

2nd International Conference on Higher Education Advances, HEAd'16, 21-23 June 2016,
València, Spain

Observing – Treating – Classifying: On the Educational Practice of Human Differentiation

Herbert Kalthoff^{a,*}

^a*Johannes Gutenberg-University Mainz, Institute of Sociology, Germany*

Abstract

Sociological research often states the reproductive function of educational systems. Thus, students are differentiated according to their cultural capital and their social background. One central dimension of the selection process is the grading performed by teachers. The paper outlines a conceptional framework of dealing with this kind of human differentiation sociologically. It therefore gives insight into the prerequisites provided by the school administration and into the empirical organization of grading – from classroom assessment to teacher meetings where final decisions are taken. Two forms of objectifying students' performances are presented: a social as well as a numerical-administrative objectification. The term 'social objectification' describes the orally conveyed assessment; the term 'numerical-administrative objectification' refers to the practice of translating student performance into mathematical and other symbols for further aggregation and editing.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the organizing committee of HEAd'16

Keywords: educational grading practices, human differentiation, social/administrative objectification.

1. Introduction

Modern societies organize the education of their offspring in specialized and diverse institutions which themselves are subject to public control. With the historical enforcement of school attendance, families are obliged to force their children to undergo the school curriculum up to a certain age. In Germany and a few other European countries, it is

* Corresponding author.

E-mail address: kalthoffh@uni-mainz.de

compulsory to get education outside of one's family in the specialized institution of a 'school' (compulsory schooling). This coercion to go to school is based on a deficiency hypothesis: The family has lost its ability to convey the increasingly refined knowledge inventories of the various disciplines and is therefore not able anymore to sufficiently qualify its children for the occupational system; therefore, a publicly organized and controlled institution is necessary, which – being placed between family and the job market – organizes the conveyance of knowledge and controls the acquisition of this knowledge through assessment (i.e. 'grades'). An alternative educational path is not provided for. Moreover, all children have to pass through the school path without exception – constant knowledge acquisition, grading, and differentiation. Furthermore, school attendance is supposed to take place well-regulated, i.e. in relatively homogeneous age groups, periodic rhythms (from one class level to the next), and with increasing difficulty (from simple to complex topics).

According to (educational) sociological conceptions, the institution of the school fulfills various functions, such as teaching and assessing the acquired positive expertise. The institutional assessment practice is sociologically interesting in that it conveys socially relevant knowledge about students through grades and school leaving certificates – a knowledge making students comparable with each other (within a subject) as well as themselves (over time) through school differentiation. The production of commensurability mainly takes place by way of numerous oral and written exams whose results are offset against each other and are documented in a ranking. An important aspect is the homogenization of the school population along age groups, which enables the school and the public school administration to operate with a twofold equality assumption:

First, public administrations ensure age-homogeneous of students by determining a biological age (six years) for their school enrollment. Before enrollment, the administration assesses whether the individual registered child has 'developed normally' and can be sent to school at all. This regulated and systematically registering enrollment of age groups – so goes the assumption – ensures relatively identical starting conditions for the children's educational careers. This equality assumption is based on the attributes 'biological age' and 'normality of development'; all other differences (gender, social or ethnic background, family socialization, body height, etc.) are disregarded, although the institution of the school is in fact being confronted with quite heterogeneous children, since their socialization experiences within their families are diverse and they go to school with various degrees of preparedness. In other words: The primacy of two characteristics corresponds with the *inhibition* of all other differences.

Second, all students undergo the same treatment program according to the iterative time of schooling and the standardized curriculum: The school path sends them from one class level to the next, from simple to complex topics. Therefore, all students hear, see, and experience the same educational topics and can thus be treated as equals in exams. This assumption corresponds with the container model of school teaching and implies permanent attention by the students, comprehensibility of the subject, and possibilities of learning through continuous interactive participation in class. Finally, the goal of exams and assessments is to present the supposedly equal students as *dissimilar*. The school principle to 'flag' fictitious equals as real unequals is in itself not a procedure that could fulfill standard quality criteria: Too inhomogeneous are the assessments between teachers, subjects, classes, and schools; too unequal are the chances of being moved or to finish school with a good graduation.

Sociology and social sciences do know little on "how teachers actually evaluate students" (Kingston 2001: 92). Therefore, the aim of the paper is twofold: Firstly, the paper gives an empirical insight into the grading practices in school as well as into the objectivation of teachers' evaluation. Secondly, it outlines some theoretical ideas by which these processes can be analyzed.

2. Contingency of Grading

There is agreement in assessment research that the teaching staff's verdict has an impact within and outside of school. It allows for a differentiation of the school population and to make decisions about the continuation of a student's school career, as well as making suggestions for students' professional post-educational careers. However, there is disagreement about the quality of teacher verdicts. Therefore, many publications concentrate on the quality criteria of school assessment – their objectivity, validity, and reliability. As Schreiber (1899) already critically asked at the end of the 19th century (sic!): Do school exams in their written and oral forms generate objective, valid, and reliable results? Afterwards, the concepts of "equality", "quality criteria", and "performance" dominated school assessment practices. Because already early publications raised the question: "How can performance ratings (...) be

generated through grades as objectively as possible?” (Lietzmann 1927: 46; own translation). This question was a reaction to three findings of US school scholars: (1) Several teaching staff do *not* correspond in the grading of one and the same written exam; (2) in a repeated assessment of the same exam text, a teacher does *not* get to the same grade as before; (3) the assessment of students by teaching staff is influenced by the local culture of each individual school (Starch/Elliott 1912; Lennes 1923; Shriner 1924). Up into the 1980s, research kept picking up the contingency of teacher verdicts by observing and criticizing it with experimental studies: Numerous investigations tried to show that the results of the educational grading procedure do not withstand the quality criteria of objectivity, reliability, and validity. Experimental combinations empirically substantiate the inexactness of the assessment means “grade”, with the teacher being the source of error and uncertainty factor. These studies largely confirm the findings – the “old complaint”, as Luhmann/Schorr (1988: 314; own translation) call it – from the first two decades of the 20th century.

The findings of assessment research (which cannot be displayed in detail here) impart an inconsistent picture. Roughly, two distinguishable theoretical frameworks can be identified: decontextualizing approaches and recontextualizing research. Decontextualization is based on the assumption that school exams are a grading practice in which the assessors are exchangeable tools that (might) get to the same and therefore valid results since these simply display the outcomes and therefore educational achievements. In other words: Decontextualizing assessment research assumes that at its core, school grading is an objective and reproducible, reliable and neutral method that is capable of principally generating the same assessment (e.g. Moss et al. 2006; Shay 2004). On the other hand, recontextualizing approaches (e.g. Filer/Pollard 2000; Torrance & Pryor 2008) take into account the practice of grading by which teachers differentiate their students. In contrast to the fictitious assumption that teacher verdicts just assess a student by measurement, these approaches are based on two premises: First, they suppose that the assessment process constitutes the student’s performance – and not vice versa (Luhmann/Schorr 1988). That is, the assessment is not preceded by an ‘achievement’, but is generated through the assessment itself. Thus, teaching staff is actively involved in the production of (correct) student replies – through respective oral hints in class, the construction of exam tasks, and the act of transforming student replies into (chargeable) credits. This involvement of teachers shows that they are tested, too, when ostensibly testing their students; for the teachers, there is also something at stake – that is, their reputation to be able to assess and adequately educate their class (Kalthoff 2013). Second, they assume that teacher verdicts (‘grades’) are principally contingent. In other words: Student performance (which is gathered orally in class and in writing in a test) is a socially attributed characteristic that can turn out differently depending on the local context and assessment culture. This means that grades do not necessarily reflect performance, but that they enable the summary of individual verdicts, their recalculation, and thus the differentiation of the students. With the contingency of the teacher’s verdict, performance objectivism and individualism are thwarted. This becomes particularly obvious in assessment situations in which explicitly other (i.e. social, cultural, ethic, etc.) differences are taken into account by teaching staff in order to render and justify their judgments.

3. The German Case

In terms of creating difference, the following can be stated about the German school system: By accepting all students, primary school collectivizes, while secondary school selects by choosing their students. Based on previous student performances, a division takes place: Usually, students proceed to the *Hauptschule* (“secondary modern school”), *Realschule* (“junior high school”), or *Gymnasium* (“high school”). This educational ‘league principle’ organizes different performance levels onto which the students are distributed. Thus, educational difference is partly determined by guiding students. The transition from primary to secondary schools is particularly coined by mutual selection processes: Parents choose their schools, and the schools choose their students and thus families. This *reciprocal casting* ends with the enrollment of students at a particular school and a distribution of students/families coordinated by the schools (such as the *Gymnasiums* in a city).

The usual course of a school career envisages a continuous, i.e. annual hike through the class levels: from fifth to sixth to seventh to eighth grade... On the part of the students, this continued perambulation presupposes the acknowledgement of performances which are in accordance with both the performance class (i.e. the respective school type) and the class level. These certificates lead to a permanent difference between the ones who made (or

didn't make) the transition and the ones who visited (or didn't visit) a particular school type. Thus, a change of school types is possible and provided for at various points in time. Furthermore, school career recommendations are made for the systematic transition points: at the end of primary school, after orientation level (at the end of sixth grade), after grade 10. This means that at different points in time, consultations take place and decisions are made on the appropriate school type for a student.

In school, the distinction of students is first determined by teaching staff along a list of differentiations by which teachers describe their students: good/bad; fast/slow; committed/not committed; discrete/requires assistance; comprehending/missing out; embracing/non-embracing; continuative/simple; thorough/negligent; etc. The common ground of these distinctions is the 'schoolability' of the child: A student is well-schoolable if she is performance-oriented, understands quickly, and is capable of formulating topic-oriented, advanced contributions, if she handles her materials with care etc., while at the same time showing social commitment (i.e. taking care of her classmates) and finally – all this presupposed – is smart on top of that: in terms of the topics in class as well as of the rules at school. For the assessment of performances, a rating scale is provided: The German grading system is as follows: "very good", "good", "satisfactory", "sufficient", "deficient" and "insufficient". The grade is expressed in Arabic numerals: "1" stands for "very good", "2" stands for "good", "3" stands for "satisfactory" etc.; "5" and "6" are failing grades. Teachers can differentiate using "plus" and "minus": a weak "very good" is a "1–", a strong "good" is a "2+" etc. In the second phase of secondary school a system of grade points (0–15) is additionally used.

Table 1. German Grading System

Grade	Insufficient			Deficient			Sufficient			Satisfactory			Good			Very Good		
	6	5–	5	5+	4–	4	4+	3–	3	3+	2–	2	2+	1–	1	1+		
Points	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		

Methodically, the grading system at school is a rating or ordinal scale; it measures differences in characteristic values (larger/smaller; better/worse; much/little; etc.) and displays them without determining a measuring unit. This means that the difference made visible by the scale – i.e. the question how strong or weak a characteristic is – is not defined exactly. Based on this scale, teachers determine the positioning of their students: The quality of student contributions (orally and in writing) is translated into credits which themselves are summed up and recalculated as (certificate) grades. This grade is thus the result of a multiple transformation: from the hardly trackable everyday routine in class along the results of written exams, documentation, and arithmetical summaries to the final result. The certificate grade itself makes a powerful appearance: All performances of a student are bundled in this single figure.

With regard to the structure of educational difference, the following dimensions (among others) can be identified: First, there are officially determined stages as well as unofficial moments of differentiation which are also recognizable; second, there are clear patterns which, however, are diffuse and imprecise; third, there is a clear classification into categories and schedules, i.e. that each student belongs to only one performance class and one ranking position per subject.

4. The process of objectivation

My research is based on these two premises, but I supplement them with two further observations: A partial oblivion of performances corresponds with the high assessment relevance in school. Even though teacher verdicts may be latent in daily class and can potentially be made relevant at any time, they are also subject to a loss in relevance; they wear out and thus lose significance. This loss in relevance points at another layer of meaning in school differentiation: Besides being an assigned social allocation function, it is firstly and primarily an internal operation that observes, selects, and distributes students for the school system. In other words: The practice of differentiation generates a distinction that can be tied in with and which allows for further differentiation – in order to produce differences which again can in turn be taken up and so forth; that is, school differentiation is a inbuilt operation and allows for external continuity operations only at specified points in time. Finally, the students are subject to a school-biographical cycle: Their time in primary school is affected by a training of communication and examination, and

secondary schools intensify this practice in order to imply and demand a stronger self-reliance in senior classes.

Based on these considerations, I conduct research – among other issues – on how school as an organization processes and frames the contingent verdicts of its teaching staff so as to legitimize and acknowledge them as objectified assessments. This takes place by way of various stabilizing, institutional procedures such as legal preconditions as well as the procession of grades in analogous or digital computing devices. Thus, I am not interested in the question about the quality of educational verdicts, but in the process of the objectification. One could ask: What does the school system do for the verdicts of its professional participants to become watertight? Second, I ask which type of human differentiation school pursues for it to generate people categories with its treatment program.

Our empirical research followed an ethnographic research design (cf. Emerson et al. 1995; Breidenstein et al. 2013). In 2014 – 2015, we made eight- to ten-week observation stays in various schools. An important starting point of our research was participant observation in lessons as a relevant reference point for the participants, which was closely connected to other stages (first writings, informal opinion-making) or from where other stages (exams, school report meetings) could be made accessible successively. Thus, we participated in lessons of a range of subjects and class levels, in informal conversations in the teachers' lounge, in formal consultations (school report meetings), in the preparation of written school-leaving exams, their execution and proofreading, as well as in the formulation of transition recommendations; we conducted numerous ethnographic and expert interviews with teachers and students and compiled relevant documents on this field.

Thus, teacher assessments are non-reliable verdicts, since any of them could always turn out differently – just like the review process in professional journals. However, school is a special case in that the contingency of an external ascription via classification through institutional procedures gets superimposed and tends to vanish. This becomes transparent in the institutional and legally framed procedures with which the educational institution gives material weight and impact to the contingent verdicts of its staff. In other words, the verdict of the teacher is objectified and thus reinforces differences. In doing so, it passes various stages of objectification:

(1) *Oral assessment during class*: The core element of communication in class is the communicative synchronization of the students. In other words: In order to be able to speak, there has to be an ability to keep silent (Hahn 1991). This rule – being silent in order to speak – establishes a separation of speakers who present their scholarly knowledge and an audience that observes (or doesn't observe) these performances. Throughout this, the communicative participation of the teacher remains stable, whereas the one of the students alters with the rhythm of changing speakers that is typical for class communication – which usually starts with the teacher asking questions. The sub-project interprets these repetitive teacher questions as a stage on which the class happening in content is performed as the solution to a puzzle. For the student, these brainteasers (“What is...?”, “When is...?”, “Who remembers...?”) contain two tasks to be solved: Which topic does the teacher's question refer to, and what is his/her point? Furthermore, the puzzle asks everybody to answer. Students who answer make themselves recognizable on this stage of class communication. They risk not having a sufficient understanding of the topic, nor may they realize what the teacher was up to; thus, their answer may not comply with the question. In their reactions to student responses, teachers orally assign knowledge (or ignorance) as commendation (or censure) as a further hint or part of an answer. This way, they mark both individualized students (knowing/not knowing) and the substance of the answer (correct/wrong). In other words: The students' individualization through the speaking procedure at school is a precondition as well as a consequence of the performative effect of the teacher comment: It describes presented knowledge and ascribes the speaker a position on the grades scale. It also becomes apparent that students expect a reaction to their answers, since they claim them to be correct. The execution of these communication rules is an important precondition for student differentiation: They allow not only for orderly communication, but also for the identification of students who act as bearers of 'achievements' with their contributions.

(2) *First textualization and documentation of difference*: We could observe how oral participation (as memorized by the teacher) was transformed into brief written protocols – pedagogical jottings. They consist of mathematic symbols (such as “+”, “↓”) and brief annotations (such as “Disturbs!”) which are written underneath the respective student's name – incl. the date. This practice of taking short notes serves the legitimization of an assessment, both for the teacher him-/herself and to the students, colleagues, and parents. It is highly selective since it is partly based on the teacher's correct memory, its completeness and repeated accomplishment. However, it also stands for an initial translation and thus dissociation of a student's performance from the assessing teacher and the oral context in

class. The written-down characters can be further processed, aggregated, and transformed into other symbols. The practice of ad-hoc assessments in class is followed by working on the grades list. We could observe that this initial textualization is carried out in various ways. Some teachers strive for regularly writing protocols or at times realizing this goal; others completely do without it. In the investigated schools and cases, we found a correlation between subjects and protocol practices: Teachers of languages, arts, or music practiced these records in a more committed way than teachers of natural sciences or mathematics tended to do without them. Thus, a pedagogical ambivalence can be asserted that potentially records every activity of a student, but weights and forgets to do so in practice.

(3) *Cooperative opinion-making*: We identified two forms of cooperative opinion-making: ad-hoc opinion-making in the teachers' lounge and institutionalized opinion-making in the school report meeting. Ad-hoc opinion-making often resembles a gossip-type conversation, with an invitation ("So?!"), a narrative (frequently trouble talk), and a pragmatic end that is connected to a recommendation ("Don't be too lenient!") (cf. Jefferson 1984). The assertion of these conversations is that everyone is dealing with the 'same' students who are (or were) taught in different subjects or grades. With their (non-)actions, (non-)expertise, and preferences, students are the continuous reference point of this type of teacher communication. It is still unclear whether these ad-hoc conversations are of any relevance that reaches beyond that very situation. In order to grasp this continuum of social ineffectiveness and consequence, I call this *reversible hardening*. This means that in the rhythmicized but uninterrupted continuation by teachers, these conversations slowly create perceptions of students that can be updated in teaching and assessment if teachers' lounge communication has been brought into class.

(4) *Written exams*: In the periodic time at school, the symptomatic daily review of knowledge through oral communication in class is repeatedly systemized in many subjects via written knowledge tests. We largely observed three models of marking – the parallel, the polar und the circular model. Teachers who use the *parallel model* of marking refer in their marking to an ideal test result (called "horizon of expectation") drawn up by themselves with which they assess the correctness of their students' answers. This orientation along 'correct answers' ignores the already accumulated assessments of students and claims to be neutral. In the assessment procedure, the ascribed number of credits is justified on the basis of the ideal test result, i.e. their own expectations. Teachers who apply the *polar model* work, too, with an ideal test but initially correct the tests of the assumed strongest and the weakest student. Firstly, by allegedly covering the whole grade width, these two tests reconfirm the teacher about having posed the right tasks. The assumed strongest student's test confirms that all tasks were solvable, while the weakest student's test confirms that they may have been challenging, but not too difficult. Secondly, correcting these two tests helps the teacher to get an impression of his students' overall performance in this test. Choosing this model thus basically serves the teacher's reconfirmation of his/her tuition as well as of the construction of test questions. The *circular method*, however, refers to student performances documented in the past. At the beginning of assessment, their tests are divided into three categories (good, medium, bad) and are then read, annotated, and assessed on the basis of this status. In either case – work with an ideal test, presorting according to former results, framing teachers' expectation by a polar constellation – , teachers reason their procedure by claiming to give grades that are in line with the individual student's performance in relation to his/her class. Without going into more empirical detail, we observed that teachers while marking the test of students shift from one model to another model, from one scale to another scale. That is to say that in front of written tests teachers can apply a 'case sensitive'-practice of marking which allow a grading that is, in their perspective, an 'objective' one (cf. Kalthoff 2013).

The assessment of a student's written performance is based on its documentation; it is a continuation of the annotation practice described for the "initial textualization" above. Teachers make grade lists for every subject. We could compile several grade lists and subjected them to a document analysis (e.g. Prior 2003). Four characteristic areas could be identified: the identification area with the students' names in an alphabetic order; the documentation area with the grades achieved with partial performances; the calculation area with the transformation of documented grades into a total number of credits with respect to the legally binding weights of partial performances; the specification area with the determined grade. The grade numbers entered into the documentation area are partly aggregated (e.g. the grade for oral participation), while other grades represent individual partial performances (e.g. the grade for a test). The step from the calculation to the specification area is not determined by a calculated total number of credits. Here, teachers secure their 'pedagogical freedom' by assigning different grades (such as "1.3" and "1.7") to students who achieved identical numerical totals. Afterwards, teachers enter their classes' grades into the central school software, may make last amendments, and sign the print-outs with their shorthand symbol. This signature

finishes the tentativeness of grades and thus their adjustability. The grades are now determined in a way that they can be negotiated in the school report meeting. This work on credits represents an independent, written reality in which grades have become a numerical object. The operative handling of individual grades generates further grade values which suggest one final grade for the school report. In this calculative practice, the numbers do not initially refer to other reference systems of meaning (such as the grades scale, normal distribution, performance record, school biography, state rankings, etc.), but only to themselves, i.e. to their arithmetic value which is symbolically manipulated by the predetermined conversion formulas. The grade in the school report does not result from the numerical operation, but from the interpretation of these value calculations.

(5) *School reports, final exams, and status passages*: Teachers repeatedly make decisions on their students' continuation of their school careers. This can be the case when deciding on moving a student up into the next class level. In this context, we investigated the obligatory school report meeting. In the cases we studied, it is a 'conference' without controversial discussions, without weighing up and consideration. It is rather a meeting in order to ratify verdicts (i.e. grades) that have already been determined and documented in writing. This refers to the grade that has been assessed prior to the school report meeting. There, teachers physically justify grades that have already been confirmed; objection is possible only within this small timeframe – these meetings often last only five to ten minutes –, whereas the cases (such as class 8c) are closed and blackboxed afterwards. After this ratification, which constitutes a 'hardening' of the grades, there is no room left for negotiations, discussions, or amendments. This also means that the school report meeting is the symbolic act of a shift: The school determines the grades in the school report – and not the respective subject teachers who are now integrated into the staff community with their assessments. This shift reveals the taxonomy of school objectification: In the case of school report and graduation, the institution of the school transforms the decision process to the top level – it actually collectivates it –, while the practice of marking has a downward effect: It individuates the student.

5. Conclusion

This paper sketches an answer to the question how sociological research can investigate the practice of academic assessment theoretically and empirically. The orientation along conventional quality criteria such as objectivity, reliability, validity is replaced by empirical observation of both school assessment practice and the objectification of the teacher judgment by the institution of the school. In this sense, the paper suggests to understand school assessment neither as an act of individual teachers nor as a neutral procedure of measuring performances, but as an organizational method that ontologically refers to the lesson, but veers away from it organizationally. The end of this grades communication is marked by a collective decision which ratifies the teacher's grade as an academic assessment. This concept of empirically studying the assessment practices at the stages of academic organizations has three implications:

1. From a situation of a mutual presence in which the participants see and hear, interact and deal with each others, one shift to an interaction in which their activities are documented and recounted selectively in their absence. One leaves the fleeting and barely controllable events and resort to situations of documentation and informal exchange.
2. One shift from an interaction in which participants represent their academic knowledge orally and in writing (to themselves as well as to others) to an interaction in which this knowledge presentation is evaluated by significant others in their absence – an extrinsic assessment that includes the teacher's own observation as a classifier. Here, a shift takes place from a social situation of school education that implies participation to a situation in which (non-)participation is used to differentiate the participants.
3. One shift from situations of individual assessment by individual teachers to situations in which these assessments are communicated and revised, transformed into software, calculated digitally, and are finally ratified organizationally. This means that one shift to situations in which individual assessments are made compatible with school procedures, their rules and deadlines, grade tables, agendas, and decisions. It is a shift into the collectivizing assessment by the school as an organization.

One can draw on a range of sociological theories to deal with these shifts more theoretically: According to system theory (e.g. Luhmann 1998), you shift from interaction to organization; according to interactionism (e.g. Goffman 1974), you move from one situation into another one with varying participants and constraints; according

to the actor-network theory (Latour 1987), you move from assessment situations to the calculative center of assessment in which other (non-)human actors are involved. With these and other theoretical lenses that complement and revise each other, the transsituative character of school assessments and their socio-material configuration can be described and analyzed in a way that focus the (organizational) practice of school grading and thus human differentiation.

References

- Breidenstein, G., Hirschauer, S., Kalthoff, H. & Nieswand, B. (2013): *Ethnografie. Die Praxis der Feldforschung*. Stuttgart: UVK.
- Emerson, R.M., Fretz, R.I. & Shaw, L.L. (1995). *Writing Ethnographic Fieldnotes*. Chicago: The University of Chicago Press.
- Filer, A. & Pollard, A. (2000). *The Social World of Pupil Assessment. Processes and Contexts of Primary Schooling*. London/New York: Continuum.
- Goffman, E. (1974). *Frame Analysis. An Essay on the Organization of Experience*. New York et al.: Harper & Row.
- Hahn, A. (1991). Rede- und Schweigeverbote. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 43(1), 86–105.
- Jeffersoon, G. (1984). On stepwise transition from talk about a trouble to inappropriately next-positioned matters: In: J.M. Atkinson & J. Heritage (Eds): *Structures of Social Action. Studies in Conversation Analysis (191-222)*. Cambridge: Cambridge University Press, S. 191-222.
- Kalthoff, H. (2013). Practices of Grading. An Ethnographic Study of Educational Assessment. *Ethnography and Education*, 8(1), 89–104.
- Kingston, P.W. (2001): The Unfulfilled Promise of Cultural Capital Theory. *Sociology of Education Extra Issue: Currents of Thought: Sociology of Education at the Dawn of the 21st Century*, 88–99.
- Latour, B. (1987). *Science in Action. How to follow scientists and engineers through society*. Cambridge: Harvard University Press.
- Lennes, N.J. (1923). *The Teaching of Arithmetic*. New York: Macmillan.
- Lietzmann, W. (1927). *Über die Beurteilung der Leistungen in der Schule. Mathematisches, Psychologisches, Pädagogisches*. Leipzig/Berlin: Teubner.
- Luhmann, N. & Schorr, K.-E. (1988). *Reflexionsprobleme im Erziehungssystem*. Frankfurt/Main: Suhrkamp.
- Luhmann, N. (1998). *Die Gesellschaft der Gesellschaft. Vol. 1 and 2*. Frankfurt/Main: Suhrkamp.
- Moss, P.A., Girard, B.J. & Haniford, L.C. (2006). Validity in Educational Assessment. *Review of Research in Education*, 30, 109–162.
- Prior, L. (2003). *Using Documents in Social Research*. New Delhi: Sage.
- Schreiber, H. (1899). Gegen Prüfungen und Noten. *Zeitschrift für Philosophie und Pädagogik*, 6, 31–38.
- Shay, S.B. (2004). The Assessment of Complex Performance: A Socially Situated Interpretive Act. *Harvard Educational Review*, 74, 307-329.
- Shriner, W.O. (1924). The reliability of teachers' marks. *Mathematical Teacher*, 17, 426–443.
- Starch, D. & Elliot E.C. (1912). Reliability of the Grading of High School Work in English. *School Review*, 20, 442–457.
- Torrance, H. & Pryor, J. (2008). The social construction of success and failure in classroom assessment in England. In H.-H. Krueger (Ed.): *Family, School, Youth Culture – International Perspectives of Pupil Research (219–237)*. Frankfurt/Main: Lang.