International Journal of Environmental & Science Education
Vol. 3, No. 3, July 2008, 95-103

Planning Experiments in Science Education Research: Comparison of a Quasi-Experimental Approach with a Matched Pair Tandem Design

Christoph Randler
University of Leipzig, Germany

Franz X. Bogner
University of Bayreuth, Germany

Received 28 April 2008; Accepted 05 May 2008

The planning of adequate control groups is a central aspect of educational research. Didactical studies, however, are often field studies encountering many problems of everyday classroom teaching? Recent research has indicated that teachers and their beliefs have an enormous influence on learning and retention. Taking this line of evidence further, a devil's advocate may emphasize that differences in treatment-control designs may be not a result of the different instructional strategies but reflect differences between different teachers. In this study, we try to sort out this kind of "teacher effect" by comparing two approaches of a complex treatment-control group design. In the first approach we compare treatment and control groups that were taught by the same teachers (labeled 'matched-pair tandem design'), and in the second approach we compare the control group with an unrelated treatment group where different teachers taught the treatment groups ('unrelated design'). When comparing the 'matched-pair tandem' design with the quasi-experimental approach, we found i) similar patterns in both educational experiments, and ii) higher effect sizes in the unrelated, quasi-experimental design.

Key words: biology teaching, ecology, educational experiment, quantitative methods, research methods

Introduction

Evidence based quantitative empirical research has a long tradition in some subjects, for example in medicine and in nature conservation (Roberts et al., 2006) while it has to be developed much farther in didactical terms (Wittrock, 1986; Tobin et al., 1994). The planning of adequate controls and of control groups is one main aspect on the way towards evidence based educational research. Didactical studies, however, are often field studies encountering many problems of everyday classroom teaching? Didactical studies focus on classes, pupils and teachers in their real school life rather than using sophisticated experimental designs as in laboratory studies. Nevertheless, such classroom evaluations are urgently needed to improve teaching and learning and to accompany the implementation of new curricula and syllabi because education is often accused of jumping on bandwagons and of implementing changes without fully exploring the impact and effectiveness of such

ISSN 1306-3065

http://www.ijese.com/

changes (Marchant and Paulson, 2001). Therefore, more evidence based investigations are needed.

The attempt to survey effective methods of teaching and analyses of learning and instructional strategies has produced a vast amount of literature although intervention studies and quasi-experimental designs with broad empirical databases are quite scarce. Within the framework of empirical research, one central aspect of evaluations is the strict planning of control groups. Recent research has indicated that teachers and their beliefs have an enormous influence on learning and retention; therefore, the teachers themselves have come in the focus of research (Pintrich et al., 1993; Bryan and Atwater, 1994; Meyer, 1994; Chin, 2006; Oh 2005; Tsai, 2006; Waters-Adams, 2006). For example, conversations in the classroom (Oh, 2005; Chin, 2006), assessment of prior knowledge (Meyer, 2004), and teachers' beliefs (Waters-Adams, 2006) have a significant influence.

Taking this line of evidence further, a devil's advocate may emphasize that differences in treatment-control designs may be not a result of the different instructional strategies, new curricula or teaching materials but rather reflect differences between different teachers. In this respect teachers can be viewed as an important factor that may influence the experimental design. In this study, we try to sort out this kind of "teacher effect" by comparing two approaches of a complex treatment-control group design. In the first approach we compare a treatment and a control group that were taught by the same teachers (labeled 'matched-pair tandem design'), and in the second approach we compare the control group with an unrelated treatment group where different teachers taught the treatment groups (unrelated design).

The latter design is usually applied in educational research (Clarke, 1969; Miller, 1984; Keeves, 1998), and treatment groups are usually assigned randomly providing a quasi-experimental approach. The aim of the study was on the methodological question of whether different designs reach the same conclusions rather than an evaluation of an educational program.

Design and Procedure

The matched-pair tandem design

A matched-pair tandem-design was applied (Figure 1) where two different classes were always matched by the same teacher, thus these two classes were considered as tandem connected by their respective teacher. These classes share many features with each other, e.g. they come from the same school and experience the same environment (e.g. rural versus urban). Each of the three teachers taught both one experimental and one control group. This structure was planned to control for the teacher effect.

In the matched-pair tandem design, treatment and control classes are matched by the same teacher. In the unrelated design, the control group is compared with an unrelated treatment. The teachers always started with their more traditional teaching approach that was outlined in a teacher-centred manner. This teaching approach was used as control group. One may argue that for an optimal experimental design the order of trials should have been randomised to avoid order effects. However, it was impossible to control for order effects, i.e. to select some of these teachers that first used the modern approach and then the traditional teacher-centred one. If we had applied the learner-centred approach first, we could not assure that the teachers would not make use of the materials, methods and ideas of this approach in their traditional lessons. Therefore, to avoid such carry-over effects, we preferred the strict schedule to begin with the teacher-centred approach.

Unrelated design

To assess the quality of the match-pair tandem design we chose a second pupil sample. These pupils also received the treatment instruction but were not matched to any control group (Figure 1). Therefore, the teachers of this instructional group were unmatched with other classes or teachers, and further, these teachers were recruited from other schools but within the same school district.

Thus, if we compare these unrelated teachers and pupils with the control group of our tandem approach, then we have a comparison which is typical for many intervention studies in experimental education. In such designs, pupils and teachers from different schools (or within schools) receive a treatment while other teachers and pupils serve as control group.

Pupil sample

Our pupil sample came from the medium stratification ("Realschule"): The German school system separates pupils at the end of the 4th grade into three stratification levels according to their cognitive abilities: "Hauptschule": lowest stratification; "Realschule": medium stratification; "Gymnasium": highest stratification). 49.4 % of the total sample were girls. N=148 pupils participated in the matched pair tandem design and filled out all three achievement tests (65 treatment, 83 control group). Additionally 140 participated in the unrelated control. The pupils were unaware of the teaching strategy, including the underlying theory. 9th graders participated in the study because ecology was an extensive part in the curriculum and we developed and implemented an ecological unit (Randler & Bogner, 2004, 2006).

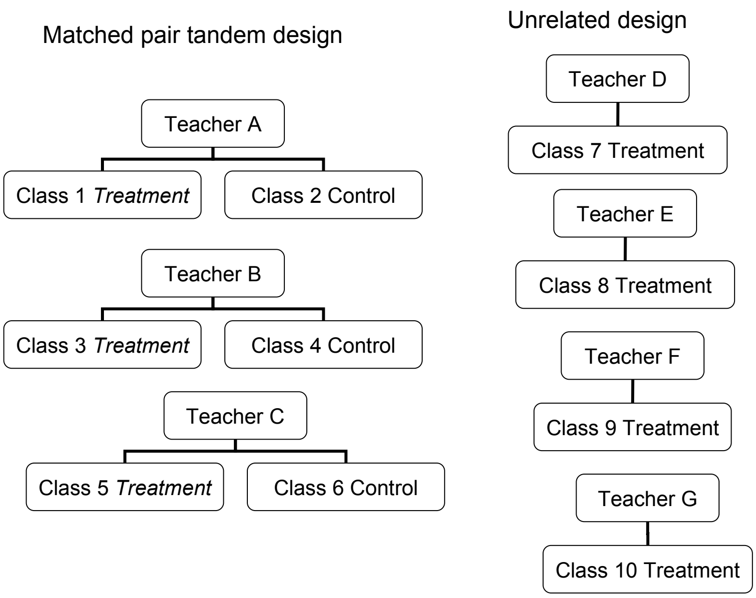

Figure 1. Experimental design.

Teachers

For both teaching approaches seven tenured biology teachers, four men and three women, participated in the study. Three teachers (two men, one woman) worked in the tandem design and the additionally four (two women, two men) taught the unrelated treatment groups. All teachers were in-service teachers with at least five to six years of teaching practice to avoid that effects may be biased by novice teachers.

Educational programme of the treatment groups

The educational programme consisted of various integrative facets including selected materials for laboratory and hands-on tasks, such as exuviae (larval remains) from dragonflies, stuffed bird specimens as well as water lilies (details see Randler & Bogner, 2004, 2006, 2008). The material was available for each participating class throughout the unit. Teachers were instructed how to use these materials (teaching methods, social forms, original objects). For the treatment group, a 25 page booklet was produced comprising supplementary materials, for example, a detailed description of the experiments and hands-on tasks. Teachers were informed about the learning goals (one sheet of paper) and of the topics of the lessons. They also received the class test in advance to properly prepare their pupils for the testing.

Educational programme of the control group

The control group received no educational material for the lessons but was advised to teach their pupils in a manner as always. This was a rather teacher-centred form of instruction. To assure that the control group does not differ from the treatment group except in materials and educational settings, teachers of the control group were informed of the learning goals (one sheet of paper) and of the lessons topics to be covered. They also had advanced knowledge of the class test so they could prepare their pupils accordingly. Therefore, both educational approaches did not differ with regard to content and duration of the lessons or with regard to the final examination (class tests; T-2).

Research design

We used a quasi-experimental approach to assign control groups and treatment groups. For the tandem design, the participating schools offered two classes of the requested grade and one biology teacher was responsible for both classes (Figure 1). Which of these classes received the treatment or the control instruction was assigned at random. Further, additional schools from the same district were asked to participate in the study and the teachers received the same materials as the matched-pair treatment group.

Research goal

The study design allows a matched pair tandem design amongst every two classes connected by one identical teacher, since one teacher first taught class-A as control group and afterwards class-B as treatment group. This design controls very strictly for the influence of the teacher. Differences between treatment and control groups should therefore reflect the difference of the instruction and not any difference between teachers. However, didactical field studies are often less strictly planned and different teachers participate in such studies and usually, treatment and control group were randomly assigned. Therefore, we compare the strictly planned tandem-design with another design where teachers were different. If these different designs yield similar results, then such randomly assigned studies should be considered as of the same quality as the tandem-design, and further, this would provide at least some evidence that such studies that are based on different teachers are also of a good validity in terms of experimental design.

The test instruments

Achievement scores were assessed by three tests: a pre-test (T-1), a class test (T-2) and a retention test (T-3) (see Figure 1). Pre-test and retention test were used without any grading, while the class test (T-2) was used for grading purposes. Retention tests (T-3) were applied 6-8 weeks after the class test (T-2) and pupils were not previously informed about the retention test. Class tests and retention tests were identical. The pre-test T-1 had a maximum score of 9, all other tests reached a maximum of 22 (T-2 and T-3). The pre-test assessed existing basic knowledge and detailed comprehension of ecological topics (Randler & Bogner 2004).

Statistics

SPSS release 13 was employed for statistical analysis. All tests were carried out two-tailed. To calculate effect sizes (Cohen's *d*) MetaWin 2.0 Calculator (Sinauer, Massachusetts, USA) was used.

Results

Comparison within the matched pair-tandem experimental design

Control and treatment groups were matched by the same teachers. Treatment and control groups scored similarly in the pre-test (T-1; treatment: 5.02 ± 0.25; control group: 4.62 ± 0.20; $T=1.237$; $df=167$; $p=0.218$; ns; Cohen's $d=0.19$). In the class test (T-2), treatment pupils performed better (14.71 ± 0.32) than the control group (13.60 ± 0.34; $T=2.349$; $df=170$; $p=0.02$; $d=0.36$). This difference remained in the retention test (T-3: treatment: 14.38 ± 0.55; control group: 12.95 ± 0.36; $T=2.189$; $df=159$; $p=0.03$; $d=0.34$). Cohen's d was similar in class test and retention test suggesting a similar effect size.

Multivariate assessment of the matched pair-tandem experimental design

Pre-test levels of prior knowledge often seem to provide one of the best predictions of learning outcomes. Therefore, we used a general linear model (GLM) to incorporate pre-test knowledge as a covariate and used gender and treatment as fixed factors. As expected, GLM revealed a highly significant influence of pre-test knowledge on subsequent tests (Table 1). Pretest was used as covariate, gender and treatment as fixed factors.

Another significant effect were differences with regard to treatment. Boys and girls benefited equally from the programme and there was no interaction between gender and treatment. Explained variances (measured as eta^2) were 0.20 for pretest (covariate) and 0.057 for treatment. Uni-variate analyses revealed a partial eta^2 of 0.037 for class test, and 0.050 for retention, suggesting a stronger treatment effect in retention. Calculation of Cohen's d after adjusting for prior knowledge revealed a d=0.39 in the class test and d=0.46 in retention which confirmed the previous results. The multivariate assessment yielded similar results as the t-tests.

Comparison of the unrelated experimental design

Control group and treatment groups were not matched by the same teachers. No differences could be found during pre-test (control group: 4.62 ± 0.20 versus unrelated treatment group: 4.32 ± 0.15; T=1.144; df=244; p=0.254; d=0.15), while difference arose in the class test (control group: 13.60 ± 0.34 versus treatment: 14.78 ± 0.21; T=-3.057; df=242; p=0.002; d=0.41) and remained stable into the retention test (control group: 12.95 ± 0.36 versus treatment: 14.80 ± 0.24; T= -4.314; df=236; p<0.001; d=0.58).

Table 1. Multivariate assessment of the strictly planned matched-pair tandem design.

Source	Wilks-Lambda	F	P	Eta^2
Constant	.278	184.198	<.0001	.722
Pretest	.800	17.758	<.001	.200
Treatment	.943	4.287	.016	.057
Gender	.988	.830	.438	.012
Gender * Treatment	1.000	.019	.981	.000

Multivariate assessment of the unrelated experimental design

The multivariate assessment revealed similar results in the unrelated design. We found a significant influence of pretest (eta^2=0.23) and treatment (eta^2=0.10), no gender effect and no interaction between gender and treatment (Table 2). Uni-variate calculations also revealed a higher effect size of treatment in retention (class test: eta^2=0.047; retention: eta^2=0.103). Calculation of Cohen's d after adjusting for prior knowledge revealed a d=0.46 in the class test and d=0.70 in retention which confirmed the previous results. The multivariate assessment yielded similar results as the t-tests. Pretest was used as covariate, gender and treatment as fixed factors.

Table 2. Multivariate assessment of the quasi-experimental design.

Source	Wilks-Lambda	F	P	Eta^2
Constant	.233	357.641	<.001	.767
Pretest	.767	33.041	<.001	.233
Treatment	.896	12.653	<.001	.104
Gender	.984	1.774	.172	.016
Gender * Treatment	.996	.422	.656	.004

Discussion

When comparing the matched-pair tandem design with a quasi-experimental approach, we found i) rather similar patterns in both educational experiments, and ii) higher effect sizes in the unrelated, quasi-experimental design. This may have implications for further studies.

The influence of teachers and their beliefs in teaching methods or prior knowledge of their pupils (Bryan and Atwater, 1994; Meyer, 1994; Waters-Adams, 2006), as well as their motivational and social engagement on learning and instruction is often demonstrated as relevant (Pintrich et al., 1993; Tsai, 2006). For example, Chin (2006) emphasized the quality of questioning (cognitive conflicts) and of feedback in science classrooms and Oh (2005) stressed the discursive role. These effects, in turn, might influence the outcome of educational research studies, and studies might be hampered by the beliefs and motivational states of the teachers as well as by their current classroom practice. However, in our first approach, the teachers remained the same, first in the control group, and then in the treatment group. Thus, general instructional methods of each individual teacher (speech, language, personality, beliefs and thinking; see e.g. Fraser et al., 1987) remained similar and differences between the approaches might be described indeed to real existing differences between the methods.

In our second approach, results were rather similar although different teachers participated in the control- and the treatment group. Nevertheless, we yielded similar results. This could be explained by two alternatives: first, despite much research, the influence of the teacher is less than previously supposed and overestimated (Pintrich et al., 1993; Bryan & Atwater, 1994; Meyer, 1994; Oh, 2005; Chin, 2006; Tsai, 2006; Waters-Adams, 2006). Second, our second approach provided a high sample size, and the influence of individual teachers is smaller than in studies with less participating teachers. We prefer the second solutions since we feel that teachers, their motivational states, believes and classroom practices definitely influence learning and retention. A higher sample size (> 200 pupils) and a randomised assignment of the classes, however, seem also to provide a sufficient control group. Nevertheless, intervention effects were smaller when the stricter main sample design was applied.

Implications and Conclusion

We suggest using both methods of experimental design. Although the matched-pair tandem designs seems to provide a more strictly planned control (as might have been revealed by lower effect sizes), we feel that the quasi-experimental approach is often the best possible, e.g. if one asks different schools for participation, one class of this school might be taught by instruction-A and the other one by instruction-B. Then, apart from the teacher, at least, environmental variables remain similar (hometown size, school environment). This could be considered some kind of a randomised block design (Clarke, 1969). The results are encouraging because they suggest that – despite an unquestioned high influence of the personality of the teachers – quasi-experimental approaches seem sufficient in educational research.

Acknowledgements

This study was an integral part of a major subject integrative research project involving researchers from Physics, German Language and Pedagogical Psychology. We, therefore, appreciate the cooperation of our colleagues P.Mayring, H.Melenk, C.v.Rhoeneck, S.Fuß, M.Gläser-Zikuda, M.Laukenmann, U.Maier and K.Metz. Additionally, we acknowledge the helpful and valuable discussion of A.Ash and L.Schottenberger. M.Wiseman helped with statistical advice. This study was supported by the Ministry of Education and Sport (MKS) the Ministry of Science, Research and Art (MWFK) [both of the state Baden-Württemberg] as well as by the University of Education Ludwigsburg (PHL).

References

Bryan, L. A., & Atwater, M. M. (2002). Teacher beliefs and cultural models: A challenge for science teacher preparation programs. *Science Education, 86(6),* 821-839.

Clarke, G. M. (1969). S*tatistics and experimental design*. London: Arnold.

Chin, C. (2006). Classroom interaction in science: Teacher questioning and feedback to students' responses. *International Journal of Science Education*, *28(11)* 1315-1346.

Fraser, B.J., Walberg, H. J., Welch, W.W., & Hattie, J.A. (1987). Synthesis of educational productivity research. *International Journal of Educational Research, 11*, 145-252.

Gläser-Zikuda, M., Fuß, S., Laukenmann, M., Metz, K., Randler, C. (2005). Promoting students' emotions and achievement – instructional design and evaluation of the ECOLE approach. *Learning & Instruction, 15*, 481-495.

Keeves, J. P. (1998). *Methods and Processes in Research in Science Education*. In: Fraser, J.B. & Tobin, K.G. (ed.). International Handbook of Science Education. Part II. Dordrecht, Kluwer.

Marchant, G.J., & Paulson, S.B. (2001). Differential school functioning in a block schedule: a comparison of academic profiles. *High School Journal, 84,* 12-20.

Meyer, H. (2004). Novice and expert teachers' conceptions of learners' prior knowledge. *Science Education, 88*(6), 970-983.

Miller, S. (1984). *Experimental design and statistics.* London & New York: Methuen.

Oh, P. S. (2005). Discursive roles of the teacher during class sessions for students presenting their science investigations. *International Journal of Science Education, 27* (15), 1825-1851.

Pintrich, P.R, Marx, R.W. & Boyie, R.A. (1993). Beyond cold conceptual change. The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research, 63,* 167-199.

Randler, C., & Bogner, F. X. (2004). Emotional and cognitive aspects of learning: the ecological unit "Lake". Pp 117-124 In: Ralle, B. & I. Eilks (Eds.): *Quality in practice-oriented research in science education*. Shaker: Aachen.

Randler, C. & Bogner, F. X. (2006). Cognitive achievements in identification skills. *Journal of Biological Education, 40(4*), 161-165.

Roberts, P. D., Stewart, G. B., & Pullin, A. S. (2006). Are review articles a reliable source of evidence to support conservation and environmental management. A comparison with medicine. *Biological Conservation, 132,* 409-423.

Tobin, K., Tippins, D.J. & Gallard, A. J. (1994). Research on instructional strategies for teaching science. In: Gabel, D. L. *Handbook of research on science teaching and learning.* New York, Macmillan.

Tsai, C. C. (2006). Reinterpreting and reconstructing science: Teachers' view changes toward the nature of science by courses of science education. *Teaching and Teacher Education, 22* (3), 363-375.

Waters-Adams, S. (2006). The relationship between understanding of the nature of science and practice: The influence of teachers' beliefs about education, teaching and learning. *International Journal of Science Education, 28*(8), 919-944.

Wittrock, M. C. (1986). *Handbook of Research on Teaching*. 3rd ed. New York: Macmillan.

Authors

Christoph Randler was Professor at the University of Leipzig, Leipzig, Germany and is currently Professor at the University of Education in Heidelberg. He has an academic background in biological education and received his PhD in 2003. **Correspondence:** University of Education, Heidelberg, Im Neuenheimer Feld 561-2, D-69120 Heidelberg, Germany; e-mail: Randler@ph-heidelberg.de

Franz X. Bogner is a Full Professor at the Department of Biology at the University of Bayreuth, University of Bayreuth, Centre of Math & Science Education, Institute of Department of Biology Didactics, Universitätsstr.30, D–95444 Bayreuth, Germany. He has an academic background in chemical ecology (PhD) and in biological/environmental education (habilitation).