OECD 《 ● CERI ↗

WHAT WORKS IN INNOVATION IN EDUCATION

FORMATIVE ASSESSMENT IN CLASSROOMS:
A REVIEW OF THE EMPIRICAL GERMAN LITERATURE

By

Olaf Köller, University of Erlangen-Nuremberg

NOT FOR CITATION

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

---

[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.

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 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.

---

[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).

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.