Research in Mathematics Education – Who Benefits?

Heinrich Bauersfeld1, Bielefeld

Abstract: For a long time mathematics education tries hard to win recognition as an academic discipline. In related classroom research and curriculum development one can find not rarely theories and methods in use that are adapted from other (and well established) disciplines. However, in many cases these adaptations do not serve the researchers’ goals, yet more, their effects can contradict the stated purposes. The article discusses a few fundamental problems related to empirical research in mathematics education (e.g. the role of the teachers in experimental/control groups), using as concretization a printed research report that is chosen deliberately (and made nameless, therefore).


ZDM-Classification: C80, D20

“... there is no agreement among leaders in the field about goals of research, important questions, objects of study, methods of investigation, criteria for evaluation, significant results, major theories, or usefulness of results ... a field in disarray.”

Lynn Arthur Steen (1999, p. 236)

Although I do not agree with Steen’s total condemnation, research in mathematics education may well find itself in a remarkable dilemma. An ethnographic observer of the scene could identify two contradictory movements. There is an increasingly loud claim for recognition as an academic discipline. The price for what is certainly growing acceptance is the adoption or adaptation of procedures and methods used by well-established disciplines – a high price, since the effect is a splitting up into schools and sub-communities. On the other hand, well-defined concepts, reliable methods and powerful predictions can meet the accompanying expectations only in highly controllable and rigidly controlled reductions of our realities, and this works all the better the sharper the reductions are. This leads to a strong tendency to move away from the complexities of an everyday classroom in order to gain scientific appreciation. The outcome is that we are hearing increasingly loud complaints about the ineffectiveness of research in math education for actual teaching practice (see, for instance, the many articles in AERA’s2 “Educational Researcher” over the last years). We find academic effort and precision on marginal issues rather than a pragmatic orientation for dealing with relevant problems. Summing up a report to the US Congress, the director of AERA’s Governmental and Professional Liaison Program, Gerald Stoufe, stated recently: “Education research does not provide critical, trustworthy, policy-relevant information about problems of compelling interest to the education public” (Stoufe 1997, p. 27). The following considerations discuss this problem, using a deliberately chosen concrete example from recent research.

In many disciplines, the discussion on research, related theoretical bases, methods and outcomes has attained a more or less professional standard (e.g. the periodical “Brain and Behavioral Sciences” and the related e-mail exchange, Michael Cole’s xfmail “chat”, or even the many sequences of pro and con articles in AERA’s “Educational Researcher”). In Mathematics Education, however, the situation seems to be rather deficient: Professional journals can and should be open for new approaches, but also for a critique of the traditional, for constructive extensions and for alternatives. Many of our journals publish one article after the other, mostly lacking coherence and following the random submissions of authors. The papers undergo covert review procedures that usually remain in the dark, and they are rewritten following anonymous recommendations. The award for finally published articles seems to be something like a scientific quality mark, like “passed”, a hidden hallmark of the published paper. All this functions more or less to the disadvantage of any real professional reflection in the community.

By chance, an article in a new journal made me think about these problems recently. I shall take this article

1I would like to thank Jonathan Harrow for native speaker advice.

2AERA: American Educational Research Association
as an example to illustrate a few general problems, although many other recent publications elsewhere could have served just as well. I am not at cross-purposes with the direction of the report nor do I doubt the authors’ best intentions. But some points in the reported research procedures and their background are worthy of our professional attention and worthwhile discussing in the hope that the following – admittedly totally subjective – remarks may elicit a more sound professional discussion.

1. Distorted competition or the miracle of prepared improvement
Assume that you are running an experimental match between two comparably good boxers. Ahead of the match, one of them prepares in a “conventional” manner. The other receives special preparation, extra physical training, psychological bolstering of his self-concept, information about the antagonist’s weaknesses and an additional special pampering before and after each round of the fight. During the match itself, you count the points of each boxer. Who do you guess will be the winner? Surely, after such differences in preparation and execution the advantage will be awarded to the specially prepared boxer. However, this is no guarantee or reliable prediction of a convincing win. Particularly a narrow win on points, therefore, would give us reason to think about interferences and other influences, for example, about the usefulness of the training design, the actual disposition of the fighters, the adequacy of the evaluative measurement and so on.

The research design of comparing a specially treated “experimental group” with a so-called “normally” or “traditionally” working non-trained “control group” has much in common with this example. With an investment of quite considerable effort, the authors prepared the teachers in their experimental classes for “implementing the learning environment”\(^3\). Before and after each lesson, “a member of the research team [...] had a 5- to 10-min preparatory or evaluative talk with the teacher”. They were also provided with “general” and “specific guidelines for the teachers”, the “preferable sequence of instructional activities”, and “all the necessary concrete material for the pupils”. Is it then a miracle, when the authors achieve positive effects in their experimental classes?

However, to the disappointment of the authors, the overall gains in the experimental group were only modest. Their ascription of possible causes was “shortcomings in the teachers’ implementation of the learning environment” and “results would have been better if we could have increased the available instructional time and if we could have integrated the learning environment better within the regular mathematics lessons”. Holding their own contributions and activities to be beyond question, they attribute the slight gain to external “factors” only, to restrictions in terms of time, to the experimental period as an unfamiliar implant into regular lessons and, not least, to the teachers’ failure. Inferences from our boxer experiment could lead us to extend the search into neglected areas: for instance, to the functioning of implicit theoretical orientations on the part of the researchers, to weaknesses in the experimental/control group design and to the researchers’ explicit and implicit assumptions about the role of the teachers.

2. Theoretical and conceptual incompatibilities
In the presentation of research outcomes, one can identify certain disadvantageous habits. Often, authors found their theoretical presuppositions on the most recent related publications that are at hand or en vogue. Competing theories then serve only as a foil for the “new” and “better” orientation, if mentioned at all. However, it is a trap to neglect the history of key concepts, of their development and of the shifts of meaning in their use by different authors. It is only too easy for the presentation of procedures and outcomes to employ a mixture of notions and concepts belonging to quite different, if not incompatible language games. Moreover, and under the pressure to seem modern and at the vanguard of scientific development, authors use bandwagon labels without sufficient analyses and without critical distance. An analysis of the presentation of research procedures and outcomes often reveals a somewhat different and covertly functioning philosophy. “Modern” façades are used to cover over and pep up methods that are questionable and dubious in the light of the stated project aims – Potemkin’s renaissance in modern research?

In our example, the simple combination of a “cognitive-rationalist perspective” with a “situative orientation” and a “social constructivist orientation” harbours some problems of compatibility. The authors argue for the “establishment of new social and sociomathematical norms” and see their own theoretical orientation as being “in line with the socioconstructivist perspective”. They stress the openness and flexibility of their approach, but their descriptions of the program and its related operationalisation reveal a rather different philosophy in action. Two examples may illustrate this:

**Example 1**
The authors claim: “the focus of teacher development and support was not on making them perform in a specific way but on preparing and equipping them to make informed decisions”. Only a few lines later, the authors present “a list of 10 general guidelines for the teachers”. Furthermore, they have developed “a specific teacher guide for each lesson” and even a complete “sequence of instructional activities”. Obviously, all this is intended to work as a detailed pre-developed frame for the teachers’ “informed decisions”. Additionally, several meetings with the research team aimed at an effective transmission of the “highly interactive teaching methods” to the experimental teachers. The teachers were “prepared for [...] implementing the learning environment” including “the creation of a classroom culture”. Such efforts are clearly not compatible with “social constructivist” models that describe classroom cultures as emerging issues but not as directly teachable, prescribed objects ready for implementation. All this contradicts the openness and the theoretical orientation claimed in the authors’ introduction.

**Example 2**
During the 20 project lessons, students in the experimental

---

\(^3\) All of the following quotations that bear no indication refer to the research report.
group were trained explicitly using a “set of eight heuristics” and, finally, they were taught “how to choose among the distinct heuristics learned so far”. A related special worksheet from the “notebook for pupils” presents the collection of rules for the “solution process”. In their final experiment, the students learned “to use the competent problem-solving model”. This, clearly, is more or less an explicit teaching of strategies and meta-rules.

Cognitive scientists discuss strategies and “competent problem solving” as knowledge about knowledge or as metaknowledge. Interactionists as well as specialists in expert/layman issues prefer to treat such orientations more as phenomena that are not thoroughly conscious and not sufficiently describable. In any case, as far as we know, it is very difficult if not impossible to teach metaknowledge directly (see, for instance, Weinert/Kluwe 1986, Resnick 1987, Mazzoni 1998, Koiriat 1998). From the perspective of “situated learning”, all attempts to teach meta-rules directly lead inevitably to situated knowledge rather than to more or less general strategies that are available free of context. Surely, meta-rules can be learned verbally as so-called general rules, but we cannot activate such mental states deliberately in new situations. There is no simple “recall from memory”, because such complex orientations are bound to the specific situation in which they were learned. As general rules, meta-strategies emerge only from the body of a rich experience related to the solving of problems of very different types and supported by the guided reflection on these activities. The long and carefully enriched preparatory courses for the teams participating in the International Mathematical Olympiads are condensed examples of the practices required (condensed, because these crash courses are practice sessions designed for highly gifted students).

3. The experimental vs. control group research design

“Researchers in mathematics education not only disagree on criteria for acceptable work, they even disagree on the need for criteria.”

Lynn Arthur Steen (1999, p. 236)

In mathematics education, this classic design has been used to provide empirically controlled confirmation of a hypothesis, to prove the expected superiority of a “new” treatment over another treatment, named the usual or traditional one. For decades, and due to several reasons, educational researchers have criticised this design as a means of obtaining reliable comparisons of teaching methods (see, for instance, the many related discussions in the Educational Researcher across the last decades, or Sierpinska/Kilpatrick 1998). To find it applied again nowadays reminds me of a discussion with a member of the former Soviet Academy for Pedagogical Sciences in Moscow in 1991. He was offering a complete series of math textbooks for an edition in the West, claiming these textbooks to be the very best to exist so far, and that he could prove this empirically. His experiment was based on very large samples from all over the Soviet Union, split into experimental and control groups. No argument was strong enough to raise doubts in his mind.

More general and stronger is the argumentation of De-landsheire and Petrosky (1998) who principally doubt “the use of numerical ratings to represent complex performances”. Reviewing the two volumes of Sierpinska and Kilpatrick (1998), Steen speaks of the “American infatuation with statistical research” and states: “Problems in mathematics education are inherently too complex for simple statistical techniques” (Steen 1999, p. 238). Unlike boxes’ punches, it is difficult to count the “use of heuristics”. Nonetheless, the authors of our example did “count the number of word problems [...] for which at least one of the explicitly taught heuristics was visibly used”! Yet, I do not want to contribute to the persistent controversy between quantitative and qualitative designs (for a more detailed methodological discussion, see, for instance, Pogrow 1998, Weinert 1998, Slavin 1997 and 1999, and Cobb/Bowers 1999). My focus here is more on the effects of research on teachers and students and on researchers’ deficits in self-reflection. The damage that this specific design produces counteracts the goals claimed in terms of improved teaching.

What is the purpose of using this design in the present study, in our example? Among other goals, the authors intended to verify the “most important hypothesis [...] that the learning environment would have a significant positive influence on pupils’ skill in the solution of non-routine mathematical application problems”. There is no serious doubt about the existence of such an impact in general. In fact, the interest here is in a “guarantee” for the authors’ specially developed curriculum. Can the dubious verification of the key hypothesis seriously serve as a scientific qualification mark for the “learning environment” produced? Who needs to be convinced through this? And finally, who can benefit from the verification claimed?

As usual, the authors close with “further research is needed”. Surprisingly, they ask for “a larger number of experimental classes”! That, they presume, would result “in more reliable and generalizable conclusions”. In the 1960s, we had already learned from Lee Cronbach that large samples will make just about everything significant. And furthermore, in large samples, the intra-group variation is also often much higher than the inter-group differences in total (as we have also learned from the “Frankfurt Project”, see below). Will an inflation of this project produce more convincing results? Will the final outcome be a more helpful orientation for teachers?

With all this in mind, it is very sad to read the authors’ excuse concerning the “lack of time” that restrained them from documenting several critical issues. This may also explain the lack of reflection on the research process itself. It would seem that the researchers’ self-reflection is somehow underdeveloped.

4. The roles of the participating teachers

My personal experiences with this research design rest upon the “Frankfurter Projekt” 1967–1972, sponsored mainly by the VW Foundation with more than a million German Marks. With 40 experimental and 42 control classes, kept together over a period of 4 years from Grade 1 through Grade 4, we tried to investigate the effects of a specially developed Mod Math curriculum (“aief”,

97
Bauersfeld 1972). The point of interest here is the role of the participating teachers: Teachers in the experimental classes were highly motivated and co-operative. They contributed to the further development and revision of the curriculum. Many of them later became tutors and advisors. Victims of the project were the teachers in the control classes. They complained heavily. They felt like guinea pigs, forced to work with their regular textbooks without access to the interesting new materials and approaches. Though we followed mainstream standards at that time (and changed methods dramatically later in the course of the project), I do regret this decision. In terms of school reform, this certainly is the weakest point of the design in question. How would you feel – as a teacher (as well as a student) – if you realised that you are being treated as an object of minor importance, excluded from all interesting issues? All your effort serves only as a negative instance, to be used as a contrast to the experimental sample that is expected to do better. And worse: in our example, the control teachers had to compete within a narrower time allowance (15 lessons instead of 20 in the experimental group)! Finally, why was it necessary to attach 7 [!] control classes to 4 experimental classes?

In the project under discussion, even the experimental teachers were treated as dependent subjects. In order “to give the teachers “co-ownership”, they were “invited to comment on first drafts of the general and specific teacher guides”. In a final interview, the experimental teachers had an opportunity to talk about their difficulties. And even half of this favoured group pleaded for a “more partnership-based way in the design of the learning environment and in the realisation of the whole research enterprise”! The authors do not discuss whether this fits in with a “social constructivist” orientation or not.

The control teachers were not listened to. The authors’ excuse: “We were not able to pursue a systematic and detailed analysis of what happened in each of the control classes.” There were only “informal discussions” with the control teachers “and an analysis of the textbooks used”. How about the motivation to do their best in these teachers? How might they feel about educational reform and their role in it? Ernst von Glaserfeld has long been pondering over a constructivist ethic; I am sure this design will not pass his standards.

How to help teachers? School reforms are doomed to failure without bottom-up support. Teachers are the only academics controlled by a whole pyramid of authorities extending above them. They have to follow syllabi. Researchers tell them “better” ways. In-service training courses present prescribed (if not decreed) curricula. And, too often, these courses are not organised among teachers themselves and designed to service their problems as close to their school as possible. One of the few valuable outcomes of TIMSS have been the documented Japanese “research sessions” (see Stigler/Hiebert 1997 and Stevenson 1998). These have indicated that the mathematics teachers’ professionalization in the Western countries is not as good as it could be. A serious partnership between researchers and teachers is more lip-service than reality. This, I think, is in need of (e.g. Lampert 1998; Brown/Ellery/Campione 1998, p. 348 “Students, teachers, and experts as researchers”).

5. Implicit models of educational reform, the “hidden agenda”

“Failure is the only possible outcome for any approach ... in which researchers hand their results to curriculum developers and teachers who are then expected to apply them in their practices.”

John Mason (quoted in Steen 1999, p. 240)

How people describe what they do is one thing; what they actually do, is another. Thus description and action, intention and realisation can drift apart, even in research projects. The authors state: “The model of teacher development adopted reflected our views about pupils’ learning by emphasising the creation of a social context wherein teachers and researchers learn from one another, rather than a model wherein the researchers transmit knowledge to the teachers”. What the authors do, in fact, contradicts this claim: Teachers are prepared and controlled when “implementing the learning environment”. They have to obey “general guidelines [...] comprising actions they should take”. Once you “give the teachers co-ownership” (italics added), you are the authority, and your orientation is top-down and hierarchical rather than a partnership. This is all the more true once teachers are allowed only to “comment” and to add “notes”. The documented practice is more like an introduction of “predetermined criteria”, and this clearly contradicts the authors’ claim to be “in line with the socioconstructivist perspective”.

With their search for “helpful suggestions for the elaboration of classroom activities and action plans that would guarantee a beneficial effect on pupils’ attitudes and beliefs” (italics added), the authors reveal their own trust in the reliability of cause-effect analyses of classroom events. This, again, contradicts the claimed constructivist “coconstruction” of “normative understandings”. The authors say (and surely want) to proceed by following an integrated theoretical orientation. What they really do, reminds me more of ancient top-down models like the RDD or RDDI (Research-Development-Dissemination-Implementation) approaches that had been questioned already in the early 1970s (for a contemporary German critique, see Deutscher Bildungsrat 1973, p. A73, and 1976, part II).

The authors presented “design experiment aimed at the development, implementation, and evaluation of a powerful environment” and by doing this, they follow these old models, even verbally. Their procedure is to develop a piece of curriculum (“learning environment”) based on “scientific literature” and interviews with experts. They transmit this curriculum to the experimental teachers who are to use it, adding minor changes. And they evaluate certain effects empirically. After that, they revise their curriculum and publish it. Finally, other teachers are expected to apply (or “implement”) the perfected, “empirically tested” curriculum. The control teachers are left to the same fate as well. They may all become inglorious followers of prescribed change, and nothing more.
6. Parallels with the economy?
Throughout the world, we are experiencing the progressing fusion of enterprises and organisations into larger and even larger companies and combines. Sometimes, I begin to think that a similar international “culture” is developing in our own Math Ed community. Increasingly, one can identify certain groups of colleagues who co-operate quite exclusively with each other like members of a club. They quote only each other. They often do not quote original work, but refer to ideas, concepts, and theories only in the ways presented by club members. They support each other’s useful arguments and mutual commendations, though they neither follow the same paradigm nor treat the same content areas or school stages. From within their own sub-culture, members criticise other sub-cultures, usually without taking account of the incompatibility of the different language games they employ. Our community is splitting increasingly into such relatively large and separated sub-cultures — like ethnic groups or cliques. Steen refers to this as ‘intellectual myopia’ and diagnoses an “intense intellectual parochialism” as its effect, “a field Balkanized by conflicting ideologies” (1999, p. 239 and 240). Damaging consequences of this schism are the rarity of constructive critique and the lack of an open professional discussion among the groups.

7. Are there alternative research strategies?

“Instead of searching desperately for secure ground on which to stand firm, it is possible to accept tension between knowing and not knowing as a productive and inescapable source of energy and security.”

John Mason (quoted in Steen 1999, p. 240)

What has been said here about the professional identity of teachers applies to the world of researchers as well. Quite simple variations in design can change the roles of researchers and teachers remarkably. Why not skip the whole design of experimental versus control group and replace it with competing groups, perhaps even more than two? If, indeed, there is no best solution to the question of teaching methods, different sets of approaches and tasks can be used, developed co-operatively by math educators and teachers rather than developed by researchers as prescribed materials only. What appears to be important is the openness towards a permanent process of revision; an intensive mutual reflection of processes, theoretical bases and decisions aiming towards an accompanying step-by-step improvement. Moreover, the problem of “integration” into regular mathematics lessons would fade away with such approaches. More importantly, serious co-operation might work towards intensifying the motivation of all collaborating teachers, because everyone would then be working “experimentally” (and competing).

Although final sessions with all developers are useful for the exchange of mutual experiences made in different groups, these meetings should not be forced into the production of one unified approach. Understood as necessarily non-perfect (and non-perfectible!) materials, they are open to further adaptation, enrichment, and change. Other teachers, who may become interested in the products of the project at a later date, will find themselves tempted and inspired much more strongly when publications offer different special approaches for trying out and for modification, illustrated with special experiences and with special case stories about student’s difficulties and successes. “Scientific quality marks” or “guarantees” are of minor importance for such decisions. Teachers from the first round and math educators can serve as tutors. The “dissemination” problem would then fade away.

By the way, math educators can document such developing projects as participant observers by, for example, finding out about conditions, effective support and constraints in the flow of activities, analysing differences among the emerging classroom cultures in the different groups and so on. A useful aid to self-reflection in co-operative teams are audiotapes of the sessions and related joint analyses. Such paths would lead us to more effective processes of change and contribute to the professionalization of teachers in more serious ways.

Admittedly, such projects would take a rather longer period than just 20 lessons. Something like a permanent reform could develop from them, leading to a more reliable and constructive competition between different theoretical bases in practice as well. Perhaps it is more important to replace the hunt for generalisations and pieces of perfected curricula by promoting many small-scale but long-term local projects in co-operation with highly motivated teachers and encouraging competition among these projects. My late friend Hans Rieckhoff used to characterise the usual short-term efforts of reform and research by comparing them with the attempt to ripen green tomatoes by using a flashlight.

To challenge the community, I would like to add a final question: Wouldn’t it be helpful to let outstanding teachers co-determine the selection and granting of research projects, at least as far as a certain proportion of the funding cake is concerned?

8. References


Author
Bauersfeld, Heinrich, Prof. Dr., Institut für Didaktik der Mathematik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld. E-mail: hbauersf@uni-bielefeld.de

Vorschau auf Analysethemen der nächsten Hefte

Für die Analysen der Jahrgänge 32 (2000) bis 33 (2001) sind folgende Themen geplant:
– Computergestütztes Lösen offener Probleme im Mathematikunterricht
– Mathematikdidaktische Forschung im Primarbereich
– Mathematik an Hochschulen lehren und lernen
– Analysis an Hochschulen
– Mathematik in der Ingenieursausbildung
– Theoretische Betrachtungen zu Schulbuchanalysen.

Vorschläge für Beiträge zu o.g. Themen erbitten wir an die Schriftleitung.

Outlook on Future Topics

The following subjects are intended for the analysis sections of Vol. 32 (2000) to Vol. 33 (2001):
– Computer-aided solution of open problems in mathematics teaching
– Research in primary mathematics education
– Teaching and learning mathematics at university level
– Calculus at universities
– Mathematics and engineering education
– Concepts and issues in textbook analyses.

Suggestions for contributions to these subjects are welcome and should be addressed to the editor.