*Nested contexts*	
	Physics	*Chemistry*	*Mathematics*	*German*
Experimental (*n*=25)	24.58	24.36	25.34	22.30
Control (*n*=25)	26.42	26.64	25.66	28.70
Z	-0.47	-0.60	-0.08	-1.76
(p)	(0.64)	(0.55)	(0.94)	(0.08)

Table 3

Indications of Different-Intervention-Functioning

A. Testing Side-Effects: Mean Scores (SD) and Reliability of Students´ Subject Interest Scale (from Hoffmann, Häußler & Lehrke, 1998)

Group	*N*	Pretest	*N*	Posttest
Experimental	24	64.00 (10.26)	22	65.32 (12.40)
Control	21	58.46 (10.15)	21	67.05 (10.87)
Cronbach´s Alpha		0.85		0.88

B. Testing Aptitude-Treatment-Interactions: Results from an ANCOVA (Bortz, 1993, p. 333) on Science Reasoning Tasks (1. Posttest, n=45)

Source	*SS*	*df*	*MS*	*F*	*p*	*Partial Eta*
Group	1.85	1	1.85	2.47	0.12	0.06
Intelligence	6.93	1	6.93	9.27	0.00	0.18
Group x Intelligence	3.00	1	3.00	4.01	0.05	0.09
Error	30.70	41	0.75			