Part III
The Literature Reviews

Changing Teaching through Formative Assessment: Research and Practice

The King's-Medway-Oxfordshire Formative Assessment Project

by
Paul Black and Dylan Wiliam, King's College, London*

INTRODUCTION

This paper is the story of a development which started with a review of what research had to say about formative assessment. The work of this review is first described. Its results led to development work with teachers to explore how ideas taken from the research could be turned into practice. A description of this work in a second section is followed by reflections on outcomes and implications in a third section. Broader reflections on how this experience throws light on the task of turning research results into practice are set out in a fourth section.

THE RESEARCH REVIEW

The story starts with our long-standing interest in formative assessment, which led us to decide that it was essential to review the literature in order to look for evidence that improving formative assessment raises standards. It also seemed necessary to look both for evidence about whether or not present practice left room for improvement, and for guidance about how to improve formative assessment.

Our survey of the research literature involved checking through many books, through the issues of over 160 journals for a period of nine years, and studying earlier reviews of research (Crooks, 1988; Natriello, 1987). This process yielded about 580 articles or chapters to study. Out of this we have

* We would like to acknowledge the initiative of the Assessment Policy Task Group of the British Educational Research Association (now known as the Assessment Reform Group) who gave the initial impetus and support for our research review. We are grateful to the Nuffield Foundation who funded the original review, and the first phase of our project. We are also grateful to Professor Myron Atkin and his colleagues in Standford University, who secured funding from the US National Science Foundation (NSF) for the last phase. Finally, we are indebted to the Medway and Oxfordshire local education authorities, their six schools and above all to their 36 teachers who took on the central and risky task of turning our ideas into practical working knowledge.

prepared a lengthy review, which used material from 250 of these sources. The review was published (Black and Wiliam, 1998a) together with comments on our work by experts from five different countries.

A ***first*** section of the review surveyed the evidence. An example was a study published in 1986, which concentrated – but not exclusively – on classroom assessment work for children with mild handicaps, and surveyed a large number of formative innovations from which 23 were selected (Fuchs and Fuchs, 1986). All in this group showed quantitative evidence of learning gains by comparing data for an experimental group with similar data from a control group. Since then, many more papers have been published describing similarly rigorous quantitative experiments. Our own review reported about 20 more such studies all of which showed that innovations which include strengthening the practice of formative assessment produced significant, and often substantial, learning gains. These studies ranged over ages (from 5-year olds to university undergraduates), across several school subjects, and over several countries.

The fact that such gains had been achieved by a variety of methods which had, as a common feature, enhanced formative assessment indicated that it is this feature which accounted, at least in part, for the successes. However, it did not follow that it would be an easy matter to achieve such gains on a wide scale in normal classrooms.

A ***second*** section covered research into current practices of teachers. The picture that emerged was depressing. In relation to effective learning it seemed that teachers' questions and tests encouraged rote and superficial learning, even where teachers said that they wanted to develop understanding. There was also evidence of the negative impact of a focus on comparing students with one another, so emphasising competition rather than personal improvement. Furthermore, teachers' feedback to students often seemed to serve social and managerial functions, often at the expense of the learning functions. Overall it seemed that formative assessment was weak in practice and that its implementation calls for rather deep changes both in teachers' perceptions of their own role in relation to their students and in their classroom practice.

A ***third*** section focused on research into the involvement of students in formative assessment. Students' beliefs about the goals of learning, about the risks involved in responding in various ways, and about what learning work should be like, were all shown to affect their motivation to take action, their selection of a line of action and the nature of their commitment to it. Other research explored the different ways in which positive action could be taken, covering such topics as study methods, study skills, and peer- and self-assessment.

A ***fourth*** section looked at ideas that could be gleaned from the research about strategies that might be productive for teachers. One feature that emerged was the potential of the learning task, as designed by a teacher, for exploring students' learning. Another was the importance of the classroom discourse, as steered by teachers' questions and by their handling of students' responses.

A ***fifth*** section shifted attention to research into comprehensive systems of teaching and learning in which formative assessment played a part. One example was mastery learning programmes. In these it was notable that students were given feedback on their current achievement against some expected level of achievement (ie the 'mastery' level), that such feedback was given rapidly; and that students were given the opportunity to discuss with their peers how to remedy any weaknesses.

A ***sixth*** section explored in more detail the literature on feedback. A notable example was the extensive review of empirical evidence by Kluger and DeNisi (1996) which showed that feedback can have positive effects only if the feedback is formulated and used as a guide to improvement. Of equal importance was the conceptual analysis which defined feedback as "… information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way" (Ramaprasad, 1983) and the development of this by Sadler (1989) to emphasise that learners must understand both the "reference level" – *i.e.* the goal of their learning – and the actual level of their understanding.

Equally important was the clear message from the research on attribution theory (for example by Vispoel and Austin, 1995) that teachers must aim to inculcate in their students the idea that success is due to internal, unstable, specific factors such as effort, rather than on stable general factors such as ability (internal) or whether one is positively regarded by the teacher (external).

Overall, the features which seem to characterise many of the studies were:

- Formative work involves new ways to enhance feedback between those taught and the teacher, ways which require new modes of pedagogy and significant changes in classroom practice.
- Underlying the various approaches are assumptions about what makes for effective learning – in particular that students have to be actively involved.
- For assessment to function formatively, the results have to be used to adjust teaching and learning – so a significant aspect of any programme will be the ways in which teachers do this.

- The ways in which assessment can affect the motivation and self-esteem of students, and the benefits of engaging students in self-assessment, both deserve careful attention.

Interpreting the research

Synthesising research cannot be an objective process – it will inevitably remain subjective. The structure of the six sections outlined above did not emerge automatically: it was our chosen way to reconceptualise, to organise, and to focus the relevant literature field. Our definition of "relevance" expanded as we went along, so we had to find ways of organising a widening field of research, and to make new conceptual links in order to be able to combine the various findings into as coherent a picture as possible. This was one reason why our review generated a momentum for work in this field: it provided a new framework that would be difficult to create in any other way. Reviewing research is not merely a derivative form of scholarship.

Publicity

Although we tried to adhere closely to the traditional standards of scholarship in the social sciences when conducting and writing our review, we did not do so when exploring the policy implications in a booklet, entitled *Inside the Black Box* (Black and Wiliam, 1998b) that we published, and publicised widely, alongside the academic review. This raised a great deal of interest and created some momentum for our project and for subsequent dissemination. While the standards of evidence we adopted in conducting the review might be characterised as those of "academic rationality", the standard for *Inside the Black Box* was much closer to that of "reasonableness" advocated by Stephen Toulmin for social enquiry (Toulmin, 2001). In some respects, *Inside the Black Box* represented our opinions and prejudices as much as anything else, although we would like to think that these are supported by evidence, and are consistent with the 50 years of experience in this field that we had between us. It is also important to note that the success of *Inside the Black Box* has been as much due to its rhetorical force as to its basis in evidence. This would make many academics uneasy – for it appears to blur the line between fact and value, but as Flyvbjerg (2001) argues, social enquiry has failed precisely because it has focused on analytic rationality rather than value-rationality (see also Wiliam, 2003).

MOVING INTO ACTION

Setting up a project

The second stage of our story followed the first almost inevitably: given that our review had shown that innovations in formative assessment could raise standards of student achievement, it was natural to think about ways to help schools secure these benefits. Our own experience of teachers' professional development had taught us that the implementation of new practices in classrooms could not be a straightforward matter of setting out a recipe for teachers to follow. For one reason, given the varied nature of the innovations and the different contexts in which they had been tried, we could not assume that they could simply be "copied" to other contexts. A second reason was that, from reading the reports of the researchers, one could not describe their work at the level of detail that would be needed to formulate advice on how to replicate them. A third reason, which would have been decisive even in the absence of the first two, was our approach to the task of turning research into practice. We believed that new ideas about teaching and learning can only be made to work in particular contexts, in our case that of teachers in (initially) UK secondary schools, if teachers are able to transform them and so create new practical knowledge relevant to their task.

So we obtained funding (from the UK's Nuffield Foundation) for a two-year development project. Six schools who taught students in the age range 11 to 18 years agreed to collaborate with us: each selected two science and two mathematics teachers willing to take on the risks and extra work involved. In second year of the project we added two teachers of English, from each of same schools, and one additional mathematics and science teacher, so that in all 48 teachers were involved. They were supported by staff from their local (district) education authorities and the project was called the King's-Medway-Oxfordshire Formative Assessment Project (KMOFAP) to highlight our close collaboration with all the other partners (Black and Wiliam, 2003).

The teachers and the researchers met in a whole day meeting every five weeks, over two years. In addition, two researchers were able to visit the schools, observe the teachers in their classrooms, give them feedback, collect interview data on their perceptions, and elicit ideas about issues for discussion in the whole day meetings. The detailed reports of our findings (Black *et al.*, 2002, 2003) are based both on records of these meetings, on the observations and records of visits to classrooms by the King's team, on interviews with and writing by the teachers themselves, and on a few discussions with student groups.

Following this project, members of the King's team have responded to numerous invitations to talk to other groups: over three years they have made over 200 such contributions. These have ranged across all subjects, and across both primary and secondary phases. In addition, there has been sustained work with four groups of primary schools. The King's team has also been involved as advisers to large scale development ventures, in several local government districts in the United Kingdom, with education ministries in Scotland and in Jersey, and in a recent exploration of classroom outcomes for a government programme which aims to improve teaching and learning practices in schools.

The quantitative evidence that formative assessment does raise standards of achievement was a powerful motivator for the teachers at the start of the project. One aspect of the KMOFAP project was that the King's team worked with each teacher to collect data on the gains in test performance of the students involved in the innovation, and comparable data for similar classes who were not involved (Wiliam *et al*. 2004). The project did not introduce any tests of its own – the achievement data used were from the tests that the schools used for all students, whether or not they were involved in the project. The analysis of these data showed an overall and significant, gain in achievement outcomes. Thus the evidence from the research review can now be supplemented by evidence of enhanced performance on the UK national and on schools' own examinations.

The practices developed

These practices will be described here under four headings: oral feedback in *classroom dialogue*, feedback *through marking*, *peer- and self-assessment*, and the *formative use of summative tests*. The account given will be brief – more detailed accounts have been published elsewhere (Black *et al.*, 2003).

For *classroom dialogue* the aim was to improve the interactive feedback which is central to formative assessment. An account of wait time research (Rowe, 1974) motivated teachers to allow longer time after asking a question so that students would have time to think out responses, and so that all could be expected to become actively involved in question and answer discussions, and to make longer replies. One particular way to increase participation was to ask students to brainstorm ideas, perhaps in pairs, for two to three minutes prior to the teacher asking for contributions. Then all answers, right or wrong, had to be taken seriously, the aim being to develop thoughtful improvement rather to evoke the expected answers. A consequence of such changes was that teachers learnt more about the pre-knowledge of their students, and about any gaps and mis-conceptions in that knowledge, so that their next moves could address the learners' real needs.

As they tried to develop this approach, teachers realised that more effort had to be spent in framing questions that were worth asking, *i.e.* questions which explored issues that are critical to the development of students' understanding. They also had to focus closely on follow-up activities to formulate meaningful responses and challenges that would help students to extend their understanding.

The task of developing an interactive style of classroom dialogue required a radical change in teaching style from many teachers, one that they found challenging, not least because it felt at first as if they were losing control. Some were well over a year into the project before such change was achieved. Subsequent work with other schools has shown that it is this aspect of formative work that teachers are least likely to implement successfully.

To address *feedback through marking*, teachers were first given an account of research studies which have established that, whilst students' learning can be advanced by feedback through comments, the giving of marks or grades has a negative effect because students ignore comments when marks are also given (Butler, 1988). These results surprised and worried the teachers, because of concern about the effect of returning students' work with comments but no marks. However, potential conflicts with school policy were resolved as experience showed that the provision of comments gave both students and their parents advice on how to improve. It also set up a new focus on the learning issues rather than on trying to interpret a mark or grade. To make the most of the learning opportunity created by feedback on written work, procedures that required students to follow up comments had to be planned as part of the overall learning process.

One consequence of this change was that teachers had to think more carefully in framing comments on written work, for it was now evident that these had to identify what had been done well and what still needed improvement, and to give guidance on how to make that improvement. As the skills of formulating and using such feedback were developed, it became more clear that the quality of the tasks set for written homework or class-work was critical: such tasks, alongside oral questioning, had to be designed to encourage students to develop and express their understanding of the key features of what they had learnt.

For *peer- and self-assessment,* the starting point was Sadler's (1989) argument that self-assessment is essential to learning because students can only achieve a learning goal if they understand that goal and can assess what they need to do to reach it. Thus the criteria for evaluating any learning achievements must be made transparent to students to enable them to have a clear overview both of the aims of their work and of what it means to complete it successfully. Insofar as they do so they begin to develop an

overview of that work so that they can manage and control it: in other words, they develop their capacity for meta-cognitive thinking. A notable example of the success of such work is the research of White and Frederiksen (1998).

For the development of self-assessment skills, the first and most difficult task is to get students to think of their work in terms of a set of goals. In practice, peer-assessment turned out to be an important stimulus to self-assessment. Peer-assessment is uniquely valuable because students may accept, from one another, criticisms of their work which they would not take seriously if made by their teacher. Peer work is also valuable because the interchange will be in language that students themselves would naturally use, and because students learn by taking the roles of teachers and examiners of others (Sadler, 1998). In particular, students appear to find it easier to make sense of criteria for their work in the context of other students' work than when looking at their own.

However, for such peer-group work to succeed, many students needed guidance about how to behave in groups, *e.g.* in listening to one another, taking turns, and offering affirmation together with constructive criticism about one another's work. A typical exercise would be on the marking of homework. Students were asked to label their work with "traffic lights", *i.e.* using red or amber if they were totally or partially unsure of their success, and green where they were confident. Then those who had used amber or green would work in mixed groups to appraise and help with one another's work, whilst the teacher would pay special attention to those who had chosen red.

Teachers developed three ways of making *formative use of summative tests*. One way was to ask students, in preparation for a test, to "traffic light" a list of key words or of the topics on which the test would be set, an exercise which would stimulate them to reflect on where they felt their learning was secure and where they needed to concentrate their efforts. One reason for doing this was that teachers had realised that many students had no strategy for preparing for a test by formulating a strategic appraisal of their learning.

A second way was to mark one another's test papers in peer groups, in the way outlined above for the marking of homework. This could be particularly challenging when they were expected to invent their own marking rubric, for to do this they had to think about the purpose of a question and about the criteria of quality to apply to responses. After peer marking, teachers could reserve their time for discussion of the questions that give particular difficulty.

A further idea was introduced from research studies (Foos *et al.*, 1994; King, 1992) which have shown that students trained to prepare for examinations by generating and then answering their own questions out-performed comparable groups who prepared in conventional ways. Preparation of test questions calls for, and so develops, an overview of the topic.

The teachers' work on summative assessments challenged our expectations that, for the context in which they worked, formative and summative assessments are so different in their purpose that they have to be kept apart. The finding that emerged was quite different – that summative tests should be, and should be seen to be, a positive part of the learning process. If they could be actively involved in the test process, students might see that they can be beneficiaries rather than victims of testing, because tests can help them improve their learning. However, this synergy could not be achieved in the case of high-stakes test set and marked externally.

REFLECTIONS ON THE OUTCOME

It was clear that the new ideas that had emerged between the teachers and ourselves involved far more than the mere addition of a few tactical tricks. Some reflection was needed to tease out more fundamental issues that seemed to be raised.

A focus on learning

One of the most surprising things that happened during the early project meetings was that the participating teachers asked us to run a session on learning theories. In retrospect, perhaps, we should not have been so surprised. We had, after all, stressed that feedback functioned formatively only if the information fed back to the learner was used by the learner in improving performance. But whilst one can work out after the event whether or not any feedback has had the desired effect, what the teachers needed was to be able to give their students feedback that they knew in advance was going to be useful. To do that they needed to build up models of how students learn.

So the teachers came to take greater care in selecting tasks, questions, and other prompts, to ensure that the responses made by students actually "put on the table" the ideas which they bring to a learning task. The key to effective learning is to then find ways to help students restructure their knowledge to build in new and more powerful ideas. In the KMOFAP classrooms, as the teachers came to listen more attentively to the students' responses, they began to appreciate more fully that learning is not a process

of passive reception of knowledge, but one in which the learners must be active in creating their own understandings.

These ideas reflect some of the main principles of the constructivist view of learning – to start where the students are and to involve the students actively in the process. It became clear to the teachers that, no matter what the pressure to achieve good test and examination scores, learning cannot be done for the student; it has to be done by the student.

Students came to understand what counted as good work through a focus on the criteria and on their exemplification. Sometimes this was done through focused whole-class discussion around a particular example; at others it was achieved through students using criteria to assess the work of their peers. The activities, by encouraging students to review their work in the light of the goals and criteria, were helping them to develop meta-cognitive approaches to learning.

Finally, the involvement of students both in whole-class dialogue and in peer-group discussions, all within a change in the classroom culture to which all four activities contributed, were creating more a more rich community of learners where the social learning of students would become more salient and effective.

A learning environment and changes of role

There are also deeper issues here. A learning environment has to be "engineered" to involve students more actively in the tasks. The emphasis has to be on the students doing the thinking and making that thinking public. As one teacher said:

> There was a definite transition at some point, from focusing on what I was putting into the process, to what the students were contributing. It became obvious that one way to make a significant sustainable change was to get the students doing more of the thinking. I then began to search for ways to make the learning process more transparent to the students. Indeed, I now spend my time looking for ways to get students to take responsibility for their learning and at the same time making the learning more collaborative.
>
> Tom, Riverside School

This teacher had changed his role, from presenter of content to leader of an exploration and development of ideas in which all students were involved. One of the striking features of the project was the way in which, in the early stages, many spoke about the new approach as "scary", because they felt that they were losing control of their classes. Toward the end of the

project, they described this same process not as a loss of control, but one of sharing responsibility for the class's learning with the class – exactly the same process, but viewed from two very different perspectives.

The learning environment envisaged requires a classroom culture that may well be unfamiliar and disconcerting for both teachers and students. The effect of the innovations implemented by our teachers was to change the rules, usually implicit, that govern the behaviours that are expected and seen as legitimate by teachers and by students. As Perrenoud (1991) put it:

> Every teacher who wants to practice formative assessment must reconstruct the teaching contract so as to counteract the habits acquired by his pupils.

For the students, they have to change from behaving as passive recipients of the knowledge offered to becoming active learners who could take responsibility for their own learning. These students became more aware of when they were learning, and when they were not. One class, who were subsequently taught by a teacher not emphasising assessment for learning, surprised that teacher by complaining: "Look, we've told you we don't understand this. Why are you going on to the next topic?".

What has been happening here is that everybody's role expectations, *i.e.* what teachers and students think that being a teacher or being a student requires you to do, have been altered. Whilst it can seem daunting to undertake such changes, they do not have to happen suddenly. Changes with the KMOFAP teachers came slowly and steadily, as experience developed and confidence grew in the use of the various strategies for enriching feedback and interaction.

Further research

In our 1998 review, we listed a number of issues for study by further research. The first issue was the extent to which the context of any study is artificial so that generalisability of the results cannot be guaranteed. This reservation was one of the reasons why we developed the KMOFAP work and now it can be applied to the generalisability of the findings of that study. Our experience of seeing other schools base their own innovations on the KMOFAP results is that a sustained commitment over at least two years is needed, that evaluation and feedback have to be built into any plan, and that any teachers involved need strong support, both from colleagues and from their school leadership.

A second research interest arose from a surprising feature – that the research we studied seemed to pay no attention to issues relating to race, class

and gender; these issues still await exploration. A third area for further enquiry is that of beliefs and assumptions about learning theory. Both the assumptions about learning underlying the curriculum and pedagogy, the beliefs of teachers about learning, about their roles as assessors and about the "abilities" and prospects of their students, will affect their interpretations of their students' learning work, and will thereby determine the quality of their formative assessment. A parallel enquiry is needed into the perceptions and beliefs held by students about themselves as learners, and into their experience of the changes that follow from innovations in formative assessment.

A fourth area is the effect on practice of the content knowledge, and the pedagogical content knowledge, that teachers deploy in their school subjects. Issues for enquiry would be the way in which these resources underlie each teacher's composition and presentation of the learning work, and the interpretative frameworks that he or she uses in responding to the evidence provided by feedback from students.

The social setting of a classroom, the community it forms, and the quality of the interactions within that community, all have strong effects in such innovations as better classroom dialogue and peer- and self-assessment. Matters to be studied here would the nature of the social setting in the classroom, as influenced both by the divisions of responsibility between learners and teachers in formative assessment, and by the constraints of the wider school system.

Two further issues now seem important. One is the tensions and possible synergies between teachers' own assessments and the assessment results and methods required by society. The other is the need to co-ordinate all of the above issues in a comprehensive theoretical framework linking assessment in classrooms to issues of pedagogy and curriculum – a task which remains to be tackled.

RESEARCH AND PRACTICE

Why did it work?

At one level, our story was now complete. A basis in research had led to a successful innovation and the publication of its outcomes proved as popular as the original report of the research (Black *et al.*, 2002, 2003). However, we were surprised that it had been so successful in promoting quite radical changes in teachers' practice, and wondered whether lessons could be learnt from it about the notoriously difficult problem of turning research into practice.

One factor that appears to have been important is the credibility that we brought as researchers to the process. In their project diaries, several of the

teachers commented that it was our espousal of these ideas, as much as the ideas themselves, that persuaded them to engage with the project: where educational research is concerned, the facts do not necessarily speak for themselves. Part of that credibility is that we chose to work with teachers in the three subjects, English, mathematics and science when, in each of these, one or two members of the team had expertise and reputations in the subject community. Thus, when specific issues, such as "Is this an appropriate question for exploring students ideas about the concept of photosynthesis?" arose, we could discuss them seriously.

A further relevant factor about the content is that the ideas had an intrinsic acceptability to the teachers. We were talking about improving learning in the classroom, which was central to their professional identities, as opposed to bureaucratic measures such as target-setting. One feature of our review was that most of it was concerned with such issues as students' perceptions, peer- and self-assessment, and the role of feedback in a pedagogy focused on learning. Thus it helped to take the emphasis in formative assessment studies away from *systems,* with its emphasis on the formative-summative interface, and re-locate it on classroom *processes*.

Linked to the previous factor is that in our choice to concentrate on the classroom processes, we had decided to live with the external constraints operating at the formative-summative interface: the failed attempts to change the *system,* in the 80s and 90s in England, were set aside. Whilst it might have been merely prudent to not try again to tilt at windmills, the more fundamental strength was that it was at the level chosen, that of the core of learning, that formative work stakes its claim for attention. Furthermore, given that any change has to work out in teachers' practical action, this is where reform should always have started. The evidence of learning gains, from the literature review and from our project, restates and reinforces the claim for priority of formative work that earlier policy recommendations (DES, 1988) tried in vain to establish. The debate about how policy should secure optimum synergy between teachers' formative, teachers' summative, and external assessments is still unresolved, but the new salience of work on formative assessment has now shifted the balance of the arguments.

The process strategy

In our development model, we attended to both the content and the process of teacher development (Reeves *et al*., 2001). We attended to the process of professional development through an acknowledgement that teachers need time, freedom, and support from colleagues, in order to reflect critically upon and to develop their practice (Lee, 2005), whilst offering also

practical strategies and techniques about how to begin the process. By themselves, however, these are not enough. Teachers also need concrete ideas about the directions in which they can productively take their practice, and thus there is a need for work on the professional development of teachers to pay specific attention to subject-specific dimensions of teacher learning (Wilson and Berne, 1999).

One of the key assumptions of the project was that if the promise of formative assessment was to be realised, traditional research designs – in which teachers are "told" what to do by researchers – would not be appropriate. We argued that a process of supported development was an essential next step. In such a process, the teachers in their classrooms had to work out the answers to many of the practical questions that the research evidence could not answer. The issues had to be reformulated in collaboration with them, where possible in relation to fundamental insights, and certainly in terms that could make sense to their peers in ordinary classrooms.

The key feature of the INSET sessions was the development of action plans. Since we were aware from other studies that effective implementation of formative assessment requires teachers to re-negotiate the "learning contract" that they had evolved with their students (Brousseau, 1984; Perrenoud, 1991), we decided that implementing formative assessment would best be done at the beginning of a new school year. For the first six months of the project (January 1999 to July 1999), therefore, we encouraged the teachers to experiment with some of the strategies and techniques suggested by the research, such as rich questioning, comment-only marking, sharing criteria with learners, and student peer- and self-assessment. Each teacher was then asked to draw up an action plan of the practices they wished to develop and to identify a single focal class with whom these strategies would be introduced at the start of the new school year in September 1999. Details of these plans can be found in Black *et al*. (2003). As the teachers explored the relevance of formative assessment for their own practice, they transformed ideas from the research and from other teachers into new ideas, strategies and techniques, and these were in turn communicated to teachers, creating a "snowball" effect. As we have introduced these ideas to more and more teachers outside the project, we have become better at communicating the key ideas.

Through our work with teachers, we have come to understand more clearly how the task of applying research into practice is much more than a simple process of "translating" the findings of researchers into the classroom. The teachers in our project were engaged in a process of knowledge creation, albeit of a distinct kind, and possibly relevant only in the settings in which they work (Hargreaves, 1999). We stressed this feature of our approach with the teachers right from the outset of the project. We

discovered later that some of them did not, at that stage, believe us: they thought that we knew exactly what we wanted them to do but were leaving them to work it out for themselves. As they came to know us better, they realised that, at the level of everyday classroom practice, we really did not know what to do.

Making research practical

Whilst we do not believe that all educational research should be useful, we do believe strongly that the majority of research in education should be undertaken with a view to improving educational provision – research in what Stokes (1997) calls "Pasteur's quadrant". And although we do not yet know everything about "what works" in teaching, we believe that there is a substantial consensus on the kinds of classrooms that promote the best learning. What we know much less about is how to get this to happen.

Researching how teachers take on research, adapt it, and make it their own is much more difficult than researching the effects of different curricula, of class sizes, or of the contribution of classroom assistants. While we do not know as much as we would like to know about effective professional development, if we adopt "the balance of probabilities" rather than "beyond reasonable doubt" as our burden of proof, then educational research has much to say. When policy without evidence meets development with some evidence, development should prevail. Thus we take issue with the stance of some policy makers who appear to want large-scale research conducted to the highest standards of analytic rationality, but the findings of which are also relevant to policy. It may often be the case that these two goals are, in fact, incompatible.

References

Black, P. and D. Wiliam (1998a), "Assessment and Classroom Learning", *Assessment in Education,* Vol. 5, pp. 7-71.

Black, P. and D. Wiliam (1998b), *Inside the Black Box: Raising Standards through Classroom Assessment*. See also *Phi Delta Kappan,* Vol. 80, pp. 139-148.

Black, P. and D. Wiliam (2003), "In Praise of Educational Research: Formative Assessment", *British Educational Research Journal*, Vol. 29(5), pp. 623-637.

Black, P. *et al.* (2002), *Working inside the Black Box: Assessment for Learning in the Classroom*, Department of Education and Professional Studies, King's College, London.

Black, P. *et al.* (2003), *Assessment for Learning: Putting it into Practice*, Open University Press, Buckingham, United Kingdom.

Brousseau, G. (1984), "The Crucial Role of the Didactical Contract in the Analysis and Construction of Situations in Teaching and Learning Mathematics" in H.-G. Steiner (ed.), *Theory of Mathematics Education: ICME 5 Topic Area and Miniconference*, Institut für Didaktik der Mathematik der Universität Bielefeld, Bielefeld, Germany, pp. 110-119.

Butler, R. (1988), "Enhancing and Undermining Intrinsic Motivation; the Effects of Task-involving and Ego-involving Evaluation on Interest and Performance", *British Journal of Educational Psychology,* vol. 58, pp. 1-14.

Crooks, T.J. (1988), "The Impact of Classroom Evaluation Practices on Students", *Review of Educational Research,* Vol. 58, pp. 438-481.

DES (1988), *Task Group on Assessment and Testing: A Report,* Department of Education and Science and the Welsh Office, London.

Flyvbjerg, B. (2001), *Making Social Science Matter: Why Social Inquiry Fails and how it can Succeed again*, Cambridge University Press, Cambridge UK.

Foos, P.W., J.J. Mora and S. Tkacz (1994), "Student Study Techniques and the Generation Effect", *Journal of Educational Psychology,* Vol. 86, pp. 567-576.

Fuchs, L.S. and D. Fuchs (1986), "Effects of Systematic Formative Evaluation: a Meta-Analysis", *Exceptional Children*, Vol. 53, pp. 199-208.

Hargreaves, D.H. (1999), "The Knowledge Creating School", *British Journal of Educational Studies*, Vol. 47, pp. 122-144.

King, A. (1992), "Facilitating Elaborative Learning through Guided Student-generated Questioning", *Educational Psychologist,* Vol. 27, pp. 111-126.

Kluger, A.N. and A. DeNisi (1996), "The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory", *Psychological Bulletin,* Vol. 119, pp. 254-284.

Lee, C. (2005), "Studying Changes in the Practice of Two Teachers", *International Journal of Teacher Development*, in press.

Natriello, G. (1987), "The Impact of Evaluation Processes on Students", *Educational Psychologist,* Vol. 22, pp. 155-175.

Perrenoud, P. (1991), "Towards a Pragmatic Approach to Formative Evaluation", in P. Weston (ed.), *Assessment of Pupils Achievement: Motivation and School Success*, Swets and Zeitlinger, Amsterdam, pp. 79-101.

Ramaprasad, A. (1983), "On the Definition of Feedback", *Behavioral Science,* Vol. 28, pp. 4-13.

Reeves, J., J. McCall and B. MacGilchrist (2001), "Change Leadership: Planning, Conceptualization and Perception", in J. MacBeath and P. Mortimore (eds.), *Improving school effectiveness,* Open University Press, Buckingham UK, pp. 122-137.

Rowe, M.B. (1974), "Wait Time and Rewards as Instructional Variables, their Influence on Language, Logic and Fate Control", *Journal of Research in Science Teaching*, Vol. 11, pp. 81-94.

Sadler, R. (1989), "Formative Assessment and the Design of Instructional Systems", *Instructional Science,* Vol. 18, pp. 119-144.

Sadler, R. (1998), "Formative Assessment: Revisiting the Territory", *Assessment in Education*, Vol. 5, pp. 77-84.

Stokes, D.E. (1997), *Pasteur's Quadrant: Basic Science and Technological Innovation*, Brookings Institution Press, Washington DC.

Toulmin, S. (2001), *Return to Reason*, Harvard University Press, Cambridge MA.

Vispoel, W.P. and J.R. Austin (1995), "Success and Failure in Junior High School: A Critical Incident Approach to Understanding Students' Attributional Beliefs", *American Educational Research Journal,* Vol. 33, pp. 377-412.

White, B.Y. and J.R. Frederiksen (1998), "Inquiry, Modeling, and Metacognition: Making Science Accessible to all Students", *Cognition and Instruction*, Vol. 16, pp. 3-118.

Wiliam, D. (2003), "The Impact of Educational Research on Mathematics Education" in A. Bishop, M.A. Clements, C. Keitel, J. Kilpatrick and F.K.S. Leung (eds.), *Second International Handbook of Mathematics Education*, Kluwer Academic Publishers, Dordrecht Netherlands, pp. 469-488.

Wiliam, D. *et al.* (2004), "Teachers Developing Assessment for Learning: Impact on Student Achievement", *Assessment in Education,* Vol. 11, pp.49-65.

Wilson, S.M. and J. Berne (1999), "Teacher Learning and the Acquisition of Professional Knowledge: An Examination of Research on Contemporary Professional Development", in A. Iran-Nejad and P.D. Pearson (eds.), *Review of Research in Education,* American Educational Research Association, Washington DC, pp. 173-209.

Formative Assessment of Learning: A Review of Publications in French

by
Linda Allal and Lucie Mottier Lopez
University of Geneva

The concept of "formative evaluation" was introduced by Scriven (1967) in an article on the evaluation of educational programmes (curricula, methods, instructional material). For Scriven, formative evaluation aims at providing data that permit successive adaptations of a new programme during the phases of its development and its implementation. Bloom (1968) quickly incorporated the idea of formative evaluation – applied to student learning – into his newly defined model of mastery learning. The characteristics of this function of evaluation were spelled out in considerable detail in subsequent publications (Bloom, 1976; Bloom, Hasting and Madaus, 1971). Over the years, an extensive literature has accumulated in English concerning formative assessment (the term "assessment" having progressively replaced "evaluation" when the object is student learning in the classroom). This literature is well-known to educational researchers in many areas of the world. On the other hand, the work carried out and published in other languages (French, German, Spanish, etc.) is relatively unknown in the English-language community. The present review is aimed at fostering international dissemination of work on formative assessment published in French over the past 25 years.[1]

Our review is based on publications by researchers and assessment specialists in France and in the French-speaking regions of Belgium, Canada, and Switzerland. To carry out the review we constructed a database composed of over 100 journal articles published in the major French-language journal in the area of assessment. We also consulted a number of key books, especially those resulting from conferences organised by the French-language associations on assessment. The review is focused on formative assessment of student learning in elementary and secondary school settings but takes into account developments in other contexts

1 We thank Janet Looney for inviting us to prepare this review in the context of an OECD/CERI project on "What works?" in the area of formative assessment of student learning. The development of the review benefited from exchanges we had in Geneva and Paris.

(particularly teacher training and higher education) that have influenced the conception and practice of formative assessment in the classroom. The first part of the review describes the material on which the review is based, its origin and coverage. The second part defines the major conceptual orientations of formative assessment in the French-language literature. The third part presents a classification of the types of empirical research that have been carried out on formative assessment.

COVERAGE OF THE REVIEW

Our database is composed of articles appearing in the journal *Mesure et évaluation en éducation* (*Measurement and Assessment in Education*).[2] The journal, initially entitled *Mesure en éducation*, was founded in 1978 by professionals in charge of school examinations in Québec. Several years later, university specialists in measurement and assessment took on a major role in the editorial board and the present title of the journal was adopted. In 1986, the editorial board was enlarged to include two sub-committees, one composed of members from universities and research centers in Québec, the other of members from European universities and research institutions in Belgium, France and Switzerland. It is worth noting that *Mesure et évaluation en éducation* is the only international, peer-reviewed journal published in French which specialises in questions of educational assessment.

From the beginning, the journal was sponsored by an active Québec association: the *Association Professionnelle de Mesure en Éducation*, which became the *Association pour le Développement de la Mesure et de l'Évaluation en Éducation*. In 1985, a parallel association was created in Europe: *Association pour le Développement des Méthodologies d'Évaluation en Éducation*. Although the two associations share the same acronym (ADMEE), their names differ in one slight but significant respect: the word *mesure* in the Canadian version is replaced by *méthodologies* in the European version. These choices are a reflection of cultural attitudes toward the concept of measurement in the research communities of the two continents. While in Canada, measurement and assessment (or evaluation) go hand in hand, in much of French-speaking Europe, there is a tendency to prefer qualitative assessment without the operations of quantification associated with measurement (for a discussion of this question, see Allal, 1997). Despite these differences, the two ADMEE associations have closely collaborated in the edition of a common journal. The annual conferences of each association attract a wide range of researchers, professionals and practitioners who work

[2] In contrast with English where the term "assessment" has replaced "evaluation" when the object is student learning, the word *évaluation* is used in French both for student assessment and for programme evaluation.

in the area of educational assessment, including participants and keynote speakers from the other side of the Atlantic. In addition, several joint conferences between the two associations have been held.

The database used for this review is composed of 105 articles published in the journal *Mesure et évaluation en éducation* between 1978 and 2002.[3] It includes articles that deal directly with formative assessment or that address issues of importance for formative assessment (*e.g.*, articles on observation methods or on new means of summative assessment that have implications for formative assessment). For each article in the database, a summary was made of the theoretical orientations that were presented and the empirical research that was reported. A coding scheme was applied to facilitate identification of various theoretical and empirical dimensions.

In addition, we examined the chapters appearing in six edited books that resulted from ADMEE conferences on assessment: Allal, Cardinet and Perrenoud (1979), De Ketele (1986), Depover and Noël (1999), Figari and Achouche (2001), Laveault (1992), Weiss, 1991. We also consulted two edited books (Grégoire, 1996a; Hivon, 1993) presenting work from symposia on assessment organised by another French-language network (*Réseau Éducation et Formation*), as well as several other well-known books in the field (Allal, Bain and Perrenoud, 1993; Bélair, 1999; Bonniol and Vial, 1997; Cardinet, 1986a, 1986b; Hadji, 1989, 1997; Huberman, 1988; Louis, 1999; Perrenoud, 1998a; Scallon, 2000).

CONCEPTUALISATION OF FORMATIVE ASSESSMENT

The initial conception of formative assessment proposed by Bloom has been enlarged in several directions by researchers working in French. After a presentation of the main orientations of this enlargement, four successive developments in French-language research on formative assessment will be described.

Enlarging the conception of formative assessment

In the initial conception of mastery learning proposed by Bloom (1968; Bloom *et al.*, 1971), an instructional unit is divided into several successive phases. First of all, teaching/learning activities are undertaken in relation with the objectives of the unit. Once these activities have been completed, a

[3] The construction of the database was facilitated by the existence of a CD-Rom which contains all issues of the journal from 1978 through 1998. This material was completed by the issues appearing between 1998 and 2002, which is the year corresponding to the most recent issues of the journal.

formative assessment, usually a paper-pencil test, is proposed to the students. The results of the test provide feedback to the teacher and students and are used to define appropriate corrective measures for students who have not yet mastered the instructional objectives. Correctives can take various forms: additional exercises, different types of instructional material (*e.g.*, verbal vs. visual representations), small-group discussions, one-to-one tutoring, computer-based tasks, but in all these cases the aim remains the *remediation of learning difficulties* identified by formative assessment. Each of the phases (teaching, testing, remediation) is planned, prepared and managed by the teacher who attempts to assure that all the students will master the objectives of the unit.

A number of publications in French have contributed to an enlargement of the conception of formative assessment. One of the earliest formulations appeared in an article by Audibert (1980) which proposed a "non-specialist's" view of formative assessment. Formative assessment, he wrote, "takes place day by day and allows the teacher and the student to adapt their respective actions to the teaching/learning situation in question. It is thus, for them, a privileged occasion for conscious reflection on their experience (*prise de conscience de leur vécu*), for objectivation in action". (p. 62)[4] Several authors (in particular, Allal, 1979, 1988; Perrenoud, 1998b) have systematically contrasted the characteristics of an enlarged perspective of formative assessment with those of the approach initially defined by Bloom. The major points of contrast are presented in Table 1.

Rather than considering formative assessment as a specific event that occurs after a phase of teaching, the enlarged perspective advocates the integration of formative assessment within each instructional activity. This integration requires a diversification of the means of assessment. In addition to paper-pencil tests, quizzes or worksheets designed to verify whether students understood the content of a lesson, assessment is carried out informally by direct teacher observation, by exchanges among students (reciprocal assessment) at various points during an instructional activity, and by whole-class discussions that allow students to present different ways of understanding a task or of carrying out an activity.

4 The French-language quotations in this paper are translated by the authors of this review. We indicate in parentheses expressions in French that are difficult to translate in a fully appropriate way.

Table 1. Bloom's initial conception vs. an enlarged conception of formative assessment (FA)

Bloom's initial conception	An enlarged conception
- Insertion of FA after a phase of teaching - Use of formative tests - Feedback + correction → remediation - Management of FA by the teacher - Mastery of objectives by all students - Remediation benefits the students who were assessed	- Integration of FA in all learning situations - Use of varied means of data collection - Feedback + adaptation of instruction → regulation - Active student involvement in FA - Differentiation of instruction and, to some extent, of objectives - Regulation at 2 levels: for the students assessed, for future students (continuing instructional improvement)

Source: Authors.

In the enlarged perspective of formative assessment developed in French-language publications, the idea of *remediation* of learning difficulties (feedback + correction) is replaced by the broader concept of *regulation* of learning (feedback + adaptation). This transformation emerged initially in a paper by Cardinet (1977) whose conception of regulation was inspired by cybernetic systems analysis. A distinction was subsequently made between three modalities of regulation associated with formative assessment (Allal, 1979, 1988):

1. *Interactive regulation* occurs when formative assessment is based on the interactions of the student with the other components of the instructional activity, that is, with the teacher, with other students and/or with material allowing self-regulated learning. The integration of different forms of interactive regulation within an instructional activity allows continuing adaptations of learning as it takes place. Interactive regulation contributes to the progression of student learning by providing feedback and guidance that stimulate student involvement at each step of instruction.

2. *Retroactive regulation* occurs when a formative assessment is conducted after completion of a phase of teaching and allows identification of the instructional objectives attained or not attained by each student. The feedback from the assessment leads to the selection of means for correcting or overcoming learning difficulties encountered by some students. It corresponds to the notion of remediation present in the initial conception of formative assessment defined by Bloom.

3. *Proactive regulation* occurs when different sources of information allow the preparation of new instructional activities

> designed to take into account differences among students. It is linked to concerns with the differentiation of instruction so as to insure enrichment and consolidation according to student needs, rather than focusing on remediation of learning difficulties.

Innovative approaches to formative assessment often combine these three types of regulation. Instructional activities are designed to include several forms of interactive regulation based on informal means of assessment (observation, discussion). More structured means of formative assessment (tests, written productions, oral examination) are introduced periodically to allow for retroactive regulation of difficulties that were not resolved by the informal interactive regulations. In addition, proactive regulation takes into account all available information so as to insure that future activities are better adapted, from the outset, to the needs of the students; in other words, differentiation of instruction is planned, rather than being just added on, after observing difficulties.

In Bloom's initial conception of formative assessment, the teacher (or sometimes, the curriculum developer) assumes responsibility for the planning and management of each assessment operation: preparation of a formative test, analysis and interpretation of the results, proposal of appropriate remediations. In an enlarged conception, external regulation (by the teacher, by the test, by remedial material) is redefined as scaffolding that assists students' development of self-regulation. This means fostering the active involvement of students in formative assessment through procedures of self-assessment, reciprocal peer-assessment, and joint teacher-student assessment (Allal, 1999).

One further point of comparison needs to be mentioned. The basic aim of mastery learning is that formative assessment, followed by feedback and correction, will allow all (or virtually all) students to attain the instructional objectives. In the perspective proposed in the French-language literature, a much greater emphasis is given to the differentiation of instruction. Although it is accepted that basic objectives (*e.g.*, learning to read) must be mastered by all students, questions are raised about the possible adaptation of the objective to better take into account student cultural experiences and personal interests. The idea is expressed, for instance, that there may be several ways of "being a reader", such as reading to act, reading to get the "gist", reading to understand in depth, reading to communicate. In this perspective, formative assessment aims at identifying qualitative differences among students that need to be taken into account in the choice of reading material, in the tasks used for assessment, in the regulations fostered in class. For example, structured activities of peer interaction about a text may allow confrontations among students who have different approaches to reading.

A final direction of enlargement has resulted from work with classroom teachers, and particularly secondary teachers who are often faced with important constraints on the time and resources available for formative assessment (Allal and Schwartz, 1996). In this context it was found useful to differentiate two complementary levels of formative assessment. Level 1 concerns formative assessment that directly benefits the students who are assessed, as proposed in the basic Bloom model. Level 2 concerns situations where formative assessment data are used to inform teacher planning of future instructional activities proposed to *new* groups of students. When teachers are unable to carry out level 1 regulations (*e.g.*, due to lack of time or other obstacles), they should nevertheless be encouraged to carry out level 2 regulations, which in the long run can lead to systemic improvement of instruction.

Since the initial publications by Bloom and his collaborators, the conception of formative assessment has of course evolved in the English-language literature. For instance, in the review by Black and Wiliam (1998), the concept of feedback is described as a "system" that operates with four components:

- Data on the student's actual level.
- Data on a reference level.
- A mechanism for comparing the levels.
- A mechanism used to alter the gap.

The concept of regulation in the French-language literature includes these four components but emphasises the importance of additional factors linked to the processes intervening in attempts to "alter the gap". These processes are reflected in:

- The actions actually carried out by the teacher and the students to alter the gap.
- The degree of active student involvement in these actions.
- The uses students make of tools and resources present in the instructional environment to adapt or enrich their learning activity.
- The meaning attributed by students and teachers to the various aspects of assessment.
- The ways in which teachers and students negotiate assessment (talk about criteria, discuss requirements, construct shared understandings about what is expected).

The conceptualisation of regulation as the essential attribute of formative assessment has benefited from the contributions of a large number of French-language publications drawing on a diversity of theoretical perspectives, which are discussed subsequently in this paper (Allal, 1979, 1988, 1993; Cardinet, 1977, 1983; Hadji, 1989; Laveault, 1999; Nunziati, 1990; Perrenoud, 1991, 1993b, 1998b; Scallon, 2000; Schneuwly and Bain, 1993; Vial, 2001; Weiss, 1993).

Four developments in the evolution of work on formative assessment

It is possible to identify four major developments in the evolution of the conception of formative assessment in the French-language literature. These developments are presented in the order of their emergence. Each new development has attempted to overcome certain limitations of prior perspectives. It is important to note, however, that new developments have led to successive re-conceptualisations of formative assessment integrating prior contributions, rather than to the disappearance of earlier viewpoints.

Focus on instrumentation

French-language researchers initially adopted the focus on instrumentation that characterised formative assessment from the outset. The *Handbook on Formative and Summative Evaluation of Student Learning*, published in 1971 by Bloom and his coworkers, served as a model for the development of instruments for formative assessment (tables of objectives coordinated with formative tests and remediation activities). Several collections of instruments were published in different subject matter areas (*e.g.*, Marchandisse and Blampain, 1974; Tourneur, Noël and Honclaire, 1975) and general guidelines for the construction of criterion-referenced tests were established (Racine, 1982). More advanced instrumentation was subsequently developed in the form of computer-based item banks and systems of "tailored testing" allowing diagnostic error analysis (*e.g.*, Dassa, 1988; De Campos, 1990; Leclercq, 1980; Séguin, 1984). The dissemination of these forms of instrumentation helped to transform the conceptions and practices of formative assessment but also raised theoretical questions. Objections emerged about a "technology" of assessment that risked being cut off from theoretical reflection about the processes of learning and teaching (see in particular, Bain, 1988, on the "instrumental illusion" of the classical approaches to formative assessment). In response, Scallon (1988) defended instrumentation of formative assessment and argued that instrument development can take into account the aims and contextual constraints of classroom instruction.

Search for theoretical frameworks

At a conference of Swiss and Belgian researchers held in Geneva in 1978, a call was formulated for more in-depth theoretical grounding of formative assessment. The search for theories that can offer conceptual orientation for conducting assessment has been pursued since then in several different directions in the French-language literature.

During the Geneva conference, Allal (1979) outlined the differences between Bloom's conception based on a neo-behaviorist model of learning and a more constructivist approach to formative assessment based on Piagetian and other cognitive theories of learning. Several conference papers and subsequent articles described the implications of a constructivist conception for specific subject matters, such as mathematics (Brun, 1979; Thouin, 1993), French (Weiss, 1979), sciences (Thouin, 1982). Further reflection on this theme was proposed by Crahay (1986) who developed the argument that a constructivist perspective is necessary but nevertheless insufficient for the definition of optimal procedures of formative assessment.

Certain preoccupations of the constructivist perspective, such as the identification of learning processes and strategies that account for observed responses, have received renewed treatment in the light of contemporary theories of cognitive psychology. Implications were drawn from these theories for two major aspects of assessment: (1) the development of diagnostic models of formative assessment based on research on learning difficulties in the areas of reading (Lété, 1996) and of mathematics (Grégoire, 1996b) and the attempt to refine diagnostic assessment by use of Anderson's ACT model of declarative and procedural knowledge (Grégoire, 1999); (2) the investigation of the role of metacognitive processes in formative assessment and in self-assessment (Allal, 1993; Laveault, 1999; Scallon, 1996).

In parallel with developments of the constructivist/cognitive perspectives, new orientations were sought in theories emphasising social and philosophical dimensions of teaching and learning. Referring to work in social psychology, Cardinet (1988) proposed looking at formative assessment as a process of successful teacher-student communication about objectives, criteria, learning difficulties, etc. Using communication theory, Ouellette (1990) defined assessment as a dialogue constructed "with reference to a process of learning, as a function of interactions within an educational relationship" (p. 13). In an eclectic approach combining philosophical, social and institutional considerations, Hadji (1989) analysed formative assessment from the viewpoint of teacher-student transactions about reciprocal expectations and interpretations of assessment outcomes.

More recently, formative assessment was examined from the viewpoint of socio-cultural theories of teaching and learning. Referring to the Vygotskian concept of social mediation of learning, Allal and Pelgrims Ducrey (2000) argued that interactive formative assessment is aimed at providing scaffolding of learning in the student's zone of proximal development. This viewpoint is especially relevant for assessment situations involving teacher interactions with small groups or with individual students. We believe, however, that the theoretical framework of situated cognition and learning offers a broader perspective for conceptualising both interactive formative assessment and use of formative assessment tools in terms of teacher and student participation in the practices of a classroom community (Allal, 2002). A situated perspective was adopted by Mottier Lopez (2002) in a detailed analysis of the influence of classroom microculture on the practice of portfolio assessment with a predominantly formative aim.

Another theoretical approach to formative assessment has been proposed by French-language researchers in the areas of "didactics" (Bain, 1988; Chevallard, 1986; Garcia Debanc and Mas, 1987). This approach analyses assessment as part of a triadic system linking the teacher, the learner and the knowledge being dealt with. Emphasis is placed on how the content structures of school disciplines determine the aims, means and functions of formative assessment. Schubauer-Leoni (1991) proposed an interpretation of assessment within the framework of the "didactical contract" linking the reciprocal expectations of teacher and learners with respect to a given content area or task. Bain and Schneuwly (1993) developed the idea that, for any given instructional activity (*e.g.*, text production), it is necessary to identify relevant scientific "reference models" (*e.g.*, theories of discourse production, of language operations, of text genre) which can inform and guide formative assessment. The relationships between formative assessment and didactics were also discussed in several chapters of a book edited by Laveault (1992).

A few authors have explicitly situated formative assessment in the intersection of several theoretical perspectives. Perrenoud (1991, 1998b) argued that it is necessary to link cognitive, communicative and didactic orientations of formative assessment in a general framework of regulation that includes but goes beyond regulation due specifically to assessment. Bonniol and Vial (1997) explored the contrasting implications of cybernetic, systemic and complexity theories for the conceptualisation of formative assessment.

It is interesting to note that several recent English-language publications on classroom assessment, in particular Shepard (2000), give an important place to the implications of constructivist, socio-cultural and situated theories of learning, thereby joining major concerns of the French-language literature.

Studies of existing assessment practices in their contexts

The search for theoretical frameworks could lead to an increasingly abstract vision of formative assessment, cut off from the realities of classroom practice. This is why it is essential to articulate theoretical work with the study of how assessment is actually practiced in the classroom. Studies in this direction have dealt with several phenomena: the interplay between instrumentation and intuition in teachers' practices of formative assessment (Allal, 1983); the fundamental incompatibility between certain instruments of formative assessment and the everyday assessment practices of teachers (Weiss, 1984); the forms of teacher-student negotiation of assessment rules and norms (Chevallard, 1986); the institutional factors affecting teachers' attitudes toward inequalities of students achievement and the effect on assessment practice (Grisay, 1988); the pragmatics of actually doing formative assessment without worrying about doctrine (Perrenoud, 1991); the systemic aspects of assessment that can foster or inhibit the development of formative assessment practices (Perrenoud, 1993a). In work on formative assessment instrumentation, such as computer-based diagnostic testing, increasing emphasis is given to taking into account classroom practices and the ways of articulating instrumentation and practice (Dassa and De Cotret, 1993). Accounts of practice by teachers and teacher educators (*e.g.*, chapters by Berset, Elliott, Wegmuller in Allal, Bain and Perrenoud, 1993) have provided concrete illustrations of different forms of regulation associated with formative assessment.

Development of active student involvement in assessment

The role of the teacher remains essential for the practice of formative assessment: it is the teacher who decides what place will be given to formative assessment and the teacher's attitudes and implicit "theories" of teaching and learning have a profound impact on how formative assessment is put into practice. There is, however, increasing recognition of the importance of encouraging active student involvement in formative assessment. Nunziati (1990) and Vial (1995) developed an in-depth conceptualisation of the student's role in the formulation of assessment goals and criteria, in the conduct of interactive assessment, and in the construction of shared understanding of what assessment means. Allal (1999) proposed three different but interrelated forms of student involvement in assessment: individual self-assessment, reciprocal peer-assessment, and co-assessment entailing confrontation of teacher and student assessments. Campanale (1997) developed a detailed model of self-assessment, including metacognitive and reflexive dimensions intervening in the transformation of pedagogical practice in the context of professional development activities. Laveault (1999) expanded the conceptualisation of self-assessment by the inclusion of

motivational regulations, in addition to cognitive and metacognitive regulations. A common theme in the French-language literature is that interactive formative assessment, between peers and between teacher and students, constitutes a framework of social mediation that fosters the student's increasing capacity to carry out more autonomous self-assessment and self-regulated learning. Frameworks for practicing various forms of self/peer/joint teacher-student assessment have been elaborated and applied in classroom settings (*e.g.*, Doyon, 1992; Doyon and Juneau, 1991). It is needs to be recognised, however, that various dilemmas and pitfalls can occur when teachers encourage student involvement in assessment and things do not turn out as planned (Allal, 1999).

EMPIRICAL RESEARCH ON FORMATIVE ASSESSMENT

This part of our review analyses the empirical research presented in French-language publications on formative assessment. It is based primarily on the journal articles in the database we constructed, but takes into account examples of research presented in the books we consulted. Publications of empirical research have been classified in three major categories: (1) experimental studies of the effects of formative assessment; (2) development of instruments and procedures of formative assessment; (3) studies of teachers' attitudes and practices of formative assessment. The classification of publications in these categories allows a rough estimation of the relative amount of research conduced in each category. It is not possible, however, to arrive at a rigorous quantification since many articles contain elements relevant to several categories.

Experimental research on the effects of formative assessment

In the English-language literature, experimental or quasi-experimental research designed to determine the effects of formative assessment on student learning is relatively widespread, as attested by existing reviews (*e.g.*, Black and Wiliam,1998) and by meta-analyses of the effects of mastery learning which includes formative assessment as a key component (*e.g.*, Block and Burns, 1976; Slavin, 1987). This type of investigation has not found an equivalent place in the French-language literature. Of the 105 articles in our database, only two present experimental vs. control group comparisons of the effects of formative assessment on student learning. One of the studies was based on a design comparing mastery learning (with formative assessment) in two history classes to traditional instruction carried out by the same teachers in two matched history classes of a Geneva high school (Huberman, Juge and Hari, 1985). The results showed a positive effect the first trimester but this effect was not maintained subsequently in the second and third trimesters. Various factors which limited the effectiveness of mastery learning – principally institutional constraints and

student tendency to make the minimum effort needed for passing a grade – are discussed in the article. The second study (Gagné and Thouin, 1991), conducted in three French-speaking Ontario high schools, concerned a formative assessment procedure focused on the correction of spelling mistakes (lexical and grammatical) in student texts. Experimental and control classes were compared with respect to pretest-posttest gains on a spelling test and on a scale measuring student attitudes with respect to assessment. The results showed a relatively small effect of formative assessment on spelling scores but a substantial improvement of student attitudes toward assessment. In addition to these two studies, there is a brief reference in an article by Dassa (1988) to a quasi-experimental study carried out in Québec which compared three ways of using computer-based diagnostic assessment tools. Positive effect sizes are reported (0.56 for achievement in French and in mathematics) but the article gives little information on the experimentation and is devoted primarily to a critical discussion of the problems linked to the integration of diagnostic technology in classroom teaching.

In the books we consulted, we identified only one experimental study of the effects of formative assessment on student learning. Del'Guidice (1999) presented an investigation in which five groups of 4th-grade students received different types of diagnostic assessment and regulation. The results of these groups were compared to those of a matched control group on several tasks of geometry (calculation of areas). The author stated that the integration of formative assessment in learning situations had a beneficial effect on immediate learning and on transfer. His master's and doctoral thesis were cited but no data were presented in the book chapter.

Development of formative assessment instruments and procedures

Articles on instrument development have appeared regularly in the journal Mesure et évaluation en éducation since its creation. Many of the articles pertain, however, to the development of measurement instruments for research or for summative assessment, or concern instruments that are ill-defined with respect to their function. We were able to identify only a limited number of articles (around a half-dozen) which present empirical evidence of the validation of formative assessment instruments. One type of instrumentation stands out because it was the object of a substantial number of studies by Canadian researchers, namely the development of diagnostic instruments for error analysis and regulation of learning in the area of mathematics. Research in this area includes a variety of approaches: research comparing different models of diagnostic test construction, including estimation of reliability, information on validity, indications about conditions of application (Bertrand et al., 1985); qualitative analysis of computer-based error diagnostics and their didactical validity (Dassa and

De Cotret, 1993; De Campos, 1990); critical reflections about the place of computerised systems of diagnostic testing, such as adaptive testing and performance-responsive drill and practice (Dassa, 1988; Dassa and Vazquez-Abad, 1992). Computer-based diagnostic instrumentation in the area of text revision has also been developed (Laurier, 1996) and extended to student self-assessment and self-regulation (Coen and Gurtner, 1999).

In addition to research on instrument validation, there are various articles (about a half-dozen) which present empirical evidence about the use and implementation of formative assessment procedures. Examples include: a study by Scallon (1985) of how students use a diagnostic assessment guide for multiplication and their attitudes toward this type of assessment; the analysis by Allal *et al.* (1987) of the self-assessment and reciprocal peer-assessment behaviors that occur in mathematics games in 2nd and 3rd grades; an investigation by Derycke (1998) comparing two types of instrumentation – a criterion-referenced checklist and a portfolio – used for student follow-up when changing teachers (*suivi pédagogique*); a study by Richard, Godbout and Picard (2000) of a team sport assessment procedure that was applied in several activities (soccer, volley ball).

The journal and the book chapters we consulted also include a sizeable number of publications (over 25) presenting formative assessment instruments or procedures that have been developed in collaborative research with teachers, either in the context of teacher education and professional development or in work on curriculum reforms. These articles include conceptual justifications and references to practice but do not offer any systematic empirical evidence regarding applications in the classroom. Examples include: the classroom assessment guide presented by Descoteaux and Lirette (1983); the kits (*trousses*) developed by Cazabon (1991) for formative assessment in language learning; the Learning portfolio (*dossier d'apprentissage*) described by Simon and Forgette-Giroux (1993).

Studies of teacher attitudes and practices of formative assessment in the classroom

Investigations of how formative assessment functions in classroom settings are based primarily on three sources of information. The first includes action-research projects involving collaboration between researchers and teachers. Projects in Switzerland showed that detailed diagnostic instruments developed by researchers were not compatible with classroom practice (Weiss, 1984) and tended to reinforce recognition of the role of interactive formative assessment in the classroom (Cardinet, 1983). Subsequent projects (*e.g.*, Schwartz and Allal, 2000) were inserted in professional development programmes designed to accompany teachers in their attempts to conceptualise and put into practice their personal versions

of formative assessment. In Canada, action-research projects were undertaken to develop formative assessment instruments in a constructivist and interactionist perspective for mathematics (Thouin, 1993) and for science instruction (Thouin, 1995). Instruments of various types were developed with teachers, tried out in their classes and shared with other practitioners. Another project allowed successive reformulations of teachers' projects for transforming their assessment practices in a more formative perspective (Desrosiers, Godbout and Marzouk, 1992).

A second source of information comes from studies based on teachers' responses to attitude scales, questionnaires or interviews. Standard instrument development methodology was used by two groups of Canadian researchers to validate scales for measuring teacher beliefs and attitudes about assessment and student learning (Gadbois *et al.*, 1991; Louis and Trahan, 1995). But, beyond the initial validation studies, investigations using the scales have not been reported in subsequent journal articles. On the other side of the Atlantic, a questionnaire survey, addressed to 113 Belgian elementary school teachers, showed that teachers were generally favorable to formative assessment but that there was often a gap between espoused beliefs and classroom practice (Van Nieuwenhoven and Jonnaert, 1994). Using questionnaires and interviews, Campanale (1997) found a positive evolution of teacher conceptions of learning and assessment during a professional development programme that gave an important place to self-assessment of practice. A less encouraging result was found in a study of student perceptions of assessment in 6th to 8th grades in Québec; responses to a questionnaire showed little evidence that students encountered formative assessment experiences (Bercier-Larivière and Forgette-Giroux, 1995).

A third source of information on assessment practice consists in detailed descriptions formulated by teachers and teacher educators of their own practices. Examples include the formative assessment procedures developed by Elliott (1993) for beginning reading, by Berset Fougerand (1993) for writing and spelling and by Wegmuller (1993) for activities of text production. Despite the anecdotal nature of these reports, they provide evidence that teachers who are interested in formative assessment can develop a wide range of procedures involving different forms of regulation and active student implication. There are also a number of books based largely on teachers' experiences with respect to formative perspectives for correcting or assessing student work (Groupe EVA, 1991; Veslin and Veslin, 1992) and the development of active student participation in assessment (Doyon and Juneau, 1991).

CONCLUSION

The French-language publications on formative assessment have contributed to a significant enlargement of the conception of formative assessment. The central idea of this conception is the regulation of teaching and learning through informal, interactive assessment and through the use of instruments that are adapted to classroom practice. The work by French-language researchers has led to a diversification and enrichment of the ways of carrying out formative assessment. Theoretical proposals have often been influenced by intensive contacts with teachers, through curriculum development projects, through teacher education programmes, through school reform movements. There has not, on the other hand, been a systematic concern for verification of the impact of formative assessment on student learning. Very little controlled experimental work has been conducted. Instrument development has not been sufficiently integrated into long-term research projects. Studies of practice are episodic and dispersed in different settings, which makes it difficult to identify patterns or trends. In summary, the theoretical promise of French-language work on formative assessment is in need of considerably more empirical grounding. This is a major challenge for the researchers of this community in the coming decades.

References

Allal, L. (1979), "Stratégies d'évaluation formative : conceptions psycho-pédagogiques et modalités d'application" in L. Allal, J. Cardinet and P. Perrenoud (eds.), *L'évaluation formative dans un enseignement différencié*, Peter Lang, Bern, pp. 153-183.

Allal, L. (1983), "Évaluation formative : entre l'intuition et l'instrumentation", *Mesure et évaluation en éducation,* Vol. 6, pp. 37-57.

Allal, L. (1988), "Vers un élargissement de la pédagogie de maîtrise : processus de régulation interactive, rétroactive et proactive" in M. Huberman (ed.), *Assurer la réussite des apprentissages scolaires ? Les propositions de la pédagogie de maîtrise*, Delachaux et Niestlé, Neuchâtel, pp. 86-126.

Allal, L. (1993), "Régulations métacognitives" in L. Allal, D. Bain and P. Perrenoud (eds.), *L'évaluation formative et didactique du français*, Delachaux et Niestlé, Neuchâtel, pp. 81-98.

Allal, L. (1997), "La mesure : variations culturelles sur le thème ADMEE", *Mesure et évaluation en éducation,* Vol. 19, pp. 1-4.

Allal, L. (1999), "Impliquer l'apprenant dans les processus d'évaluation : promesses et pièges de l'autoévaluation" in C. Depover and B. Noël (eds.), *L'évaluation des compétences et des processus cognitifs : modèles, pratiques et contextes*, De Boeck, Brussels, pp. 35-56.

Allal, L. (2002), "L'évaluation dans le contexte de l'apprentissage situé : peut-on concevoir l'évaluation comme un acte de participation à une communauté de pratiques ?", Conference for the 15th symposium of ADMEE-EUROPE, Université de Lausanne, September.

Allal, L., E. Baeriswyl, M. Tra Bach and E. Wegmuller (1987), "Le jeu comme situation d'auto-évaluation", *Mesure et évaluation en éducation,* Vol. 10, pp. 47-64.

Allal, L., D. Bain and P. Perrenoud (eds.) (1993), *L'évaluation formative et didactique du français*, Delachaux et Niestlé, Neuchâtel.

Allal, L., J. Cardinet and P. Perrenoud (eds.) (1979), *L'évaluation formative dans un enseignement différencié*, Lang, Bern.

Allal, L. and G. Pelgrims Ducrey (2000), "Assessment *of* – or *in* – the zone of proximal development", *Learning and Instruction,* Vol. 10, pp. 137-152.

Allal, L. and G. Schwartz (1996), "Quelle place pour l'évaluation formative dans l'enseignement au cycle d'orientation ?", *CO Infos,* No. 178, pp. 5-8.

Audibert, S. (1980), "En d'autres mots … l'évaluation des apprentissages !", *Mesure et évaluation en éducation,* Vol. 3, pp. 59-64.

Bain, D. (1988), "L'évaluation formative fait fausse route", *Mesure et évaluation en éducation,* Vol. 10, pp. 23-32.

Bain, D. and B. Schneuwly (1993), "Pour une évaluation formative intégrée dans la pédagogie du français : de la nécessité et de l'utilité de modèles de référence" in L. Allal, D. Bain and P. Perrenoud (eds.), *L'évaluation formative et didactique du français*, Delachaux et Niestlé, Neuchâtel, pp. 51-79.

Bélair, L. (1999), *L'évaluation dans l'école : nouvelles pratiques*, ESF, Paris.

Bercier-Larivière, M. and R. Forgette-Giroux (1995), "L'évaluation des apprentissages telle que perçue par des élèves de 6, 7 et 8e années", *Mesure et évaluation en éducation,* Vol. 18, pp. 37-58.

Berset Fougerand, B. (1993), "Écrire … à haute voix : intégration de l'orthographe dans la production écrite" in L. Allal, D. Bain and P. Perrenoud (eds.), *L'évaluation formative et didactique du français*, Delachaux et Niestlé, Neuchâtel, pp. 171-196.

Bertrand, R., M. Tremblay-Desrochers, M. Morin and J. Roberge-Brassard (1985), "Analyse comparative de trois modèles d'instruments diagnostiques en mathématiques", *Mesure et évaluation en éducation,* Vol. 8, pp. 5-41.

Black, P. and D. Wiliam (1998), "Assessment and Classroom Learning", *Assessment in Education*, 5(1), pp. 7-74.

Block, J. and R. Burns (1976), "Mastery Learning", *Review of Research in Education,* 4, pp. 3-49.

Bloom, B.S. (1968), "Learning for Mastery", *Evaluation Comment,* 1(2), pp. 1-12.

Bloom, B.S. (1976), *Human Characteristics and School Learning*, McGraw-Hill, New York.

Bloom, B.S., J.T. Hasting and G.F. Madaus (1971), *Handbook on Formative and Summative Evaluation of Student Learning*, McGraw-Hill Book Co, New York.

Bonniol, J.-J. and M. Vial (1997), *Les modèles de l'évaluation*, De Boeck, Brussels.

Brun, J. (1979), "L'évaluation formative dans un enseignement différencié de mathématiques", in L. Allal, J. Cardinet and P. Perrenoud (eds.), *L'évaluation formative dans un enseignement différencié*, Peter Lang, Bern, pp. 170-181.

Campanale, F. (1997), "Auto-évaluation et transformation de pratiques pédagogiques", *Mesure et évaluation en éducation,* Vol. 20, pp. 1-24.

Cardinet, J. (1977), *Objectifs éducatifs et évaluation individualisée* (second edition, Report No. R77.05), Institut Romand de Recherches et de Documentation Pédagogiques, Neuchâtel.

Cardinet, J. (1983), "Quelques directions de progrès possibles pour l'appréciation du travail des élèves", *Mesure et évaluation en éducation,* Vol. 6, pp. 5-35.

Cardinet, J. (1986a), *Évaluation scolaire et mesure*, De Boeck, Brussels.

Cardinet, J. (1986b), *Évaluation scolaire et pratique*, De Boeck, Brussels.

Cardinet, J. (1988), "La maîtrise, communication réussie" in M. Huberman (ed.), *Assurer la réussite des apprentissages scolaires ? Les propositions de la pédagogie de maîtrise*, Delachaux et Niestlé, Neuchâtel, pp. 155-195.

Cazabon, B. (1991), "L'évaluation formative de la communication : l'intégration des composantes", *Mesure et évaluation en éducation,* Vol. 14, pp. 5-22.

Chevallard, Y. (1986), "Vers une analyse didactique des faits d'évaluation" in J.-M. De Ketele (ed.), *L'évaluation : approche descriptive ou prescriptive ?*, De Boeck, Brussels, pp. 31-59.

Coen, P.F. and J.L. Gurtner (1999), "Processus cognitifs en jeu dans une tâche d'écriture assistée par le logiciel AutoéVal" in C. Depover and B. Noël (ed.), *L'évaluation des compétences et des processus cognitifs : modèles, pratiques et contextes*, De Boeck, Brussels, pp. 239-254.

Crahay, M. (1986), "Évaluation formative et théorie constructiviste du développement" in J.-M. De Ketele (ed.), *L'évaluation: approche descriptive ou prescriptive ?*, De Boeck, Brussels, pp. 135-187.

Dassa, C. (1988), "L'intégration du diagnostic pédagogique aux apprentissages scolaires : de la théorie à la pratique", *Mesure et évaluation en éducation,* Vol. 11, pp. 7-26.

Dassa, C. and S.R. De Cotret (1993), "Validation d'un système informatisé de diagnostic en mathématiques au secondaire : une approche centrée sur l'analyse didactique", *Mesure et évaluation en éducation,* Vol. 16, pp. 5-26.

Dassa, C. and J. Vazquez-Abad (1992), "De l'évaluation informatisée à l'intervention pédagogique", *Mesure et évaluation en éducation,* Vol. 15, pp. 17-24.

De Campos, M. (1990), "Outil diagnostique et enseignement assisté par ordinateur", *Mesure et évaluation en éducation,* Vol. 13, pp. 55-69.

De Ketele J.-M. (1986) (ed.), *L'évaluation : approche descriptive ou prescriptive ?*, De Boeck, Brussels.

Del'Guidice, J. (1999), "L'évaluation-régulation, vecteur des transferts d'apprentissage" in C. Depover and B. Noël (eds.), *L'évaluation des compétences et des processus cognitifs : modèles, pratiques et contextes*, De Boeck, Brussels, pp. 99-114.

Depover, C. and B. Noël (eds.) (1999), *L'évaluation des compétences et des processus cognitifs : modèles, pratiques et contextes*, De Boeck, Brussels.

Derycke, M. (1998), "Suivi pédagogique ; grille critériée et dossier d'apprentissage", *Mesure et évaluation en éducation*, Vol. 20, pp. 33-53.

Descoteaux, M-S. and N. Lirette (1983), "Présentation du guide d'évaluation en classe", *Mesure et évaluation en éducation*, Vol. 6, pp. 123-133.

Desrosiers, P., P. Godbout and A. Marzouk (1992), "Des pistes pour soutenir la transformation des pratiques évaluatives" in D. Laveault (ed.), *Les pratiques d'évaluation en éducation*, Éditions de l'ADMÉÉ, Montreal, pp. 139-150.

Doyon, C. (1992), "Une pratique d'autoévaluation des apprentissages au primaire" in D. Laveault (ed.), *Les pratiques d'évaluation en éducation*, M Editeur, Montmagny, Québec, pp. 75-86.

Doyon, C. and R. Juneau (1991), *Faire participer l'élève à l'évaluation de ses apprentissages*, Beauchemin, Laval, Québec.

Elliott, N. (1993), "En observant l'apprenti-lecteur" in L. Allal, D. Bain and P. Perrenoud (eds.), *L'évaluation formative et didactique du français*, Delachaux et Niestlé, Neuchâtel, pp. 145-160.

Figari, G. and M. Achouche (2001), *L'activité évaluative réinterrogée*, De Boeck, Brussels.

Gadbois, L., R. Burelle, C. Parent and S.P. Séguin (1991), "Un instrument de mesure des croyances et attitudes des enseignants à l'égard des pratiques d'évaluation formative des apprentissages de leurs élèves", *Mesure et évaluation en éducation*, Vol. 14, pp. 5-24.

Gagné, F. and M. Thouin (1991), "L'évaluation formative des apprentissages en orthographe et attitude des élèves à l'égard de l'évaluation", *Mesure et évaluation en éducation*, Vol. 14, pp. 5-16.

Garcia-Debanc, C. and M. Mas (1987), "Évaluation des productions écrites des élèves", *Enjeux*, Vol. 11, pp. 108-122.

Grégoire, J. (ed.) (1996a), *Évaluer les apprentissages. Les apports de la psychologie cognitive*, De Boeck, Brussels.

Grégoire, J. (1996b), "Quelle démarche d'évaluation diagnostique des troubles d'apprentissage en mathématique ?", *Évaluer les apprentissages. Les apports de la psychologie cognitive*, De Boeck, Brussels, pp. 19-37.

Grégoire, J. (1999), "Que peut apporter la psychologie cognitive à l'évaluation formative et à l'évaluation diagnostique?" in C. Depover and B. Noël (eds.), *L'évaluation des compétences et des processus cognitifs : modèles, pratiques et contextes*, De Boeck, Brussels, pp. 17-33.

Grisay, A. (1988), "La pédagogie de maîtrise face aux rationalités inégalitaires des systèmes d'enseignement" in M. Huberman (ed.), *Maîtriser les processus d'apprentissage. Fondements et perspectives de la pédagogie de maîtrise*, Delachaux et Niestlé, Paris, pp. 235-265.

Groupe EVA (1991), *Évaluer les écrits à l'école primaire*, Hachette, Paris.

Hadji, C. (1989), *L'évaluation, règles du jeu, des intentions aux outils*, Éditions ESF, Paris.

Hadji, C. (1997), *L'évaluation démystifiée*, ESF, Paris.

Hivon, R. (ed.) (1993), *L'évaluation des apprentissages. Réflexions, nouvelles tendances et formation*, Éditions du CRP, Université de Sherbrooke, Québec.

Huberman, M. (ed.) (1988), *Assurer la réussite des apprentissages scolaires ? Les propositions de la pédagogie de maîtrise*, Delachaux et Niestlé, Neuchâtel.

Huberman, M., P.A. Juge and P.A. Hari (1985), "La pédagogie de maîtrise : une évaluation instructive au niveau gymnasial", *Mesure et évaluation en éducation*, Vol. 8, pp. 43-82.

Laurier, M. (1996), "Pour un diagnostic informatisé en révision de texte", *Mesure et évaluation en éducation*, Vol. 18, pp. 85-106.

Laveault, D. (ed.) (1992), *Les pratiques d'évaluation en éducation*, Éditions de l'ADMÉÉ, Montreal.

Laveault, D. (1999), "Autoévaluation et régulation des apprentissages" in C. Depover and B. Noël (ed.), *L'évaluation des compétences et des processus cognitifs : modèles, pratiques et contextes*, De Boeck, Brussels, pp. 57-79.

Leclercq, D. (1980), "Computerised Tailored Testing: Structured and Calibrated Item Banks for Summative and Formative Evaluation", *European Journal of Education*, Vol. 15(3), pp. 251-260.

Lété, B. (1996), "La remédiation des difficultés de lecture par la rétroaction verbale sur ordinateur" in J. Grégoire (ed.), *Évaluer les apprentissages: les apports de la psychologie cognitive*, De Boeck University, Brussels, pp. 133-155.

Louis, R. (1999), *L'évaluation des apprentissages en classe: théorie et pratique*, Éditions Études Vivantes, Laval.

Louis, R. and M. Trahan (1995), "Une mesure des croyances des enseignants titulaires du primaire relative à trois approches d'évaluation des apprentissage", *Mesure et évaluation en éducation,* Vol. 17, pp. 61-88.

Marchandisse, G. and D. Blampain (1974), *Techniques d'évaluation formative en langue maternelle*, ministère de l'Éducation nationale de la Culture française, Brussels.

Mottier Lopez, L. (2002), "Interroger la pratique du portfolio en situation scolaire dans une perspective 'située' de l'apprentissage", Paper for the 15th symposium of ADMEE-EUROPE, Université de Lausanne, September.

Nunziati, G. (1990), "Pour construire un dispositif d'évaluation formatrice", *Cahiers Pédagogiques*, No. 280, pp. 48-64.

Ouellette, L.-M. (1990), "La communication comme support théorique à l'évaluation", *Mesure et évaluation en éducation*, Vol. 13, pp. 5-22.

Perrenoud, P. (1991), "Pour une approche pragmatique de l'évaluation formative", *Mesure et évaluation en éducation*, Vol. 13, pp. 49-81.

Perrenoud, P. (1993a), "Touche pas à mon évaluation ! Une approche systémique du changement", *Mesure et évaluation en éducation*, Vol. 16, pp. 107-132.

Perrenoud, P. (1993b), "Vers des démarches didactiques favorisant une régulation individualisée des apprentissages" in L. Allal, D. Bain and P. Perrenoud (eds.), *L'évaluation formative et didactique du français*, Delachaux et Niestlé, Neuchâtel, pp. 31-50.

Perrenoud, P. (1998a), *L'évaluation des élèves. De la fabrication de l'excellence à la régulation des apprentissages. Entre deux logiques*, De Boeck, Brussels.

Perrenoud, P. (1998b), "From Formative Evaluation to a Controlled Regulation of Learning Processes. Towards a Wider Conceptual Field", *Assessment in Education*, 5(1), pp. 85-102.

Racine, S.B. (1982), **"**La validité et la fidélité dans la mesure critériée", *Mesure et évaluation en éducation*, Vol. 5, pp. 92-110.

Richard, J-F., P. Godbout and Y. Picard (2000), "La validation d'une procédure d'évaluation formative en jeux et sports collectifs", *Mesure et évaluation en éducation*, Vol. 23, pp. 43-67.

Scallon, G. (1985), "La participation des élèves au diagnostic pédagogique : exploration avec des élèves de 4e secondaire en mathématiques", *Mesure et évaluation en éducation*, Vol. 8, pp. 5-44.

Scallon, G. (1988), "Plaidoyer pour une méthodologie instrumentée d'évaluation formative", *Mesure et évaluation en éducation*, Vol. 11, pp. 43-55.

Scallon, G. (1996), "Évaluation formative et psychologie cognitive : mouvances et tendances" in G. Grégoire (ed.), *Évaluer les apprentissages. Les apports de la psychologie cognitive*, De Boeck, Brussels, pp. 159-173.

Scallon, G. (2000), *L'évaluation formative*, Éditions du Renouveau Pédagogique Inc, Saint-Laurent, Québec.

Schneuwly, B. and D. Bain (1993), "Mécanismes de régulation des activités textuelles : stratégies d'intervention dans les séquences didactiques" in L. Allal, D. Bain and P. Perrenoud (eds.), *L'évaluation formative et didactique du français*, Delachaux et Niestlé, Neuchâtel, pp. 219-238.

Schubauer-Leoni, M.-L. (1991), "L'évaluation didactique: une affaire contractuelle" in J. Weiss (ed.), *L'évaluation : problème de communication*, DelVal, Cousset, pp. 79-95.

Schwartz, G. and L. Allal (2000), *Vers une pratique de l'évaluation formative dans le secondaire I: Analyses d'expériences menées au cycle d'orientation de Genève*, Développement et innovation pédagogique au cycle d'orientation, Geneva.

Scriven, M. (1967), "The Methodology of Evaluation", *AERA Monograph Series on Evaluation*, 1, pp. 39-83.

Séguin, S.P. (1984), "L'utilisation des micro-ordinateurs pour l'évaluation des apprentissages : quelques perspectives d'avenir", *Mesure et évaluation en éducation*, Vol. 7, pp. 53-64.

Shepard, L. (2000), "The Role of Assessment in a Learning Culture", *Educational Researcher*, 29(7), pp. 1-14.

Simon, M. and R. Forgette-Giroux (1993), "Vers une utilisation rationnelle du dossier d'apprentissage", *Mesure et évaluation en éducation*, Vol. 16, pp. 27-40.

Slavin, R.E. (1987), "Mastery Learning Reconsidered", *Review of Educational Research*, 57, pp. 175-213.

Thouin, M. (1982), "La définition des objectifs et l'évaluation dans le domaine affectif, quelques éléments de réflexion", *Mesure et évaluation en éducation*, Vol. 5, pp. 31-34.

Thouin, M. (1993), "L'évaluation des apprentissages en mathématiques : une perspective constructiviste", *Mesure et évaluation en éducation,* Vol. 16, pp. 47-64.

Thouin, M. (1995), "Le développement d'instruments de mesure des apprentissages en sciences de la nature au primaire", *Mesure et évaluation en éducation*, Vol. 16, pp. 95-123.

Tourneur, Y., Y. Noël and B. Honclaire (1975), *Liste des objectifs, épreuves d'évaluation et outils de rattrapage en mathématiques*, ministère de l'Éducation nationale et de la Culture française, Brussels.

Van Nieuwenhoven, C. and P. Jonnaert (1994), "Une approche des représentations des enseignants du primaire à propos de l'évaluation", *Mesure et évaluation en éducation*, Vol. 16, pp. 41-79.

Veslin, O. and J. Veslin (1992), *Corriger des copies: évaluer pour former*, Hachette, Paris.

Vial, M. (1995), "Nature et fonction de l'auto-évaluation dans le dispositif de formation", *Revue française de pédagogie*, No. 112, pp. 69-76.

Vial, M. (2001), "Évaluation et régulation" in G. Figari and M. Achouche (eds.), *L'activité évaluative réinterrogée*, De Boeck, Brussels, pp. 68-78.

Wegmuller, E. (1993), "Et s'ils apprenaient à écrire en connaissance de cause! L'ovniprésence de l'évaluation formative" in L. Allal, D. Bain and P. Perrenoud (eds.), *L'évaluation formative et didactique du français*, Delachaux et Niestlé, Neuchâtel, pp. 197-218.

Weiss, J. (1979), "L'évaluation formative dans un enseignement différencié du français: une conception de la formation à dépasser" in L. Allal, J. Cardinet and P. Perrenoud (eds.), *L'évaluation formative dans un enseignement différencié*, Peter Lang, Bern, pp. 194-202.

Weiss, J. (1984), "Heurs et malheurs d'un instrument d'évaluation", *Mesure et évaluation en éducation*, Vol. 7, pp. 31-42.

Weiss, J. (ed.) (1991), *Évaluation : problème de communication*, DelVal, Cousset, Suisse.

Weiss, J. (1993), "Interaction formative et régulation didactique" in L. Allal, D. Bain and P. Perrenoud (eds.), *L'évaluation formative et didactique du français*, Delachaux et Niestlé, Neuchâtel, pp. 113-122.

Formative Assessment in Classrooms: A Review of the Empirical German Literature

by
Olaf Köller, University of Erlangen-Nuremberg

INTRODUCTION AND DATABASES

Germany has a long tradition of philosophers and educational reformers who proposed alternative education (so-called *Reformpädagogik*) as a more appropriate approach to teaching that meets students' needs for competence, autonomy and self-determination. Beyond other features, alternative education has emphasised that teachers should be aware of how they provide feedback to students, as feedback indicating personal growth to students will foster their learning and motivational development. Although there has been growing consensus across centuries and decades in Germany that the kind of feedback determines whether students achieve cognitive, emotional and motivational growth, systematic research on this issue has been conducted in relatively few German studies. In particular, there has been very little systematic empirical research on formative assessment in Black and Wiliam's (1998) sense. These authors interpret formative assessment "as encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged". (Black and Wiliam, 1998, p. 7-8)

Databases

The literature review covers the time period from 1980 until 2003. The search was conducted by several means. The first approach was to search using key words in two German databases, PSYNDEX and FIS-BILDUNG. While the first one contains the more psychologically-oriented literature (similar to PsychInfo), the second one mainly encompasses work in the fields of education or pedagogy (like ERIC). This search was of limited success because formative assessment is not a common concept in the German literature. More general descriptors (*e.g.*, assessment, feedback) resulted in more data that could be handled for this review. In addition, contents of several German journals that publish empirical studies in the field of education and/or instruction were scanned. These journals were (translations in parentheses):

- *Zeitschrift für Pädagogik* (Journal of pedagogy).

- *Zeitschrift für Erziehungswissenschaft* (Journal of educational science).
- *Unterrichtswissenschaft* (Research on instruction).
- *Zeitschrift für Pädagogische Psychologie* (Journal of educational psychology).
- *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* (Journal of developmental and educational psychology).
- *Psychologie in Erziehung und Unterricht* (Psychology in education and instruction).

In addition, a citation search of relevant articles in the above-mentioned journals was conducted. The resulting literature yielded more than 150 articles and book chapters. The body of this paper reviews selected theoretical papers and empirical studies of outstanding relevance to this report (rather than reviewing all articles located in these three search modes).

HISTORICAL ROOTS OF FORMATIVE ASSESSMENT IN GERMANY

Concepts of alternative education (*Reformpädagogik*) have been the most important historical roots of formative assessment in German classrooms. Hellmich and Teigler (1992) argue that particularly the works by Montessori, Freinet, Kerschensteiner and Steiner have been very influential.[1] In contrast to traditional teacher-directed approaches, these authors have highlighted students' needs for autonomy and self-determination.[2]

In *Montessori's pedagogy*, the teacher acts more or less in the background and becomes a careful observer and individual counsellor of the students, providing help to optimise their knowledge acquisition. The principles of autonomy, self-action and self-control which encourage students to assess their learning progress are of particular importance. *Freinet's pedagogy* places a strong emphasis on self-assessment. Students should learn to define their own projects, to assess their learning progress and whether they have reached their goals in these projects. Tools for

1 Particularly Steiner's approach has led to the foundation of the so-called Waldorf-schools. These schools belong to private educational sector but all school leaving certificates are equivalent to those provided by public schools.

2 For the purpose of this review, however, we will concentrate on the impact of alternative education on feedback processes in classrooms. More general descriptions of alternative education, particularly of the work by Freinet, Montessori, and Steiner can be found in Hellmich and Teigler (1992).

formative assessment in this sense are student week plans, diaries and working materials that allow students to assess and correct their own work.

The *Waldorf-pedagogy*, based on *Steiner's work*, has called for the abolishment of marks. Proponents of this approach have also argued against the German practice of requiring students who have received poor grades to repeat school years. *Kerschensteiner* proposed the advantages of self-assessments not only for the evaluation of final results but also for each working or learning step in school. Students from Waldorf-schools do not get any marks until the end of lower secondary level (grade 10) and remedial measures are conducted for poor achieving students so that they do not have to repeat a school year.

After World War II, concepts of alternative education felt into desuetude and it was not until the 1960s that alternative education was rediscovered and brought into the debate on educational reforms (*Bildungsreform*). Furthermore a strong critique of grades emerged in this period, because several empirical studies demonstrated that the psychometric properties (objectivity, reliability, and validity) of grades were quite poor (see Ingenkamp, 1971 for an overview). Educational reformers called for:

- The abolishment of grades.
- More standardised tests as measures of summative assessment instead of grades.
- More individualised feedback.
- Process-oriented instead of product-oriented diagnosis.
- More remedial measures for low-achieving students.
- A stronger emphasis on encouraging and motivating teaching.

Consequently several alternative tools for student assessment were proposed, all of which had a more formative as opposed to a summative character.

MEASURES OF ALTERNATIVE ASSESSMENT IN GERMAN SCHOOLS

The term "alternative assessment" is used here to illustrate that some of the measures presented below are important assessment tools beyond marks but are not really formative.

Diagnostic forms

Diagnostic forms (*Diagnosebögen*, cf. Ingenkamp, 1985) provide detailed information about learning success and allow a much more differentiated assessment than grades. Interestingly they were firstly introduced in German classrooms in 1915 and rediscovered in the last 30 years. Teachers in comprehensive schools have used these measures of formative assessment.

Major goals of using diagnostic forms are (cf. Winter, 1991):

- Assessment of social learning outcomes.
- Differentiated feedback information for both students and parents.
- Awareness of individual learning progress and growth in ability.
- Information which helps to optimise knowledge acquisition and to initiate remedial measures for low-achieving students.

Proponents have argued that teachers should use diagnostic assessment after each instruction unit for each student. However, teachers as well as school administrators have declined diagnostic testing that is too time-consuming. Consequently these measures disappeared from German classrooms in the 1980s.

Learning reports

Teachers typically complete learning reports (*Lernberichte*, cf. Lübke, 1996) twice a year. They are alternative form of summative assessment and combine information about social and cognitive learning outcomes. The learning reports contain both individual assessments and evaluations of the total class. Each student and the whole class receive advice on how to optimise motivational and cognitive development.

Diaries on learning success

Diaries (*Lerntagebücher*, Herrmann and Höfer, 1999) provide opportunities for students to reflect on their own learning processes and to detect and correct deficits over time. Diaries thus serve as a tool for autonomous and self-regulated learning. The advantages of diaries include:

- Opportunities for individual reflections.
- Opportunities for communication among students about achievement or learning goals.
- Help in preparing for final examinations (cf. Herrmann and Höfer, 1999).

Student week plans

Student week plans (*Wochenarbeitspläne*) are based upon Freinet's work. Typically, the week plans are used in elementary schools where teachers have more degrees of freedom with respect to their assessment practice. The week plan allows students to check whether they have reached their goals and solved all problems during the previous lessons across one week. The idea is that students become much more aware of their achievement levels and learn to be open to criticism (if they have not reached their aims). The week plan always includes an individual growth curve demonstrating the achievements during the week.

Portfolio

Portfolios are particularly useful in co-operative learning settings (cf. Herold and Landherr, 2001) because they allow students to evaluate their own impact on group-results. Students not only rate their behaviour within the group behaviour but also have to justify their ratings. Typically the ratings are discussed among all group members.

Some empirical evidence for the effectiveness of measures of alternative assessment

In recent years, a few German researchers have conducted empirical studies on assessment (see Grunder and Bohl, 2001 for an overview). Köller and Trautwein (2003) examined the use of alternative assessment measure in five comprehensive schools. They compared math and science achievement of 8th graders from these schools with 8th graders who had been tested with the same instruments in the TIMS study. Achievement scores of these five schools were above average (compared to the nationally representative TIMS study), suggesting that strategies of alternative assessment might have had positive effects on learning outcomes.

MARKS VS. VERBAL REPORTS AS ASSESSMENT MEASURES

In 1970, the Conference of Federal Ministers of Education (*Kultusminister-Konferenz*) decided that marks should be substituted by verbal reports in primary schools, at least in grades 1 and 2. This decision was intended to individualise education.

Again, major goals of this reform were:

- Avoiding pressure to achieve.
- Promoting cooperation instead of competition.

- Reducing social disparities and preventing declines in the achievement levels of disadvantaged students.
- Individual support.
- Assessment based on individual progress instead of social comparisons.

Empirical studies of the implementation and practice of verbal reports in elementary schools, however, showed that the reform was not working as hoped. For example, Benner and Ramseger (1985) conducted a content analysis of about 450 verbal reports. Four different types of verbal reports could be identified:

- *Normative reports* assessed the students based upon criteria defined in curricula and text books.
- *Nice reports* were highly encouraging but failed to obtain any information on the real achievement level, deficits and developmental potential of the student.
- *Descriptive reports* provided a clear picture of the students' achievement levels but ignored any information of students' progress in the different subjects.
- Finally, *developmental reports* had a truly formative character in that they described progress and deficits and how these deficits could be eliminated. Note that only this type represents a measure of formative assessment to any extent.

Valtin (cf. Valtin, 2002; Wagner and Valtin, 2003) analysed the effects of different types of assessment (marks vs. verbal reports) on the development of educational outcomes in elementary school. Her panel comprised 241 children from East and West Berlin who were tested several times, individually or in groups, from grade 2 to grade 4. Outcomes were attitude toward learning and toward school subjects, academic self-concept, achievement motivation, test anxiety, intelligence, and academic achievement in mathematics and German. Contrary to her prediction students did not profit notably from verbal reports.

One reason for these disappointing findings might be that the teachers in Valtin's study only practiced formative assessment when writing the reports but not in everyday situations in the classroom. The work of Rheinberg in particular (cf. Rheinberg and Krug, 1999) has demonstrated that formative assessment during ordinary lessons can have huge effects on motivation. His approach is described in the next section.

ADDITIONAL STUDIES IN GERMANY ON FORMATIVE ASSESSMENT

Some studies have systematically investigated effects of feedback processes on student characteristics. Interestingly, this research has been mainly carried out by psychological researchers who have been strongly influenced by American researchers on motivation such as Atkinson and McClelland. Major proponents in Germany included Heckhausen (1989), Rheinberg (Rheinberg and Krug, 1999), and Meyer (Meyer and Plöger, 1979). Heckhausen and Rheinberg established the concept of teacher's frame of reference (individual vs. social). In their studies, teachers using an individual frame of reference provided temporal feedback to students and emphasised improvement, whereas teachers with a social frame of reference assessed their students' accomplishments on the basis of comparisons with others. Meyer's research focused on the paradoxical effects of praise and blame, that is, he investigated situations in which teacher's praise (blame) led the student to think that he or she must be stupid (bright).

Teachers' reference norms: the work by Rheinberg

There is a long international research tradition investigating the effects of different types of feedback based on individual or social comparisons. Ames (1992) noted that social comparisons are encouraged by the frequent allocation of grades that rank-order students along a single continuum based on performance in the same task, by the public announcement of results, and by competitive learning environments that emphasise the importance of outperforming other students. In a strong critique of such competitive environments, Covington (1992) argued that competition reduces levels of academic achievement and undermines self-worth. Marsh (1991) further argues that competition and social comparison processes are likely to be stronger in highly selective school settings, thus exacerbating the negative effects on variables like academic self-concept or self-esteem.

In order to establish alternative frames of reference in the classroom, teachers can emphasise improvement, effort, and learning (individual frame of reference), rather than grades, ability differences, and outperforming classmates (social frame of reference). Concerning the important role of different types of comparisons, the German motivational psychologist Rheinberg (1980, 1999; also see Rheinberg and Krug, 1999) has established the concept of teachers' reference norms which has substantial theoretical overlap with major ideas of goal theory as proposed by Nicholls (1984). Based on research in motivation conducted by McClelland (cf. McClelland *et al.*, 1953) or Heckhausen (1989), Rheinberg defined teacher's reference norm as a standard to which individual achievements are compared. Such standards can be based upon different frames of reference. Comparing

individual achievements with prior achievements constitutes an individual reference norm, while comparing students' achievements with those of their class mates defines a social reference norm. The advantage of an individual perspective is that students directly register any improvement in their achievements, and can thus bolster their academic self-concept.

An important aspect of Rheinberg's work is that he not only distinguishes between the two types of teacher feedback, but that he also argues that teachers with a social reference norm typically present tasks of the same difficulty level to all students to obtain valid information about inter-individual differences. Holding the difficulty levels constant allows teachers to attribute students' achievement differences to ability. Furthermore teachers with a social reference norm believe that ability differences among students are highly stable across time. Therefore, poor achieving students will always show poor accomplishments, while bright students will always perform well in school.

Teachers with an individual reference norm prefer a quite different perspective, in that they judge their students based on prior achievement levels. Achievement gains over time are praised, stagnation or regression is blamed. There is no doubt, that an individual reference norm can be easily applied in everyday lessons, when students work on tasks by themselves. Table 1 summarises the differences between teachers with a social reference norm and those with an individual reference norm (see Rheinberg, 1980, p. 123 and Rheinberg, 1999, p. 44).

Table 1. Differences between teachers with an individual (IRN) and a social reference norm (SRN)

Variable	SRN	IRN
Comparisons	Cross-sectional, among students	Longitudinal, within students
Individualisation	Individualised instruction, assigning different task to students with different achievement levels	Longitudinal, within students
Causal attributions	More frequent, primarily time-constant factors (*e.g.*, ability), internal attributions of success and failure	Less frequent, preference for time-variant causes (persistence, concentration, attention); internal attributions of success, external or at least internal and variable attributions of failure
Feedback	Based on social comparison, emphasising the rank of each student within a class	Based on temporal comparisons, emphasising individual progress and growth

Source: Taken from Rheinberg (1980), p. 123 and Rheinberg (1999), p. 44 (slightly modified).

Rheinberg and colleagues have conducted many experimental studies investigating the effects of different reference norms on student outcomes, two of which are presented subsequently (see Mischo and Rheinberg, 1995 and Köller, 2004, for more complete overviews of studies investigating effects of reference norms on educational outcomes). Additionally an article by Lüdtke and Köller (2002) is described since these authors provided evidence for the effectiveness of an individual reference norm on students' academic self-concepts based upon two large German field studies with samples sizes of N = 3 992 and N = 2 150 students from grades 7 and 8, respectively.

Krug and Lecybyl (1999a)

These authors conducted an experiment on the effects of different reference norms (individual vs. social). Participants included 44 students from two classes of a vocational school. Students in both classes had the same teacher in social sciences. In one class, however, this teacher used an individual reference norm over a period of eight weeks, while she used a social reference norm in the other class. Dependent variables included observer ratings of students' understanding of the content taught, achievement tests, the teacher-students-relationship, students' participation, and how much students liked the lessons. The findings were quite mixed, that is, students in the individual reference norm condition had higher values on some of the outcome measures, while no differences occurred on the other measures. Note, however, that no dependent variable had a higher mean in the social reference norm condition.

Krug and Lecybyl (1999b)

Krug and Lecybyl conducted a second study similar to the first, but distinguished between low, middle and high-achieving students. Again the sample included students (17 in class 1 and 19 in class 2) from two classes of a vocational school and the teacher was the same in both conditions. Again, positive effects of an individual reference norm on several outcome measures were observed. These effects, however, were largest for poor achieving students.

Lüdtke and Köller (2002)

The two studies of these authors were inspired by Marsh's (1987) work on the big-fish-little-pond-effect (BFLPE). The BFLPE describes the phenomenon that equally able students have lower academic self-concepts in classes or schools where the average achievement level is higher than in classes or schools where the average achievement level is lower. Social

comparison theory (Festinger, 1954) provides a theoretical framework explaining the BFLPE: students are inherently more likely to make social comparisons with higher-achieving students – thus leading to lower academic self-concepts — in high-ability classes than in low-ability classes. In their study, Lüdtke and Köller investigated the effects of teacher feedback on the BFLPE in large samples of secondary level students (see above). The basic assumption was that the BFLPE would be smaller in classes in which teachers strongly emphasise improvement, effort, and learning (individual reference norm). The authors, however, found that the negative BFLPE was observable in all classes but that there was an additional positive effect of an individual reference norm on academic self-concept.

Paradoxical effects of praise and blame: the work by Meyer

It is a common belief that positive teacher feedback (praise) during regular lessons has positive rather than negative effects on student characteristics such as motivation, self-esteem and learning. Negative feedback (blame) is usually expected to have the opposite effects. However, Meyer (1982, 1992; also see Meyer *et al.*, 1979), a German researcher in the field of motivation, has conducted a series of experiments showing that praise and blame can have counter-intuitive effects on students self-evaluations, meaning that praise can, under some special circumstances, reduce ones self-perceptions of ability, whereas blame can increase such self-perceptions. From his attributional point of view, the effects of teacher praise depend on a student's interpretation. If praise is attributed to ability, the student's self-perceptions of ability may increase. If praise is attributed to effort, the student's self-perception of ability may even decrease (if the perception of high effort is perceived as an indicator of low ability, particularly after simple tasks). Thus, praise does not always lead to a perception of high ability, and blame does not necessarily lead to a low estimation of ability. Such findings were first reported by Meyer and colleagues (Meyer *et al.*, 1979). Effort attributions were assumed to be the intervening variables (Meyer, 1992). The general method has been to present participants with a scenario in which two students receive feedback for an identical outcome. One student is praised (or criticised), the other receives neutral feedback, for instance: "Peter and Paul have each got 7 out of 10 problems right. The teacher gives Paul neutral feedback, 'You've got seven problems right, Paul.' However, he praises Peter: 'Well done, Peter!'". In the failure conditions, praise is usually replaced by blame: "Well, that wasn't very good, Peter!". Participants are then asked to judge the ability of both protagonists (see Meyer *et al.*, 1979).

It has to be admitted that this scenario method tends to assess rather unrealistic interaction sequences. However, some studies with more realistic

settings, either experimental (Meyer, Mittag and Engler, 1986) or field studies (Tacke and Linder, 1981), have also shown paradoxical effects of praise and blame (see Pikowsky, 1988). Rheinberg and Weich (1988) were able to show that paradoxical ability attributions were even made spontaneously when identical achievements were sanctioned in different ways. Meyer *et al.* (1986) showed that paradoxical inferences are not restricted to ability attributions in scenario studies but even have effects on students' self-concept of ability. In their study, students who were praised inferred lower task-specific competence than students who received neutral feedback.

The level of cognitive development seems to be a moderator of such paradoxical effects: Barker and Graham (1987) found that the apparently paradoxical effects of praise and criticism occur more frequently as a function of increasing age. Whereas 4- to 5-year-olds always inferred that praise indicated high ability and high effort, paradoxical effects began to appear among 11- to 12-year-olds.

To summarise the research on praise and blame has clearly shown that teachers' feedback can have paradoxical effects in that praise has negative effects, while the consequences of blame could be positive. These findings do not necessarily devaluate such feedback as a helpful formative measure but argue for caution in daily situations in which feedback is provided.

SUMMARY AND SOME REMARKS ON FUTURE DIRECTIONS IN RESEARCH ON FORMATIVE ASSESSMENT IN GERMANY

The previous sections of this literature report have shown that there is not very much German research on effects of formative assessment on educational outcomes. This is surprising to some extent, because there are many approaches of formative assessment described in the German literature. These approaches have not been sufficiently evaluated. Despite this lack of research, there are currently some very interesting videotape studies for several subjects (*i.e.*, English, math, and science) that may facilitate insight into the assessment practices of German teachers and the way in which they affect learning. Within the TIMS study (Stigler *et al.*, 1996) 100 German math lessons were videotaped. All these videos can be coded with respect to teachers' assessment practices. Similar studies are currently conducted for physics (project head: Prof. Dr. Manfred Prenzel from the Institute for Science Education) and English (project head: Prof. Dr. Eckard Klime, German Institute for International Educational Research). All studies collect not only video data but also achievement as well as motivation, social and other data. Therefore it will be possible to analyse the relationships between assessment styles and all educational outcomes.

References

Ames, C. (1992), "Classroom Goals, Structures, and Student Motivation", *Journal of Educational Psychology*, Vol. 84, pp. 261-271.

Barker, G.P. and S. Graham (1987), "Developmental Study of Praise and Blame as Attributional Cues", *Journal of Educational Psychology*, Vol. 79, pp. 62-66.

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

Benner, D. and J. Ramseger (1985), "Zwischen Zifferenzensur und pädagogischem Entwicklungsbericht. Zeugnisse ohne Noten in der Grundschule", *Zeitschrift für Pädagogik*, Vol. 31, pp. 151-174.

Black, P. and D. Wiliam (1998), "Assessment and Classroom Learning", *Assessment in Education*, Vol. 5, pp. 7-74.

Covington, M.V. (1992), "Making the Grade: A Self-worth Perspective on Motivation and School Reform", Cambridge University Press.

Deutsches PISA-Konsortium (ed.) (2001), "PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich", Leske + Budrich, Opladen.

Deutsches PISA-Konsortium (ed.) (2002), "PISA 2000. Die Länder der Bundesrepublik Deutschland im Vergleich", Leske + Budrich, Opladen.

Deutsches PISA-Konsortium (ed.) (2003), "PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland", Leske + Budrich, Opladen.

Festinger, L. (1954), "A Theory of Social Comparison Processes", *Human Relations*, 7, pp. 117-140.

Grunder, H.-U. and T. Bohl (eds.) (2001), *Neue Formen der Leistungsbeurteilung*, Schneider, Hohengehren.

Heckhausen, H. (1989), *Motivation und Handeln*, Springer, Berlin.

Hellmich, A. and P. Teigler (eds.) (1992), *Montessori-, Freinet-, Waldorfpädagogik*, Beltz, Weinheim.

Herold, M. and B. Landherr (2001), *Selbstorganisiertes Lernen: SOL; ein systemischer Ansatz für Unterricht*, Schneider, Hohengehren.

Herrmann, J. and C. Höfer (1999), *Evaluation in der Schule – Unterrichtsevaluation*, Bertelsmann Stiftung, Gütersloh.

Ingenkamp, K. (1971), *Die Fragwürdigkeit der Zensurengebung*, Beltz, Weinheim.

Ingenkamp, K. (1985), *Lehrbuch der Pädagogischen Diagnostik*, Beltz, Weinheim.

Köller, O. (2004), *Konsequenzen von Leistungsgruppierungen*, Waxmann, Münster.

Köller, O. and U. Trautwein (eds.) (2003), *Schulqualität und Schülerleistung. Evaluationsstudie über innovative Schulentwicklung an fünf hessischen Gesamtschulen*, Juventa, Weinheim.

Krug, S. and R. Lecybyl. (1999a), "Die Wirkung experimentell variierten Lehrerverhaltens auf Unterrichtswahrnehmung, Lernbereitschaft und Leistung von Schülern" in F. Rheinberg and S. Krug (eds.), *Motivationsförderung im Schulalltag*, Hogrefe, Göttingen, pp. 81-94.

Krug, S. and R. Lecybyl, R. (1999b), "Die Veränderung von Einstellung, Mitarbeit und Lernleistung im Verlauf einer bezugsnormspezifischen Motivationsintervention" in F. Rheinberg and S. Krug (eds.), *Motivationsförderung im Schulalltag*, Hogrefe, Göttingen, pp. 95-114.

Lübke, S.-I. (1996), *Schule ohne Noten*, Leske + Budrich, Opladen.

Lüdtke, O. and O. Köller (2002), "Individuelle Bezugsnormorientierung und soziale Vergleiche im Mathematikunterricht: Der Einfluss unterschiedlicher Referenzrahmen auf das fachspezifische Selbstkonzept der Begabung", *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, Vol. 34, pp. 156-166.

Marsh, H.W. (1987), "The Big-fish-little-pond Effect on Academic Self-concept", *Journal of Educational Psychology*, Vol. 79, pp. 280-295.

Marsh, H.W. (1991), "The Failure of High-ability High Schools to Deliver Academic Benefits: The Importance of Academic Self-concept and Educational Aspirations", *American Educational Research Journal*, Vol. 28, pp. 445-480.

McClelland, D.C., J.W. Atkinson, R.A. Clark and E.L. Lowell (1953), *The Achievement Motive*, Appleton-Century-Crofts, New York.

Meyer, W.-U. (1982), "Indirect Communications about Perceived Ability Estimates", *Journal of Educational Psychology*, Vol. 74, pp. 888-897.

Meyer, W.-U. (1992), "Paradoxical Effects of Praise and Criticism on Perceived Ability", *European Review of Social Psychology*, Vol. 3, pp. 259-283.

Meyer, W.-U., M. Bachmann, U. Biermann, M. Hempelmann, F.-O. Plöger and H. Spiller (1979), "The Informational Value of Praise and Blame on Perceptions of Ability", *Journal of Educational Psychology*, Vol. 71, pp. 259-268.

Meyer, W.-U., W. Mittag and U. Engler (1986), "Some Effects of Praise and Blame on Perceived Ability and Affect", *Social Cognition*, Vol. 4, pp. 293-308.

Meyer, W.-U. and F.-O. Plöger (1979), "Scheinbar paradoxe Wirkungen von Lob und Tadel auf die wahrgenommene eigene Begabung" in S.H. Filipp (ed.), *Selbstkonzept-Forschung*, Klett, Stuttgart, pp. 221-236.

Mischo, C. and F. Rheinberg (1995), "Erziehungsziele von Lehrern und individuelle Bezugsnormender Leistungsbewertung", *Zeitschrift für Pädagogische Psychologie*, Vol. 9, pp. 139-151.

Nicholls, J.G. (1984), "Achievement Motivation: Conceptions of Ability, Subjective Experience, Task Choice, and Performance", *Psychological Review*, Vol. 91, pp. 328-346.

Pikowsky, B. (1988), "Lob im Unterricht: Lehrer- und Schülerkognitionen", *Zeitschrift für Pädagogische Psychologie*, Vol. 2, pp. 251-258.

Rheinberg, F. (1980), *Leistungsbewertung und Lernmotivation*, Hogrefe, Göttingen.

Rheinberg, F. (1999), "Trainings auf der Basis eines kognitiven Motivationsmodells" in F. Rheinberg and S. Krug (eds.), *Motivationsförderung im Schulalltag*, Hogrefe, Göttingen, 2, pp. 36-52.

Rheinberg, F. and S. Krug (eds.) (1999), *Motivationsförderung im Schulalltag*, Hogrefe, Göttingen, 2.

Rheinberg, F. and K.W. Weich (1988), "Wie gefährlich ist Lob? Eine Untersuchung zum 'paradoxen Effekt' von Lehrersanktionen", *Zeitschrift für Pädagogische Psychologie*, Vol. 2, pp. 227-233.

Stigler, J.W., P. Gonzales, T. Kawanaka, S. Knoll and A. Serrano (1996), "The TIMSS Videotape Classroom Study: Methods and Preliminary Findings", Prepared for the National Center for Education Statistics, U.S. Department of Education, Los Angeles, CA.

Tacke, G. and F. Linder (1981), "Der Einfluss individualisierenden Lehrerverhaltens auf das Selbstkonzept von Schülern", *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, Vol. 13, pp. 190-193.

Valtin, R. (2002), *Was ist ein gutes Zeugnis? Noten und verbale Beurteilungen auf dem Prüfstand*, Juventa, Weinheim.

Wagner, C. and R. Valtin (2003), "Noten oder Verbalbeurteilungen? Die Wirkung unterschiedlicher Bewertungsformen auf die schulische Entwicklung von Grundschulkindern", *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, Vol. 35, pp. 27-36.

Winter, F. (1991), *Schüler lernen Selbstbewertung*, Peter Lang, Frankfurt.

Also available in the CERI collection

Internationalisation and Trade in Higher Education *– Opportunities and Challenges*
250 pages • September 2004 • ISBN: 92-64-01504-3 • 50 euros

Quality and Recognition in Higher Education *– The Cross-border Challenge*
203 pages • July 2004 • ISBN: 92-64-01508-6 • 30 euros

Innovation in the Knowledge Economy *– Implications for Education and Learning*
Knowledge Management series
96 pages • May 2004 • ISBN: 92-64-10560-3 • 21 euros

Equity in Education *– Students with Disabilities, Learning Difficulties and Disadvantages*
165 pages • May 2004 • ISBN: 92-64-10368-6 • 40 euros

Disability in Higher Education
168 pages • December 2003 • ISBN: 92-64-10505-0 • 24 euros

Measuring Knowledge Management in the Business Sector: First Steps
Knowledge Management series
219 pages • December 2003 • ISBN: 92-64-10026-1 • 40 euros

New Challenges for Educational Research
Knowledge Management series
146 pages • August 2003 • ISBN: 92-64-10030-X • 21 euros

Networks of Innovation *– Towards New Models for Managing Schools and Systems*
Schooling for Tomorrow series
182 pages • June 2003 • ISBN: 92-64-10034-2 • 25 euros

Understanding the Brain *– Towards a New Learning Science*
115 pages • July 2002 • ISBN: 92-64-19734-6 • 23 euros

www.oecd.org/bookshop

Table of Contents

PART III. THE LITERATURE REVIEWS