

Observation in Teaching: Toward a Practice of Objectivity

JACQUELIEN BULTERMAN-BOS

Vrije Universiteit Amsterdam, The Netherlands

JAN TERWEL

Vrije Universiteit Amsterdam, The Netherlands

NICO VERLOOP

Leiden University, The Netherlands

WIM WARDEKKER

Vrije Universiteit Amsterdam, The Netherlands

Because most informal classroom assessment is based on the observation of students by teachers, the purpose of the present study is to construct a view of this assessment that is consistent with the practitioners' perspective. To find out how teachers frame observation, we analyzed stories of 25 Dutch teachers about teaching diverse learners and focused on what they said about the observation of students. Most attempts to improve observation have recommended the separation of subject and object in an effort to eliminate teachers' personal frameworks. The present study, however, shows that observation is embedded in the action of teaching. Thus, the realities of the classroom do not allow teachers to separate themselves from what they are observing. An alternative view of objectivity is presented. This transactional view shows that the quality of teachers' personal frameworks is important and should not be eliminated: The supposition of being neutral and detached hampers the teaching of diverse learners. To prevent self-fulfilling prophecies from influencing student achievement in a negative way, pedagogical craftsmanship is essential. Starting from this view, the reliability and the validity of observation for classroom assessment are reframed in a nonstatistical way. According to this view, social practice plays an important role in deciding whether observation deserves a quality warrant. Apart from an alternative view of concepts such as objectivity, a practice of objectivity requires new ways of constructing theories and mentoring teachers to sharpen their perceptive faculty.

Modern views of instruction stress the diversity of learners and call for customization (Reigeluth, 1999). Before customization can take place, teachers need to discover the differences among students. Apart from formal (classroom) assessment, teachers use informal assessment such as observation to get to know their students (Airasian, 1991; Salmon-Cox, 1980; Shavelson & Stern, 1981; Stiggins & Bridgeford 1985). Informal aspects of observation, however, have been suspect for years. Research on teacher expectations was said to show that teachers, by imposing low expectations on some groups of students, failed to provide the best educational experience for them. Therefore, teachers were taught to become objective in the traditional empiricist sense and were encouraged to eliminate prior frameworks and past experiences as much as possible. Our contribution—based on the analysis of 25 stories about teaching diverse learners—shows that teachers who are observing are also maintaining the flow of activities in the classroom. We therefore speak about observation in teaching. This implies that teachers influence what is being observed. The separation of subject and object, advocated by the traditional view of objectivity is not possible. In line with a more current view of objectivity, our study shows that a detached and neutral view paralyzes teacher craftsmanship and the teaching of diverse learners. We therefore developed an alternative path to improving the quality of teachers' observation of students. This approach acknowledges that teachers' frameworks cannot be eliminated, even if one tries to do so. It starts instead from the importance of developing teacher craftsmanship to improving the quality of observations. Starting from this view, the reliability and the validity of observation for classroom assessment are reframed in a nonstatistical way, and a way to develop a practice of objectivity is shown.

OVERVIEW OF LITERATURE ON INFORMAL ASSESSMENT THROUGH OBSERVATION

The present study focuses on the informal aspects of classroom assessment, namely, those aspects that are embedded in instructional events. Cizek (1999) distinguished informal aspects of classroom assessment from classroom assessment that is dissociated from instruction for the purpose of evaluation. In the older literature, the two aspects of assessment are denoted by summative and formative evaluation (Scriven, 1967; Bloom, Hastings, & Madaus, 1971). The information-gathering process we focused on takes place during the learning process and is necessary for teachers to understand their pupils, monitor their instruction, and establish a viable classroom culture. According to Airasian, much of this information gathering is done by means of informal observation of pupil behavior and performance (Airasian, 1991, 1994).

Various authors have complained about the research community's lack of attention to informal aspects of classroom assessment. Airasian (1991) found that the textbooks used in teacher education and measurement courses overwhelmingly concentrate on formal types of classroom measurement. These textbooks rely on instruments for producing written records of pupil performance, which are used to make judgments about pupils. In this respect, Airasian believes that most education books and courses do not reflect the daily activities of teachers in classrooms. He states that there is much more to educational measurement than formal number-producing techniques. "Teachers do not deal with problems and decisions in general or in the abstract; they deal with Marie, who is different from Paul, who is different from Jerome and who has a problem that needs attention right now" (p. 14). Airasian not only calls for attention to classroom assessment but also suggests that the informal aspects of classroom assessment, such as observations, be examined. Salmon-Cox (1980) also stressed the importance of observation. She concluded that teachers, when talking about how they assess their students, most frequently mention observation.

The types of observational activities examined in the present study start at the beginning of each year, when teachers are faced with the task of getting to know a new group of students with disparate interests, abilities, backgrounds, school experiences, and affects. The focus is on more student characteristics than achievement alone (Airasian, 1991; Shavelson & Stern, 1981). Teachers arrive at these judgments by instinct, intuition, and practical knowledge. Therefore, Airasian, referring to this process of information gathering, uses the expression "sizing up" students. He identifies seven characteristics of this process: It is done early in the school year; it is student centered; much of the information is gathered by means of informal observation of pupil behavior and performance; the impressions are rarely written down; there is a broad range of characteristics that teachers attend to; teachers are very confident about the accuracy of their initial perceptions and, as a consequence, these perceptions are difficult to change; and they remain stable throughout the school year.

Many authors have raised concerns about the quality of observation. Stiggins, Conklin, and Bridgeford (1986) indicate the complexity of the task. Terwel (1993) is aware that teachers who use observation might size up students incorrectly. Verloop and Zwarts (1987) and Verloop and Van der Schoot (1995), though acknowledging that informal assessments such as observations are an essential part of teaching, are also concerned about the quality of this method: Teachers are not sufficiently aware of how their observations are distorted by their initial impressions. Concerns about the quality of observations have been colored by the bulk of research on teacher expectations. For a long time, the research community believed that teachers' expectations of students were based not on honest observations but on

prejudices that teachers imposed on students. According to Merton (1957), a self-fulfilling prophecy occurs when “a false definition of the situation evokes a new behavior which makes the originally false conception come true” (p. 423). In *Pygmalion in the Classroom*, it was suggested that expectations for individual pupils influenced their actual achievement (Rosenthal & Jacobson, 1968). This notion was combined with ideas on the educational reproduction of social inequality (Hausser, 1980; Jungbluth, 1984; Rist, 1970). As Hall and Merkel (1985) put it, it was generally believed that low-expectancy children, typically thought to be children of minority groups or of low social status, were being harmed by teachers acting on their low expectations of these children. Much of the teacher expectation literature was designed to show that teachers used their own value systems to select both favored and unfavored students and thus were responsible for wide variation in achievement. It was even believed that student intelligence was affected by teacher expectations (Marburger, 1963; Rosenthal & Jacobson, 1968).

Such claims stemming from research on teacher expectations have been vigorously challenged. As Wineburg (1987) reminds us, authors such as Richard Snow (1969), Robert Thorndike (1968), and N. L. Gage (1966, 1971) questioned *Pygmalion* after publication, criticizing the setup of the study and demonstrating that it lacked empirical evidence. However, the *Pygmalion* claims were hailed by the popular media (such as *The New York Times*). Wineburg depicts how, in the course of time, reports of *Pygmalion* in the press came to stand for whatever people wanted, regardless of the original research questions asked. In the *Scientific American*, Rosenthal and Jacobson (1968) went as far as to question the wisdom of special programs to overcome the educational handicaps of disadvantaged children. They argued that such programs rested on the assumption that disadvantaged children had some problem or deficit that must be remedied; they maintained that such thinking was, at best, misguided.

Notwithstanding the harsh criticism from within the educational community, the notion of the self-fulfilling prophecy emerged as the common coin of educational research. According to Wineburg (1987), few ideas have influenced educational research and practice as much as this notion. Meyer (1985) estimated that since the original study, there have been between 300 and 400 published reports related to the self-fulfilling prophecy in education. Brophy and Good (1970) were among the first to study naturally occurring expectations. They concluded that many “teacher expectation effects” were best understood as student effects on teachers, rather than the other way round as many studies had suggested. Dusek, Hall, and Meyer (1985) attempted to consolidate and integrate theoretical and empirical research on teacher expectations. Like Brophy and Good, they concluded that many of the studies are correlational studies, whose weakness lies in

the uncertainty of causal inferences. If teacher expectations appear to be related to student outcomes, "it is inappropriate to conclude that the teacher behavior mediated the students' performance. Obviously, student behavior could well have mediated teacher behaviors" (Dusek et al., 1985; Meyer, 1985, p. 355). Evidence that did seem to support the existence of self-fulfilling prophecies was mostly derived from experimental studies. These studies, however, have only limited generalizability to natural classrooms (Brophy, 1985; Meyer, 1985; Mitman & Snow, 1985). The review by Dusek et al. (1985) supported the conclusions that Brophy (1983, p. 634) had drawn earlier, saying that "most differential teacher expectations are accurate and reality based, and most differential teacher interaction with students represents either appropriate, proactive response to differential student need, or at least understandable reactive response to differential student behavior." These conclusions are associated with naturalistic studies, dealing with real teachers and real information about real students. Yet, the potential for teachers' expectations to function as self-fulfilling prophecies always exists. Brophy (1985) assumes that 5% of the variance in achievement can be accounted for in terms of the self-fulfilling prophecy.

We have presented the research on teacher expectation and reviews of this research at greater length because these conclusions have not been widely accepted in all branches of education. These conclusions show that the prejudice with which the whole professional group of teachers has been associated is inappropriate. The reviewers of the teacher expectation research did not conclude that teacher observations were perfect but showed that the possible effects of teacher expectation have been over-rated. The question how observation of students can improve is posed in the knowledge that improving the quality of observation is important, but should not be considered as the key to educational equality in and of itself.

How can we frame the characteristics of quality observations? This is an important question for those who are engaged in teacher education. Authors such as Stiggins, Conklin, and Bridgeford (1986) and Airasian (1991) stress the importance of classroom realities. They emphasize that to really assist teachers, researchers must understand that the demands of the classroom are different from what is relevant in research. Along with this realization, Stiggins et al. still believes that teachers need help from researchers. The kind of help needed remains unspecified. Airasian also believes that better teacher education is important for improving the quality of observations: Because teachers are often unaware of the extent to which they rely on informal, unsystematic assessment for decision making, more information on sizing up students and on instructional assessment should be included in courses and texts. The need for valid and reliable information should be stressed in these courses, but in a commonsense, nonstatistical manner. Stiggins et al. and Airasian realize that in practical situations, a traditional

scientific view of validity and reliability does not suffice. They therefore compare favorably with authors who recommend that teachers apply instruments similar to those researchers use (see De Corte, 1981, and Wajnrub, 1992). The question of what a commonsense view of validity and reliability looks like has recently attracted the attention of several researchers. Moss (1992, 1994), for instance, presents such a view in the context of formal, high-stakes assessment. The present study, which is a combination of empirical research and theoretical research, focuses on an alternative definition of validity and reliability concerning informal assessments, which mainly take place through observation.

This study starts from the way in which teachers frame observation. Next, it uses this view to define the concepts of validity and reliability in a nonstatistical way and elaborates on the role that researchers can play in improving observations. In this way, we hope to contribute to a practice of objectivity.

OBSERVATION OF STUDENTS FROM A TEACHER'S PERSPECTIVE

METHOD OF INQUIRY: COLLECTING AND ANALYZING NARRATIVES

During the past 20 years, the character of research on teacher education has changed. In contrast to the past, much more attention is now being paid to both the process of teaching as it occurs in natural situations and to the cultural, historical, social, and institutional contexts in which teachers operate. Much of the behaviorism-driven research of the 1960s and 1970s concentrated on various discrete, decontextualized teaching behaviors: Consequently, the results were poorly anchored in the practical, day-to-day problems of teachers and were easily washed out when preservice teachers entered practice or in-service teachers returned to the classroom (Brown & McIntyre, 1993; Zeichner, 1999; Zeichner & Tabachnick, 1981). Moreover, the positivistic framework prevalent in those days—also characterized as the model of technical rationality (Schön, 1983) or the top-down approach (Hargreaves & Fullan, 1992)—emphasized the failings of teachers and ignored their special skills and personal qualities (Elbaz, 1983, 1997). Nowadays, many researchers are trying to understand what teachers know and how they learn. The value of teachers' practical knowledge is acknowledged, and many researchers realize that teachers' interpretive frames for understanding and improving their own classroom practice should be addressed. (Cochran-Smith, 1990; Elbaz, 1983; Feiman-Nemser & Floden, 1986; Fenstermacher, 1994; McDonald, 1988; Putnam & Borko, 2000; Shulman, 1986; Verloop, 1992; Wardekker, 1989).

When we started the present research project, we found ourselves in the middle of this shift. For years, behavioristic literature had been imposed on

teachers, prescribing individual learning trajectories for students, which were supposed to be based on the ongoing measurement of both entry situation and subsequent performance. Teachers, however, did not implement such prescriptions. Their practice was characterized as traditional (Alberts, 1987, 1991). The negative characterization of teachers' practice (in general as well as concerning assessment specifically), along with the suggestion that teachers contributed to social inequality by not imposing the right expectations on diverse learners, had led to a climate in which teaching diverse learners had become a touchy subject, even among preservice teachers. As teacher educators, we felt we could not remain aloof from this situation. However, we also believed that whatever ideology one attends to, the subject of teaching diverse learners cannot be ignored by teacher educators: Diversity of learners is everywhere, even in so-called homogeneous (or tracked) classes. Moreover, we assumed that teachers, especially those with mixed classes, had somehow developed their own approach to dealing with diversity. Therefore, we realized we needed to cooperate with teachers to develop our course on teaching mixed-ability classes. In the Netherlands, such classes commonly occur in the first years of secondary education. To ascertain how teachers understand this work, we collected their stories (Carter, 1995; Connelly & Clandinin, 1990; Doyle, 1997) on teaching diverse learners.

Noddings (1986) emphasized the importance of collegiality in research on teaching. In narrative inquiry, a sense of equality between participants is particularly important (Hogan, 1988). However, in research-practitioner relationships, teachers have long been silenced through being used as the object of study. Therefore, we realized that giving teachers a voice is an active process to which the setup of our study had to contribute. We considered role taking as an essential criterion for the objectivity of qualitative research, both as a process of self-insertion in the other's story, as a way of coming to know the other's story as well as giving the other a voice (Connelly & Clandinin, 1990; Elbow, 1986; Smaling, 1990). We chose an interviewer who could relate to the teachers in various ways and who had been trained extensively in interview techniques developed in client-centered therapy and in behavior therapy. Moreover, she was a teacher trainer with recent teaching experience. She was familiar with the literature on teaching diverse learners, as well as with the doubts among many teachers about formal decontextualized knowledge on this subject. In her own teaching practice, she experienced the teaching of diverse learners as problematic. Compared to the average age of the teachers interviewed, she was relatively young. In this way, she represented the beginning teachers who were supposed to learn from the practical knowledge of the more experienced teachers. Thus, the interviewed teachers were pictured as more experienced and were also addressed as such in our letter, in which we invited the

teachers to tell their stories about teaching diverse learners to help us develop a course on this subject.

To offer the teacher the maximum opportunity to approach our subject from their own angle, we chose an interview format that was as open as possible. However, we felt we needed to define beforehand the aspects of teaching that needed to be addressed. For this, we used a very commonly used planning model, consisting of entry situation, learning goals/content, evaluation, learning processes, lesson format, and media—the didactical analysis or DA model (De Corte, Geerligs, Lagerweij, Peters, & Vandenberghe, 1981). We did not introduce the DA model as such to the teachers, realizing that such models are not framed in the language used by practitioners (Clark & Yinger, 1978; Peterson, Marx, & Clark, 1978; Shulman, 1980; Yinger, 1977; Zahorik, 1975). Instead, we introduced its components in a commonsense way. In this way, the interviewer did not introduce *entry situation*, but she made sure that the teachers, telling their story about teaching classes containing diverse learners, addressed the way they discovered individual differences. If the conversation (Florio-Ruane, 1991) did not turn to this subject spontaneously, she explicitly asked how the teacher discovered individual differences. For the rest of the time, she mainly listened to the teachers' stories. Moreover, she invited the teachers to give concrete form to their experience and asked them to comment on the experiences of other teachers.

We choose our schools randomly from a list of the schools to which our students were sent for in-service training. Most studies of teaching diverse learners have concentrated on the teaching of Dutch and mathematics; therefore, we decided to choose different subjects and asked the schools to contact us with teachers of English and biology. Both subjects are considered to be important, although English is taught more hours a week. This is why schools employ more teachers of English than of biology. We did not set any teacher-quality constraints. Apart from the fact that these qualities are both hard to define and hard to measure, we believe that research should represent not only teachers' successes but also their problems. Because, however, we did not want our information to be overshadowed by the typical problems of beginning teachers, we only interviewed those who had more than 4 years experience. Our only other restriction was that the teachers taught classes in the first years in which student ability was mixed to a certain extent. In the end, 25 teachers participated in our project, 9 teachers of biology and 16 of English. Schools that attracted many rural students, schools in medium-sized cities, as well as schools with inner-city students were represented in our sample.

Before meeting the teachers personally, we sent each a letter in which we expressed our view of the value of practical knowledge. We explained that since this knowledge tends to be tacit (Feiman-Nemser & Floden, 1986;

Schon, 1983), it would be valuable if they used a notebook to record some events and reflect on them before the interviews took place. We initially asked each teacher to plan 1 hour for each interview. In many cases, the teachers themselves took more time to tell their story. The interview was ended once all the topics had been addressed and both parties had the feeling that everything had been said that needed to be said. Along with the invitation letter, we sent each teacher a short questionnaire, asking for some formal facts, such as age, the number of students and number of classes they had, and the kind of follow-up training they had followed after becoming a qualified teacher.

During the interviews, some of the teachers—being wary of researchers—appeared to be hesitant to express themselves along lines of thought other than those developed in the literature. Then the interviewer said something like, “That’s what they all say, but does it also comply with your own experience?” Gradually, these teachers learned that none of their opinions or experiences were considered to be incorrect and went on to express themselves more freely. All conversations were recorded on audiotape and transcribed.

The conversations yielded beautiful stories and very useful material for our course on teaching diverse learners. They offered students, teachers, and teacher trainers the possibility to relate their personal experience to that of the interviewed teacher. Interestingly enough, among the body of stories, each of us—depending on our own personality or teaching style—had his or her favorite teacher who appealed most to us. Thus, the criterion that a narrative, if it is a good narrative, constitutes an invitation to participate, and can be read and lived vicariously by others seemed to be fulfilled (Conelly, 1978; Connelly & Clandinin, 1990; Guba & Lincoln, 1989). Our colleagues who only participated indirectly in our research commented positively on the warm atmosphere and the collegial tone of the conversations. Many of the teachers interviewed came across as being very committed to teaching diverse learners and seemed very capable of managing (completely) heterogeneous classes. After the interviews, some of the teachers reacted by saying “they were not aware they knew that much.” This showed that the conversation had raised their consciousness and had helped them to make their tacit knowledge more explicit. A minority of those interviewed portrayed themselves as bluntly cynical. Nevertheless, even their stories yielded a genuine picture of a true cynic, which made their position more understandable for others.

Initially, we only used the spoken version of the interviews as discussion material during the course on teaching diverse learners. Gradually, however, as we got more familiar with the conversations, we surmised that patterns could be distinguished that would summarize the many hours of experience into more manageable material. Summarizing stories by formu-

lating patterns, however, has a drawback: It depersonalizes the material. It removes the individual idiosyncrasies that make a story appealing. Such a story creates a unique opportunity for participation for *some* individuals. On the other hand, the formulation of patterns within a whole body of conversations increases its generative power (Wardekker, 2000) or transferability (Guba & Lincoln, 1989) for a larger group of people. (Both concepts are introduced to avoid the concept of generalizability.)

Analyzing stories is a complicated process. Because stories are multi-dimensional, one cannot simply place segments into a priori-formulated categories. (In our case, these a priori-formulated categories consisted of the component of the planning model out of which, in this study, we focused on entry situation.) Many segments refer to different topics at the same time. The categories are often interrelated; the motivation for what is expressed in a particular category often lies in what is expressed in other categories. Moreover, within the same categories, teachers often address the same topics from different angles. We intended to do justice to as many relations as possible, which is why we could not analyze the conversations automatically. Instead, we had to rely on a human being, who by listening repeatedly to the stories, had become very familiar with them and was aware of the context in which the segments belonged. As a memory aid, an "outline map" of each interview was created. In this map, all expressions concerning the main categories were summarized. (Of course, various expressions appeared in more than one category.) The summarized expressions were indexed with the page number of the transcript conversation for quick reference to the original fragment. This made detailed comparison possible without losing the context of each fragment. The outline map, moreover, gave us a quick overview of the interview and enabled us to see segments in one category in the contexts of other categories. Making outline maps by using both the spoken and written version is very intense work and again offered the opportunity to become familiar with the original conversations, which is necessary to do justice to the material.

By using the outline map, we discovered that the two most frequently mentioned ways in which teachers identified individual differences were (1) observations (which we categorized under entry situation) and (2) all student activities and work graded by teachers, such as teacher-made tests and text-embedded tests (which we categorized under evaluation of students). In this article, we only focus on informal assessment for sizing up students. However, the outline map showed that formal assessment also plays a role in discovering differences. In talking about observation, we thus should keep in mind that teachers also assess students in a more formal and systematic way.

For the present study, we primarily collected and analyzed all the teachers' remarks about observation of students. Formulating patterns can be

compared to focusing a camera lens: One should not get too close to the subject (otherwise, only a few details are captured), nor too far (otherwise, detail is lost; the summary becomes meaningless). Formulating a pattern that captures a large group is an inductive process in which the material itself, as well as the inventiveness of the researcher, plays a role. After having formulated our patterns, we checked whether they applied to all the teachers. The first three patterns were ranked on a dimension of increasing intensity. It seemed likely that teachers who had expressed themselves in the third pattern also had experience in the preceding two categories, although they had not always been explicit about it. It also seemed obvious that the fourth pattern also applied to everyone, although not all the teachers had been explicit about it.

RESULTS

Before we present the patterns that show a practitioners' perspective of observation, we will supply some information about general classroom practice in the Netherlands. The classes of our teachers have an average number of 25 students, with a variation from 18 to 31 students. English and biology are taught 3 and 2 hours, respectively, a week, which means that teachers with a full-time biology job of 28 hours generally meet with 14 classes. Most lessons consist of both a centralized part (in which the teacher leads a discussion, gives instruction), as well as a decentralized part, in which students work in small groups. Although both teaching styles generally occur in most lessons, some teachers rely on the decentralized style, while others prefer a centralized style. The teaching style also depends on the subject and the material available.

Following the stories of teachers, we developed four patterns of observations, of which the first three consist of two dimensions. These dimensions refer to (1) the mode of teacher observation (active/passive) and (2) the student stimulus for that observation (active/passive). Observing actively refers to a mode in which teachers are aware they are observing students: It happens intentionally and consciously and is more or less planned. If teachers observe passively, they are primarily focused on other activities; nonetheless, they pick up information on the students. When the students contribute actively to what the teacher observes, they intentionally try to draw the teacher's attention. If they contribute passively, they are not intentionally drawing the teacher's attention to something special, but they are nonetheless conveying information. Three patterns were distinguished that describe the process of short-term observation: (1) triggered observation (the teacher is passive but is drawn to something by the student), (2) incidental observation (both teacher and student are passive), and (3) intentional observation (the teacher is active; the student is passive or active). A

fourth pattern, long-term observation, covers those fragments that showed that prior experience with other students plays a role in identifying differences between current students.

The four patterns of observation derived from our surveys are defined and explained in the following sections. They are presented and illustrated with quotes from the original conversations, using fictitious names. To reproduce all of the (multidimensional) fragments obtained in the interviews would require a great deal of space and would not enhance this article. Therefore, we have included only a limited set of examples.

TRIGGERED OBSERVATION

In this pattern, a student or a group of students actively demands the teacher's attention. Given a choice, the teacher would have disregarded the student (or a group), but under triggered observation the teacher is compelled to take notice. The teacher's mode is passive—he or she is engaged in other things. However, the student or group of students draw(s) her or his attention:

Mr. Smit:

A few bright students can usually start their assignments after a single standard instruction, while the rest require a second explanation—but even after this second round, some still can't follow the material and say (to my irritation): "I don't understand this." As a result, I have to work like an octopus: some of the group is finished, while others have yet to start! This makes correcting assignments difficult and requires a great deal of energy.

The disturbance of the teacher's mindset may evoke (negative) emotions. Mr. Smit uses the word "irritation." Another teacher, Mr. Singel, when referring to a similar situation, uses the word "tiring." Mr. Spiegel refers to a learning process in this respect: He is still learning that some of his students (who would have been placed in the lowest ability track if a moderate mixed-ability policy had not been present) have lower levels of ability than he had come to expect. Here also, external signals (e.g., the student's attitude) have reminded him of the need to pay attention to things that otherwise would have been disregarded. By doing this, he shows that he has revised his perspective of a group of students.

INCIDENTAL OBSERVATION

In this pattern, information about student characteristics comes naturally from daily classroom interaction. As is the case with triggered observation, the information is gathered while other activities are going on, such as

correcting homework, engaging in instruction, and doing assignments in class. Unlike the first pattern, this type of observation is not provoked by students. The teacher is not actively engaged in gathering information either. Nevertheless, information is exchanged. This type of observation is unavoidable; even teachers who do not dwell on differences come across this pattern.

Mr. De Hond:

The extremes are the easiest to place: someone who stands out because he is very good, or someone who is obviously very weak. You know right away that someone is bright if he always gives the right answers in class. You give someone a passage to read, for example, and it just doesn't work. There is someone like this in the group I just had, a real smart cookie. He simply radiates this in everything he does. I don't even have to look at his grades, I just have to listen to his answers. If I ask a question, he immediately raises his hand and gives the correct answer. In this way, you build an impression. You run into this automatically.

The word “automatically” is a key word in this pattern. During regular activities, teachers unintentionally observe a great deal. Probably, the information is processed unconsciously. Even teachers who have not planned to observe their students come across the way they work and who they are. It does, however, depend on the teaching style to what extent information about students automatically comes to teachers. Mr. Bloem, who often works with individual students, illustrates this.

Mr. Bloem:

During whole class instruction—which happens regularly—you work at the front of the class. Then you have the teacher on one side of the room and a block of students on the other. In these cases, you work with a group, not with individuals. In a group, some of the students will make themselves noticeable by making comments or asking questions, but the bulk of the group remains a blur. These students are present and—you hope—they are paying attention, but that is all the contact you have. When you work with students individually, you interact with each person one-on-one. This lets you get to know them better.

Moreover, the content being addressed can be more or less immediate to observe. If students are learning how to use certain concepts (which is often the case in biology), the extent to which they do this correctly does not appear as easily from their engagement in the subject as it does when students have to learn how to pronounce or how to spell a word (which is often the case in English). Thus, the immediacy of the content being taught as well as the way in which the subject is taught is a variable in the amount

of information that comes naturally from daily interaction between teacher and student. This accounts for the fact that teachers of English more often say they automatically obtain useful information on how students are acquiring the subjects than teachers of biology do.

Various teachers indicated that they do not remember all the information they observe. Thus, if this information is retained anyway, it is probably retained in the subconscious.

INTENTIONAL OBSERVATION

In this pattern, the teacher is actively engaged in information gathering. Thus, he or she is focused on informal assessment. This either takes place during a moment specifically planned for informal assessment or during instruction or exercise. Two teachers described how they systematically give turns.

Ms. Vogel:

I give everybody a turn in each lesson. When we go over their homework, everyone gives an answer, so you get a good picture of who knows the material and who doesn't.

Mr. Visser:

I want to hear from everyone at least once a week, if possible, either as a part of reading aloud or via direct questioning.

In this way, these teachers make sure they do not skip individual students, and they pay attention to every individual at least one time per lesson or week. However, this does not mean they actually observe systematically because students do not take turns for the same tasks.

Ms. Koning described how she checks whether her instruction has come across as intended.

Ms. Koning:

You sometimes have students who simply don't understand anything and don't ask any questions. You can find them by asking questions during class. After I've explained something, I always try to get some feedback from the group. For example, I'll ask: "What is the function of the possessive pronoun?" Then I try to get the answer. For the non-native kids in class, Dutch is a foreign language. Within this group especially, you first have to see if they know what a possessive pronoun is. Also, when they study vocabulary words, they may encounter Dutch words that they don't run across in their daily environment. In these cases, I'd say: "The Dutch word "overwerk" is "overtime" in English. Can you tell me what "overtime" is, and can you use it in a sentence? When

do you have to work overtime? Does your father ever have to work overtime?" It is only after they understand the concept that they can use the word effectively in English.

Various teachers gather information by discussing students' notebooks while other students work independently. Ms. Wolf illustrates this by showing how she tries to find out what a student has missed, referring to what she observes in their notebook.

Ms. Wolf:

An experiment, for instance, has the title: determining the threshold value. There are students who can precisely complete the steps of the experiment, but run into trouble with the conclusion. They've tasted various concentrations of water and sugar. The conclusion? What conclusion? The mixture was either tasty or terrible! This kind of student gives "yes" as a conclusion. As a teacher, you then have to see what's gone wrong. "Try to explain what the threshold value is!," but the student has no idea. You then grab the textbook and say: "Do you understand what this says?" We go through the examples. Then I ask them to come up with other examples. I try to relate that back to the experiment, at which point a light occasionally goes on.

Some teachers praised the teaching style in which students work individually or in small groups because it allows the teacher, who walks around, to get to know their students better. To observe intentionally, however, does not guarantee that the information obtained makes sense to the teacher. Mrs. Akkermans said that she often reflects on her observations later on while she washes the dishes at home.

I always wonder what it is that goes on in the head of such a child. Sometimes, at home while washing the dishes, I think: darned, what is it with Marco? Why can't he get it?

Teachers rarely mention keeping record of their observations. Mrs. Tulp is an exception.

Mrs. Tulp:

Along with the usual grades, I also make a note of how the homework is done, if they do it regularly. I note whether someone does sloppy work, or is insecure, or whether someone is deaf or can't see well. For example, there are kids who don't wear their glasses and sit in the back of class. I know that they can't see what I've put up on the overhead projector. If I didn't jot a quick note on this, I'd forget it.

Along with grades, which are written down, most teachers make mental notes about their observations of individual students. Mrs. Akkermans said

she believes she remembers her observations. She adds to this, however, that she would not be able to do this if she were not a part-timer. We analyzed the stories on whether the teachers believed they knew their students well and concluded that part-timers generally were more positive on this than full-timers (who often commented they believed they did not know their students well enough).

LONG-TERM OBSERVATIONS

During their careers, teachers observe how students behave and how they react to the material that is presented to them. This information, gathered in the past, is significant for assessing differences in the current situation. This occurs on a small scale when teachers work with the same textbook over a period of time. Such teachers have a historical reference on how students from similar groups reacted to the textbook, which enables teachers to anticipate the behavior of their current students.

Mr. Langen:

When you hand out assignments, you know ahead of time that certain kids will have problems. In order to know this, you have to use the same textbook several years in a row, so that you know where the problems are. This is the third or fourth year that I've used this book, so I've gotten pretty familiar with it. When I pass out an assignment, I know beforehand which kids will have trouble, so I make sure that I walk past them as soon as possible. Doing so lets you keep an eye on them.

Also on a larger scale, images from the past alert teachers to situations in the present. Discovering differences is a matter of "experience," according to Ms. Bouw. Mr. Messen asserts that one develops a "sense," unlike student teachers, who still have to develop observational skills.

Mr. Messen:

The group that I'm mentor for—those are great kids. I know them very well: I got to know them in their first year when I was their study coordinator for about five hours per week. As a result, you know all their comings and goings, and interacting with them is not a problem. I also have a second group where things are not so simple, since this our first year. So you have to play it by ear. But you develop an extra sense. I regularly supervise student teachers, and it is clear that at the beginning of their careers, they have no feeling for this sort of thing. Once you have done it a number of years, though . . . it seems that you just build up a sixth sense for that sort of thing.

This quote reflects Schön's theory (1983) describing that an experienced practitioner builds up a repertoire of examples, images, understandings,

and actions which help him or her give meaning to new situations. Often, images of student behavior in the past do not match current situations precisely, but teacher reflection helps adjust the image to make sense of them.

CONCLUSIONS

Our work allowed us to identify seven characteristics of the process of observation as an aspect of classroom assessment:

1. *Teachers differ in the extent to which they are actively engaged in observation.* Some teachers mainly observe passively; others need students who draw their attention.
2. *The process is interpersonal.* Rather than viewing observation as a means of measurement in which an object (a student characteristic such as achievement or motivation) is measured by means of a standardized instrument, two subjects are involved: Teachers have to get to know their students. This process of observation helps build relationships. As Stiggins (1991, p. 9) puts it, "In classrooms assessment is virtually never a detached, scientific, objective laboratory act. Rather, it is virtually always an interpersonal act with personal antecedents and personal consequences."
3. *Teachers who are observing are playing two roles at the same time.* Because observation is embedded in the action of teaching, it is inevitable that observation also happens passively or unconsciously. But even if observation is intentional, it is the observer who at the same time initiates and controls the flow of activities (Shulman, 1980) to be observed. If no activities are taking place, no information about the learning process can be observed. Teachers thus influence their observations by the way in which they control the activities to be observed. Therefore, it is more accurate to talk about observation in teaching rather than observation.
4. *Observation in teaching is a process of mutual influence.* Students also contribute actively to what can be observed. Under triggered observation, they even compel teachers to take notice. By interacting with students, teachers find out whether their frames of reference are adequate. Adaptation of frames is part of a negotiation process with students, which is required to avoid miscommunication. This process evokes a range of emotions, for instance, stress and irritation or bonding.
5. *How results should be attributed is uncertain.* Because observation takes place under the influence of both parties, it is not clear who should

be held responsible for what is happening. What teachers observe in their classrooms partly mirrors their own influence. For instance, if a student is not motivated, teachers can never be sure that they have motivated their students well enough; if students do not achieve enough, teachers can never be sure they have instructed their students well enough. The process of observation in teaching thus reflects the intrinsic uncertainties of the teaching craft (cf. McDonald, 1992).

6. *Information obtained by observation is never complete.* Though a lot of raw data is collected both actively and passively, and may seem to form a coherent picture, new fragments may continually change the existing picture. New bits and pieces of information are continually captured, and old bits are lost (forgotten) or replaced. This process continues as long as the teacher–student interaction lasts. Reflection is needed to put the fragments of information together into a more or less meaningful picture of an individual student. The mental room available for this process also depends on the number of students of a teacher has.

7. *Images of former students play a role in discovering differences between current students.* They enable teachers to recognize certain “types” of students more effectively. Because images from the past will not match precisely those of the present, reflection is needed.

TOWARD QUALITY OBSERVATIONS

The seven points discussed previously tell us about the process of information gathering; however, they do not address the quality of teachers’ observations. The issue of how to frame the quality of observation will be explored in the following sections. Does the interpersonal character of observation imply a merit or a flaw? Should the interpersonal character of observation be encouraged, or will biased information be the result? The literature generally relates teachers’ prior assumptions to distorted observations. But could prior experience also play a positive role?

Using the characteristics of observation in practice, we investigated whether this practice can survive the scrutiny of fair comment. Following the line of reasoning developed by Newell (1986) and Eisner (1992), we developed a transactional view of objectivity that is used to reframe standards such as validity and reliability. According to this view, social practice plays a decisive role in distinguishing between sound and shaky beliefs. This implies a new way of constructing valid and reliable theories, which are the lens that is used for observation. We will explain that a practice of objectivity also requires a new way of mentoring teachers to sharpen their perceptive faculty.

TWO VIEWS OF OBJECTIVITY

As our review of the literature showed, little attention is generally paid to informal observation, and the attention paid to practical instructions on this is even more scarce. Good and Brophy (1978), in their book *Looking in Classrooms*, however, gave practical instructions about how teachers should conduct observational activities. They explain that what “we think we see is not congruent with reality. . . . Teachers on occasion react not to what they physically hear but to their *interpretation* of what the student said. Their past experiences with a student often influence their interpretation of what the student seems to be saying. This is not to suggest that teachers should not interpret student comments, but to argue that they should be *aware* when they do so. . . . Expectation should be appropriate. Teachers shouldn’t react to a label (low achiever, low potential, slow learner) but to the student as he is” (p. 88).

Two tendencies, although not formally defined or even labeled as a view of objectivity, can be detected in the work referred to previously. One tendency stresses the importance of eliminating prior assumptions, past experiences, and personal convictions. Observed behavior is regarded as being different per se from interpretations of observed behavior, which are considered as a source of error. To achieve objectivity, formal instruments, such as observation forms, are used. But a different perspective on objectivity can also be discerned: Instruments are not only used to assess behavior, but they are also presented to enhance awareness about prevailing interpretations. The emphasis is no longer on separating values from “true” observations; the discussion is now focused on the desirability of the values being used. Good and Brophy implicitly concede that perceptual frameworks play a role in observation and cannot be eliminated.

In the following sections, these two tendencies play a crucial role in our discussion of how quality observations should be framed.

THE CONCEPT OF OBJECTIVITY

The first interpretation of objectivity in which the role of personal judgments is reduced is central to the concept of objectivity as it is most commonly used in the literature. Therefore, we refer to it as the traditional view. Because beliefs about an objective world must hold independent of the observer, in this interpretation, judgments are dissociated from any connection with the opinions and experiences of persons. The aim of research is knowledge (episteme), not beliefs (doxa). According to this view, subjectivity is in the mind; what is objective is outside the mind. Scientific methods are used to minimize personal judgment. They consist of procedures that need to be applied to attain an understanding of events and objects.

The assumption that this approach leads to pure understanding of events and objects has been criticized. It is argued that human observation is always framework dependent (Eisner, 1992; Newell, 1986). Perception of the world is perception influenced by skill, point of view, focus, language, and framework. This does not imply that the creation of procedures that eliminate personal judgment is impossible. One of the most common examples of a method that excludes personal judgment is an achievement test: once the multiple-choice test has been constructed, it can be scored objectively, for instance, by an optical scanner. Yet consensus achieved through procedural objectivity does not demonstrate that a pure view of the outer world has been attained; it actually demonstrates that people can agree: they agree on the way in which the procedures should be designed and applied, which reveals their common frameworks. In essence, such procedures do not really eliminate personal judgment but draw upon the assumption that everybody judges in the same way (Eisner, 1992; Newell, 1986).

As we have shown in the previous section, methods based on procedural objectivity have also been developed for classroom observation. They do not always draw on common frameworks, although some do. An example of a common framework used in classroom observation is counting. If there are indications that a certain teacher favors boys over girls, one can count how many turns the teacher gives to boys and how many to girls. Owing to a common framework—we all know how to count; we also agree about what boys and girls are—this observational procedure works well. If more complex events are to be observed, however, common frameworks could well be absent. In such situations, judgment is required. But because the investigators intend to eliminate personal judgment, they develop procedures defining how judgment has to take place. Such procedures prescribe in what categories observers should classify their findings. These procedures thus favor a particular way of looking at the phenomenon, while overlooking other aspects. If no common frameworks exist, which is often the case in complex situations, there is a dilemma: either complexity is acknowledged and personal judgment and disagreement are accepted, or complexity is not acknowledged and a strict way of looking at the phenomenon is prescribed, forcing everybody to judge in the same way. In neither of these cases, however, are we able to know the world in its, as Eisner calls it, “pristine state,” a state in which our frameworks are eliminated. Procedures themselves are a way of framing the world.

The bottom line is not that procedures in general are odious or suspect; it is that the presupposition that we can eliminate human judgment is naive. Observation is loaded with human theories. What we can do to go beyond the idiosyncrasies of different views is try to use reason to judge these theories. It is from this assumption, developed by Newell (1986), that an alternative view of objectivity starts, which we refer to as the transac-

tional view. In this view, knowledge that is independent of human frameworks is regarded as an unattainable commodity. Even within these human constraints, Newell still considers the concept of objectivity useful. He contrasts objective judgments with prejudiced, biased, or dogmatic judgments. Objectivity goes together with respect for certain norms, including standards of evidence and argument-related ways of resolving disputes, settling issues, and deciding beliefs (p. 18). The distinctive characteristic of this view is that objectivity attaches to persons through their actions. Thus, what makes a judgment objective is not particular to outer objects, but is some particular practice of people (p. 17). The question "How can one be objective in one's own view?" poses no special problems beyond the question "How can one act with reasonableness and impartiality in one's own view?" (p. 32). Objectivity is associated with impartiality, detachment, disinterestedness, and a willingness to submit to standards of evidence. The transactional view tries to identify the human actions that ensure objectiveness (p. 23). Objectivity becomes a product of a proper method. Nevertheless, Newell continues, linking objectivity with the practice of objective methodology has its dangers: the methodology traditionally prescribed is geared to preventing the intrusion of a subject-related bias with an "overreaching zeal." Newell explains why this assumption is, at best, a liability: Removing the interpretive categories of observers is removing their capacity to classify and describe what they observe. Observational checks are risked, not rescued, by neutrality, and attempts to eliminate the contribution of the observer without eliminating the observation are self-defeating. If the action of observing is refined to a point where it ceases to be discernible as the product of individual perceptiveness, it becomes detached from any actively alert agent; it ceases, in short, to be observation (p. 28). Thus, objectivity is not to be contrasted with that trivializing sense of subjectivity in which judgments are "subjective" simply because they are judgments or expressions of the point of view of some individual agent. What matters for objectivity is not whether a person's opinions steer his judgments but whether the opinions embodied in her or his judgments can survive the scrutiny of fair comment (p. 31).

Thus, in this view, objectivity is something a person can learn: for example, by trying to free oneself from the bias of one's beliefs. What bias is is agreed upon in the social practice in which the person operates. In a loose sense, Newell remarks, the objective person is the rational person, but the sense is loose because rationality's scope is indeterminate in our ways of speaking and needs the qualification that the rational person respects the reasonable criticism and the reasonable demands of others. He or she may or may not be the self-interested person. The rational pursuit of one's interest can clash with objectiveness by disregarding unwelcome evidence and justified opposition; equally, objectivity is not surrendered by holding on to one's interests on the strength of a reasonable case. Strong self-

regard threatens objectivity—not through self-interest but by overriding claims to an impartial weighing-up (p. 36). Objectivity is therefore linked with personal action and responsibility. The transactional view sees it as normative, as a virtue.

This brings objectivity back to a human scale. Who decides what is fair comment and reasonable criticism? No one other than human beings, operating in a community and sharing frameworks. They will judge opinions as good, not by a correspondence with reality (we cannot determine this) but because certain theories that are part of our frameworks make sense and are supported by reason; as Eisner (1990), following Toulmin, puts it “because they are sound doxa instead of shaky ones.”

This transactional view on objectivity will be used in the following sections in which quality observations are reframed.

CONCLUSIONS AND DISCUSSION

REFRAMING QUALITY OBSERVATIONS

A transactional view shows that quality observations cannot be discussed in terms of correspondence with reality. Eliminating personal judgment yields a different picture, though not necessarily a better one. A way of reacting to a student “as he or she is,” as Good and Brophy suggested, simply does not exist. People will never be able to go beyond their own view of somebody else; at the same time, however, a transactional approach shows that all possible views of somebody else are not equally valuable.

When viewing observation in teaching from a transactional perspective, some characteristics of quality observation can be inferred:

1. We have shown that the observations of students are interpersonal. According to a transactional view, this is not a threat to objectivity in and of itself. It asks whether the observer is disposed to standing back and considering the state of his or her feelings and interpretations.
2. We indicated that teachers negotiate meaning while interacting with students who understand the world differently. As human beings are able to reflect on themselves, interaction with people from other (sub)cultures may result in mutual understanding. Thus, common frameworks may develop (Procee, 1991). The transactional view on objectivity stresses that actively alert agents, open to discovery, are needed. It addresses the issue of observation by asking whether teachers and students have achieved some mutual understanding.
3. As we have noted, observations are theory loaded. A transactional approach asks whether the theories being used make sense.

4. Knowing many academic and practical theories and having broad experience with students improves the observational skills of teachers. Equally important is the insight that this baggage may be insufficient for sizing up a new student accurately. Strong self-regard may threaten objectivity. A judgment about a student is always provisional, and the teacher should be open to revising his or judgment. Uncertainty remains a feature of both observation and of quality observation. This brings us to the interesting paradox that knowing a great deal about people is essential, but that, at the same time, it is useful to relinquish this knowledge to remain open to new insights.

Quality observations thus cannot be isolated from the other aspects of the values, skills, knowledge, and attitude that make teachers craftspeople. What is needed to become a quality observer is, for instance, a command of (practical) theories on teaching and learning, pedagogical content knowledge, familiarity with theories on human development, management skills, a reflective attitude on the feelings evoked by teaching, and attitudes such as openness and awareness. Sound beliefs on this are not only products of academic training but are also the products of experience and (self-) reflection. The more teachers understand about students, learning, classes, and about themselves, the better their observations will be.

Our narratives illustrate how observational skills are related to teacher craftsmanship. Ms. Wolf, for instance, indicated that students sometimes conduct biological experiments mechanically: They do everything as expected, but without understanding. This is why she intentionally checks whether students can come up with a conclusion. Her theory of learning makes her sensitive to mechanical learning. Similarly, Ms. Koning's experience is that students who speak Dutch as a second language may have problems understanding Dutch words. This made her alert to their special problems while learning English. Thus, pedagogical content knowledge plays a role in observation. We also quoted a teacher who was swayed by his feelings of irritation about students who did not meet his expectations. Another teacher stated that he had come to terms with such feelings. By reflecting on his own feelings and asking himself whether it was appropriate to become irritated for this reason, the latter teacher shows he is on his way to learning objectivity.

REFRAMING THE RELIABILITY AND THE VALIDITY OF OBSERVATION

Now that we have indicated some general characteristics of quality observations, we can elaborate on the reliability and validity of observation in teaching. As Airasian suggested, we describe this in a conceptual way.

In terms of the reliability of teacher observation, we refer to the extent to which people are open to observation (are aware of what is happening),

which enables them to obtain fragments of observation and put them together into a (more or less) coherent picture that appears to fit during teacher–student interaction. An arbitrary impression does not suffice. Real attention at appropriate and diverse moments seems to be the way to prevent disregarding essential parts of the picture. In our narratives, we came across teachers who were busy with other things, thus ignoring (and not becoming aware of) existing student differences. The reliability of teacher observation, moreover, is partly defined by the situation. A human being can only pay attention to a limited number of things at one time. Having too many students to observe jeopardizes reliable observations. In stating that no teacher should have direct responsibility for more than 80 students, Sizer (1992), without using measurement jargon, indicated that the reliability of observation is related to the organization of the school. At the same time, the reliability of teacher observation is also a matter of being open for and being aware of what happens with students at various times during teacher–student interaction.

The concept of validity can also be described in a commonsense, non-statistical way. Research has shown that teachers attend to a broad range of student characteristics (Airasian, 1991; Clark & Peterson, 1986; Shavelson & Stern, 1981). They are inundated with information during lessons. Consciously or unconsciously, people form fragments of observation into theories that will direct future actions. The success of the action becomes a warrant of the validity of the working interpretation (Moss, 1994). The validity of observation refers to the extent to which the theory being used for interpreting fragments of observation appears to fit during teacher–student interaction.

Taking it broadly, reliability refers to the extent to which people are open to observation and are aware of what is happening. Validity refers to their observational frames that help people interpret the observation. Although these two warrants of quality can be separated in a conceptual way, they are inseparable in a practical way. As Eisner (1992, p. 12) puts it: “Percepts without frameworks are empty, and frameworks without percepts are blind. An empty mind sees nothing.” Or to paraphrase the title of Pamela Moss’s article (1994): there is no reliability without validity. Numerous encounters with students and years of experience by teachers who only register what they see but do not understand it do not lead to quality observations in and of itself. And understanding various students without being aware of what is happening with present students, likewise, does not lead to quality observations.

THE DECISIVE ROLE OF SOCIAL PRACTICE AND THE POSITION OF RESEARCHERS

Our definitions of validity and reliability indicate that guarantees of quality cannot be given outside of concrete social practice. Practitioners find out

whether observational frameworks/theoretical assumptions make sense because, as Schön (1983) puts it, "the situation talks back." Rather than academic research playing the decisive role in determining the quality of assumptions as well as the validity and reliability of measurement (as is the case for a traditional approach), the transactional approach assigns this decisive role to practitioners.

Is there still a need for researchers? Only if they are part of a teaching practice. The transactional view points out that we can bring theory and practice closer together than was previously possible. In the traditional line of thought, in which it was assumed that a view should not be distorted by any kind of human interference, a remote position for researchers was logical. However, assuming that personal judgment cannot be eliminated and should instead be qualified, a distant position is in no way an asset. If knowledge is indeed a transaction between the world and the frameworks we bring to it, people who want to learn new things and wish to distinguish between sound and shaky ideas should position themselves as close as possible to the phenomena they are interested in.

The traditional division of labor in education between researchers and practitioners has been criticized by Terwel (1993), who introduced the concept of "overlapping group memberships," referring to teachers who become members of the community of researchers. Cochran-Smith and Lytle (1990) have also proposed teacher research. Rather than suggesting an extra task for teachers in addition to teaching, however, we propose a different role for researchers. Their main task should still be to construct theories, thus feeding the theoretical component of practitioners' frameworks with ecologically valid insights. To develop these theories, however, researchers need to operate in the ecology to which their theories refer. As our definitions of validity and reliability show, a social practice is necessary to determine whether theories deserve these warrants of quality. Just as professors of medicine and other physicians are engaged in the same practice and use the same language, researchers and teachers also need to be engaged in the same practice to develop a common professional language. If the way in which researchers approach practical problems shows that their theories indeed make sense of concrete situations, these theories become part of the shared frameworks of both researchers and practitioners. In this way, an ecologically valid professional language emerges. Such a language, developed by researcher-teachers demonstrating in practice how this language makes sense, can help other teachers grow in their profession.

EVALUATION OF THE USE OF FORMAL INSTRUMENTS FOR OBSERVATION

The absence of a common language between researchers and practitioners for sizing up students makes the ecological validity of formal instruments

questionable. If the makers of the instrument and the teachers do not share the same goals, the results produced by the instrument will not correspond with the practical classroom situation. Thus, the teacher, judging differently than the instrument, will simply see the results of the instrument as distorted or as somebody else's view. If common frameworks do exist, as they do, for instance, for spelling and computing, the use of formal instruments (such as tests) can be effective. In the case of observation in teaching, which is a very complex task for which no common language exists, the use of instruments is problematic.

The use of observational instruments is questionable for other reasons as well. These instruments require the roles of the observer and the teacher to be split. This is impractical if only one person is available. As a consequence of these split roles, instruments measure what the observer sees but fail to measure what the teacher-observer sees. Therefore, they do not measure observation in teaching.

The final drawback is even more significant. Although people can negotiate meaning and learn to understand other cultures, instruments cannot. Instruments are not conscious agents, reflecting on themselves. Hence, an instrument will never recognize its own cultural bias. Human encounter, however, might be accompanied by cultural bias but might also lead to awareness of this bias and getting beyond it.

WORKING AGAINST SELF-FULFILLING PROPHECIES THAT NEGATIVELY INFLUENCE STUDENT ACHIEVEMENT

Pygmalion has affected schools and teachers, sometimes dramatically. Mr. Langen's story reflects his school's attempt to prevent teacher expectations from influencing student achievement negatively. Therefore, completely heterogeneous classes were introduced.

Mr. Langen:

We don't like to pigeon-hole our students too quickly. Thus, when the education ministry stimulated heterogeneous classes, we were among the first to participate. The idea was: in heterogeneous classes, your past performance wouldn't really matter. That was a historic blunder. In principle, it means that you don't use the fact of a heterogeneous pupil population. In our group, we instead denied the presence of heterogeneity. Simply put, that's what happened: simply ignore the differences, and pretend they don't exist. Naturally, that doesn't work.

Mr. Langen shows how *Pygmalion* stopped teachers from taking the diversity of learners into account. We also noticed this during other conversations. Various teachers used euphemisms to talk about low achievers to avoid

labeling. Others explicitly excused themselves by explaining a terminology of low achievers was necessary to handle the situation in a mixed-ability class. These teachers, familiar as they are with the assumptions on teacher expectations, did not want to impose low expectations on students, which actually deprived them of a vocabulary to address the diversity of learners.

How should this be interpreted? The euphemisms indicate a climate in which the mere observation that students differ in achievement had become politically incorrect. Oddly enough, although teachers were challenged to observe more objectively, the situation arose in which honest observation in itself became somehow subverted. A transactional analysis explains why things have developed this way: other than the traditional view of objectivity, it acknowledges that people need frameworks. To see differences between students, teachers need frameworks—expectations!—in which differences exist. If teacher expectations in itself are put in an unfavorable light, however, differences between students are either not worth the attention or a contradiction has been created: How can a teacher, while taking the diversity of learners seriously, not have different expectations of students?

In our conversations, all the teachers including those who used euphemisms, were positive about the existence of the diversity of learners. The framework that teachers should have equal expectations of all students clearly did not serve to make sense of the classroom situation. In the following quote, Mr. Langen illustrates he almost has to go beyond his observations to help his low-achieving students along.

Mr. Langen:

If a child sees that the teacher thinks he can do it, he will show that he can do it. If the teacher makes it clear that the child can't perform, then the child won't perform. As a result, the teacher needs to maintain his faith in the child and to instill the child with self-confidence. This can be very difficult, since some children achieve next to nothing in class. In these cases, it is a challenge to give the child some degree of self-confidence. I always try, but doing so is not easy.

This fragment reflects a theory about the development of human beings. The need to have a positive self-concept is part of this theory; hope is one of the values this teacher adheres to. He realizes that what he observes may not be the final truth about his student; at the same time, he does not deny the present situation but uses it to stimulate the development of the child. A traditional view on objectivity would have been relentless: this student achieves “next to nothing.”

We conclude that expectations that influence student achievement in a negative way are not undermined by simply taking student results at face value. Instead, pedagogics are needed. An objective teacher is not some-

body who eliminates values and personal commitment but one who uses them as a starting point.

IMPROVING OBSERVATION

Observation in teaching is an important aspect of classroom assessment. Although it is natural that teachers obtain information about students passively, they should also obtain information intentionally and reflect on both their intentional and unintentional observations. Teachers not only need sufficient opportunity and awareness to form fragments of observation into a (more or less) coherent picture, but they should also be given work time to reflect on the way they have formed these fragments into theories. Ecologically valid theories help them recognize relevant aspects of student behavior. The development of this kind of theory is only in its initial stage.

Quality observations thus cannot be realized by simply organizing teacher training courses. A social practice is necessary to decide whether a particular interpretation of fragments fits. Observation is learned in concrete settings with real students and real subjects. Analysis of one's own assumptions and feelings in these settings, becoming conscious of the creative process of tailoring general ideas to concrete situations, testing interpretations, and seeing the results of a different approach, are essential. Schools should create a forum for this.

What would this forum entail? We indicated earlier that the teachers' impression of an individual student reflects not only the behavior of students but the behavior of the teacher as well. We also concluded that the quality of observation in teaching cannot be isolated from the other qualities needed to develop their skills. Discussions on individual students between "good" teachers and "bad" teachers can be hard because both types of teachers know the students in a different way. Problems with students can be attributed to the teacher's personal failure or to a lack of professionalism. Being open about one's problems means running the risk of lowering one's status. Therefore, the forum should involve more than a discussion component.

In bringing about openness on the (idiosyncratic) assumptions used in observation, a safe situation is needed in which discussions are not overshadowed by issues of status and competence. Teachers who want to improve their observational skills need mentoring by a competent (specially licensed) teacher who can demonstrate that a certain view of a particular student makes sense. The expertise of senior teachers should also include skills to help their fellow teachers analyze whether the results of their observations are connected with "bias"; reflection on personal feelings should be stimulated; they should help fellow teachers put tacit frameworks of observation, tacit practical knowledge, and tacit pedagogical content knowledge

into words. Moreover, senior teachers should challenge fellow teachers to test whether newly formed insights make sense in practice.

The role of the researcher-teacher and the senior teacher would ideally merge. By contributing to an ecologically valid professional language, and by demonstrating the value of these theories in practice, researcher-practitioners would enable others to observe more than they were able to in the past, to sharpen their perceptive faculty, and thus to broaden their expertise. Together they would be working toward a practice of objectivity.

References

- Airasian, P. W. (1991). Perspectives on measurement instruction. *Educational Measurement: Issues and Practice*, 19(1), 13–16.
- Airasian, P. W. (1994). *Classroom assessment* (2nd ed.). New York: McGraw-Hill.
- Alberts, R. (1987). Professionalisering van docenten op het vlak van evaluatie [Professional development of teachers concerning evaluation]. In T. Bergen, J. Giesbers, & C. Morsch (Eds.), *Professionalisering van onderwijsgeevenden* [Professional development of teachers]. Lisse, The Netherlands: Swets & Zeitlinger.
- Alberts, R. (1991). Professionalisering van de evaluatiepraktijk [Professional development of evaluation practice]. *Tijdschrift voor lerarenopleiders*, 12(3), 44–48.
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, 75, 631–661.
- Brophy, J. E. (1985). Teacher-student interaction. In J. B. Dusek, V. C. Hall, & W. J. Meyer (Eds.), *Teacher expectancies* (pp. 303–328). Hillsdale, NJ: Lawrence Erlbaum.
- Brophy, J. E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology*, 75, 631–661.
- Brown, S., & McIntyre, D. (1993). *Making sense of teaching*. Philadelphia: Open University Press.
- Carter, K. (1995). Teaching stories and local understandings. *Journal of Educational Research*, 88, 326–330.
- Cizek, G. J. (1999). Learning, achievement and assessment: Constructs at a crossroad. In G. D. Phye (Ed.), *Handbook of classroom assessment, learning, adjustment and achievement* (pp. 1–32). San Diego, CA: Academic Press.
- Clandinin, D. J., & Connelly, F. M. (1987). Teachers' personal knowledge: What counts as personal in studies of the personal. *Journal of Curriculum Studies*, 19(6), 487–500.
- Clark, C., & Yinger, J. (1978). *Research on teacher thinking* (Research series No. 12). East Lansing: Michigan State University, Institute for Research on Teaching.
- Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 255–296). New York: MacMillan.
- Cochran-Smith, M., & Lytle, S. L. (1990). Research on teaching and teacher research: The issues that divide. *Educational Researcher*, 19(2), 2–11.
- Connelly, F. M. (1978). How shall we publish case studies of curriculum development? *Curriculum Inquiry*, 8(1), 78–82.
- Connelly, F. M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational Researcher*, 19(5), 2–14.
- De Corte, E., Geerligs, C. T., Lagerweij, N. A. J., Peters, J. J., & Vandenbergh, R. (1981).

- Beknopte didaxiologie. Vijfde volledig herziene druk [Concise theory of teaching methods]* (5th ed.). Groningen, The Netherlands: Wolters-Noordhoff.
- Doyle, W. (1997). Heard any really good stories lately? A critique of the critics of narrative in educational research. *Teaching and Teacher Education*, 13(1), 93–99.
- Dusek, J. B., Hall, V. C., & Meyer, W. J. (1985). *Teacher expectancies*. Hillsdale, NJ: Lawrence Erlbaum.
- Eisner, E. (1992). Objectivity in educational research. *Curriculum Inquiry*, 22(1), 9–15.
- Elbaz, F. (1983). *Teacher thinking: A study of practical knowledge*. London: Croom Helm.
- Elbaz, F. (1997). Narrative research: political issues and implications. *Teaching and Teacher Education*, 13(1), 75–83.
- Elbow, P. (1986). *Embracing contraries: Explorations in teaching and learning*. Oxford, UK: Oxford University Press.
- Feiman-Nemser S., & Floden, R. E. (1986). The cultures of teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 505–526). New York: MacMillan.
- Fenstermacher, G. D. (1994). The knower and known: The nature of knowledge in research on teaching. *Review of Research on Teaching*, 20, 3–56.
- Florio-Ruane, S. (1991). Conversations and narrative in collaborative research. In C. Witherell & N. Noddings (Eds.), *Stories lives tell. Narrative and dialogue in education*. New York: Teachers College.
- Gage, N. L. (1966, September). *Discussion of the symposium on teachers' expectations as an unintended determinant of pupils' intellectual reputation and competence*. Paper presented on the program of Divisions 15 and 8 of the American Psychological Association, New York.
- Gage, N. L. (1971). Preface. In J. D. Elashoff & R. E. Snow (Eds.), *Pygmalion reconsidered* (pp. 4–5). Worthington, OH: Jones.
- Good, T. L., & Brophy, J. F. (1978). *Looking in classrooms* (2nd ed.). New York: Harper & Row.
- Guba, E. G., & Lincoln, Y. S. (1989). *Personal communication*. Beverly Hills, CA: Sage.
- Hall, V. C., & Merkel, S. P. (1985). Teacher expectancy effects and educational psychology. In J. B. Dusek, V. C. Hall, & W. J. Meyer (Eds.), *Teacher expectancies* (pp. 67–92). Hillsdale, NJ: Lawrence Erlbaum.
- Hargreaves, A., & Fullan, M. G. (1992). *Understanding teacher development*. New York: Teachers College Press.
- Hausser, K. (1980). *Die Einteilung von Schülern; Theorie und Praxis schulischer Differenzierung [Categorizing pupils: Theory and practice of educational differentiation]*. Weinheim und Basel, Germany: Belz.
- Hogan, P. (1988). *A community of teacher researchers: A story of empowerment and voice* (unpublished manuscript). University of Calgary.
- Jungbluth, P. (1984). *Verborgen differentiatie. Leerlingbeeld en onderwijsaanbod op de basisschool [Hidden differentiation: Teachers' perspectives on pupils curriculum]*. Nijmegen, The Netherlands: ITS.
- Marburger, C. L. (1963). Considerations for educational planning. In A. H. Passow (Ed.), *Education in depressed areas* (pp. 298–321). New York: Teachers College of Columbia University, Bureau of Publications.
- McDonald, J. P. (1988). The emergence of the teacher's voice: Implications for the new reform. *Teachers College Record*, 89(4), 471–486.
- McDonald, J. P. (1992). *Teaching: Making sense of an uncertain craft*. New York: Teachers College.
- Merton, R. K. (1957). *Social theory and social structure*. New York: Free Press.
- Meyer, W. J. (1985). Summary, integration and prospective. In J. B. Dusek, V. C. Hall, & W. J. Meyer (Eds.), *Teacher expectancies* (pp. 353–371). Hillsdale, NJ: Lawrence Erlbaum.
- Mitman, A. L., & Snow, R. E. (1985). Logical and methodological problems in teacher expectancy research. In J. B. Dusek, V. C. Hall, & W. J. Meyer (Eds.), *Teacher expectancies* (pp. 93–134). Hillsdale, NJ: Lawrence Erlbaum.
- Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62(3), 229–258.
- Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher*, 23(2), 5–12.

- Newell, R. W. (1986). *Objectivity, empiricism and truth*. London: Routledge and Kegan Paul.
- Noddings, N. (1986). Fidelity in teaching, teacher education, and research for teaching. *Harvard Educational Review*, 56(4), 496–510.
- Peterson, P. L., Marx, R. W., & Clark, C. M. (1978). Teacher planning, teacher behavior, and student achievement. *American Educational Research Journal*, 15(3), 417–432.
- Procee, H. (1991). *Over de grenzen van culturen. Voorbij universalisme en relativisme* [Over the borders of cultures. Beyond universalism and relativism]. Meppel, The Netherlands: Boom.
- Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, 29(1), 4–15.
- Reigeluth, C. M. (1999). *Instructional design theories and models. A new paradigm of instructional theories* (vol. 2). Mahwah, NJ: Lawrence Erlbaum.
- Rist, R. C. (1970). Student social class and teacher expectation: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review*, 40, 411–451.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupil intellectual development*. New York: Holt, Rinehart & Winston.
- Rosenthal, R., & Jacobson, L. (1968). Teacher expectation for the disadvantaged. *Scientific American*, 218, 19–23.
- Salmon-Cox, L. (1980, April). *Teachers and tests: What's really happening?* Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.
- Schön, D. A. (1983). *The reflective practitioner. How professionals think in action*. New York: Basic Books.
- Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum Evaluation* (American Educational Monograph Series on Evaluation, no.1). Chicago: Rand McNally.
- Shavelson, R. J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research*, 51(4), 455–498.
- Shulman, L. S. (1980). Test design: a view from practice. In E. L. Baker & E. S. Quellmalz (Eds.), *Educational testing and evaluation* (pp. 63–73). Los Angeles: Sage.
- Shulman, L. S. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 505–526). New York: MacMillan.
- Sizer, T. R. (1992). *Horace's school. Redesigning the American high school*. Boston: Houghton Mifflin.
- Smaling, A. (1990). Objectiviteit en rol-neming [Objectivity and Role-taking]. In I. Maso & A. Smaling (Eds.), *Objectiviteit in kwalitatief onderzoek [Objectivity in qualitative research]*. Amsterdam/Meppel (the Netherlands): Boom.
- Snow, R. E. (1969). Unfinished Pygmalion (review of Pygmalion in the classroom). *Contemporary Psychology*, 14, 197–200.
- Stiggins, R. J. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice*, 10(1), 7–12.
- Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22(4), 271–286.
- Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective education. *Educational Measurement: Issues and Practice*, 5(2), 5–17.
- Terwel, J. (1993). Het bevorderen van authentiek leren [Promoting authentic learning]. In B. van Oers & W. Wardekker (Eds.), *De leerling als deelnemer aan de cultuur [The student as participant of the culture]*. Delft, The Netherlands: Eburon.
- Thorndike, R. L. (1968). Review of Pygmalion in the classroom. *American Educational Research Journal*, 5, 708–711.
- Verloop, N. (1992). Praktijkkennis van docenten: een blinde vlek van de onderwijskunde [Practical knowledge of teachers: A blind spot of educational theory]. *Pedagogische Studien*, 69, 410–423.
- Verloop, N., & Van der Schoot, F. (1995). Didactische evaluatie [Evaluation as an aspect of teaching strategy]. In J. Lowyck & N. Verloop (Eds.), *Onderwijskunde. Een kennisbasis voor*

- professionals* [Educational theory. A knowledge base for professionals]. Groningen, The Netherlands: Wolters-Noordhoff.
- Verloop, N., & Zwarts, M. A. (1987). Evalueren [Evaluation]. In P. Span, J. M. C. Nelissen, H. F. Pijning, & C. Dietvorst (Eds.), *Onderwijzen en leren* [Teaching and learning] (pp. 223–248). Groningen, The Netherlands: Wolters Noordhoff.
- Wajnrub, R. (1992). *Classroom observation tasks*. Cambridge: University Press.
- Wardekker, W. L. (1989). Praktijktheorieën van leraren [Practical theories of teachers]. *Pedagogisch Tijdschrift* (Speciale editie vierde landelijke pedagogendag, 27 mei 1989). Amersfoort, The Netherlands: Acco.
- Wardekker, W. L. (2000). Criteria for the quality of inquiry. *Mind, Culture and Activity*, 7(4), 259–272.
- Wineburg, S. S. (1987, December). The self-fulfillment of the self-fulfilling prophecy. *Educational Researcher*, 6(6), 28–37.
- Yinger, R. J. (1977). *A study of teacher planning: description and theory development using ethnographic and information processing methods* (Research series No. 18). East Lansing: Michigan State University, Institute for Research on Teaching.
- Zahorik, J. A. (1975). Teachers planning models. *Educational Leadership*, 33(2), 134–139.
- Zeichner, K. (1999). The new scholarship in teacher education. *Educational Researcher*, 28(9), 4–15.
- Zeichner, K., & Tabachnick, B. R. (1981). Are the effects of teacher education washed out by school experience? *Journal of Teacher Education*, 32(3), 7–11.

JACQUELIEN BULTERMAN-BOS is completing a dissertation about teaching diverse learners. Her research interests include the relation between theory and practice in education. She is a former teacher and a teacher trainer.

JAN TERWEL is professor of educational psychology at the Faculty of Psychology and Education, Vrije University, Amsterdam. His research interests include adaptive learning and instruction, social interaction and group composition. A recent publication by Dr. Terwel is “The Effects of Integrated Social and Cognitive Strategy Instruction on the Mathematics Achievement in Secondary Education,” *Learning and Instruction*, 9(5), coauthored with D. Hoek and P. Van den Eeden.

NICO VERLOOP is professor of education and dean of ICLON Graduate School of Education at Leiden University, The Netherlands. He is immediate past president of the Dutch Educational Research Association. His research interests include research on teacher education and teacher cognition. His most recent publication is “Student Teachers’ Beliefs About Mentoring and Learning to Teach During Teaching Practice,” *British Journal of Educational Psychology*, 71, coauthored with J. D. Vermunt.

WIM WARDEKKER is assistant professor at the Faculty of Psychology and Education at Vrije Universiteit. His research interest is in developing theories and practices of education based on a sociocultural viewpoint. His most recent publication is “Criteria for the Quality of Inquiry,” *Mind, Culture, and Activity*, 7(4).