Quality development in mathematics education by focussing on the outcome: new answers or new questions?
Introduction

Andreas Büchter, Dortmund (Germany) &
Timo Leuders, Freiburg (Germany)

1. The paradigm shift: from input to outcome

At present the German education system undergoes a drastic change, introducing new steering instruments which are expected to influence teaching practice considerably. As a consequence of the mediocre results of German students in international large scale assessment studies the governments of the sixteen Federal States unanimously decreed that quality improvement of schools should henceforth be steered by the "outcome" instead of the "input" (cf. article by Blum et al. in this issue). The education system has for a long time nurtured the false hope that quality could be achieved and secured by making detailed prescriptions with regard to teachers' education, by equipping schools and by defining the obligatory content in great detail. After seeing that this approach has not proven sufficiently successful, the new hope grows that quality improvement is better reached by defining expected competencies and by assessing them centrally. In central guidelines that appear radically narrow teachers and schools receive increased responsibility. Since this steering system combines central and decentral strategies it is often referred to as "integrated quality management" (Rolff 2004).

The change we just outlined is the most fundamental reform in Germany since the education reform at the end of the 1960s or even since the restoration of the diversified school system after the Second World War (cf. Baumert, Cortina & Leschinsky 2003). The enormous speed and the wide extent of the recently initiated change can be better understood when seen in the historical context:

(1) Since the debate on the school system in the seventies the political parties' different education policies blocked any major change regarding the structure of the school system (cf. Baumert, Cortina & Leschinsky 2003). During these thirty years preceding the "PISA-shock" many comparable countries have realised a fundamental modernisation of their school system.
(2) The German education system is – like German educational science – still influenced by the tradition of the humanities. Thus the so-called "empirical turn" in education policy (Lange 1999), which now shifts the focus measuring the outcome, encounters a theory and a practice that are mostly unprepared.

Being healthfully sceptical one has to concede that it is most uncertain whether the aforementioned orientation on the outcome (some German authors use the term "output-orientation") yields the desirable quality improvement. It can be even considered too early to give a preliminary estimate of the effects of the outlined reform. Even protagonists of systematic assessment strategies call for caution and dissuade from excessive euphoria:

„In practice such an output-orientation can prove treacherous. It is too alluring to consider measuring the output already as a system optimisation. […] … an output-orientation demands for a transparent causal attribution of any output-success or output-failure. […] It is important to state that it is not measuring the output that is inherently problematic but the lack of specific feedback, that is the lack of effective output-orientation." (Leschinsky & Cortina 2003, p. 47)

2. Mathematics as a key subject

Why is it that a journal for *research in mathematics education* dedicates an entire issue on this topic? There were no subject-specific aspects in the preceding analysis. Nevertheless, the subject mathematics plays a central role in the depicted development:

– In the majority of large scale assessments like TIMSS-II (vgl. Baumert, Lehmann et al. 1997) and PISA 2000 (vgl. Deutsches PISA-Konsortium 2001) students' mathematical performance was examined.
– In pilot-projects like the federal program SINUS (acronym for „increasing the efficiency of teaching mathematics and science", cf. BLK 1997), which were established after TIMSS-II, a majority of participating schools engage in improving the mathematics education.
– In the seminal counselling paper on developing national educational standards (Klieme et al. 2003) mathematics education researchers had a substantial influence. Mathematics is among the first subjects for which national educational standards were developed.
– In the scaling process of standards mathematics will act as a precursor when along with the collection of data in PISA 2006 mathematics items will be evaluated for the substantiation of the national standards (cf. article by Blum et al. in this issue).
– Finally mathematics can be considered among the central domains of an internationally accepted concept of "literacy" that is emerging in a process of "convergence and agreement overarching the cultures" (cf. Baumert 2002).

In spite of these good reasons suspicion cannot be entirely discarded that mathematics may have deserved this "pole position" because it proves to be a "test-friendly" subject. Commonly used mathematical tasks are relatively "obliging" when it comes to objective evaluation. Hence reliable testing is more easily achieved than in the domain of foreign language for example. Reconsidering this doubtful benefit thoroughly one realises that this hold only when testing is restricted to rather basic techniques. Assessing competencies related to mathematical processes, such as "reasoning", "modelling" or "problem solving" similar difficulties as in the language domains could arise. But since the emphasis on these competencies is regarded as a substantial progress in comparison with former curricula (cf. articles by Blum et al. and by Leuders, Barzel & Hußmann in this issue) it is imperative

to find ways to adequately include these competencies in any newly introduced assessment programmes.

3. Contingency of the new steering instruments

Looking at the recent development in education policy and especially at the rapid creation of steering instruments (standards, assessment) one may have the impression of a consequent and necessary evolution with hardly any alternative. Seen from a political perspective, the decision of education policy (represented by the standing conference of the ministries of culture of the 16 federal states) to implement the measures described in the preceding section can be regarded as a historical consensus. Though different approaches are conceivable (cf. article by Leuders, Barzel & Hußmann in this issue) education policy had to favour the described one for presumably two reasons:

– The example of many countries considered successful after analysing the results of PISA 2000 gave rise to an increased interest in a system based on autonomy and accountability of schools. This was already a branch of innovation that was moderately investigated in pilot projects.
– The scarcity of financial resources in an economically difficult period made other more costly approaches appear less feasible.

Looking closer at the actual steering instruments created in this process, as the authors of this issue will do from different perspectives, the evolution and its products (standards and tests) appear much more contingent. We will briefly illustrate this contingency by referring to the aspect of defining standards:

– Educational standards which have recently been developed in Germany (e.g. KMK 2003) only define competence expectations and restrain from giving guidelines on how to organise learning processes – these decisions are entirely left to the teachers and schools. Thus national standards and some (but not all) of the curricula of the federal states of Germany follow the counselling paper by Klieme et al. (2003). Nevertheless educational standards for mathematics can look quite different as the example of the highly appreciated "principles and standards" of the American NCTM (2000) demonstrate: These standards are not mandatory for the states and often the education policy in a state considerably diverges from the recommendations given in the standards. In contrast to the German standards which focus on the outcome the NCTM standards extensively refer to theoretically and empirically sound teaching and learning principles and thus try to give a picture of "ideal mathematics learning environments" (cf. Reiss 2004, p. 637f.).
– National educational standards in Germany (e.g. KMK 2003) are largely intended and considered as *average standards*, meaning that they describe the average level of expected achievement of students at the end of a certain schooling period. This approach is comparable to the approach of Dutch curricula (cf. article by van den Heuvel-Panhuizen & Wijers in this issue). On the other hand the national curriculum in Sweden (Skolverket 2004) decrees and expects minimum of competencies, they are conceived as *minimum standards* (cf. article by Kjellström & Pettersson in this issue). The difference in effect on the school system is a much discussed topic.
– The structure underlying the description of competencies, the so-called "competence-area model" is – due to its highly normative character – nationally and internationally varying among the standards (cf. article by Leuders, Barzel & Hußmann in this issue, see esp. fig. 3). Still one can find many similarities that reflect the state of theory in mathematics education. On the superficial level this can already be seen by the number of competence areas varying e.g. from eight (Niss 2003) through six (KMK 2003) to four "process-related" competences (cf. article by Leuders, Barzel & Hußmann in this issue).
– In the counselling paper by Klieme et al. (2003) it was recommended that the federal states should create curricula which should contain a description of core content, a chronology of topics and ideas for teaching thus supplementing the strictly outcome-oriented national standards. But already the first two federal curricula that were created did not follow this recommendation and adhered to a strict outcome paradigm (cf. article by Leuders, Barzel & Hußmann in this issue). Nevertheless further curricula which are recently prepared for implementation appear to take over the "input perspective" at least partially.

Looking at the newly created instruments of assessing the standards one can recognize a similar level of contingency. It should be the task of research in school improvement and mathematics education to explicate the specific differences and to enlighten their respective effects.

4. New answers or new questions?

The creation of new steering instruments is accompanied by projects of school improvement (e.g. SINUS, see above), by support in implementation given by administration institutions and by teacher training.

These measures can be regarded as answers to the unsatisfactory performance of the German education system. Whether these answers can effectively contribute to improve students' performance in the course of the next years is still an open question. To realise the chances inherent in this process and to avoid or at least control the risks one has to consider several key questions:

a) With regard to the concrete examples of contingency of the new steering instruments, one should ask:
   – Do outcome standards give sufficient orientation to teachers? Which further support with regard to creating content chronologies and adequate learning arrangements is needed? Can "opportunity to learn" standards be helpful?
   – When standards are used for examinations or for selecting students for the different school types, which

standards are more favourable: average or minimum standards?
- How do competence (area) models look like that are psychologically and empirically valid and at the same time useful and intelligible for teachers in practice?
- What is the sensible extent of a sustainable core in a so called core-curriculum?

b) In the initial phase of the new curricular paradigm many resources have to be used in creating standards and instruments for assessing students' achievement. On the other hand to improve the quality of teaching and learning it is necessary to focus on improving teaching as the pivotal element of innovation. Therefore this area should not be neglected in distributing resources into the education system. The question is: What kind of and how many resources are needed for developing standards and assessment strategies on one hand, and what and how many resources should be given for improving teaching on the other hand?

c) In the US there is much experience with so called "high-stakes tests" which steer the issue of certificates. The results of the test system are – not only from the point of view of mathematics education – regarded as problematic. The hopes of improving teaching quality have widely been deceived. Instead many undesirable side effects like „teaching to the test", reduction of motivation or increase of drop-out rates have emerged (cf. Amreiner & Berliner 2003). Especially the implicit redefinition and simplification of the curriculum by reducing teaching efforts on the testable competencies and on the actually tested content is a highly problematic consequence of any test instruments which are connected with gratifications or sanctions. The question is: Can such instruments still be considered adequate for quality improvement? How can such „washback"-effects probably be used in positive, productive way?

Questions like these (though the list is certainly not exhaustive) are important to investigate thoroughly and with empirical methods, so we can achieve reliable knowledge about the effect of political and didactical key decisions, some of which are considered by the authors in this issue.

5. Description and analysis of recent German experience – comparison with other European countries

Though at the moment there are hardly any empirically sound results with respect to the models, the processes and the effects of change in the German education system there should be an intensive discussion and thorough theoretical reflection of the steps already taken. This is even more important since the development of new instruments has not finished – many federal states are still in the process of creating new curricula and assessment structures and can profit from experience in other parts of the country and of other comparable countries.

Therefore this issue is dedicated to giving mathematics education researchers which have personally participated in the creation of new instruments the opportunity to report their experience and explicate their views and intentions. The German point of view is contrasted and supplemented by international experience from Sweden and the Netherlands – two countries that have received much attention with respect to their education system.

Three of the following analyses will deal with the definition of standards, the other three with the assessment of standards especially with questions of comparative tests and of feedback of results as a key element for school improvement.

Blum et al. report about the role of national standards in quality development. Leuders, Barzel & Hußmann describe the creation of standards on the level of a federal state – these standards are the texts the teachers work with and thus play a central role in initiating innovation. Van den Heuvel-Panhuizen & Wijers supplement this with the Dutch perspective on standards and curricula which is especially important due to its long tradition and its founding on the Freudenthal concept of „realistic mathematics education".

The second part of the analyses starts with the contribution of Kjellström & Pettersson who give a picture of the Swedish concept of assessment and diagnosis in mathematics and its reference to the curricular framework in Sweden. The recent concepts for comparative tests in the context of the German development are discussed by Lorenz (with respect to grade 4) and Büchter & Leuders (for grade 9). In the latter contribution there is an additional focus on the question of the quality of feedback instruments.

References

Amreiner, A. L. & Berliner, D. C. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership, 60 (5)*, p. 32-38.

Baumert, J. (2002). Deutschland in internationalen Bildungsvergleich. In N. Kilius, J. Kluge & L. Reisch (Hrsg.), *Zukunft der Bildung* (p. 100-150). Frankfurt a. M.: Suhrkamp.

Baumert, J., Cortina, K. S. & Leschinky, A. (2003). Grundlegende Entwicklungen und Strukturprobleme im allgemein bildenden Schulwesen. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* (p. 52-147). Reinbek bei Hamburg: Rowohlt.

Baumert, J., Lehmann, R. et al. (1997). *TIMSS. Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde.* Opladen: Leske+Budrich.

BLK (1997). *Gutachten zur Vorbereitung des Programms „Steigerung der Effizienz des mathematisch-naturwissenschaftlichen Unterrichts"*. Bonn: Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung. (http://www.blk-bonn.de/papers/heft60.pdf)

Deutsches PISA-Konsortium (Hrsg.) (2001). *PISA 2000 – Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich.* Opladen: Leske & Budrich.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. (2004). *The Development of National Educational Standards. An Expertise.* Bonn: Bundesministerium für Bildung und Forschung. (http://www.bmbf.de/publikationen/2713.php)

KMK (2003). *Bildungsstandards im Fach Mathematik für den mittleren Schulabschluss. Beschluss vom 4.12.2003.* Bonn: Kultusministerkonferenz. (http://www.kmk.org)

Lange, H. (1999). Qualitätssicherung in Schulen. *Die Deutsche Schule, 91 (2),* p. 144-159.

Leschinsky, A. & Cortina, K. S. (2003). Zur sozialen Einbettung bildungspolitischer Trends in der Bundesrepublik. In K. S. Cortina, J. Baumert, A. Leschinsky, K. U. Mayer & L. Trommer (Hrsg.), *Das Bildungswesen in der Bundesrepublik Deutschland. Strukturen und Entwicklungen im Überblick* (p. 20-51). Reinbek bei Hamburg: Rowohlt.

NCTM (2000). *Principles and Standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Niss, M. (2003). Mathematical Competencies and the Learning of Mathematics: The Danish KOM Project. In A. Gagatsis & S. Papastavridis (Eds.), *3rd Mediterranean Conference on Mathematical Education*. Athens – Hellas 3-5 January 2003. Athens: The Hellenic Mathematical Society, p. 115-124.

Reiss, K. (2004). Bildungsstandards und die Rolle der Fachdidaktik am Beispiel der Mathematik. *Zeitschrift für Pädagogik, 50 (5),* p. 635-649.

Rolff, H.-G. (2004). Zwei Linien der Steuerung der Qualität von Schulen? In H.G. Holtappels, K. Klemm, H. Pfeiffer, H.-G. Rolff & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung. Band 13. Daten, Beispiele und Perspektiven* (p. 97-123). Weinheim/München: Juventa.

Skolverket (2004). *Syllabuses fort he compulsory school.* Stockholm: Skolverket & Fritzes.

Authors

Büchter, Andreas, Institut für Entwicklung und Erforschung des Mathematikunterrichts, Universität Dortmund, D-44221 Dortmund.
E-Mail: andreas@buechter.net

Leuders, Timo, Prof. Dr., Institut für Mathematik und Informatik und ihre Didaktiken, Pädagogische Hochschule Freiburg, Kunzenweg 21, D-79117 Freiburg.
E-Mail: leuders@ph-freiburg.de