



Quality of teaching in four European countries: a review of the literature and application of an assessment instrument

Wim van de Grift

To cite this article: Wim van de Grift (2007) Quality of teaching in four European countries: a review of the literature and application of an assessment instrument, Educational Research, 49:2, 127-152, DOI: [10.1080/00131880701369651](https://doi.org/10.1080/00131880701369651)

To link to this article: <http://dx.doi.org/10.1080/00131880701369651>



Published online: 16 May 2007.



Submit your article to this journal [↗](#)



Article views: 733



View related articles [↗](#)



Citing articles: 22 View citing articles [↗](#)

Quality of teaching in four European countries: a review of the literature and application of an assessment instrument

Wim van de Grift*

Netherlands Inspectorate of Education, Utrecht, The Netherlands

Background

From 2002 onwards, initiatives and first steps for the project International Comparative Analysis of Learning and Teaching (ICALT) have been taken by the inspectorates of education in England, Flanders (Belgium), Lower Saxony (Germany) and The Netherlands. The inspectorates of education in these European countries reviewed the results of research on the basic characteristics of good and effective teaching and selected standards and indicators for an observation instrument that could be used to evaluate the quality of teaching. The inspectorates from these countries jointly developed an instrument to observe and analyse the quality of learning and teaching in primary schools.

Purpose

The observation instrument was piloted for reliability and inter-rater reliability, and for validity, in the four countries.

Sample

Mathematics lessons in England, Flanders (Belgium), Lower Saxony (Germany) and The Netherlands were observed in 854 classrooms, with children who were about 9 years old when they started the school year.

Design and methods

Inspectors in the four countries were trained in the proper use of the observation instrument, and used the instrument during their own inspections or evaluations.

Results

This study shows that the quality of teaching in the four countries can be compared in a reliable and valid way as regards five aspects: 'efficient classroom management', 'safe and stimulating learning climate', 'clear instruction', 'adaptation of teaching' and 'teaching-learning strategies'.

Conclusions

It is found that only a few percentage points of difference between teachers are due to differences existing in the four countries. Furthermore, it may be concluded that the five aspects of quality of teaching are positively and significantly correlated with pupil involvement, attitude, behaviour and attainment.

Keywords: *Teaching; Classroom management; Learning climate; Instruction; Adaptation of teaching; Learning strategies; Pupils' involvement; Observation*

*Netherlands Inspectorate of Education, PO Box 2730, 3500 GS Utrecht, The Netherlands.

Email: W.vandeGrift@owinsp.nl

ISSN 0013-1881 (print)/ISSN 1469-5847 (online)/07/020127-26

© 2007 NFER

DOI: 10.1080/00131880701369651

Introduction

There are already many publications on comparative education data. For instance, the OECD publishes *Education at a glance* every year, and the European Commission publishes annually its *Key data on education in Europe*. These publications inform us about the international differences in teachers' salaries, contact hours, class sizes, and so on, but there are no publications that provide reliable and valid information on international differences in the quality of learning and teaching strategies.

Since 1960, many observation instruments have been developed which can be used to evaluate the quality of teaching (Flanders, 1961, 1970; Stallings & Kaskowitz, 1974; Tricket & Moos, 1974; Stallings *et al.*, 1979; Capie *et al.*, 1980; Florida Coalition for the Development of a Performance Measurement System, 1983; Stringfield *et al.*, 1985; Veenman *et al.*, 1986; Evertson, 1987; Slavin, 1987; Virgilio, 1987; Evertson & Burry, 1989; Virgilio & Teddlie, 1989; Teddlie *et al.*, 1990; Schaffer & Nesselrodt, 1992; Booij *et al.*, 1995; Ofsted, 1995; Houtveen & Overmars, 1996; The Netherlands Inspectorate of Education, 1998; Houtveen *et al.*, 1999a). These classroom observation instruments differ in quality and scope, and some instruments borrowed a lot from earlier published tools. Several instruments had been developed originally for teacher training rather than research purposes, but were very useful for selecting items that are in agreement with the results of the review of research literature presented below. The mentioned observation instruments are seldom or never used in international comparisons. There are, however, some interesting exceptions (Anderson *et al.*, 1989; Postlethwaite & Ross, 1992), but these are only small-scale studies or studies based on information from questionnaires and not from observations of independent observers.

The Inspectorates of Education in several countries in Europe worked together to develop an observation instrument for use, in an international context, in evaluating the quality of teaching in elementary education. This paper reports about the construction of this observation instrument.

Theoretical framework

Several high-quality reviews of research on the results of studies on the relationship between the basic characteristics of teaching and the academic achievements of pupils are available (Purkey & Smith, 1983; Levine & Lezotte, 1990, 1995; Scheerens, 1992; Walberg & Haertel, 1992; Creemers, 1994; Ellis & Worthington, 1994; Cotton, 1995; Sammons *et al.*, 1995). According to these reviews, several categories of variables are associated with the effectiveness of the educational process: opportunity to learn, learning and instruction time, classroom management, learning climate, instruction, adaptive teaching, teaching-learning strategies, monitoring pupils' results, special measures for struggling learners and pupil engagement.

Our objective was to construct an event sampling observation instrument that could be used in classrooms every time an inspector visits a classroom. An important characteristic of such an observation instrument is that the standards and indicators must be observable in (almost) each lesson. Some events do not happen every time, in

every classroom, and do not have to happen every time either. With some categories of variables, some problems will arise when we stick to the use of an observation instrument: opportunities to learn, monitoring pupils' results and special measures for struggling learners. It is not to be expected, in every lesson, that important reductions or alterations in the curriculum take place, or pupils' results are monitored, or special measures taken for struggling learners observed. Further, on these three issues, the most important decisions are taken at school level, not by individual teachers. For these categories other kinds of instruments seem to be more suitable, which was not the aim of this study.

Another problem has to do with learning time. A great deal of research shows that the amount of learning time is a good predictor for the effectiveness of teaching (Rosenshine & Berliner, 1978; Berliner, 1988; Karweit, 1989). Time, however, has no meaning in itself: the way time is used is the important thing (Stallings, 1980). Time is an instrument to measure, for instance, the opportunity pupils get to learn the curriculum, or to measure the efficiency of classroom management, or to give struggling learners better opportunities to master the basics of the curriculum. Therefore, we have decided to use no extra observation category for the use of learning time as such. For the remaining categories we studied the research literature in detail in order to identify standards and indicators of good and effective teaching.

Efficient classroom management

Bloom (1976), Carroll (1963), Harnischfeger and Wiley (1978) and Wyne and Stuck (1982) identified several ways of increasing allocated time, namely beginning and ending lessons on time, reducing transition time and minimizing wasted time. Other important factors are ensuring there are no queues waiting at the teacher's desk, well-structured lessons and orderly lesson progression.

Several researchers developed instruments for observing efficient use of learning and teaching time in the recent past. Evertson and Anderson (1978) used time logs; time sampling instruments were used by Flanders (1961), Veenman *et al.* (1986), Roelofs (1993), The Netherlands Inspectorate of Education (1998); and Houtveen *et al.* (1999b) used event sampling instruments. The connection between time spent and pupils' results was established in a large number of empirical research projects (Scheerens & Bosker, 1997). Several researchers stress the optimal use of time in terms of classroom management, as well as time spent on explicit instruction of skills and integration of skills (Carnine *et al.*, 1998; Dixon *et al.*, 1998). The average teacher in primary education spends 43% of the lesson time on instruction, 50% of the time on working on assignments and 7% on management and organization (Lam, 1996). Research has suggested that effective teachers spend 15% less time on management and organization and 50% more time on instruction and interactive activities, such as questioning and answering, and providing corrective feedback or explanations. The findings of several studies indicate that more academically effective teachers had generally better-organized classrooms and fewer behaviour problems with pupils (Brophy & Evertson, 1976; Anderson *et al.*, 1979).

Not only is the efficient use of time highly important, but also the balance of activities. When individual work is excessive, pupil engagement may decrease (Rosenshine & Berliner, 1978; Rosenshine, 1980). Effective teachers organize their time such that they can spend at least some time with the total group, in small groups and with individuals (Borg, 1980; Kindsvatter *et al.*, 1988; Creemers, 1994).

Safe and stimulating learning climate

Several reviews on the relationships between the characteristics of the learning climate and pupils' achievements have already been published and some date from more than 20 years ago (Moos, 1974; Andersen, 1982; Fraser, 1985). A good learning climate consists of two main elements: it is safe for children, and it stimulates them to learn.

Safe and orderly climate. A safe and orderly climate is not only good for pupils' health and well-being, but also for pupils' results. Schweitzer (1984) found, in a study covering 16 primary schools, that an orderly and safe environment had a correlation of no less than 0.59 with the achievement of pupils. Vermeulen (1988) found, in a study of 17 primary schools, that an orderly and secure educational climate was significantly correlated (0.34) with achievement. Evertson *et al.* (1980) found, in a study of 86 teachers, that pupil obedience to teachers was significantly related to achievement in mathematics (0.14) but not to achievement in mother tongue instruction (0.06). They found similar results for the consistency of enforcement of rules and effective teacher organization and control. Consistency of enforcement of rules was significantly related to achievement in mathematics (0.15) but not to achievement in mother tongue instruction (0.02), while effective organization and control rules were significantly related to achievement in mathematics (0.19) but not to achievements in mother tongue instruction (0.03). These results concerning the relationship between a safe and orderly climate with pupil achievements were also found in secondary schools (Spade *et al.*, 1985; van de Grift *et al.*, 1997).

Stimulating learning environment. Several studies carried out between 1967 and 1980 make clear that specific characteristics of a learning environment have an important influence on the academic achievement of pupils (McDill *et al.*, 1967, 1969; Weber, 1971; Office of Educational Performance Review, 1974; Brookover & Schneider, 1975; Wellish *et al.*, 1978; Brookover *et al.*, 1978; Rutter *et al.*, 1979, 1980). The findings of some of these studies also indicate that 20% to 40% of the differences between pupils' achievements could be explained by school climate factors. Not all these studies were carried out in primary education, however, and not all had a very precise definition of educational climate. Sometimes instruction, the monitoring of pupils' achievement and educational management were included in the definitions used. Nevertheless, these early studies inspired many researchers to carry out more precise research on large samples and with improved statistical techniques such as multi-level regression analyses. This more recent research, performed using more effective techniques in better-quality samples and by means of careful quasi-experiments, indicates that the expectations in the 1970s were somewhat optimistic.

However, these more recent studies have found that certain differences between the achievements of pupils in different classrooms can be explained by specific characteristics of the learning climate. We now examine these results in greater detail.

Self-confidence through positive expectations by teachers. Evertson *et al.* (1980) found, in a study of 86 teachers, that teacher praise of pupil responses was significantly related to achievement in mathematics (0.19) and to achievement in mother tongue instruction (0.09). Schweitzer (1984) found, in a study of 16 primary schools, that teachers' high expectations of learning achievement had a correlation of 0.79 with pupil achievement. Teddlie and Stringfield (1984) found, in a study of 76 primary schools, that teachers' expectations had positive correlations (0.22 to 0.54) to results on norm-referenced test results in basic skills. Mortimore *et al.* (1988) found, in a study of 50 primary schools, that a positive climate was positively related to learning achievement. Houtveen *et al.* (2004) found, in a quasi-experiment in primary education, that achievement in mathematics increased significantly when teachers stimulated pupil self-confidence. These findings were confirmed also in secondary education (Madaus *et al.*, 1979; Van Marwijk Kooij-Von Baumhauer, 1984; Scheerens *et al.*, 1989; Van der Werf & Tesser, 1989; van de Grift *et al.*, 1997).

It is vital to give initially less successful pupils a second chance to demonstrate success after corrective feedback (Guskey, 2003). Successful pupils believe self-improvement is possible and are continually motivated to achieve this goal (Ellis & Worthington, 1994). There is considerable evidence that there is a positive correlation between high success rates and pupils' learning outcomes while the correlation with low success rates is negative (Anderson *et al.*, 1979; Fisher *et al.*, 1980). In addition to increasing academic achievement, successful experiences as regards tasks relates positively to internalized pupil attributions of success (Anderson *et al.*, 1988). Pupils who experience frequent failure tend to attribute their success to other external factors (such as luck, task ease) and may, over a period of time, exhibit behavioural characteristics associated with 'learned helplessness' and engage in task avoidance behaviour (Thomas & Pashley, 1982; Adelman & Taylor, 1983). Therefore, the rate of success at which a pupil completes a task should be regarded as a critical instructional principle (Ellis & Worthington, 1994). It is assumed that all pupils can master a subject given sufficient time and appropriate instruction (Block, 1980). The same study from Block (1980) provides some evidence that learning increases when pupils experience high rates of success during instructional activities.

Emphasis on basic skills and an achievement-oriented attitude. Schweitzer (1984) found, in a study of 16 primary schools, that teacher emphasis on basic skills had a correlation of 0.12 with the achievement of pupils. These results are consistent with the results of Bosker and Hofman (1987), in a study of 72 primary schools. They found that an achievement-oriented attitude and teacher experience explained 28% of the variance between schools in mathematics and 8% of the variance between schools in language subjects.

These findings in primary education were not fully consistent with the results of an early study in 1976 by Brimer *et al.* They found, in their study of 44 secondary

schools, that there were positive and negative relationships of pressure on examination achievement. In later studies, this problem of negative relationships seems to disappear. Madaus *et al.* (1979) found, in a study of 52 secondary schools, that achievement pressure explained some variance in exam results and in school achievement measured with a standardized test. Seven years later, Tesser (1986) found, in a study of 335 secondary schools, that an achievement-oriented school policy was positively related (0.13) to the education level attained. We might conclude that teachers' emphasis on basic skills and an achievement-oriented attitude of teachers has a moderate but significant relationship with pupils' achievements.

Self-regulated learning. There is good reason why classrooms should be organized in a way that invites pupils to regulate and monitor their own learning behaviour and to assist pupils in becoming independent and self-regulatory. Research shows clear relationships between pupil self-management and learning results (Brown, 1978; Ellis & Friend, 1991; Ellis & Worthington, 1994; Boekaerts, 2002). Effective learners differ from ineffective learners in their ability to regulate and monitor their own behaviour in terms of motivation, socialization, and academic and cognitive demands. Effective learners have, for example, an internal locus of control; they actively use prior knowledge and skills to gain new knowledge and skills; and they work actively to self-regulate their thoughts and actions (Ellis & Worthington, 1994; Boekaerts *et al.*, 2000).

An intellectually challenging teaching climate. Instruction can only last as long as classroom layout supports it. A good 'infrastructure' is required that could be described as an explorative learning environment. Apart from organizational reasons, an explorative learning environment has value in itself, because it contributes to school success and the intrinsic motivation of pupils (Carver & Scheier, 2000; Ryan & Deci, 2000). Mortimore *et al.* (1988) found, in a study of 50 primary schools, that an intellectually challenging teaching climate and a work-centred setting was positively related to learning achievement. Houtveen *et al.* (2004) found, in a quasi-experiment in primary education, that achievement in mathematics increased significantly when teachers created an explorative learning environment.

Clear instruction

In the literature on the effectiveness of instruction, the following categories are found: clear instruction with clear objectives, well-structured lessons, activating pupils and, where needed, 'direct' instruction.

Clear objectives. Clear specification of lesson objectives to pupils (Melton, 1978) and the use of clear and simple language (Land, 1987) contribute to instructional quality. Mortimore *et al.* (1988) found, in a study of 50 primary schools, that a limited number of focal points in each lesson were positively related to learning achievement. Other highly important aspects were presenting information in an orderly manner (Kallison, 1986), noting transitions to new topics (Smith & Cotton, 1980), the use of

a wide variety of vivid images and examples (Mayer & Gallini, 1990; Hiebert *et al.*, 1991) and frequent restatement of essential principles (Maddox & Hoole, 1975).

Well-structured lessons. Mortimore *et al.* (1988) found, in a study of 50 primary schools, that structured lessons had a positive effect on learning achievement. Van der Werf and Tesser (1989) also found, in their study of 184 primary schools, that structured teaching had significant effects on the advice that primary school pupils were given as regards the choice of type of secondary school.

It is important to note transitions to new topics (Smith & Cotton, 1980); a well-structured lesson has the right balance of activities. Confrey and Good (1981) found that low-ability classes spent much of their time on repetition and drill activities and less time on instruction activities by the teacher. Effective teachers organize their lessons, so they can spend some time with the total group, small groups and individuals (Borg, 1980; Kindsvatter *et al.*, 1988).

Activating pupils. Anderson *et al.* (1979) found that the percentage of academic interactions in which the pupil gave the correct answer was positively related ($R = 0.49$) to achievement gain. Evertson *et al.* (1980) found, in a study of 86 teachers, that the use of class discussion by teachers was significantly related to achievement in mathematics (0.19) but not to achievement in mother tongue learning (0.03). They also found that there was a significant correlation between teacher acceptance of pupil ideas and contributions and achievements in mathematics (0.21 to 0.28), but not always achievements in mother tongue learning (-0.03 to 0.09). Rosenshine (1980) found that a pupil's task engagement rate appears to increase when the teacher directs activities actively. Task engagement during seatwork may be optimized when the teacher interacts substantively with pupils (Rosenhine, 1980). The most important aspect of instructional quality is, however, the degree to which the lesson makes sense to the pupils. Lessons should be related to pupils' background knowledge, using such devices as advanced organizers (Pressley *et al.*, 1992; Nunes & Bryant, 1996), or simply reminding pupils of previously learned material at relevant points in the lesson. Use of modern media and other visual representations can also contribute to quality of instruction (Kozma, 1991). The same is true for the use of many vivid images and examples (Mayer & Gallini, 1990; Hiebert *et al.*, 1991).

Direct instruction. Rosenshine (1979) found that effective teachers incorporate an instructional sequence, called direct instruction, into their lessons. Direct instruction is characterized by beginning the lesson with a short statement of goals, reviewing previous learning, presenting new material in small steps, allowing pupils practice time after each step, giving clear and detailed instructions/explanations, providing active and ample practice, asking questions, checking for understanding and obtaining responses from all pupils, providing guided practice and explicit instruction. Rosenshine emphasized that these steps are especially appropriate when material is new, difficult or hierarchical, or when pupils are young or experiencing learning

difficulties. The Netherlands Inspectorate of Education found that 82% of the teachers started their arithmetic lesson with a short statement of goals. Some 89% of the teachers presented new arithmetic material in small steps, 84% of the teachers gave clear and detailed instructions or explanations and 74% of the teachers asked questions, checked for understanding and obtained responses from all pupils (Van de Grift, 1994). The 'direct instruction model' has been proven to be effective, especially for young children and children with less academic abilities (Becker & Carnine, 1981; Rosenshine, 1986; Baumann, 1988; Kameenui & Carnine, 1988; Dixon *et al.*, 1992, 1998; Veenman, 1992; Muijs & Reynolds, 2003; Houtveen *et al.*, 2004; Houtveen & van de Grift, 2006).

Adaptive teaching

There has been much debate concerning the effectiveness of mixed-ability groups. Spade *et al.* (1985) found, in a study of 4000 pupils in secondary schools, a low and positive relationship (0.10) between streaming of pupils and mathematics achievement. Brimer *et al.* (1976) found, in a study involving 44 secondary schools, low positive and high negative relationships between mixed-ability groups and exam results. The findings seem to be more consistent in elementary education. In elementary schools heterogeneous groups appear to provide the best opportunity to learn for both low-achieving pupils and average pupils (Slavin, 1987, 1996; Gamoran, 1992; Oakes *et al.*, 1992; Reezigt, 1993; Hallam & Toutounji, 1996; Houtveen & van de Grift, 2001). High-quality instruction given to the whole class is essential. Heterogeneous grouping is, however, certainly not enough to help those pupils who are at risk of school failure. These pupils require extended learning and instruction time. In all cases, the extension of instruction time for struggling learners demands a classroom organization in which the remainder of the pupils are able to manage their own learning process. Houtveen *et al.* (2004) found, in a quasi-experiment in primary education, that achievement in mathematics increased significantly when teachers diagnosed pupils' academic problems through testing and implemented prescribed plans for pupils identified at risk. In a study by Houtveen and van de Grift (2006) a significant positive relationship was found between pupils' achievements on initial reading and frequently diagnosing pupils' academic problems through testing and implementing prescribed plans for pupils identified at risk.

Teaching learning strategies

Due to emerging research on cognition and information processing, so-called cognitive strategies have been developed in a number of subject areas that pupils can use to help perform higher-level operations (Van Parreren, 1988; Dixon *et al.*, 1992; Carnine *et al.*, 1998). A cognitive strategy is a heuristic that serves to support the learner, facilitating the development of internal procedures that enable him/her to perform the higher-level procedures. In teaching less-structured tasks, the teacher uses 'scaffolds' to support the pupils as the pupils learn the cognitive strategy and then the cognitive strategy supports pupils while they attempt to complete the

less-structured task. Scaffolds are forms of support provided by the teacher (or another pupil) to help pupils bridge the gap between their current abilities and the intended goal. They can be seen as adjustable and temporary support that can be removed when no longer necessary (Palinscar & Brown, 1984). Scaffolding procedures reduce the complexities of problems, breaking them down into manageable chunks that the child has a real chance of solving (Bickhard, 1992). Examples of teachers' scaffolds include: providing simplified problems, modelling procedures and thinking aloud as they solve the problem. Scaffolds may also be tools such as cue cards or checklists. Scaffolds are gradually withdrawn or faded as learners become more independent, although pupils may continue to rely on scaffolds or periodically request them when they encounter particularly difficult problems (Rosenshine & Meister, 1997; Carnine *et al.*, 1998). Teachers who explicitly model, scaffold, explain strategies, give corrective feedback and ensure that children master the material taught contribute highly to the academic success of their pupils (Evertson *et al.*, 1980; Good & Brophy, 1986; Rosenshine & Stevens, 1986; Dixon *et al.*, 1992, 1998; Veenman, 1992; Ellis & Worthington, 1994; Slavin, 1996; Carnine *et al.*, 1998).

Involvement of pupils

The degree of pupil engagement during an allocated period of time is of major concern (Rosenshine & Berliner, 1978; Denham & Lieberman, 1980; Fisher *et al.*, 1980). Average pupil engagement rates during an instructional activity are 60% to 75%, but may range from 30% to 90% (Kindsvatter *et al.*, 1988). The Dutch Commissie Evaluatie Basisonderwijs (1994) found, in a large sample, that in 88% of the lessons no more than three pupils were not engaged for a period of several minutes. In 1% of the lessons, at least one-third of the pupils were not attentive during instruction. Van der Meer *et al.* (1986), Veenman *et al.* (1986) and Houtveen *et al.* (1999b) found average engaged learning time to be 75%. They found ranges of engaged learning time between 60% and 86% during reading lessons and ranges of between 45% and 89% during arithmetic lessons.

Engaged learning time varies considerably from classroom to classroom (Kindsvatter *et al.*, 1988). There is evidence that teachers can influence pupils' engaged learning time by avoiding excessive seatwork and by making lessons more interactive (Rosenshine & Berliner, 1978; Rosenshine, 1980).

Denham and Lieberman (1980) found empirical evidence that the degree to which pupils are engaged during allocated periods of time was directly and positively related to their learning outcomes.

Development of an observation instrument

In the previous sections, we have opted for the use of an observation instrument. A reasonable alternative would be the questionnaire method; a good deal of experience has been acquired with this method over time. The great advantage is, of course, that working with questionnaires is cheap and efficient. The disadvantage, however, is that direct questioning of teachers regarding their own teaching makes it necessary to

introduce correction for socially desirable responses; after all, we are solely interested in the actual strategies adopted by teachers and not in their knowledge of what constitutes a 'good' response. While some experience has been gained with correction for social desirability, correction techniques used have seldom been able to fully convince the critics.

Another alternative is indirect questioning, by asking pupils and head teachers to complete questionnaires about the teaching strategies deployed by their teachers. The study we have in mind, however, focuses on the teaching of relatively young children. There must be some doubt as to whether these children are yet able to make objective and stable assessments. There is the additional consideration that they would simply not yet be able to assess some of the factors involved here. Even if they were able to confirm the quality of the teaching accurately, they would still have difficulty in detaching this from their own direct experience, in which they may (and are allowed to) differ from their fellow classmates. This leaves us with indirect questioning of head teachers. We are not great supporters of this method, since many school heads hardly ever visit lessons taught by their teachers, and if they do, it is seldom for the purpose of systematic observation of teaching strategies. A 'halo' effect is likely to emerge with the questioning of head teachers, where the idea of the overall performances of their teachers can highly influence the answers they give to the questionnaires. Head teachers could be requested, of course, to make observations in classes taught by their teachers, but this brings us back to the topic of observation instruments.

Available observation instruments

The observation instruments we have previously mentioned can be divided into several categories: time logs or time sampling instruments versus event sampling instruments; narrative recordings; low-inference versus high-inference instruments; and one-dimensional versus multidimensional instruments.

Time sampling instruments versus event sampling instruments. One of the oldest time sampling instruments is probably an instrument developed by Flanders in 1961. He made a distinction between student talk and teacher talk with direct and indirect influence. Time sampling instruments are often used for pupil engagement ratings and teacher behaviour. Time logs, a special case of time sampling instruments, are used to obtain information on how long transitions took and how much time was spent on each type of activity. Time sampling instruments were used by Stallings and Kaskowitz (1974) and Evertson and Anderson (1978). Evertson and Anderson (1978) asked observers to count every 15 minutes how many pupils could be classified in each of the following categories of engagement: on-task academic, on-task procedural, off-task sanctioned, off-task unsanctioned and 'can't tell'.

Narrative recording instruments were used by Evertson and Anderson (1978) and the HM Inspectorate in England. Narrative recording generally presupposes several categories like organization and management, or teaching and learning. Observers are

asked to keep written records of all the occurrences which involved these categories. Sometimes narrative recordings are quantified according to component ratings, or scored on the basis of judgements such as sufficient or insufficient. For example, the English inspectorate used a narrative recording instrument with four categories: teaching, response, attitude and behaviour.

Event sampling instruments were used by Stringfield *et al.* (1985), Virgilio (1987), Schaffer and Nesselrodt (1992) and The Netherlands Inspectorate of Education (1998). Event sampling instruments are used to score behaviour, including: presenting information clearly, developing positive attitudes, managing pupil behaviour, etc. Event sampling instruments are used, for instance, by The Netherlands Inspectorate of Education for reasons of convenience, a good 'fit' with the inspectors' job and standardization.

Low- and high-inference instruments. Many low-inference instruments are simply objective counts of discrete behaviours. Low-inference observation instruments are often called coding systems or rating systems. Time logs and time sampling instruments are examples of low-inference instruments. In these instruments, the variables are computed by dividing the total number of times that an event is observed during an observation period of a lesson by the total observation time. The variables of time logs and time sampling instruments are rate variables representing the average frequencies, proportions or ratios of an event per observation period. High-inference instruments are more vulnerable to subjectivity. Nevertheless, we have chosen an instrument with a high-inference standard, and tried to solve the problem of subjectivity with the help of 'good practice indicators' in combination with thorough procedures of inter-rater reliability.

One-dimensional versus multidimensional instruments. At first sight, it seems that in different parts of the world different processes of teaching are practised. The most impressive example of this is the so-called Chinese paradox: Watkins and Biggs (1996) painted a research-based picture of how Chinese pupils and their teachers see the context of their learning. Their research is focused on the question: 'how can Chinese learners be so successful academically, when their teaching and learning seem to be so orientated to rote memorization?' How is it possible that Chinese learners often outperform their Western peers, when their learning style seems so passive? They conclude that this paradox might be explained by the cross-cultural differences in the very process of teaching and learning. These differences concern the relationship between memorizing and understanding, and even the nature of motivation. We can learn from the Chinese case that it does not seem wise to adopt a one-dimensional view of teaching. If it is true that different teaching strategies are effective in different cultures, contexts and circumstances, then it might be true that effective teaching means looking for and working on the right balance in a range of activities and teaching strategies. From this point of view, we regard teaching as a multidimensional construct. Teaching will be defined as a range of behaviours that enable pupils to learn effectively, not as a one-dimensional trait.

Construction of an observation instrument

For the construction of the observation instrument, two methods were used to identify the indicators of good teaching and learning. The first was a study into the research that has been done on the quality of teaching. The second was a consensus-seeking discussion between the central inspectorates in England, Flanders (Belgium), Lower Saxony (Germany), The Netherlands and later with North-Rhine Westphalia (Germany), Scotland, Ireland and the Czech Republic. One significant finding has been that a great deal of agreement exists among different European inspectorates as to the basic elements of good and responsible teaching.

The discussions between the inspectorates fed with results of research formed the base for the development of an observation instrument. Working on the assumption of the internationally constant basic elements of teaching, the teaching strategies adopted by teachers were put into an operational format in an observation instrument under six quality characteristics, which together comprise 24 indicators. For every item in this scale, several examples of 'good practice' have been formulated. These examples of good practice may help the observer to focus attention on the same entities as other observers while making the observations. Inspectors can score these indicators: (1) predominantly weak; (2) more weaknesses than strengths; (3) more strengths than weaknesses; and (4) predominantly strong. This instrument is brief and straightforward and can be completed easily by an inspector during a lesson period of approximately 40 minutes. This instrument has been tested on reliability and validity.

Samples. During 2003, 854 mathematics lessons were evaluated with the instrument in order to evaluate the quality of teaching in England, Flanders (Belgium), Lower Saxony (Germany) and The Netherlands. The mathematics lessons were observed in classrooms with children about 9 years old when they started the school year. The Flemish, Lower Saxony and Dutch schools in this project were selected as simply the next school to be visited by an inspector, and there was no other selection criterion. This procedure is essentially the same as the procedures of drawing a simple random sample, so these samples are representative for their respective populations. The English schools were not selected by a random sampling technique. These schools were schools visited in the final year of 'the second inspection cycle'. This means that the English sample is not representative. The results in the English schools are likely to be better than the population mean in England.

Internal consistency. We computed the internal consistency of each of the scales with Cronbach's coefficient alpha. The set of items of each scale should be correlated with one another, the interrelationship of the set of items in a scale is sufficient when the coefficient exceeds 0.70. The results of our analysis are presented in Table 1. All scales were internally consistent ($\alpha > 0.70$) in all four countries.

Inter-rater reliability. Inter-rater reliability relates to the behaviour of persons making the observations, the key question being: do different observers working in identical

situations come to identical assessments? This reliability has been studied in a number of small bilateral partnership projects of the inspectors of The Netherlands Inspectorate of Education working with inspectors of the English, Flemish and Lower Saxony inspectorates respectively. Pairs of inspectors visited classrooms in The Netherlands, England, Lower Saxony and Flanders (Belgium). The dual observations took place at a stage at which the instrument was not fully developed, hence some data are missing. Consensus figures between 75% and 91% were encountered between Dutch and English inspectors in both Dutch and English lessons; with Flemish and Dutch inspectors in Flemish lessons; and with Lower Saxony and Dutch inspectors in German lessons. Overall, we found consensus in 83% of the cases between two independent observers working in identical situations. We can conclude from this that the English, Lower Saxony, Flemish and Dutch inspectors rated identical situations in an identical way.

Construct validity. It is vital, of course, that the observation instrument measures what we want it to measure—i.e. the quality of teaching. Important indications for this are the correlations found between the scales of the observation instrument and other instruments designed for the same purpose. The English inspectorate works with a relatively open observation instrument, in which the quality of the teaching is measured against two observation categories: ‘teaching’ and ‘response’. These categories are scored on a seven-point scale. The English inspectors not only used the international observation instrument, but also their own instrument. This offers the possibility to compute correlations between both instruments. These correlations are presented in Table 2. For the five scales of the instrument correlations around 0.70 are found with ‘teaching’ and ‘learning’, which is fairly high.

Another means of judging construct validity is by computing on correlations between the five scales an overall judgement about the quality of teaching; the correlations are presented in Table 3. The five scales of the instrument have correlations between 0.59 and 0.72. We may conclude that the scales ‘safe and stimulating learning climate’, ‘clear instruction’, ‘adaptation of teaching’, ‘teaching–learning strategies’ and ‘classroom layout’ have sufficient construct validity for use in international comparisons.

Table 1. Internal consistency (Cronbach’s α) of five scales for observing teaching

| Scale no. and name | Items | Lower | | | | |
|---------------------------------------|-------|--------------------------|----------------------------|--------------------------|-----------------------------------|-----------------------|
| | | England <i>N</i> = 87 | Flanders <i>N</i> = 122 | Saxony <i>N</i> = 402 | The Netherlands <i>N</i> = 243 | All <i>N</i> = 854 |
| Efficient classroom management | 4 | 0.88 | 0.82 | 0.78 | 0.79 | 0.80 |
| Safe and stimulating learning climate | 7 | 0.89 | 0.84 | 0.86 | 0.76 | 0.85 |
| Clear instruction | 7 | 0.90 | 0.79 | 0.83 | 0.81 | 0.84 |
| Adaptation of teaching | 2 | 0.82 | 0.88 | 0.92 | 0.75 | 0.88 |
| Teaching–learning strategies | 3 | 0.81 | 0.78 | 0.77 | 0.70 | 0.77 |

Table 2. Inter-rater reliability (percentage of consensus) of five scales for observing teaching

| | The Netherlands England N = 46 | England The Netherlands N = 50 | The Netherlands Lower Saxony N = 26 | The Netherlands Flanders N = 29 |
|---------------------------------------|--------------------------------------|--------------------------------------|---|---------------------------------------|
| Efficient classroom management | 89 | 92 | 81 | 79 |
| Safe and stimulating learning climate | 85 | 91 | – | 83 |
| Clear instruction | 84 | 87 | 93 | 80 |
| Adaptation of teaching | 78 | 85 | 73 | 75 |
| Teaching–learning strategies | 76 | 85 | 79 | 76 |

Table 3. Correlations between five scales for observing teaching with related concepts

| | N = 87 Teaching | N = 87 Learning | N = 854 Overall quality of teaching |
|---------------------------------------|--------------------|--------------------|--|
| Efficient classroom management | 0.70 | 0.68 | 0.65 |
| Safe and stimulating learning climate | 0.70 | 0.68 | 0.72 |
| Clear instruction | 0.68 | 0.66 | 0.72 |
| Adaptation of teaching | 0.66 | 0.64 | 0.59 |
| Teaching–learning strategies | 0.72 | 0.72 | 0.68 |

Quality of teaching in England, Flanders (Belgium), Lower Saxony (Germany) and The Netherlands

The results of the research on the quality of teaching in mathematics lessons with 9-year-old children in the four European countries are presented in Table 4. All scale scores are standardized by dividing the sum score by the product of the number of items and the number of response categories. As a result of this standardization, the results can be interpreted as: the average teacher satisfies #% of the indicators of the scale.

About 80% of the teachers in the four European countries score positively as regards ‘efficient classroom management’; about three-quarters of the teachers in the four countries score positively as regards ‘safe and stimulating learning climate’ and ‘clear instruction’; and about two-thirds of the teachers in the four countries score positively as regards ‘adapting instruction and assignments’ and ‘teaching–learning strategies’.

There appear to be no significant differences in quality of teaching between teachers in Flanders (Belgium), Lower Saxony and The Netherlands. However, the English teachers observed had better results than the teachers in all three countries on

Table 4. Quality of teaching in maths lessons for 9-year-old children in four countries

| | England <i>N</i> = 87 | Flanders <i>N</i> = 122 | Lower Saxony <i>N</i> = 402 | The Netherlands <i>N</i> = 243 | International average <i>N</i> = 854 | % between country variance |
|---------------------------------------|--------------------------|----------------------------|-----------------------------------|-----------------------------------|--|----------------------------------|
| Efficient classroom management | 82 | 82 | 80 | 79 | 80 | 0.9 |
| Safe and stimulating learning climate | 80 | 77 | 75 | 73 | 75 | 2.7 |
| Clear instruction | 80 | 70 | 74 | 69 | 72 | 8.9 |
| Adaptation of teaching | 75 | 57 | 58 | 65 | 62 | 7.6 |
| Teaching-learning strategies | 74 | 64 | 65 | 63 | 65 | 3.4 |
| Involvement of pupils | 80 | 79 | 80 | 77 | 79 | 0.01 |

the European continent as regards ‘adaptation of teaching’ and had better results as regards ‘clear instruction’ and ‘teaching learning strategies’ than their Flemish and Dutch colleagues. This might be due to the special characteristics of the English sample.

An important question is: how great are the differences in the quality of teaching in the four countries? Using an analysis of variance, we can compute the aspects of the differences that can be attributed to differences in countries. It was found that the differences in quality of teaching in the four countries varied between 0% (for ‘involvement of pupils’ and ‘efficient classroom management’) and less than 9% (for ‘clear instruction’). From data gathered by The Netherlands Inspectorate of Education we know that 17% to 40% of the variance in teaching can be attributed to differences in schools. So the differences in quality of teaching are relatively small in the four countries. Column seven in Table 4 presents an overview.

Quality of teaching and pupils’ involvement, attitude and attainment

It is important that the scores on the observation instrument accord with indicators for the results teachers want to accomplish; hence we computed the correlations between the scales of the observation instrument and the involvement of pupils. These correlations are presented in Table 5. It is observed that correlation between the scales with the involvement of pupils varies by between 0.36 and 0.60.

The English inspectorate uses, in their own inspections for the behaviour of pupils, the observation categories of pupils’ attainment and pupils’ attitude and behaviour. This offered an opportunity to compute the correlations between the scales of the observation instrument and ‘attainment’ and ‘attitude and behaviour’ of pupils. These correlations varied for the five scales by between 0.32 and 0.68.

We may conclude that the scales of the observation instrument explain an important amount of the variance in the involvement, attitude, behaviour and attainment of pupils, which is also important evidence for the predictive validity of this observation instrument.

Table 5. Correlations between five scales for observing teaching with pupils' involvement, attitude and behaviour and attainment

| | <i>N</i> = 854 Involvement of pupils | <i>N</i> = 87 Attitude and behaviour | <i>N</i> = 87 Attainment |
|---------------------------------------|--|--|-----------------------------|
| Efficient classroom management | 0.54 | 0.56 | 0.39 |
| Safe and stimulating learning climate | 0.60 | 0.68 | 0.51 |
| Clear instruction | 0.60 | 0.48 | 0.39 |
| Adaptation of teaching | 0.36 | 0.47 | 0.32 |
| Teaching-learning strategies | 0.50 | 0.61 | 0.56 |

Conclusion

We conclude that this research project has shown that the quality of teaching in England, Flanders (Belgium), Lower Saxony (Germany) and The Netherlands can be compared in a reliable and valid way as regards aspects such as 'safe and stimulating learning climate', 'clear instruction', 'adaptation of teaching', 'teaching-learning strategies' and 'efficient classroom management'. We think it useful also to extend the observation instrument with interview modules for 'opportunity to learn', and the 'monitoring of pupils' achievements'. We conclude that only a few percent of the difference between teachers could be explained by differences between the four countries.

Further, we conclude that a 'safe and stimulating learning climate', 'clear instruction', 'adaptation of teaching', 'teaching-learning strategies' and 'efficient classroom management' are positively and significantly related with 'pupils' involvement', 'attitude and behaviour' and 'attainment'. Of course, it is important to study in more detail the relationships between pupils' achievements and the aspects of quality of teaching. This can be done, for instance, in the forthcoming studies into the international comparison of mathematics and comprehensive reading—see TIMSS: <http://nces.ed.gov/timss/>; PIRLS: <http://timss.bc.edu/pirls2001.html>; IEA: <http://www.iea.cc/>; PISA: <http://www.pisa.oecd.org>; and the OECD: <http://www.oecd.org>.

References

- Adelman, H. S. & Taylor, L. (1983) Enhancing motivation for overcoming learning and behavior problems, *Journal of Learning Disabilities*, 16(7), 384–392.
- Anderson, C. S. (1982) The search for climate, *Review of Educational Research*, 52(3), 368–420.
- Anderson, L. M., Evertson, C. M. & Brophy, J. E. (1979) An experimental study of effective teaching in first grade groups, *Elementary School Journal*, 79(1), 193–223.
- Anderson, L. M., Stevens, D. D., Prawat, R. S. & Nickerson, J. (1988) Classroom task environments and students' task related beliefs, *Elementary School Journal*, 88(3), 281–295.
- Anderson, L. W., Ryan, D. W. & Shapiro, B. J. (1989) *The IEA classroom environment study* (New York, Pergamon).
- Baumann, J. F. (1988) Teaching third-grade students to comprehend anaphoric relationships: the application of a direct instruction model, *Reading Research Quarterly*, 21(1), 70–79.

- Becker, W. C. & Carnine, D. W. (1981) Direct instruction: a behavior theory model for comprehensive educational intervention with the disadvantages, in: S. Bijou (Ed.) *Contributions of behaviour modification in education* (Hillsdale, NJ, Lawrence Erlbaum), 1–106.
- Berliner, D. C. (1988) The half-full glass: a review of research on teaching, in: E. L. Meyen, G. A. Vergason & R. J. Whelan (Eds) *Effective instructional strategies for exceptional children* (Denver, CO, Love).
- Bickhard, M. H. (1992) Scaffolding and self-scaffolding: central aspects of development, in: L. T. Winegar & J. Valsiner (Eds) *Children's development within social context*, Vol. 2 (Hillsdale, NJ, Lawrence Erlbaum).
- Block, J. H. (1980) Success rate, in: C. Denham & A. Lieberman (Eds) *Time to learn* (Washington, DC, National Institute of Education), 95–106.
- Bloom, B. S. (1976) *Human characteristics and school learning* (New York, McGraw-Hill).
- Boekaerts, M. (2002) Bringing about change in the classroom: strengths and weaknesses of the self-regulated learning approach, *Learning and Instruction*, 12(6), 589–604.
- Boekaerts, M., Pintrich, P. R. & Zeidner, M. (Eds) (2000) *Handbook of self-regulation* (San Diego, CA, Academic Press).
- Booij, N., Houtveen, A. A. M. & Overmars, A. M. (1995) *Instructie bij Begrijpend Lezen [Reading comprehension instruction]* (Utrecht, ISOR).
- Borg, W. R. (1980) Time and school learning, in: C. Denham & A. Lieberman (Eds) *Time to learn* (Washington, DC, National Institute of Education), 33–72.
- Bosker, R. J. & Hofman, W. H. A. (1987) Dimensies van schoolkwaliteit: de algemene en milieuspecifieke invloed van scholen op de prestaties en het keuzegedrag van leerlingen [Dimensions of school quality], in: J. Scheerens & W. G. R. Stoel (Eds) *Effectiviteit van onderwijsorganisaties [Effectiveness of school organisations]* (Lisse, Swets & Zeitlinger), 51–70.
- Brimer, A., Madaus, G. F., Chapman, B., Kellaghan, T. & Woodroff, R. (1976) *Differences in school achievement* (Windsor, NFER-Nelson).
- Brookover, W. B. & Schneider, J. M. (1975) Academic environments and elementary school achievement, *Journal of Research and Development in Education*, 9(1), 82–91.
- Brookover, W. B., Schweitzer, J. H., Schneider, J. M., Beady, C. H., Flood, P. K. & Wisenbaker, J. M. (1978) Elementary school social climate and school achievement, *American Educational Research Journal*, 15(2), 301–318.
- Brophy, J. E. & Evertson, C. M. (1976) *Learning from teaching: a developmental perspective* (Boston, MA, Allyn & Bacon).
- Brown, A. (1978) Knowing when, where and how to remember, in: R. Glaser (Ed.) *Advances in instructional psychology*, Vol. 1 (Hillsdale, NJ, Lawrence Erlbaum).
- Capie, W., Johnson, C. E., Anderson, S. J., Ellett, J. C. & Okey, J. R. (1980) *Teacher performance assessment instruments* (Athens, GA, University of Georgia).
- Carnine, D. W. (1979) Direct instruction: a successful system for educationally high-risk children, *Journal of Curriculum Studies*, 11(1), 29–45.
- Carnine, D. W., Dixon, R. C. & Silbert, J. (1998) Effective strategies for teaching mathematics, in: E. J. Kameenui & D. W. Carnine (Eds) *Effective teaching strategies that accommodate diverse learners* (Englewood Cliffs, NJ, Prentice-Hall).
- Carrol, J. B. (1963) A model of school learning, *Teachers College Record*, 64(8), 723–733.
- Carver, C. S. & Scheier, M. F. (2000) On the structure of behavioural self-regulation, in: M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds) *Handbook of self-regulation* (San Diego, CA, Academic Press), 41–84.
- Commissie Evaluatie Basisonderwijs [Commission for the Evaluation of Primary Education] (CEB) (1994) *Inhoud en opbrengsten van het basisonderwijs [Curriculum, process and output of elementary schools]* (De Meern, Inspectie van het Onderwijs).
- Confrey, J. & Good, T. (1981) *Academic progress: student and teacher perspectives* (Lansing, MI, Michigan State University, Institute for Research on Teaching).
- Cotton, K. (1995) *Effective schooling practices: a research synthesis* (Portland, OR, Northwest Regional Educational Laboratory).

- Creemers, B. P. M. (1994) *The effective classroom* (London, Cassell).
- Creemers, B. P. M. & Reezigt, G. J. (1997) School level conditions affecting the effectiveness of instruction, *School Effectiveness and School Improvement*, 7(3), 197–229.
- Denham, C. & Lieberman, A. (1980) *Time to learn* (Washington, DC, National Institute of Education).
- Dixon, R., Carnine, D. W. & Karmenui, E. J. (1992) *Research synthesis in mathematics: curriculum guidelines for diverse learners. Monograph for the National Center to impose the tools of educators* (Eugene, OR, University of Oregon).
- Dixon, R., Carnine, D. W., Lee, D. W. & Wallin, J. (1998) *Review of high-quality experimental mathematics research* (Austin, TX, University of Texas).
- Ellis, E. S. & Friend, P. (1991) Adolescents with learning abilities, in: B. Y. L. Wong (Ed.) *Learning about learning disabilities* (New York, Academic Press), 505–561.
- Ellis, E. S. & Worthington, L. A. (1994) *Research synthesis on effective teaching principles and the design of quality tools for educators* (Technical Report No. 5) (Eugene, OR, University of Oregon).
- Evertson, C. (1987) *Classroom observation record: observation record for project STAR* (Nashville, TN, Vanderbilt University).
- Evertson, C. & Burry, J. (1989) Capturing classroom context: the observation instrument as lens for assessment, *Journal of Personnel Evaluation in Education*, 2(4), 297–320.
- Evertson, C. M. & Anderson, L. M. (1978) *Interim progress report: the classroom organization study* (Austin, TX, University of Texas).
- Evertson, C. M., Anderson, C. W., Anderson, L. & Brophy, J. E. (1980) Relationships between classroom behaviors and student outcomes in junior high mathematics and English classes, *American Educational Research Journal*, 17(1), 43–60.
- Fisher, C., Marliane, R. & Filby, N. N. (1979) Improving teaching by increasing ‘academic learning time’, *Educational Leadership*, 37(1), 52–54.
- Fisher, C. W., Berliner, D. G., Filby, N. N., Marliave, R., Cahen, L. S. & Dishaw, M. (1980) Teaching behaviors, academic learning times, and student achievement: an overview, in: C. Denham & A. Lieberman (Eds) *Time to learn* (Washington, DC, National Institute of Education), 7–32.
- Flanders, N. (1970) *Analyzing teaching behavior* (Reading, MA, Addison-Wesley).
- Flanders, N. A. (1961) *Interaction analysis: a technique for quantifying teacher influence* (Minneapolis, MN, University of Minnesota, College of Education, Bureau of Educational Research).
- Florida Coalition for the Development of a Performance Measurement System (1983) *Domains: knowledge base of the Florida performance measurement system* (Tallahassee, FL, Office of Teacher Education, Certification and In-service Staff Development).
- Fraser, B. J. (1985) *The study of learning environments* (Salem, OR, Assessment Research).
- Gamoran, A. (1992) Is ability grouping equitable: synthesis of research, *Educational Leadership*, 50(1), 11–17.
- Good, T. & Brophy, J. (1986) School effects, in: M. C. Wittrock (Ed.) *Handbook of research on teaching* (New York, Macmillan), 570–602.
- Guskey, T. R. (2003) Assessments learning, *Educational Leadership*, 60(5), 7–11.
- Hallam, S. & Toutounji, I. (1996) *What do we know about the ability grouping of pupils by ability? A research review* (London, University of London Institute of Education).
- Harnishfeger, A. & Wiley, D. E. (1978) Conceptual issues in models of school learning, *Curriculum Studies*, 10(3), 215–231.
- Hiebert, J., Wearne, D. & Taber, S. (1991) Fourth grades’ gradual construction of decimal fractions during instruction using different physical representations, *Elementary School Journal*, 91(4), 321–341.
- Houtveen, A. A. M., Booij, N., de Jong, R. & van de Grift, W. J. C. M. (1999a) Adaptive instruction and pupil achievement, *School Effectiveness and School Improvement*, 10(2), 172–192.
- Houtveen, A. A. M., de Graaf-Haalboom, A. G. & van de Grift, W. J. C. M. (1999b) *Instructie bij spelling [Instruction in spelling]* (Utrecht, ISOR).

- Houtveen, A. A. M., Mijs, D., Vernooy, K., van de Grift, W. J. C. M. & Koekebacker, E. (2003) *Risicoleerlingen bij technisch lezen [Pupils at risk. Evaluation of the technical reading and handling of diverse needs programme]* (Utrecht, ISOR).
- Houtveen, A. A. M. & Overmars, A. M. (1996) *Instructie bij rekenen en wiskunde [Instruction in mathematics education]* (Utrecht, ISOR).
- Houtveen, A. A. M. & van de Grift, W. J. C. M. (2001) Inclusion and adaptive instruction in elementary education, *Journal of Education for Students Placed at Risk*, 6(4), 389–411.
- Houtveen, A. A. M. & van de Grift, W. J. C. M. (2006) *Reading instruction for struggling learners* (Utrecht, ISOR).
- Houtveen, A. A. M., van de Grift, W. J. C. M. & Creemers, B. P. M. (2004) Effective school improvement in mathematics, *School Effectiveness and School Improvement*, 15(3–4), 337–376.
- Kallison, J. M. (1986) Effects of lesson organization on achievement, *American Educational Research Journal*, 23(2), 337–347.
- Kameenui, E. J. & Carnine, D. W. (Eds) (1998) *Effective teaching strategies that accommodate diverse learners* (Columbus, OH, Merrill-Prentice-Hall).
- Karweit, N. (1989) Time and learning: a review, in: R. E. Slavin (Ed.) *School and classroom organization* (Hillsdale, NJ, Lawrence Erlbaum), 69–95.
- Kindsvatter, R., Wilen, W. & Ishler, M. (1988) *Dynamics of effective teaching* (New York, Longman).
- Kozma, R. (1991) Learning with media, *Review of Educational Research*, 61(2), 179–211.
- Lam, J. F. (1996) *Tijd en kwaliteit in het basisonderwijs [Time and quality in primary education]* (Enschede, PrintPartners Ipskamp).
- Land, M. L. (1987) Vagueness and clarity, in: M. J. Dunkin (Ed.) *International encyclopedia of teaching and teacher education* (New York, Pergamon).
- Levine, D. U. & Lezotte, L. W. (1990) *Unusually effective schools: a review and analysis of research and practice* (Madison, WI, Center for Effective Schools Research and Development).
- Levine, D. U. & Lezotte, L. W. (1995) Effective schools research, in: J. A. Banks & C. A. M. Banks (Eds) *Handbook of research on multicultural education* (New York, Macmillan), 525–547.
- Madaus, G. F., Kellaghan, Th., Rakow, E. A. & King, D. J. (1979) The sensitivity of measures of school effectiveness, *Harvard Educational Review*, 49(2), 207–230.
- Maddox, H. & Hoole, E. (1975) Performance decrement in the lecture, *Educational Review*, 28(1), 17–30.
- Mayer, R. E. & Gallini, J. K. (1990) When is an illustration worth ten thousand words? *Journal of Educational Psychology*, 82(4), 715–726.
- McDill, E. L., Meyers, E. D. & Rigsby, L. C. (1967) Institutional effects on the academic behavior of high school students, *Sociology of Education*, 40(3), 181–199.
- McDill, E. L., Rigsby, L. C. & Meyers, E. D. (1969) Educational climates of high schools: their effects and sources, *American Journal of Sociology*, 32(2), 567–586.
- Melton, R. F. (1978) Resolution of conflicting claims concerning the effects of behavioural objectives on student learning, *Review of Educational Research*, 48(2), 291–302.
- Moos, R. H. (1974) *The social climate scales: an overview* (Palo Alto, CA, Consulting Psychologists Press).
- Mortimore, P. (1998) *The road to improvement: reflections on school effectiveness* (Lisse, Swets & Zeitlinger).
- Mortimore, P., Sammons, P. L., Stoll, L., Lewis, D. & Ecob, R. (1988) *School matters: the junior years* (Wells, Open Books).
- Muijs, D. & Reynolds, D. (2000) School effectiveness and teacher effectiveness: some preliminary findings from the evaluation of the mathematics enhancement programme, *School Effectiveness and School Improvement*, 11(3), 247–263.
- Muijs, D. & Reynolds, D. (2003) Student background and teacher effects on achievement and attainment in mathematics: a longitudinal study, *Educational Research and Evaluation*, 9(3), 289–314.
- Nunes, T. & Bryant, P. (1996) *Children doing mathematics* (Oxford, Blackwell).

- Oakes, J., Gamoran, A. & Page, R. N. (1992) Curriculum differentiation: opportunities, outcomes and meanings, in: P. W. Jackson (Ed.) *Handbook of research on the curriculum* (Washington, DC, AERA), 570–609.
- Office for Standards in Education (Ofsted) (1995) *Guidance on the inspection of nursery and primary schools* (London, Ofsted).
- Office of Education Performance Review (1974) *School factors influencing reading achievement: a case study of two inner city schools* (New York, Office of Education Performance Review).
- Palinscar, A. S. & Brown, A. L. (1984) Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities, *Cognition and Instruction*, 1(2), 117–175.
- Postlethwaite, T. N. & Ross, K. N. (1992) *Effective schools in reading* (The Hague, IEA).
- Pressley, M., Wood, E., Woloshyn, V. E., King, M. A. & Menke, D. (1992) Encouraging mindful use of prior knowledge: attempting to construct explanatory answers facilitates learning, *Educational Psychologist*, 27(1), 91–109.
- Purkey, S. L. & Smith, M. S. (1983) Effective schools: a review, *Elementary School Journal*, 83(4), 427–452.
- Reezigt, G. J. (1993) *Effecten van differentiatie op de basisschool [Effects of differentiation in primary education]* (Groningen, RION).
- Roelofs, E. (1993) *Teamgerichte nascholing en coaching [Training and coaching of teams]* (Nijmegen, Vakgroep Onderwijskunde).
- Rosenshine, B. (1979) Content, time and direct instruction, in: P. L. Peterson & H. J. Walberg (Eds) *Research on teaching* (Berkeley, CA, McCutchan), 28–56.
- Rosenshine, B. (1980) How time is spent in elementary classrooms, in: C. Denham & A. Lieberman (Eds) *Time to learn* (Washington, DC, National Institute of Education), 107–126.
- Rosenshine, B. V. (1986) Synthesis of research on explicit teaching, *Educational Leadership*, 43(7), 60–69.
- Rosenshine, B. V. & Berliner, B. C. (1978) Academic engaged time, *British Journal of Teacher Education*, 4(1), 3–16.
- Rosenshine, B. V. & Meister, C. (1997) Cognitive strategy instruction in reading, in: S. A. Stahl & D. A. Hayes (Eds) *Instructional models in reading* (Hillsdale, NJ, Lawrence Erlbaum), 85–107.
- Rosenshine, B. & Stevens, R. (1986) Teaching functions, in: M. C. Wittrock (Ed.) *Handbook of research on teaching* (3rd edn) (New York, Macmillan), 85–107.
- Rutter, M. (1980) School influences on children's behaviour and development, *Pediatrics*, 65(2), 208–220.
- Rutter, M., Maughan, B., Mortimore, P., Austin, J. & Smith, A. (1979) *Fifteen thousand hours: secondary schools and their effects on children* (Wells, Open Books).
- Ryan, R. M. & Deci, E. L. (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being, *American Psychologist*, 55(1), 68–78.
- Sammons, P., Hillman, J. & Mortimore, P. (1995) *Key characteristics of effective schools: a review of school effectiveness research* (London, Ofsted).
- Schaffer, E. C. & Nesselrodt, P. S. (1992) The development and testing of the special strategies observation system. Paper presented at the *AERA Meeting*, San Francisco, CA, 21–23 April.
- Scheerens, J. (1992) *Effective schooling: research, theory and practice* (London, Cassell).
- Scheerens, J. & Bosker, R. (1997) *The foundations of educational effectiveness* (Oxford, Pergamon).
- Scheerens, J., Vermeulen, C. J. A. J. & Pelgrum, W. J. (1989) Generalizability of school and instructional effectiveness indicators across nations, *International Journal of Educational Research*, Special Issue: Development in school effectiveness research (eds B. P. M. Creemers & J. Scheerens), 13(7), 685–825.
- Schweitzer, J. H. (1984) *Characteristics of effective schools* (New Orleans, AER).
- Scriven, M. (1967) The methodology of evaluation, in: R. W. Tyler, R. M. Gagné & M. Scriven (Eds) *Perspectives of curriculum evaluation* (Chicago, Rand McNally), 102–120.
- Slavin, R. E. (1987) Ability grouping and achievement in elementary schools: a best evidence synthesis, *Review of Educational Research*, 57(3), 293–336.
- Slavin, R. E. (1996) *Education for all: contexts of learning* (Lisse, Swets & Zeitlinger).

- Smith, L. R. & Cotton, M. L. (1980) Effect of lesson vagueness and discontinuity on student achievement and attitudes, *Journal of Educational Psychology*, 72(5), 670–675.
- Spade, J. Z., Vanfossen, B. E. & Jones, E. D. (1985) *Effective schools: characteristics of schools which predict mathematics and science performance* (Chicago, AERA).
- Stallings, J. (1980) Allocated learning time revisited, or beyond time on task, *Educational Researcher*, 9(11), 11–16.
- Stallings, J. & Kaskowitz, D. (1974) *Follow through classroom observation evaluation, 1972–1973* (Menlo Park, CA, SRI International).
- Stallings, J., Needels, M. & Stayrook, N. (1979) *How to change the process of teaching basic reading skills in secondary schools* (Menlo Park, CA, SRI International).
- Stringfield, S., Teddlie, C. & Suarez, S. (1985) Classroom interaction in effective and ineffective schools: preliminary results from phase III of the Louisiana School Effectiveness Study, *Journal of Classroom Interaction*, 20(2), 31–37.
- Teddlie, C. & Stringfield, S. (1984) *The Louisiana School Effectiveness Study* (New Orleans, AERA).
- Teddlie, C., Virgilio, I. & Oescher, J. (1990) Development and validation of the *Virgilio teachers' behavior instrument*, *Educational and Psychological Measurement*, 50(2), 421–430.
- Tesser, P. (1986) *Sociale herkomst en schoolloopbanen in het voortgezet onderwijs* [Social background and school careers in secondary education] (Nijmegen, ITS).
- The Netherlands Inspectorate of Education (1998) *Schooltoezicht primair Onderwijs* [Inspection of primary education] (De Meern, Inspectie van het Onderwijs).
- Thomas, A. & Pashley, B. (1982) Effects of classroom training on LD students' task persistence and attributions, *Learning Disability Quarterly*, 5(2), 133–144.
- Tricket, E. J. & Moos, R. H. (1974) *The Classroom environment scale (CES)* (Palo Alto, CA, Consulting Psychologists Press, Inc).
- Van Marwijk Kooij-Von Baumhauer, L. (1984) *Scholen verschillen* [Schools differ] (Groningen, Wolters-Noordhoff).
- Van de Grift, W. J. C. M. (1994) *Technisch rapport van het onderzoek onder 386 basisscholen ten behoeve van de evaluatie van het basisonderwijs* [Report on the evaluation of 386 schools for primary education] (De Meern, Inspectie van het Onderwijs).
- Van de Grift, W. J. C. M., Houtveen, Th. & Vermeulen, C. (1997) Instructional climate in Dutch secondary education, *School Effectiveness and School Improvement*, 8(4), 449–462.
- Van der Meer, T. A. M., Appelhof, P. N., ten Doesschate, J. J. & Kramer, G. B. (1986) *Het onderwijs in de klas tussen 9 en 12* [Education for children between 9 and 12 years of age] (Utrecht, School Advies Centrum/Vakgroep Onderwijskunde).
- Van der Werf, G. & Tesser, P. (1989) The effects of educational priorities on children from lower income families and ethnic minorities, in: B. P. M. Creemers & C. F. Van Parreren *Ontwikkelen onderwijs* [Developmental education] (Amersfoort, ACCO).
- Veenman, S., Lem, P., Voeten, B., Winkelmolen, B. & Lassche, H. (1986) *Onderwijs in combinatieklassen* [Education in multi-graded classrooms] ('s-Gravenhage, SVO).
- Veenman, S. (1992) Effectieve instructie volgens het directe instructie-model, *Pedagogische Studiën*, 69(4), 242–269.
- van Velzen, B., Miles, M., Ekholm, M., Hameyer, U. & Robin, D. (1985) *Making school improvement work: a conceptual guide to practice* (Leuven, ACCO).
- Vermeulen, C. J. A. J. (1988) The effectiveness of seventeen elementary schools in Rotterdam, *Western European Education*, 19(4), 5–21.
- Virgilio, I. (1987) *An examination of the relationships among school effectiveness in elementary and junior high schools*. Doctoral dissertation, University of New Orleans.
- Virgilio, I. & Teddlie, C. (1989) *Technical manual for the Virgilio teacher behavior inventory*. Unpublished ms., University of New Orleans.
- Walberg, H. J. & Haertel, G. D. (1992) Educational psychology's first century, *Journal of Educational Psychology*, 84(1), 6–19.
- Watkins, D. A. & Biggs, J. B. (1996) *The Chinese learner: cultural, psychological and contextual influences* (Hong Kong, CERC/ACER).

- Weber, G. (1971) *Inner city children can be taught to read: four successful schools* (Washington, DC, Council for Basic Education).
- Wellish, J. B., MacQueen, A. H., Carriere, R. A. & Duck, G. A. (1978) School management and organisation in successful schools, *Sociology of Education*, 51(3), 211–226.
- Wyne, M. D. & Stuck, G. B. (1982) Time and learning: implications for the classroom teacher, *Elementary School Journal*, 83(1), 67–75.

Appendix 1. Lesson observation form for evaluating the quality of teaching

(Please fill in the answer, or circle the correct answer)

| | | |
|---|------------------------------------|-----|
| School name/number: _____ | teacher: _____ | m/f |
| Group: _____ | age of teacher (yrs): _____ | |
| Activity/subject-matter: _____ | mixed age group: _____ | y/n |
| Name inspector: _____ | # of pupils in classroom: _____ | |
| % pupils from socially deprived families _____ | # of pupils in school: _____ | |
| % pupils for which the language of instruction is not their native language _____ | # of residents in community: _____ | |

| Indicator: The teacher ... | Rate ¹ | Good practice examples: <i>The teacher ...</i> | Observed ² |
|---|---|---|-----------------------|
| Efficient classroom management | 11 ... gives a well-structured lesson | 1 2 3 4 ... ensures clearly recognizable components in the lessons (lesson structure) | 0 1 |
| | 12 ... ensures the orderly progression of the lesson | 1 2 3 4 Entering and leaving the classroom takes place in an orderly manner | 0 1 |
| | | ... intervenes in a timely and appropriate way to any order disruptions | 0 1 |
| | | ... acts as a 'watchdog' for agreed codes of behaviour and rules | 0 1 |
| 13 ... uses learning time efficiently | 1 2 3 4 There is no loss of time at the start, during or at the end of the lesson | 0 1 | |
| | There are no 'dead' moments | 0 1 | |
| | The children are not left waiting | 0 1 | |
| 14 ... ensures efficient classroom management | 1 2 3 4 ... makes clear which lesson materials should be used | 0 1 | |
| | The lesson materials are ready to use | 0 1 | |

(continued)

Appendix 1. (Continued)

| Indicator: The teacher ... | | Rate ¹ | Good practice examples: <i>The teacher ...</i> | Observed ² |
|--|---|--|---|-----------------------|
| | | | <i>The lesson materials are adapted to the level and experience of the pupils</i> | 0 1 |
| Safe and stimulating learning climate | 21 ... ensures a relaxed atmosphere | 1 2 3 4 | ... addresses the children in a positive manner | 0 1 |
| | | | ... reacts with humour, and stimulates humour | 0 1 |
| | | | ... allows children to make mistakes | 0 1 |
| | 22 ... promotes mutual respect | 1 2 3 4 | ... encourages children to listen to one another | 0 1 |
| | | | ... intervenes when children are being laughed at | 0 1 |
| | | | ... takes (cultural) differences and idiosyncrasies into account | 0 1 |
| | 23 ... supports the self-confidence of pupils | 1 2 3 4 | ... feeds back on questions and answers from pupils in a positive way | 0 1 |
| | | | ... expresses positive expectations to pupils about what they are able to take on | 0 1 |
| | 24 ... shows respect for the pupils in behaviour and language use | 1 2 3 4 | ... allows pupils to finish speaking | 0 1 |
| | | | ... listens to what pupils have to say | 0 1 |
| | | | ... makes no role-confirming remarks | 0 1 |
| | 25 ... ensures cohesion | 1 2 3 4 | ... honours the contributions made by children | 0 1 |
| | | ... ensures solidarity between pupils | 0 1 | |
| | | ... ensures that events are experienced as group events | 0 1 | |
| 26 ... stimulates the independence of pupils | 1 2 3 4 | ... allows pupils to work independently on another assignment or to take up an individually selected task after completing an assignment | 0 1 | |

(continued)

Appendix 1. (Continued)

| Indicator: The teacher ... | | Rate ¹ | Good practice examples: <i>The teacher ...</i> | Observed ² |
|----------------------------|---|-------------------|--|-----------------------|
| | | | <i>... allows pupils to work with self-correcting materials</i> | 0 1 |
| | | | <i>... has pupils working on daily and weekly tasks</i> | 0 1 |
| | 27 ... promotes cooperation between pupils | 1 2 3 4 | <i>... provides opportunities for pupils to help one another</i> | 0 1 |
| | | | <i>... gives assignments that incite cooperation</i> | 0 1 |
| | | | <i>... gives pupils the opportunity to play together or to carry out assignments together</i> | 0 1 |
| Clear instruction | 31 ... clarifies the lesson objectives at the start of the lesson | 1 2 3 4 | <i>... informs pupils at the start of the lesson about the aims of the lesson</i> | 0 1 |
| | | | <i>... clarifies the aim of the assignment and what the pupils will learn from it</i> | 0 1 |
| | 32 ... evaluates whether the objectives have been achieved at the end of the lesson | 1 2 3 4 | <i>... verifies and/or evaluates whether the aims of the lesson have been achieved</i> | 0 1 |
| | | | <i>... checks the pupils' achievements</i> | 0 1 |
| | 33 ... gives clear instructions and explanations | 1 2 3 4 | <i>... activates the children's prior knowledge</i> | 0 1 |
| | | | <i>... explains in sequential stages</i> | 0 1 |
| | | | <i>... asks questions that are understood by the pupils</i> | 0 1 |
| | | | <i>... summarizes the lesson materials from time to time</i> | 0 1 |
| | 34 ... gives clear explanations of the learning materials and the assignments | 1 2 3 4 | <i>... ensures that every child knows what he/she has to do</i> | 0 1 |
| | | | <i>... clearly indicates the materials that can be used as learning aids</i> | 0 1 |
| | 35 ... gives feedback to pupils | 1 2 3 4 | <i>... checks whether pupils have understood the lesson materials when instructing the class</i> | 0 1 |
| | | | <i>... checks whether pupils are completing the assignments correctly</i> | 0 1 |

(continued)

Appendix 1. (Continued)

| Indicator: The teacher ... | Rate ¹ | Good practice examples: <i>The teacher ...</i> | Observed ² |
|---|-------------------|---|-----------------------|
| | | ... gives feedback on the way pupils arrive at their answers | 0 1 |
| | | ... gives feedback on the social functioning involved in the completion of the tasks (group work) | 0 1 |
| 36 ... involves all pupils in the lesson | 1 2 3 4 | ... gives assignments that stimulate pupils into active involvement | 0 1 |
| | | ... poses questions that initiate reflection | 0 1 |
| | | ... ensures that pupils listen carefully and keep on working | 0 1 |
| | | ... waits sufficiently long to allow children to reflect after posing a question | 0 1 |
| | | ... gives the opportunity to respond to pupils who don't put their hands up | 0 1 |
| 37 ... makes use of teaching methods that activate the pupils | 1 2 3 4 | ... makes use of conversational forms and discussion forms | 0 1 |
| | | ... provides graduated exercises | 0 1 |
| | | ... permits working in groups/corners | 0 1 |
| | | ... makes use of ICT | 0 1 |
| Adaptation of teaching 41 ... adapts the instruction to the relevant differences between pupils | 1 2 3 4 | ... allows pupils who need less instruction to commence with the work | 0 1 |
| | | ... gives extra instruction to small groups or individual pupils | 0 1 |
| | | ... does not direct himself exclusively to the middle bracket | 0 1 |
| 42 ... adapts the assignments and processing to the relevant differences between pupils | 1 2 3 4 | ... makes a distinction in the scope of the assignments between individual children | 0 1 |
| | | ... does not give all children the same time to complete the assignment | 0 1 |

(continued)

Appendix 1. (Continued)

| Indicator: The teacher ... | | Rate ¹ | Good practice examples: <i>The teacher ...</i> | Observed ² |
|------------------------------|----|-------------------|---|-----------------------|
| | | | <i>... allows some children to make use of auxiliary materials</i> | 0 1 |
| Teaching learning strategies | 51 | 1 2 3 4 | <i>... ensures that the teaching materials are orientated towards transfer</i> | 0 1 |
| | | | <i>... teaches pupils solution strategies or search and reference strategies</i> | 0 1 |
| | | | <i>... teaches pupils the use of organization resources</i> | 0 1 |
| | | | <i>... promotes the conscious use of what has been learned in other (different) areas of learning</i> | 0 1 |
| | 52 | 1 2 3 4 | <i>... stimulates the use of control activities</i> | 0 1 |
| | | | <i>... gives attention to estimatory calculation/ anticipatory reading</i> | 0 1 |
| | | | <i>... has solutions relate to the context</i> | 0 1 |
| | | | <i>... stimulates the use of alternative solutions</i> | 0 1 |
| | 53 | 1 2 3 4 | <i>... provides interactive instruction and activities</i> | 0 1 |
| | | | <i>... facilitates mutual interaction between pupils</i> | 0 1 |
| | | | <i>... ensures interaction between pupils and the teacher</i> | 0 1 |
| Involvement of pupils | 61 | 1 2 3 4 | <i>There is good individual involvement by the pupils</i> | 0 1 |
| | | | <i>Pupils actively listen to the instructions</i> | 0 1 |
| | | | <i>Pupils take part in learning/ group discussions</i> | 0 1 |
| | | | <i>Pupils work on the assignments in a concentrated, task-focused way</i> | 0 1 |
| Final judgement | 71 | 1 2 3 4 | <i>The overall quality of teaching is judged as:</i> | |

¹Please circle the correct answer: 1 = predominantly weak; 2 = more weaknesses than strengths 3 = more strengths than weaknesses; 4 = predominantly strong. Score only 3 when all good practice examples (if applicable) are really observed.

²Please circle (voluntary) the correct answer: 0 = no, I didn't observe this; 1 = yes, I have observed this.