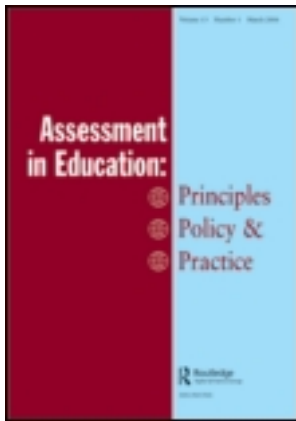


This article was downloaded by: [Professor Linda Allal]

On: 27 January 2014, At: 01:01

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Assessment in Education: Principles, Policy & Practice

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/caie20>

### Teachers' professional judgement in assessment: a cognitive act and a socially situated practice

Linda Allal <sup>a</sup>

<sup>a</sup> Faculty of Psychology and Educational Sciences , University of Geneva , Geneva , Switzerland

Published online: 05 Nov 2012.

To cite this article: Linda Allal (2013) Teachers' professional judgement in assessment: a cognitive act and a socially situated practice, *Assessment in Education: Principles, Policy & Practice*, 20:1, 20-34, DOI: [10.1080/0969594X.2012.736364](https://doi.org/10.1080/0969594X.2012.736364)

To link to this article: <http://dx.doi.org/10.1080/0969594X.2012.736364>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

## **Teachers' professional judgement in assessment: a cognitive act and a socially situated practice**

Linda Allal\*

*Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland*

*(Received 31 January 2012; final version received 1 October 2012)*

This paper presents a study of teachers' professional judgement in the area of summative assessment. It adopts a situated perspective on assessment practices in classroom and school settings. The study is based on interviews with 10 sixth-grade teachers and on the assessment documents they used when determining end-of-term grades in students' report cards. The main findings from qualitative data analysis highlight both the individual cognitive and the socially situated aspects of teachers' judgements. The findings are discussed with respect to three levels of teacher judgement and the implications for activities of social moderation.

**Keywords:** teacher judgement; summative assessment; socially situated practice; grading in primary school; social moderation

Teachers' professional judgement intervenes in all areas of their activity: when they plan and prepare learning activities; when they conduct lessons and decide which students to call on or how to adapt their initial lesson plan; when they interact with individual students and offer various forms of assistance; when they meet with parents to discuss a student's progress; when they carry out formative and summative assessments in the classroom. We propose to qualify teachers' judgement as 'professional' when it takes into account the resources and constraints of their work setting and is informed by professional knowledge acquired through experience and through training. In this paper, we will examine professional judgement in the area of summative assessment, and, in particular, the ways in which teachers determine the grades recorded in students' report cards. After introducing the concept of professional judgement in assessment and the situated perspective that orients our research, we will present a study carried out with sixth-grade teachers working in public primary schools in the canton of Geneva, Switzerland.

### **Professional judgement in summative assessment**

There is a long-standing tradition of looking at teacher assessment practices through the lenses of measurement theory and of considering these practices as largely defective. In particular, teacher judgement as expressed in grades assigned to student work and recorded in student report cards has been criticised for failing to

---

\*Email: [linda.allal@unige.ch](mailto:linda.allal@unige.ch)

meet sufficient standards of reliability, objectivity, and validity. These criticisms have reinforced the reliance in many school systems on the use of external, standardised tests to assess student achievement. A different perspective has been proposed by researchers who attempt to understand how teachers' assessment practices function in classroom and school settings and why their practices necessarily differ, in their inherent logic and rationale, from standardised measurement. An early expression of this perspective appeared, over 60 years ago, in an article by Scates (1943) who argued that teachers' assessments are based on 'interplay between objectivity and judgement' (6) and respond to criteria that differ from those of scientific measurement. A growing body of research on classroom assessment, conducted primarily since the mid-1980s, has provided insights into the processes involved in various aspects of assessment, from informal formative assessment integrated in teaching/learning activities to formal reporting of summative assessment results, typically in the form of grades in student report cards. Studies of grading by Brookhart (1994) and others have shown that grades generally express a composite judgement based on evidence about student achievement, effort and attitude. In our research on grading practices in primary school (Allal 1988), it was found that teachers devised various strategies for combining quantitative evidence – principally scores on teacher-made tests – with qualitative information drawn from informal observations, interactions with students, and examination of students' daily work folders.

Several researchers have proposed new conceptualisations of the qualities to be sought in classroom assessment. As an alternative to the psychometric definition of reliability, Smith (2003) has formulated the concept of 'sufficiency of information', which implies that the teacher seek a sufficient amount of relevant information to be able to assign the grade that best represents the student's level of learning in the area under consideration. Moss (2003) has proposed a reconceptualisation of validity based on hermeneutics and sociocultural theory as frames of reference for assessment, entailing confrontation and interpretation of qualitatively different sources of information about student learning. Other researchers have retained classical measurement terminology but have adapted its usage to better describe teachers' summative assessments: for example, in a systematic review of the dependability of teachers' assessments for summative purposes, Harlen (2005) prioritised the validity component of dependability over that of reliability.

Teacher judgement in classroom assessments is obviously subject to error and bias, as is judgement in all areas of professional expertise (medical, financial, etc.). The teaching profession, school administrators and researchers working in the area of assessment have a collective responsibility to look for ways of improving the quality of teachers' judgement. One important approach is the development of intervention projects designed to improve the consistency of teacher judgements through their participation in activities of social moderation (i.e. meetings where teachers discuss, confront, and negotiate grades assigned to student work or other assessment decisions). A project carried out by Black et al. (2011) allowed secondary school teachers to explore and develop their conceptions of validity in summative assessments through the construction of portfolios and participation in social moderation exercises. A study conducted by Wyatt-Smith, Klenowski, and Gunn (2010) researched social moderation based on the introduction of explicit standards, accompanied by annotated work samples and grading guidelines, as a way of improving the consistency of middle-school teachers' judgements about grades. Both studies

documented the importance of social moderation as a means of constructing shared assessment practices, but also revealed the complexity of the factors that intervene in teachers' judgements. For example, in the Wyatt-Smith, Klenowski, and Gunn study, tensions were observed between official standards and the tacit, 'in the head' standards that teachers had acquired through their classroom experience. Our research presented in this paper aims at understanding how teacher judgement functions in its natural setting, both in the classroom and in the school, given an institutional context that allows teachers considerable latitude in determining end-of-term grades. We are interested in the degree to which teachers voluntarily engage in concerted assessment activity and develop shared practices. By adopting a situated perspective on teacher judgement, we seek to investigate the relationship between the cognitive and the social dimensions of assessment practice.

### **A situated perspective on teacher judgement in assessment**

The process by which teachers determine the grades to record in report cards typically takes place in isolation, outside direct interaction with students and other teachers, and is generally seen as an activity in which teachers exercise fully their autonomy (Kain 1996). In a situated perspective, however, even individual cognitive acts carried out by a professional in isolation and with autonomy are considered to be profoundly shaped by the collective practices of a professional community (Wenger 1998). Using qualitative data from research on teacher judgement during social moderation activities, Klenowski and Adie (2009) showed how teachers' discourse and actions illustrate each of the components – practice, meaning, identity, community – of Wenger's model of situated learning. Our approach to situated cognition has been influenced by the conceptualisation formulated by Cobb et al. (1997). They consider individual psychological processes and shared social practices to be joined in a reflexive, mutually constitutive relationship. However, in contrast with Wenger, they believe that it is possible to focus, both theoretically and empirically, on each plane of human activity: on individual construction of repertoires of action and on practices of enculturation in educational settings. One important concept developed by Cobb et al. is that meaning (for example, in our study, the meaning of a grade given by a teacher, or the meaning of fairness in grading) is never identical in the minds of all actors (students, teachers) but that the interactions among actors allow the emergence of 'taken-as-shared' meaning that guides their activity.

We believe that teachers' judgement in assessment is analogous to clinical judgement in the medical professions in that it implies establishing a relationship between the singular (everything the evaluator knows about a particular individual) and the general (formal and tacit professional knowledge, as well as institutional norms and rules) in order to formulate the most appropriate course of action possible. Although a clinical judgement may be expressed in one of several discrete categories (e.g. stages of a disease, a set of grades), it is through the construction of an underlying 'narrative' that the professional makes sense out of various pieces of information, considers the potential risks and benefits associated with different options, and formulates a judgement that orients subsequent action (White and Stancombe 2003). This view is congruent with research on the functioning of organisations showing that 'sensemaking' by a professional is a cognitive process that entails detecting relevant cues, attributing retrospective meaning, integrating them in a plausible narrative that allows understanding of a situation and sets the

stage for action and, at the same time, it is a socially situated process shaped by the exchanges among actors and by the systemic features of an organisational and cultural context (Weick, Sutcliffe, and Obstfeld 2005). One main finding of organisational research is that, ‘sensemaking is driven by plausibility rather than accuracy’ (415). Accuracy implies that a competent professional is able to apprehend correctly a true state of events and act accordingly; plausibility implies that a judgement or decision is socially constructed in a manner that is persuasive for all concerned parties. In the case of teachers’ judgements in summative assessments, this would mean that the plausibility of the message conveyed by a grade – in the transactions between teachers, students and their parents – may be more significant than the accuracy of the grade (i.e. its correspondence to some hypothetical true measure of student competency, as postulated in measurement theory).

### Context of our study

Three characteristics of the Swiss school system that differ from many other systems need to be briefly mentioned. Although there are mechanisms of coordination between the education ministries of the 26 Swiss cantons, each canton’s school system has broad administrative and financial autonomy. One recent outcome of inter-canton cooperation was the adoption in 2011 of national standards of education; a system of monitoring is being developed but will not be fully operational until 2015. The orientation of students to tracks of secondary school studies (academic, vocational, or other) begins quite early (at age 11–12 in most cantons) and is based primarily on grades given by teachers, although results on external examinations prepared by the canton or school district are often taken into account.

In the canton of Geneva, where our research was conducted, a primary school reform movement was initiated in 1994 and has continued to have a substantial impact on teaching practices in many schools. This movement emphasised the importance of teachers working together in their school setting and undoubtedly reinforced many forms of collaboration among teachers who had formerly worked in a more individualistic manner. The reform also encouraged innovative approaches to classroom assessment. Emphasis was placed on formative assessment and student self-assessment. Several transformations of summative assessment were introduced: in particular, the development of portfolios and the use of new report card formats using qualitative assessment categories for each subject matter, rather than numerical grades. A wide range of school-based professional development activities were implemented to support the practices advocated by the reform movement. Approximately two-thirds of the canton’s primary schools were actively involved in the reform movement, and an even larger number participated in professional development activities that reflected the ideas of the reform.

Protests led by a ‘back to basics’ type of lobby soon emerged, however. One consequence was the reintroduction in 2005 of the requirement, in the third through sixth grades (students’ ages 9–12), that teachers record numerical grades, on a scale of 1 to 6, for each major subject-matter area, at the end of each term, in the student’s report card. A grade of 4 (‘objectives attained’), approximately equivalent to a grade of C in the Anglo-American grading system, was set as the minimum requirement in the areas of French and mathematics for student orientation at the end of sixth grade to the academic track of secondary school (although derogations could be requested by parents under certain circumstances). By studying teachers’

professional judgement in sixth grade, we knew we would be entering an area of assessment fraught with dilemmas because of the consequences for the students' future: students who are not oriented toward the academic track in seventh grade rarely manage to undertake these types of studies at a later date.

At the time our study was conducted in 2006, the official regulations required that teachers apply the grading scale only in the establishment of the end-of-term grades in the report cards, but allowed teachers broad latitude to decide how to carry out assessments during the term, whether or not to assign grades on tests and work done in class, and how to establish the grades recorded in the report cards. In particular, there was no obligation to base end-of-term grades on an average of classroom test grades. Two official references were provided to the teachers: (1) the curriculum objectives to be attained in each subject-matter area at the end of sixth grade, as well as examples of assessment tasks corresponding to key objectives; (2) the external, canton-wide examinations taken by all sixth-grade students in three subjects (French, German, mathematics).

### **Aim of our study**

The overall aim of our study was to collect information on teachers' summative assessment practices in sixth grade in order to elucidate the role of professional judgement in the determination of the end-of-term grades recorded in students' report cards. The collection of data, based on interviews and analysis of assessment documents used by the teachers, was designed to identify:

- the cognitive acts intervening in grading with respect to the choice of sources of information, the interpretation and ways of combining information, the reasoning and rationale leading to the formulation of a judgement;
- the socially situated dimensions of assessment as elaborated through the teacher's transactions (direct interactions and exchanges mediated by tools and artefacts) with students in the classroom, with colleagues in the school, and with resources and constraints provided by the school system.

### **Procedures**

We chose to carry out a study that would document teachers' practices and their reflections about assessment in a qualitative manner. A relatively small number of teachers were involved, but a substantial body of information was collected regarding each teacher who was interviewed at two points in the school year and provided all the assessment documents used to determine end-of-term grades.

### **Participants**

Ten teachers of sixth-grade classes (students' ages 11–12) from 10 different schools participated in our study. Contacts were made with teachers who had participated in professional development activities on assessment and who were teaching in schools that had taken part in the primary school reform movement. The selected teachers included six women and four men, with a teaching experience ranging from 6 to 40 years. They were working in a cross-section of schools with student populations representative of the cultural and socio-economic diversity of the canton. Our study of

professional judgement was thus restricted to experienced sixth-grade teachers, who had had the opportunity to reflect on assessment, but did not entail any significant restriction with respect to student characteristics. In a small-scale qualitative study, both teacher and student characteristics can be relevant when interpreting the interview data of any given teacher, but do not provide a basis for identifying systematic relationships between these characteristics and the teachers' assessment practices.

### ***Data collection***

Two interviews, lasting 1½ to 2 hours, were conducted with each teacher in his or her classroom after school hours, the first one at the end of the second term in March and the second one at the end of the school year in June. In order to prepare the interview, we asked the teacher to select two students for whom the teacher had hesitated, when completing the student's report card, between the grades of 3 ('objectives nearly attained') and 4 ('objectives attained') and to bring to the interview all relevant information concerning these students: their report cards, their work folders, the tests they had taken, etc., as well as any material (records, lists, tables, graphs, etc.) the teacher used to determine the end-of-term grades. We asked the teacher to select one student for whom the hesitation between the grades of 3 and 4 concerned mathematics and another student for whom it concerned 'environment' (geography, history, natural science). Our decision to focus the interview on two students with borderline results had two justifications. First, we believed that professional judgement would be more clearly evidenced when a teacher is confronted with a dilemma, with a decision that may have important consequences for the student, and therefore needs to reflect on the best course of action. Second, we felt that the material brought to the interview would help the teachers to re-enact their grading practices and reasoning in two concrete situations, rather than engaging in overly general discourse about assessment or in retrospective reconstruction disconnected from reality. It is difficult to know, however, if our request to select two borderline students and bring material concerning them to the interview influenced, in one way or another, the process of assessing these students.

An interview format was developed around the following topics: the information on which the student's end-of-term grade was based; the process of determining the grade; the means of communication that preceded or accompanied the grade in the student's report card; and the teacher's recommendations regarding the student's orientation to a track of secondary school. In addition, some general questions on the evolution of the teacher's assessment practices were asked at the end of the interview. Each topic and sub-topic was approached by an open-ended question: for example, 'How did you take into account this information [reference to a document] to decide on the grade to give [student's name]?' This question was followed, as needed, by more specific questions (e.g. 'What was most important in your decision to give a grade of 3?' 'What could have led you to give a 4 rather than a 3?'), as well as by questions probing the frames of references that underlie or justify the teacher's assessment judgements (e.g. 'To what extent is your way of grading different or similar to that of your colleagues?' 'Did the student's result on the external mathematics examination influence your own end-of-year grade?' 'How does this decision fit in with your ideas about pedagogy?'). The exact phrasing of the questions was adapted to the discourse used by each teacher when talking

about assessment. The expression ‘professional judgement’ was not used by the interviewer.

### ***Data analysis***

Each audio-recorded interview was transcribed and coded. Using the qualitative approach proposed by Miles and Huberman (1984), data were extracted from the transcript and inserted, summarised or categorised in a series of data formats. Information from the documents the teacher brought to the interviews was also integrated into the data formats. This initial classification of the data from the two interviews of each teacher was carried out by the person who had conducted the interviews. A second member of the research team then read the interview transcripts, examined the information in the formats and annotated the documents with questions and comments about points that appeared unclear or for which there could be a divergence of interpretation. These annotations were discussed in meetings of the research team, which decided on the final formulations to be retained. Subsequently, other formats were developed in order to compare similarities and differences across teachers. In addition, significant passages from the interviews were selected to exemplify teachers’ narratives about their resolution of assessment dilemmas or other important aspects of their assessment practice.

### **Major findings**

This paper presents major findings regarding the teachers’ grading practices in mathematics. Practices in the area of environment entailed very similar operations, but often included more informal observation. More extensive presentations of the data, for both subject matters, are provided elsewhere (chapters 7 to 10 in Lafortune and Allal 2008; Allal and Mottier Lopez 2009).

### ***Professional judgement as a cognitive act***

The data we collected showed that teachers’ judgements about end-of-term grades involved two main operations: (a) gathering information from a variety of sources; and (b) combining information in an interpretative synthesis that sometimes included but was not limited to an arithmetic algorithm. All teachers considered students’ results on classroom tests as their primary source of information for determining end-of-term grades (although one teacher gave equal importance to more open-ended classroom activities). At the end of the year, most teachers also took into account the student’s result (score compared to a criterion-referenced standard) on the external mathematics examination. When a student’s results on these formal assessments fluctuated and/or when teachers were not convinced that they provided a complete and valid picture of a student’s understanding of the subject matter, all teachers examined other, more qualitative sources of information. These sources included: student productions (e.g. reports on a problem-solving activity), observations of student participation in classroom activities, folders containing homework or worksheets, portfolios of student work accompanied by reflective comments or self-assessments, conferences with the student and/or with his or her parents, and discussions with other teachers or with other professionals working in the school (e.g. school nurse, psychologist assigned to the school district). For some teachers,



the process of combining the information to determine the end-of-term grade included a calculation, in particular the averaging of scores or grades obtained on several tests, but the process was never reduced to the simple application of an arithmetic algorithm. Quantitative elements, whether a profile or an average of scores, were always combined with qualitative information to determine which of two end-of-term grades under consideration (3 or 4) would be the most appropriate indicator of student grades understanding and competency. This process of combining quantitative and qualitative information entailed an interpretative synthesis in which the teacher decided how much importance would be given to various types of information and which pieces of information appeared to be most relevant in the case of a given student.

Although all the teachers ensured a certain standardisation of their assessment procedures, particularly with respect to the administration of classroom tests, they all introduced adaptations in their procedures to take into account individual student specificities. When determining a student's end-of-term grade, teachers sometimes decided not to take into account the result on one test, particularly if the result was not considered valid because of a major problem in the student's personal life (e.g. mother's hospitalisation). In other cases, teachers decided to have the student take an extra test, or to do an extra activity not requested of other children in the class, either as a replacement for the test that was dropped from consideration, or as a means of collecting additional information on the student's competency. When hesitating between two grades, teachers sought out information from additional sources not examined for all students. Examples included: carrying out a systematic analysis of the student's worksheets and/or homework; observation and interaction with the student while he or she carried out a task; and discussion of the student's situation with other teachers or professionals.

Although many teachers tended to construct their judgement methodically by accumulation of information, some started with an initial impression, based on their tacit knowledge of the student's work, which they then confronted with several types of evidence. Pietro, for example, followed the second approach when deciding about the mathematics grade to assign to Angelo for the second term (all names are pseudonyms). Pietro's initial impression was that Angelo's work during the term was 'worth 3'. When he examined Angelo's results, objective by objective, on three classroom tests and on four more open-ended activities carried out in class, Pietro saw that Angelo's test results were compatible with a grade of 3, but that his work in the activities provided evidence of conceptual understanding that could justify a grade of 4. For Pietro, the activities showed: 'what Angelo is capable of doing ... when he is not in a situation of stress ...; on that basis, ... I tilted from a 3 to a 4'.

### ***Professional judgement as a socially situated practice***

When the teachers described their assessment practices and their judgements about student grades, they did not necessarily explain how their approach to assessment had developed. However, when we asked questions probing for more elaboration, all of the teachers referred to aspects of their professional work setting that had shaped their personal practices and their ways of formulating judgements. Nearly all the teachers spoke of close collaboration with other teachers in their school, either with other sixth-grade teachers or with a group of teachers in charge of a multi-grade cohort of students. Collaboration regarding summative assessment

focused primarily on the construction and application of common tools (mathematics tests and other tasks). A few teachers described discussions with another teacher concerning the determination of end-of-term grades, particularly for students with borderline or low achievement. Pietro, for example, said that discussions about grades with his sixth-grade colleague provided useful information because each of them had the opportunity to observe and interact with students in both classes during mathematics workshops mixing the two classes.

When explaining their practices and grading judgements, many teachers referred to discussions in their school about the importance of ‘coherence’ in communication concerning assessment with students and their parents. This concern often led to the elaboration of a policy regarding summative assessment which was followed on a school-wide basis. For example, it was decided in some schools that test results would be expressed as scores or by qualitative rubrics without conversion into numerical grades and that grades on the official 1 to 6 scale would be given only at the end of term when the report card was filled out.

The assessment documents brought to the interviews showed evidence of the teachers’ appropriation of references provided by the school system (in particular, the curriculum objectives) and of principles and procedures advocated and illustrated during school-based professional development activities. For example, it was decided in most schools that summative tests in mathematics would have a ‘header’ based on one of several formats presented during professional development. Although certain aspects of the headers varied between schools, they all included the objectives being assessed, the corresponding items, the student’s result for each objective, as well as a space for teacher comments and suggestions, the aim being to help students focus on a profile of learning outcomes with respect to objectives, rather than on the summarisation of achievement in a single numerical score or grade. When the teachers referred to their beliefs about summative assessments, they often evoked ideas that had been debated during professional development. One of the most prominent was the idea that the grade given to a student provides a ‘message’ that should help the student move forward in his or her learning. Teachers varied, however, in their opinions about the type of message to provide: many argued that if available evidence does not allow a clear-cut choice between two grades, it is preferable to assign the higher grade as a message of encouragement, whereas others felt that it is sometimes necessary to assign the lower grade as a sort of ‘wake-up call’ that will stimulate student investment in school work.

In order to illustrate and contextualise the main findings of our study, we will present a portrait of a teacher’s assessment practices and judgements formulated during grading and of the transactions within the classroom and within the professional community of the school that shaped her practices.

### **A portrait: Alice’s assessment practices**

Alice has taught upper primary classes for 29 years in the same school located in a neighbourhood with a relatively high percentage of families of lower socio-economic status. Alice’s description of her assessment practices and the documents she brought to the interview allowed us to identify several features of the cognitive activity that characterises her professional judgement. To a larger extent than many other teachers, she tends to collect quite detailed documentary evidence of students’

work. In addition to the classroom tests the students take each term in mathematics, Alice has developed several personal tools for the assessment of student progress:

- A notebook containing records of students' daily work (worksheets, homework), outcomes of problem-solving activities in small groups, various annotations, as well as symbols and colours that help her follow each student's progression;
- Observation sheets on which she records brief notes regarding students who encounter difficulties (their errors, questions, behaviours).

In cases where students' test results fluctuate and Alice hesitates between two end-of-term grades, she carries out a very systematic review of the available information, including checking the errors made on the tests with respect to the students' daily work on similar tasks. If needed, she collects additional information from discussions with the students, their parents, her colleagues or other professionals (e.g. school nurse). She tries to situate the student's work during the term in a learning progression, linking past work, current evidence of competency and future schooling opportunities. She believes you have to have a 'long view' of learning over time in order to formulate an appropriate assessment. Although she applies the same basic assessment procedures to all students in her class, she thinks it is important to adapt her procedures, when needed, to take into account individual students' specific needs or situations.

Her approach can be illustrated by the process followed to determine the mathematics grade given to Khalil at the end of second term. Knowing that Khalil is confronted with family difficulties linked to the illness of his mother, Alice has organised regular conferences with him and with his father since the beginning of the year. Khalil's results in mathematics were very good during the first term (grade of 5), but have steadily declined on the three tests given during the second term (16/20, 24/36, 5/15). These results lead Alice to hesitate between a grade of 3 or 4 for the second term. Her observations show that Khalil often understands mathematics concepts quickly, but her records indicate that he has been doing his homework less and less carefully and has not mastered several key operations concerning division. When discussing with Khalil about the future, Alice has discovered that 'he has dreams, he says he'd like to become a doctor', but at the same time he admits that he does not invest much in his schoolwork. Alice is concerned that if Khalil has a grade of 4, the downward spiral will continue and his chances of entering the academic track in seventh grade will be diminished. She finally decides to give Khalil a grade of 3 and has an in-depth discussion with him about the investment that will be needed during the third term so that he can enter and stay in the course of studies he aspires to follow. This discussion concerns questions such as: 'Is there a way to set aside time in school to get ahead on homework?' 'How important is math if you want to be a doctor?' The strategy pays off. Khalil's work improves during the third term: he obtains a grade of 4, has a good result on the external mathematics exam, and is able to enter the academic track of seventh grade.

Alice's reasoning about the grade to give Khalil entails a succession of cognitive acts that constitute her manner of collecting and combining information so as to formulate a judgement. At the same time, these acts are integrated in assessment practices shaped by the transactions between Alice and her students. The conferences

with Khalil and his father are not merely ‘sources of information’; they are events during which Alice constructs with Khalil and his father a taken-as-shared understanding about Khalil’s capacities, about the type of studies he wants to undertake, about the difficulties he encounters, about possible ways of overcoming obstacles and supporting learning both in school and at home. The construction of classroom assessment culture results, however, not only from individualised encounters between the teacher and each student, but also from experiences that involve the entire class. One example in Alice’s class concerns the notion of ‘fairness’ in assessment. Every week or two, a ‘classroom forum’ takes place during which the students and teacher discuss problems that have occurred and how to solve them. Each class member can submit anonymously (via a letter box) questions or problems to discuss. Having overheard one student telling another that the grade he received shows that he is the teacher’s ‘favourite’, Alice decides to submit the question ‘Does the teacher have favourites?’ to the classroom forum. The discussion that ensues allows exploration of the notion of fairness in assessment, especially with respect to grades given in the students’ report cards. Questions discussed include: ‘What are the signs of favouritism?’ ‘Are there different ways of understanding what appears to be favouritism?’ ‘Can the teacher take into account differences among students while still assuring fair and equitable assessment?’ As this example suggests, the elaboration of taken-as-shared understanding about the goals of assessment and the notion of fairness in assessment contributes to the development of the classroom assessment culture within which Alice’s practices and judgements take on meaning.

Alice’s assessment practices are also shaped by the assessment culture that has emerged in her school. Alice works very closely with another sixth-grade teacher: they prepare most of their summative tests together, administer them the same day and use the same correction criteria. Alice also has exchanges with other professionals on issues regarding assessment: for example, she asks the school nurse to participate in some of the conferences with Khalil and his father; she consults the psychologist attached to the local secondary school about Khalil’s scores on a standardised test used for orientation. These exchanges are situated, moreover, in a broader framework of school-wide teacher collaboration regarding assessment. The teachers in Alice’s school have developed several means of coordinating their practices. In the teachers’ meeting room, there is a filing cabinet with all the summative assessment tools used in each classroom, so that each teacher can see what is done by other teachers and what types of assessments students have encountered in previous years or will encounter in subsequent years. The teachers have developed a common policy regarding numerical grades: grades are not given on tests and classroom work so that each teacher will be able to reflect on the most appropriate end-of-term grade in a subject-matter area, given all available information, and so that students and parents will not focus on calculating averages. Because this policy has been discussed and adopted collectively, each teacher is able to explain its rationale to both students and parents. Given the long tradition, in Geneva as in other school systems, of basing report card grades on averages of several test grades, it is not always easy to convince parents that a focus on learning progression is more important than calculating averages, but the teachers believe that the chances of doing so are increased when they adopt shared collegial responsibility for their assessment policy. In Alice’s school, the teachers also conduct periodic meetings (attended by the school director, school nurse, and sometimes a psychologist from the canton referral service) to discuss the situation of students who are encountering learning

and/or other difficulties (medical, psychological, family problems, etc.). At these meetings, it is possible for different adults who have worked with a given student to share their observations and to plan ways of providing support for the student's learning and well-being. The meetings allow a form of collective validation of teacher judgements that take into account student specificities and can also encourage realignment of teachers' judgements with respect to the taken-as-shared norms of the school's assessment culture.

### **Summing up**

To start with, certain limitations of the research reported here need to be mentioned. We have explored only a small part of teachers' summative assessment practices: the professional judgement of experienced sixth-grade teachers when assigning end-of-term mathematics grades recorded in students' report cards. Although the assessment documents the teachers brought to the interviews often provided concrete support for what they said, interview data are never exempt from some degree of distortion. An additional limitation is the lack of direct observation of the teachers' transactions with their classes and colleagues.

We nevertheless believe that our research makes a valuable contribution to the understanding of teachers' professional judgement. The main findings of our research, as well as the portrait of Alice's assessment practices, provide substantial support for a situated conception of professional judgement as both an individual cognitive act and a socially situated practice. The grading procedures developed by each teacher reflected personal choices about sources of information, about how to combine information and arrive at a judgement, and about ways of adapting their assessment procedures to take into account student specificities. But underlying these cognitive acts, it was generally possible to identify the socially situated nature of the practices that the teacher had elaborated through interactions with students, through collaboration with colleagues, both informally and during professional development activities, and through appropriation of institutional rules and resources.

Our study sheds light, in particular, on teachers' grading practices in situations where they encountered dilemmas. When confronted with ambiguous or conflicting evidence regarding a student's learning outcomes (e.g. a curriculum objective appeared to be attained in one test or task, but not in another), the teachers sought additional information. In most cases, they were concerned with reducing uncertainty by finding enough information to 'tip the balance' toward one grade or the other, which is an approach close to the idea of sufficiency of information proposed by Smith (2003). In some cases, however, they sought new information of a different nature than the already available evidence, information that would 'fill out the picture', or allow a new perspective, which is nearer to the hermeneutic reconceptualisation of validity proposed by Moss (2003).

### **Teacher judgement and social moderation**

Our concluding remarks will concern the development of professional judgement in assessment through activities of social moderation. We find it useful to differentiate three levels of judgement that may have different implications for social moderation. The first level concerns the judgements teachers make when constructing or

choosing instruments for classroom assessment. It includes judgements about the relative importance of curriculum objectives, about the types of tools to use (tests, open-ended tasks), and about the criteria for correcting and grading. The teachers we interviewed had all developed this type of collaboration with colleagues. A second level of judgement concerns the assessment of a piece of student work (test, task) and the degree of agreement among teachers about the grade or other assessment annotation to assign to the work; this level is often the main focus of the research on social moderation (e.g. Wyatt-Smith, Klenowski, and Gunn 2010). Our study provided little evidence regarding this level of judgement. Although teachers often prepared and administered tests in unison, the correcting and grading of the test papers was generally done individually. In other words, collaboration regarding the construction of assessment instruments did not usually extend to assuring consistency of assessment outcomes. The main focus of our study was on a third level of judgement, namely the process of assigning end-of-term grades, to be recorded in the student's report card, on the basis of several different sources of evidence. Several of the teachers we interviewed gave examples of what might be called 'informal' social moderation: principally, consultation of another teacher about the grade to give a student and school meetings where teachers discussed the situation of students encountering difficulties with respect to learning and/or other aspects of their lives. For the large majority of students, however, end-of-term grades were decided by each teacher without interaction with colleagues.

When developing activities of social moderation, the question needs to be raised as to the aim of these activities. Although consistency of judgement across teachers when developing assessment instruments and grading a piece of work is a useful goal, this is less evident in the case of grades reported at the end of a term or school year. Our research suggests that teachers generally aspire to construct and use classroom tests that will give dependable (valid, reliable) results, but the combining of test results with other sources of information for reporting end-of-term grades remains largely a clinical, interpretative process. This process is linked to an intuitive conception of what might be called 'differentiated equity' (as contrasted with formal equity).<sup>1</sup> When assigning an end-of term grade, teachers use information that is formally equivalent for all students (primarily test results), but for students at and below the borderline, they also seek and use additional information that varies in its nature and its interpretation from one student to another. For example, several considerations led one teacher who hesitated between the grades 3 and 4 to opt finally for a grade of 4 that would allow a boy to go into the academic track of secondary school. By systematically comparing the student's classroom test results to his results on the canton math exam and to the activities in his math portfolio, the teacher saw that several topics in the area of geometry were not well understood, but that most major objectives were attained, particularly with respect to numerical operations. In addition, the teacher took into account the close collaboration that had been established between the boy's parents and his various teachers throughout primary school: 'The parents are very supportive, they will see he keeps up, does his homework'. From the teacher's viewpoint, you have to use *all* the information you have about a student to formulate the judgement that is the most appropriate for that *individual*. Although you may lack equivalent information about other students, it does not diminish your responsibility to use judiciously the information you have. Obviously, from the point of view of measurement theory, taking into account the support offered by the student's parents would be considered as a

bias, but many teachers would consider it as important information for making a sound judgement in the best interest of the student. The situation is similar when a doctor takes into account not only standardised lab results but also information (of variable quality) on the patient's home situation (e.g. degree of family support for compliance with treatment).

Although improving consistency among teachers may be the aim of activities of social moderation for teacher judgement at levels one and two, another approach may be needed for teacher judgement at level three. This approach would seek to reduce major discrepancies between teachers, as well as bias founded on misconception and prejudice, but its focus would be more on the coherence and transparency of each teacher's judgement than on strict consistency between teachers. This would imply that each teacher is able to explain the rationale of his or her grading procedure to students, parents, and other teachers, and is willing to participate in critical discussions that foster development of well-grounded professional judgement in summative assessment. In this perspective, standardised external assessments and teachers' judgements are considered to be two different but complementary forms of assessment, each of which is legitimate in its own right, each of which needs to meet high, but non-identical, standards of quality.

### Acknowledgements

The study presented here was carried out by a team that included Linda Allal, Jamila Dorner, Lucie Mottier Lopez, Walther Tessaro, Edith Wegmuller (University of Geneva), Filippo Cattafi, Ariane Favre Marmet, Jean-Marc Hohl and Bernard Reidweg (Assessment Unit, Department of Primary Education, Geneva). We sincerely thank the teachers who participated in the study and the school system for allocating two half-days of release time to each teacher.

### Note

1. Differentiated equity has as its goal to promote learning and achievement at the highest possible level by all students. It can include elements of standardisation (e.g. all students take the same classroom tests under the same conditions), but uses qualitatively different sources of information that are appropriate for understanding and supporting each student's learning trajectory. In the French-language world of education, this conception has its roots in the work of the sociologist Bourdieu (1966) who wrote (our translation): '... by treating all learners, however unequal they may in fact be, as equal in rights and duties, the school system is led in fact to give its sanction to initial inequalities with respect to the culture [of schooling] ... Formal equality that regulates pedagogical practice serves in fact to mask and justify indifference regarding real inequalities with respect to schooling and the culture taught ...' (336–7).

### Notes on contributor

Linda Allal is professor emeritus in the Faculty of Psychology and Educational Sciences of the University of Geneva. Her research focuses on classroom assessment, on processes of co-regulation of learning and on writing activities in primary school.

### References

- Allal, L.K. 1988. Quantitative and qualitative components of teachers' evaluation strategies. *Teaching and Teacher Education* 4, no. 1: 41–51.

- Allal, L., and L. Mottier Lopez. 2009. Au coeur du jugement professionnel en évaluation: Des démarches de triangulation [At the heart of professional judgement: Procedures of triangulation]. *Les Dossiers des Sciences de l'Éducation* 22: 25–54.
- Black, P., C. Harrison, J. Hodgen, B. Marshall, and N. Serret. 2011. Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice* 18, no. 4: 451–69.
- Bourdieu, P. 1966. L'école conservatrice: L'inégalité sociale devant l'école et devant la culture [The conservative school: Social inequality with respect to schooling and with respect to culture]. *Revue française de sociologie* 3: 325–46.
- Brookhart, S.M. 1994. Teachers' grading: Theory and practice. *Applied Measurement in Education* 7, no. 2: 279–301.
- Cobb, P., K. Gravemeijer, E. Yackel, K. McClain, and J. Whitenack. 1997. Mathematizing and symbolizing: The emergence of chains of signification in one first-grade classroom. In *Situated cognition: Social, semiotic, and psychological perspectives*, ed. D. Kirshner and J.A. Whitson, 151–233. Mahwah, NJ: Laurence Erlbaum Associates.
- Harlen, W. 2005. Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal* 16, no. 2: 207–23.
- Kain, D.L. 1996. Looking beneath the surface: Teacher collaboration through the lens of grading practices. *Teachers College Record* 97, no. 4: 569–87.
- Klenowski, V., and L. Adie. 2009. Moderation as judgement practice: Reconciling system level accountability and local level practice. *Curriculum Perspectives* 29, no. 1: 10–28.
- Lafortune, L., and L. Allal, eds. 2008. *Jugement professionnel en évaluation: Pratiques enseignantes au Québec et à Genève* [Professional judgement in assessment: Teaching practices in Quebec and in Geneva]. Québec: Presses de l'Université du Québec.
- Miles, M.B., and A.M. Huberman. 1984. *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills, CA: Sage.
- Moss, P.A. 2003. Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice* 22, no. 4: 13–25.
- Scates, D.E. 1943. Differences between measurement criteria of pure scientists and of classroom teachers. *Journal of Educational Research* 37, no. 1: 1–13.
- Smith, J.K. 2003. Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice* 22, no. 4: 26–33.
- Weick, K.E., K.M. Sutcliffe, and D. Obstfeld. 2005. Organizing and the process of sense-making. *Organization Science* 16, no. 4: 409–21.
- Wenger, E. 1998. *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.
- White, S., and J. Stancombe. 2003. *Judgement in the health and welfare professions*. Maidenhead: Open University Press.
- Wyatt-Smith, C., V. Klenowski, and S. Gunn. 2010. The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice* 17, no. 1: 59–75.