Human Studies
https://doi.org/10.1007/s10746-025-09824-y

ANALYSIS

The Evaluation of Performance in the Study of Engineering

Herbert Kalthoff[1] · **Kornelia Engert**[2]

Received: 6 March 2025 / Accepted: 18 November 2025

Abstract
Using engineering as an example, this article empirically examines the assessment of students at universities. To this end, it presents two phases of the university course: practicing and applying specialist knowledge in workshops and test laboratories, as well as presenting the results of group work. The evaluation following these two phases are analyzed as continuous and explicit forms of evaluation. Theoretically, the article takes a recognition theory perspective, replacing decontextualizing research through a contextualizing perspective, and argues that performance only comes about through the assessment and evaluation itself, that its outcome is contingent, and that it simultaneously has strong biographical effects.

Keywords Sociology of evaluation · Qualitative research of higher education · Training · Engineering sciences · Germany

Introduction

So far, the sociologies of scientific knowledge have analyzed research activities either in natural or social sciences. This paper turns the perspective to a somehow forgotten part of the professional academic field: the teaching of students in the respective subjects. In doing so, it conceives of university teaching as an institutionally organized program of systematic training of its students, endowing them with academic degrees and titles (compare Mannheim, 1935; Bourdieu, 1996). The concept

✉ Herbert Kalthoff
herbert.kalthoff@uni-mainz.de

Kornelia Engert
kornelia.engert@th-ab.de

[1] Johannes Gutenberg University Mainz, Mainz, Germany

[2] Aschaffenburg University of Applied Sciences, A-Burg, Germany

Published online: 23 December 2025

of training implies that students undergo an "epistemological enculturation" (Decker & Hoffmann, 2023: 403), which exposes them to instruction, assessment, and evaluation that is intended to shape their mindsets, perceptions, practices, and bodies. Even if everyday university life tends to hide or de-actualize this effect, its teaching, therefore, aims to (trans-)form students' subject-specific and cultural knowledge as well as their self-image. It is a process in which students are not only human objects of treatment by others (the teaching staff) but also subjects of their own education, as they actively participate in it. Within this context, students are ascribed individual worth and quality—values, that are assumed to be determined solely based on their performance.

This is where the paper comes in: using the example of engineering education at German universities, it provides empirical insight into various phases of academic teaching, examination, and assessment. In engineering sciences, teaching is characterized by a continuous alternation between cognitive forms of instruction (e.g., lectures or seminars) and practice-oriented exercises in basic techniques (e.g., learning to solder). University students acquire theoretical background knowledge, learn the thinking and calculation skills required for the subject, train their hand-eye coordination, learn to use objects for their intended purposes, learn to assess the quality of their own work, and finally undergo examination and assessment of their knowledge. In this way, the essay brings together perspectives from the sociology of education and the sociology of e/valuation. The latter explores practices of e/valuating (im)material goods, art, or institutions, while the sociology of education examines school teaching and student assessment but has tended to ignore assessment and evaluation of students in universities.[1] In the following, we will first outline the state of research and our theoretical considerations (2.). We then describe our methodological approach as well as our empirical research (3.). In the empirical analysis section, we focus, firstly, on the practical part of the study program, which takes place in workshops and test labs at the beginning and end of the program (4.). Secondly, we analyze oral examinations in which students' knowledge or projects are tested and evaluated, either in individual or group exams (5.) Finally, we outline the role of university instructions, exams, and evaluation and sum up our results (6.).

State of Research and Theoretical Perspective

The evaluation of people by actors who have an institutionally assigned power to do so is a phenomenon that can be found in many social areas, such as law, confession, medicine, and peer review in science (e.g., Hamann & Velarde, 2023). In schools, teaching, assessment, and evaluation of students are almost indispensable parts of everyday life.[2] Their accomplishment has been studied primarily by the sociology of education, investigating these practices in schools intensively. There are two main

[1] For a conceptualization of Social Studies of Teaching and Education, see Kalthoff (2011).

[2] The paper uses *assessment* to refer to formal situations of testing (for example, a written exam); and *evaluation* when the paper refers to situations of grading a written or oral exam (Wiggins, 1998; Krüger & Reinhart, 2016).

findings that made early studies well known: Firstly, different teachers do not arrive at the same result when assessing the same exam paper; secondly, when re-evaluating the same exam paper, a teacher does not award the same grade as before (Starch/Elliot, 1912; Lennes, 1923: 470 ff.). These results indicate that school evaluation cannot be reduced analytically to a representation of previously achieved performance and that evaluating must therefore be assumed to be contingent. For many years, further research has therefore focused on whether, and if so, how it is feasible to assess student performance as objectively and validly as possible, and how teachers' procedures should be understood. Without going into detail, it can be said that, from 1930 onwards, the evaluation process in schools has been scrutinized repeatedly.[3] In addition, research over the past 30 years has broadened its focus to include assessment and evaluation, taking a broader perspective on both the processes and social impacts of evaluation in schools, empirically as well as theoretically (e.g., Filer & Pollard, 2000; Zaborowski et al., 2011; Stommel, 2023; Kalthoff, 2019). This is how Kain (1996) sees grading practices embedded within the teacher community and teachers' cooperation. Brookhart (2004) expands on the idea of embedding and argues for a re-embedding of teachers' practices of grading within the local teaching and learning cultures, and Verkuyten (2000) analyzes the strategies of teachers to legitimize their grading and save their own face in teachers meetings. Experienced teachers strengthen their position by introducing their norms of grading and former decisions to novice ('young') teachers. While these studies focused on the teacher-teacher and teacher-learning culture relationship, Pryor/Torrance (2000) show how assessment is a result of the interplay between teachers and students. In an ethnographic study, Kalthoff (2013) specifies this point and shows that when teachers evaluate their students, they are also evaluating their own performance, i.e., their contribution to the students' performance.

In contrast, and despite some classical work, such as Bourdieu's (1996) – analyzing the treatment of *Élèves* in French elite universities as the continuous delivery of performance and its evaluation – little is known about the respective practices in academic subjects (Kingston, 2001: 92). Given the differences between school and university teaching, the outlined findings of educational research cannot be readily transferred to university teaching. The differences relate to the question of access to the institutions, the (prescribed) course of study, the way in which teaching, examinations, and evaluations are performed, and the form in which interaction and social relations between students and teachers/university staff take place. For example, German universities do not follow a fixed curriculum or a system of promotion to the next class, but offer subjects, some of which are very selective, others are organized in a school-like manner, and still others require self-guided learning (Kalthoff & Engert, 2024; Kalthoff & Kölsch, 2025). Depending on the subject, the relationship between university lecturers and students ranges from anonymous (visual) contact (in lectures) to comparatively fleeting encounters (in seminars) and more intensive individual relationships (when supervising theses or colloquia). A key difference between school teaching and university teaching lies in the communicative function

[3] Among other studies, see Eells (1930); Hartog and Rhodes (1934); Ulshöfer (1949), and Finlayson (1951).

of the teaching staff. While the role of schoolteachers is closely linked to the framing of oral communication as an organized search for errors or right answers[4] this situational, immediate feedback is rather frowned upon and suppressed in university seminars with their (fictitious) ideal of an exchange between (more or less) equal peers (Engert, 2022). In this context, the nature of negative feedback has been studied, emphasizing the role of laughter and 'failure talk' in securing the non-offensive character of academic teaching (Lymer et al., 2025).

Like schools, university evaluation is other-evaluation in the first place, and likewise the university staff award – behind closed doors – grades based on student work. But unlike schools, university professors might delegate the grading to their research assistants or to computer programs. The results are published in a written, de-individualized form of a list (e.g., for lectures, serial exercises) or in oral, individualized form (for term papers, oral final exams). University assessment practices thus vary between de-individualizing, standardized, and highly selective forms of assessments and individualized examinations, for which other attributions in the social dimension are asserted. As with the evaluation of student work in schools, we also assume that university evaluations are contingent. While self-observation has been empirically observed among teachers (see above), we assume that it does not play a very significant role in university teaching due to the high degree of anonymity of mass universities as well as to the high individualization of university staff (especially professors). The question of worth or quality of their students seems to play only a minor role for university staff in the first few semesters, while this attitude changes when it comes to final theses, master's degrees, or PhDs.

Although there are some studies on the conditions of university teaching (e.g., Felten, 2013), little research has been done on teaching practices in university seminars or workshops. With regard to the implementation of practical exercises, Lynch et al. (1983: 210 ff.) report on studies that document how "didactic experiments" in chemistry challenge students insofar as they do not have the physical routines for handling the objects and must first practice these in the course of the "experiment". As early as the 1970s, Atkinson/Delamont (1976) used the example of university teaching in the clinical phase of medical studies to demonstrate various strategies for combining teaching and practical contexts, such as bedside teaching, which is not part of patient treatment, or "waiting nights," during which students could observe real patient treatment and were partially involved in it. Current studies show, among other things, how the university teaching staff organize their courses in such a way that they try to select 'good' students, what attitude they expect from students, and how they moderate seminar discussions (compare Meyer et al., 2022). Greiffenhagen (2014) analyzes how teachers use the blackboard in a university mathematics lecture to mark steps in the proof and how the teacher's writing and speaking body interacts with the blackboard. It becomes clear how the slowness of writing on the blackboard

[4] This has been analyzed in numerous studies, see only for example McHoul (1978, 1990), Mehan (1974), already Simmel (1922: 64 ff.) and Gadamer (1960: 368 ff.). The paradoxical structure of the teacher question as a communication device of immanent referencing is staged in the question-answer-comment structure of the school lesson as follows: Those who ask the questions know the answers, and those who answer, but do not necessarily know the answers, do not ask the question to which they are supposed to know the answer (see, too, Lynch et al., 1983: 216 f.).

makes the proof cognitively comprehensible. Other studies have shown that school or university teaching relies on the active role of students, who in turn interpret and implement the instructions given by teachers (see Lynch & Jordan, 1995, Lynch & Lindwall, 2023). This also applies to ethnographic studies that analyze in detail the sequential organization and embodied work of teaching, for example, at the beginning and end of seminars (Eglin, 2009; Tyagunova & Greiffenhagen, 2017).

The few studies available on students' evaluations differ in terms of their theoretical approach and understanding of the subject under investigation. In the perspective of some studies following a measurement theory, university exams generate objective and valid results and thus accomplish a 1:1 representation of the students' performance (Ioannidou, 1997; Stumm et al., 2011; for an overview, see Nieminen et al., 2023). This research assumes that evaluators are neutral instances of this institutional procedure and thus somehow unmarked actors. Other studies dismiss this normative approach and focus more on the fabrication of grades, the practices, and artefacts involved, and the social effects related to it. It is shown, for example, how the self-understanding of university subjects as well as their (everyday) theories of student learning affect their teaching and evaluation practices (Tsarouha, 2019; Stratton & Mannix, 2005). In turn, these doings affect the self-image of students (Nieminen, 2021; Campbell, 2015) whose socio-structural background already influences the way they can interact with their professors (Jack, 2016). Other studies show that intensive exchange between students and tutors promotes socialization into academic culture and that (very) good students, too, can shape the studies of their fellow students (Ricken, 2011; Winston, 1999). Furthermore, it is shown how assessment formats (e.g., multiple-choice exams) or past test papers direct students' attention to (supposedly) important course content (Greving et al., 2023; Tyagunova, 2019). Finally, Forde-Leaves et al. (2023) argue that university examinations should not be understood solely as situational events, but rather as references to external factors. For example, university teachers legitimize their assessments with non-university discourses in which expectations regarding the quality of university assessments are articulated.

The research outlined here illustrates that research on the assessment and evaluation of students in universities either follows a decontextualizing normative approach of a neutral measurement or considers the local practices of evaluation and its circumstances and conditions. Decontextualization is linked to the idea that an exam is a measurement procedure in which those who measure are interchangeable instruments that (should) produce identical and therefore valid and objective results, as they merely reflect the performance of the students. From this perspective, the assessment (of a student's work, for example) seems to take place without the assessors themselves, as they merely objectively reflect the students' performance as it is. Thus, evaluation would just fix the worth and quality of a student's performance, a case purged of sociality. In contrast, a contextualizing approach assumes that no performance by the students exists as long as it has not been recognized through local practices of testing and assessment (Boud et al., 2018). Performance thus only emerges in the act of evaluation, namely through the recognition of students' oral statements or written work by the evaluators. In other words, through evaluative recognition, the act of evaluation transforms what students have said or written into a performance

that confirms or changes their status on the grading scale. For their part, students acquire their professional self-image through this symbolic recognition by an institutionally authorized other.

In this paper, we follow this contextualizing approach of performance with its recognition theoretical perspective (in detail see Honneth, 1995). First, we assume that the basis for evaluating work, practice, events, persons, etc. is its recognition by third parties who, in turn, are endowed with the symbolic power of an institution to do so. In the case of higher education, these evaluations are consequential, as they help determine the biographical futures of students. Secondly, we assume that performance is not an entity that precedes evaluation. Instead, performance is generated and simultaneously determined by the very act of an evaluation, i.e., by the (negative) recognition of the evaluators. The results of this process (such as grades, numerical scores, etc.) are highly contingent, which the evaluators document in their flexible use of standards (see Banerjee et al., 2025). That is to say that no performance precedes the evaluation, but evaluation constitutes performance, not the other way around (compare Luhmann/Schorr, 1988). Therefore, the paper empirically examines the training of students and the evaluation of their knowledge, as well as the evaluators' conception of a successful exam. In doing so, we assume that teaching and examination are interdependent, i.e., they relate to each other and form a unity, even though they are performed separately.

Methodological Remarks

The empirical database of this manuscript comprises participant observations, interviews, and the collection of field-specific documents related to teaching and assessment in Engineering. We adopt the notion of 'Engineering' as an umbrella term for a very heterogeneous field, which integrates many courses of study. Considered to be more structured and industry-oriented than other academic subjects, its aim is, above all, to impart and develop technical and applied knowledge. Core competencies include mathematical, scientific, technical, and economic expertise, which students must acquire and apply. Qualification theses are dedicated, for example, to mechanical Engineering, modeling of processes, questions of navigation, and planning of urban quarters; technical objects are created, but also planning drafts, which are the subject of evaluation.

We conducted field research at four (technical) universities across three Engineering fields ('Environment and Water,' 'Transport and Automotive,' and 'Architecture and Urban Planning'), participated in fifteen courses (lectures, project seminars, exercises, and laboratory projects), and twenty-six exams (written and oral presentations). In addition, we collected documents (Bogner et al., 2009) and conducted nine guided interviews (so-called semi-structured interviews) and twenty-one ethnographic interviews (Spradley, 1979) with lecturers from various status groups (professors, postdocs, etc.) on teaching, exam design, and assessment. When selecting the interviewees, we made sure to include representatives of different genders and national origins. Guided interviews (in German: *Leitfadeninterviews*) are understood as interviews conducted by social scientists themselves using a previously created

guideline; it represents a more formal, traditional interview situation. Such a guideline consists of a list of topics that are addressed in the interview, although the interviewee's statements must be continuously checked in terms of their relevance to the research interest; this includes potential deviation from the guideline (compare Hopf, 2016; Fontana & Frey, 2000). We conducted ethnographic interviews, understood as "friendly conversation" (Spradley, 1979: 55), before and after classes or exams, in the participants' offices, in the hallway during a coffee break, or in administrative offices. Ethnographic interviews, therefore, leverage the researcher's mobility and allow addressing their actual observations to the participants (compare Spradley, 1979). The interviews were transcribed, ethnographic protocols revised, and open-coded using the principles of Grounded Theory (Strauss, 1987; Emerson et al., 1995). Collected documents were also analyzed in terms of their design and impact (Atkinson & Coffey, 2011). We understand open coding as a process of translating empirical observations into analytical categories. Such an intensive, fine-grained analysis aims to overcome familiarity with the subject, and sociological explication is freed from the participants' practical concerns.

Practicing and Exercises

Within their studies, students in Engineering pass through a modularized track of different stages framed by the curriculum. They go from introductory lectures to advanced seminars, from practical exercises to project work, and from written examinations to oral presentations. These stages presuppose different spatial conditions, provide for distinct social forms of participation, and are designed as mandatory: the introductory lecture, an internship, and a final project. A lecture (with examination) organizes a different form of participation than a laboratory exercise in small groups; an introductory lecture bringing together various disciplines focuses more on general questions and topics; in contrast, an internship sorts students according to their disciplines, as it focuses on subject-specific tasks and exercises.

Through this shift from general instruction in lectures to exercises, students are confronted with the practical handling of material and technical aspects of the subject, which they are introduced to and trained in. We will show that this happens in different ways. Our first example refers to a practical exercise on 'Electrical Energy Conversion' (see Fig. 1). It is intended as an initial introduction to practical work for

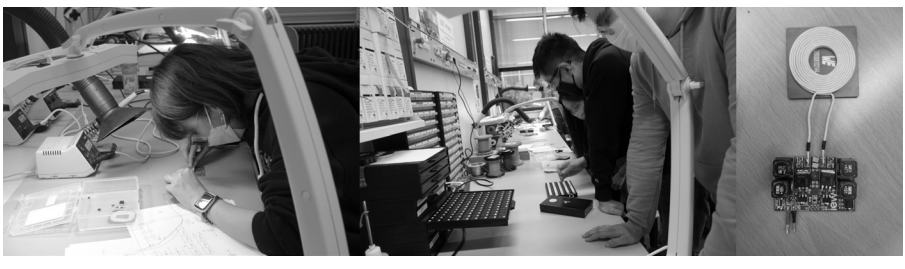

Fig. 1 Soldering workshop

bachelor's students and takes place in a room that is organized like a workshop. Each workbench is specifically equipped with a soldering device and required accessories, an LED light with a desk magnifier, solder fume extraction, and an electrostatic discharge pad. Other specific components (capacitors, transistors, diodes, wires, various cables, etc.) are sorted into drawers and folders.[5]

During their sessions, students are supposed to sit at the workbenches and to perform their task. During construction, the solder is repeatedly taken from a piece of wire and heated with the soldering iron. All the components are fused with the contacts of the board so that they permanently bond as they cool down. Soldering calmly and quickly and using the right amount of tin is key. Long exposure to heat might destroy semiconductor components, and disturbances during cooling can lead to 'cold' solder joints that break open easily ('loose contact'). Some special components (e.g., SMD ceramic capacitors) are so tiny that they have to be positioned with tweezers. After each mounting, the soldering iron is wiped wet (sponge) or dry (brass sponge). In between, the instructor assists in selecting the required components and refers to a prototype model.

Even if they approach the workstations intuitively, their, at times, careless handling of the materials and their hunched and tense postures account for their novice status. Their approach to soldering is casually tested to see if they can read the graphemic signs (the circuit diagram) and use them for their material practice. As they continue soldering, they repeatedly refer to a diagram showing the circuit board. This serves as a visual reference for the desired result, without showing how this can be achieved in detail. Later, the instructor comments upon inspection that some solder joints 'don't look good.' Some components have been damaged by too much heat. Additionally, two solder joints are too close and will lead to a short circuit. As another student checks her own circuit board with the oscilloscope, the desired curve is not displayed at first. After examining it, the instructor finds that a diode is in the wrong position: the anode and cathode must be swapped. In another case, further troubleshooting follows: One student must replace almost all major components before the voltage amplitude takes the expected course. Even though all students were ultimately able to complete their circuit boards, the cases observed nevertheless show that engineering science places value on a practice whose practical significance has been lost considering the automation of soldering. In other words: Students thus have experienced a symbolic lesson of their subject holding a self-made component in their hands.[6] For their part, the teaching staff accompanies, assists, and supports these (introductory) practices and thus trains students in their competences. However, even the lecturer is not always able to identify problems but merely narrows down where to look. The attribution of (non-)ability is often de-dramatized: not only the student activity but also the material is susceptible to failure. Therefore, the essential parts of the exercises are repairs and external verification, either by the teaching staff

[5] Energy is transferred without contact via two superimposed coils and AC voltage is converted into DC voltage via the secondary coil.

[6] Soldering is not only an (elementary) practice but constitutes an imaginatively charged setting of hardware Engineering, rooted in powerful identity narratives (e.g., the Apple garage in which Steve Wozniak soldered the first Mac prototype; see Fuller, 2015).

or by instruments (here: the oscilloscope). This is essentially the second lesson of the soldering workshop: an exercise in identifying (technical) problems, dealing with them, and solving them.

For our second empirical case, we move to another subdiscipline, namely urban water management. The topics covered in the lecture are supplemented by practical exercises in a so-called test lab.[7] These test labs, some of which are very large, are equipped with a variety of instruments (including hydraulic flumes, flumes for precision measurements and sediment transport, etc.) where students can test their knowledge in practice. In the cases we observed, small groups of Master students carried out exercises on thematically structured test setups (such as the "Fluid Friction Apparatus"). These setups visualize flow characteristics and allow students to observe them, measure flow resistance, regulate diameters, etc. The exercises are designed for Master students, working out their findings in working groups and presenting them together in oral examinations. We now analyze how students are trained in the use of the "Fluid Friction Apparatus" (FFA) under the supervision and precise guidance of a tutor. Under this guided instruction, they handle materials, substances, and media: water (in the pipes), formulas (on the blackboard), spreadsheets (on the computer), and tutorials (on paper) (Fig. 2).

The FFA is commercially available and permanently installed on a wall in the test lab. It consists of several horizontally and parallel arranged pipes whose inner surface changes the water flow due to its rough or smooth surface. To prepare their

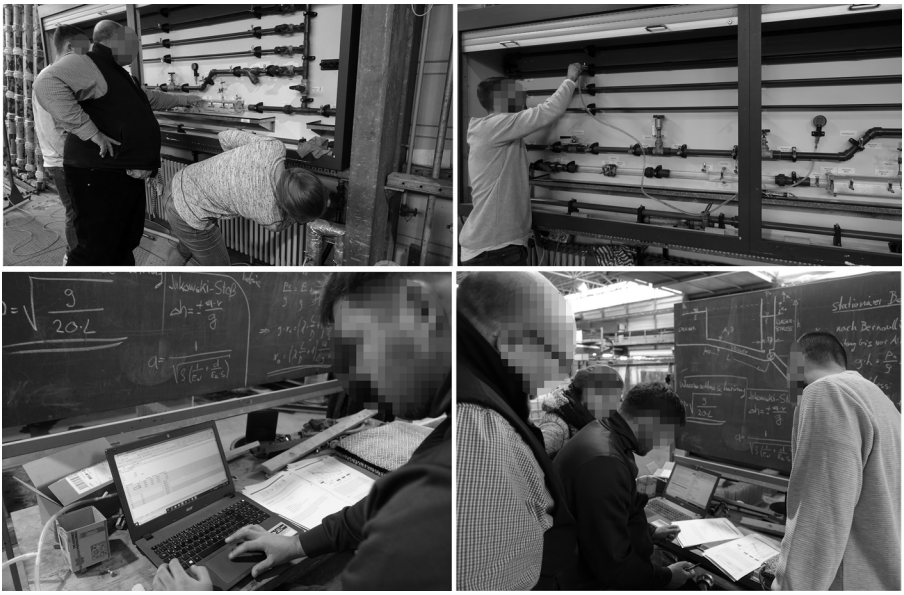

Fig. 2 Urban water management

[7] The participants we observed referred to it as a research lab. We prefer the term test lab to indicate that it is not researchers who are conducting experiments, but novices who are performing written and oral exercises under "a guided discovery instruction" (Atkinson & Delamont, 1976: 133).

exercises, students use a 23-page printed script that explains the FFA's test setup and how it works and sets out tasks. One task, for example, was to determine the friction losses and flow rates of two different types of pipes. That is to say that these differences in flow should be observed, measured, and calculated by the student group. But before the test run can begin, the instructor explains, among other things, how the experiment works: explaining where the inlet is, that the pumps must be turned on and the valves opened, that the pipes must be free of air, and where the water leaves the system again. Even when the group begins its work to "get the experiment up and running," the instructor participates from time to time in two ways: First, he tells the students what to do and what to pay attention to; second, he tests their propositional knowledge ("What is a millibar?"). While soldering requires individual performance, the water management case shows that the tutor provides students with more support and guidance up to a certain point: an intensive briefing with the aim of ensuring the test run is successful. An extract of the dialogue:[8]

T: … I would still open the other valves first to see if any air is coming out. Here, for example, there's definitely no air coming out, right? Water is coming out.

S: But isn't air going in there …?

T: No, there's already overpressure there. Yes, otherwise it would be stupid if air was still coming in up there, but there's already excess pressure (…) If I blow into it now ((blows into the pipe)), then of course it goes up. It shows on the ((board)) to three decimal places, but what you see there is bar.

S: Mhh.

T: Meaning? Accuracy? It's still fluctuating a bit here, but let's not get confused, there's still a bit of air in the pipe.

S: It's one millibar.

T: One millibar? Meaning? If we convert it to height?

S: I can't do that anymore.

T: It's important, I need it here, because they want to calculate loss HEIGHTS. ….

The excerpt illustrates how the tutor draws the students' attention to various aspects (such as the air being released from the pipes) or highlights meanings (such as in the case of conversions). It also shows how he clarifies questions or uncertainties raised by the stu-

[8] In this paper, we use the following transcription signs: T: tutor; S: student; E: examiner; A: assessor; (…): omissions; ((board)): comment from us; HEIGHTS: loud and emphasized pronunciation.

dents by referring to what is happening in the pipe ("overpressure"). After this detailed briefing, the students begin their test run. They open valves, observe the water flow, leaf through the script, take measurements (e.g., the temperature of the water flowing into a tub), enter them into their electronic spreadsheet, and write formulas (for calculating the loss height, for example) and draw diagrams (e.g., on resistance behavior for hydraulically smooth or rough surfaces) on the blackboard. On the one hand, every now and then, the students turn to the tutor with questions.

S: … Is it enough if we measure it ((the temperature)) once?

T: No, I wouldn't do that. I would always check it every now and then, when there's a brief pause. I would just take another look and measure it. Does it remain constant?

S: Okay, okay.

….

S: I was just wondering, what if it's open ((the pipes))?

T: Then turn it off (…) Well, it's only a little water.

S: Yeah, sure.

It becomes clear here that exercises not only make the instruments tangible and demonstrate what it means to work in engineering, but also that the objects themselves (in this case, the shape of the flow, the pipes, etc.) and their observation raise questions that students seek answers or clarification for. On the other hand, it is the tutor who, throughout the test, provides students with guidance by asking questions. An example:

T: … What does the MID say now? (Magnetic Inductive Flow Meter).

S1: Oh, I see.

S2: (reads) 0.58.

T: Let's start with 0.3 or so ….

Even though the students developed a diagram and performed calculations using the values they determined, their test remains a tutor-assisted test; in other words, practicing under guidance. Therefore, these exercises in the hydraulic test lab exhibit different practical competencies of the students as well as their willingness and ability to act cooperatively. These embodied competencies imply a disciplinary view and practical understanding of the discipline. For the case of the hydraulic lab, a sense of measurement is required that stands for a *form of empiricism* that may contradict

or irritate theoretical models. Dealing with variation in measurements, in particular, demands experience as well as qualified judgment and processing skills on behalf of the students. In other words, the treatment program of the lab puts the students to the test, trains their technical skills, and develops their ability to cooperate in the search for solutions and their individual efforts.

Finally, the lab sessions also facilitate periods of time in which students' ways of practicing Engineering are observed. The tutor explains:

> "… For example, the two ladies ((looks at a working group in the lab internship)) have helped out before, and that's a great thing. A lot sticks with them, and then they already know how to handle such a device. You can also see this HERE: they are very structured. Quickly, they measured their results, asked a few more specific questions, and then said: See you then! ….

A distinction is made between students who have had experience and those who have not; those who have learned to handle devices adequately and those who cannot (yet) do so; and those who work in a structured way, ask specific questions, and finish quickly, and those students who cannot (yet) do so. A selected group of students is categorized as those who can be 'used' because they need little help: their way of working and their performance indicate preparation and personal effort, which sets them apart from other students.[9] In contrast to the de-individualized lectures, students clearly show their skills through their handling of the materials and tools.

Oral Exams and Deliberative Evaluation

This section analyzes oral exams and the subsequent evaluation by the university teaching staff. Our paper thus moves from analyzing teaching as a form of incidental commenting, correcting, and evaluating of students' work to analyzing situations of explicit assessment and evaluation. Up to this point, it has become clear that the study of Engineering implies practical exercises in workshops and test labs. They represent an introduction into the practical fundamentals of the subject, but also the distinction between technical objects and substances as well as the human participants who use or process them. In the following, we document and analyze three oral exams and deliberation of the grades. Our first case involves an oral exam that students take to complete a lecture they have attended. However, only those students who have successfully completed a report on a selected topic from the lecture are admitted to this oral exam. The situation is performed as follows:

> *Oral exam: German waterways (extract)*
>
> T: … My first question for you is: name some typical elements of German waterways.

[9] It goes without saying that these exercises are also used in test laboratories to observe students and select the (very) best ones for other tasks. We will not explore this aspect further in this essay.

S: Okay, on the one hand there are ports where ships can dock and unload their cargo. Then there are locks for regulating water levels and overcoming differences in elevation. Canals can also be included, as they are man-made. Yes, basically those three groups.

T: Okay, that fits. Then I have a photo with me. (…) The question is to name the components of the facility that can be seen in the photo.

S: Okay, so you can see a lock. I assume that's the lower head (*Unterhaupt*). This part up here looks higher. That's the water level. That's a lift gate ((*Hubtor*)).

T: Correct.

S: And, yes. That's basically the guide wall that leads to the lower head.

T: What do you call this structure here, where the lift gate is basically suspended and guided?

S: Hmm, is that the mask ((*Maske or Quermauer*))?

T: Correct, great, perfect. That's the mask. ….

Two features characterize this oral exam, which is only documented in part here: the question/task-answer structure and the immediate feedback on the quality of the answer by the examiner ("okay," "correct," "great"). From a conversational analysis perspective, the feedback corresponds to the third turn – the comment. In school exams, it has been observed that teachers hold back their comments and thus leave the relevance of the statements unmarked (compare Zaborowski et al., 2011). In the present case, the examiner deviates from this school practice and immediately communicates the adequacy of the student's answer. In doing so, he refrains from ambiguous conversation, which initially withholds any hint of judgment. At the same time, these comments give the student an idea of how her answers are valued by the examiner and what grade she can expect.

After the exam, the student must leave the room. This small, almost self-evident action creates a new constellation with two different characteristics: *first*, those who are evaluated are kept out, and therefore, the asymmetry between the evaluators and the evaluated is reinforced. *Second*, this evaluation differs from the ad hoc comments made during the question-answer part of the exam, even though they do have a social impact (see above). *Third*, when discussing grades, it is not only the correctness of the answer that is evaluated, but also the way in which students formulate them. In the present case, the examiner (E) and assessor (A) discuss how they want to evaluate the student's answers. To this end, they mutually confirm that this oral examination is a clear-cut case:

Deliberation: Nice exam (extract)

A: … Then we don't need to talk for long, right?

E: We don't need to talk for long. That was—.

A: I'd say that's a 1.0.

E: Yes, a 1.0, yes. We don't need to talk for long, he knew everything.

A: Great. He did really well. (…)

E: That was a direct hit.

A: Very nice.

E: Yes, what are we going to do now?

A: ((laughs))

E: Nice exam, he knew everything. (…)

A: Yes, he just knew everything. ….

The extract of the deliberation shows that the actors repeatedly assure each other that they will not need much time to determine the grade, that the student knew everything, and that the exam therefore met aesthetic criteria ("nice"). Their problem is that they allow time to pass because it is assumed that the determination of a grade requires more time since various aspects must be considered. In this case, the actors shorten the process by agreeing in advance on the clarity of the case. This means that the time required to determine the grade legitimizes the examiners' judgment. The fact that they can agree on this is also due to the fact that the examiner has already framed the grade deliberation through his third-turn comments (see above).

The second case takes place in the context of student projects (see Sect. 3) that address topics or issues from the 'real world out there.' Students then analyze local conditions and develop, present, and justify possible measures or concepts. This is also where academic staff come into play in different ways. On the one hand, they evaluate student projects in whose conceptualization they were involved with feedback and recommendations. On the other hand, they can disagree on how they value different aspects of each concept. During the term, the interim status of the students' designs is presented and discussed in the form of intensive sessions of 'correction' (participants) in which the teaching staff adjust the designs to their own expertise. In this way, their approach is also included in the drafts (Fig. 3).

The teaching staff sitting in the front row like a "jury" (participants) want to convey a sense of how professional award juries or experts evaluate design drafts outside the university teaching context. The subsequent rows are occupied by stu-

Fig. 3 The presentation

dents, who will take turns presenting their work. What is special about this evaluation constellation with evaluators, examinees, and audience (Waibel et al., 2021) are the planned role changes. Some participants switch from the role of audience to the role of examinee, while others switch from the role of observing university teaching staff to the role of the main mentor. Even though the final grades are settled afterward, the students witness a variety of comments from the teaching staff (see above). These range from praise ("strong concept") to relativizing comments ("better leave out this hand-drawing") and detailed criticism ("well, the nursery looks a bit puny") to the disqualification of larger parts of the design ("basically a dismal spatial situation, really bad"). These evaluations are not only directed at the students, but they can also be reactions to some colleague's judgment ("With all the praise so far, I would like to emphasize these critical points"). Students can also take a stand on the criticism or clarify something ("I think there is a misunderstanding here").

After the public discussion, the jurors withdraw after three to four presentations each and privately come up with their grade proposals. This deliberation covers both the (special) features of the project work and its oral presentation by the students. It is observed whether they are able to communicate the purpose of their project work and whether the evaluators are also given a chance to speak. Using the following example, we will analyze how the project work of student groups is compared and evaluated.

Peter: … in direct comparison, I tend to have better grades for the other groups now.

Marco: I would also suggest a 2.0 for A01, definitely.

Sophie: These scissors? Well, I would think this is a 2.3. I had a really hard time with these scissors that ended up in this well.

Marco: (*Directed to Leana*) You supervised the first proposal. What is your suggestion for A01?

Leana: Well, I'm having a hard time with this one. I think they really put in the work; they came up with a wholly remodeled concept, and of course, this was a lot of work because they started from scratch again. It's their problem that they were worse before, but anyway, I think they've improved a lot. I don't know. When I compare them to B02, how strongly may one take that into account? (…) Of course, I think their concept is better than B02's, but whether this suffices to say that they receive a 2.0 in the end, I'm not sure.

Peter: We had zero-minus in there. (*Looking at a sheet with preliminary results*)

Leana: Yeah, that's why.

Peter: But they revised it again.

Leana: They set it up completely differently, yes.

Marco: Leana, now at level A01, is the proposal for your group 2.0 or 2.3?

Leana: Well, it must be 2.3.

Marco: Really?

Leana: Yeah, the concept is good. I just liked it with this tension. Of course, there are also weaknesses. But the set-up was good, except for the corner with the single-family houses. We had a lot of discussions about how to implement the new plots, but yeah.

Marco: Well, then it's 2.3. It's your group. ….

Two of the supervisors (Peter and Marco) communicated that they would give a 2.0 to the group presentation from that morning (A01). Another colleague (Sophie) is still critical of this draft, which she calls 'scissors,' and states that she is sticking to her suggestion of a 2.3.[10] The right to formulate the first suggestion for a grade belongs to the supervisor of the A01 group (Leana), who first explains the cooperation and the late progress of the group. She attributes the poor initial performance to the group and questions whether a 2.0 may be justified. It becomes clear that two colleagues (Leana and Sophie) form an evaluation pair or an "opposition alliance" (Hirschauer, 2015: 51) insofar as Leana essentially follows her colleague's sugges-

[10] Hirschauer (2015) provides empirical insights into the discussion and decision-making strategies of a science journal, which bear a certain resemblance to the evaluation process discussed here. In the case of the journal, it becomes clear, among other things, how the editors evaluate submitted manuscripts in very different ways, support each other's arguments, reject objections to a manuscript, or align their efforts with the predictable decision.

tion and is able to secure this evaluation to the surprise of her colleagues ("Really?"), who suggested upping the grade. In addition, it becomes clear that Leana defends her judgment with further explanatory statements, while other colleagues opt for a grade forthrightly (and may only justify when necessary). In this way, however, she also stands as a judge who does not identify with those she is judging but who tries to be 'fair' and independent, even though the small difference she makes is contested among the students. Finally, consensus is achieved through shortcut procedures, such as the right of the first proposal or evaluation pairs. In the case discussed here, the absolute grade (2.3) also represents the lowest grade achieved in this course. On the one hand, such locally pronounced non-exploitation of the evaluation scale de-differentiates evaluation and feeds all actors—university, teaching staff, students, families, etc.—with the illusion that there are only good students in this university course of that university, and thus only good, proven teaching by the teaching staff. On the other hand, it increases the symbolic weight of rather subtle grade differences. In the end, the symbolic distance between 1.0 and 2.3 appears worlds apart.

In our third case, we document how different assessments are handled, what role disappointment with student behavior plays, and how comparisons with other students factor in. An extract from our minutes:

> After the oral exam, the co-chair is surprised at how much the student knew. However, he also adds that without her index cards, the candidate gave the impression of being "lost". The examiner notes that the student was quite "stuck" at one point, which was also evident from her voice. The co-chair then points out that he nevertheless gave her mostly "ac" or "a" comments, which adds up to a grade of 1.6 for the question section. The examiner interpreted that she did not ask difficult questions but rather formulated questions in such a way that the answers were practically predetermined. Both the co-chair and the examiner agree that the presentation should be graded between 2.0 and 2.3. The examiner tends toward a 2.3 but does not want to be unfair. She explains that she had already criticized the calculation in the interim report, but that it had not been corrected. She interprets this as a lack of understanding on the part of the candidate, as she only understood today that the amount must be deducted from the expiration date. Furthermore, she draws a comparison to a previous candidate who simply did not finish drawing the potential network. In her opinion, this is just as serious, and she wonders why she is bothering to give the students feedback if it is not considered. On this basis, they award a grade of 2.3. Finally, the examiner remarks in a low voice that the candidate should play less with her cell phone. She had already attracted attention during the lecture.

The comments made by university teaching staff are revealing in three respects: the examiner's surprise, the comparison, and the differences in grading. As the extract shows, the examiner is surprised by the student's presentation. What is surprising is not so much her being surprised, but the fact that this (positive) emotional reaction does not need to be justified and, at the same time, communicates an (implicit) expectation that indicates a grade value. Furthermore, the comparison: as a rule, examinees are compared with others or with themselves. In this case, the examiner draws a

comparison by pointing out missing work or a lack of accuracy: an inaccurate calculation in one case is to be evaluated in the same way as an incompletely drawn network in another case. Furthermore, the case illustrates that the grade calculated by the co-chair based on his use of the evaluation sheet provided by the university organization can itself become the subject of discussion and clarification. The examiner's comments are crucial here, as they document disagreement with the co-chair's assessment. First, her exam questions had directed and enabled the students' answers; second, the group to which the student belonged had not understood the subject matter well; third, the student had used aids ('cards') and had not always argued clearly. These deliberations about the examinee's knowledge are supplemented in the social dimension by inappropriate media behavior: the student is subsequently criticized for her behavior in the lecture. This means that the examiner uses her arguments to correct the calculated grade and, thus, the entries in the evaluation sheet—downwards.

Finally, the evaluation of students by the lecturers who taught them in courses is often the rule. Therefore, lecturers may be surprised by a student research paper or presentation; accordingly, their questions may also be adapted to the perceived level of the student. In addition, the process of grading can be accelerated, shortened, or determined by an explicit right of proposal on the part of the examiner or by alliances between teachers, regardless of whether university administration forms categorize student answers, assign numerical values, and make the final grade calculable. As shown, this type of numerical objectivation is framed and, if necessary, corrected by its social counterpart.

Conclusion

This article examined the practice of assessment in engineering sciences at German universities. It focused on two forms of assessment: firstly, ad hoc correction and assessment in the context of university teaching, and secondly, explicit assessment in the context of oral examinations. With regard to the first form, it should be noted that, although it can be observed repeatedly, it has no immediate consequences for the students. It does occur, and it creates mental images in the minds of teachers of good and less good, prepared and unprepared students, without these impressions being documented in writing and systematically incorporated into further assessment, but they can be mobilized in oral final examinations. These more or less fleeting encounters and impressions do not remove the anonymity that is characteristic of university teaching. The second form we examined takes place repeatedly in a predefined period, namely at the end of the semester, i.e., it concludes the lecture or project seminar that takes students to the workshops and experimental laboratories of the universities for practical exercises, among other things, where the paper observed the first form of assessment as an element of teaching. The hallmark of this second form is its multiple asymmetry, which is inscribed in the functional roles of the examination system: subjects of assessment vs. objects of assessment; questioners vs. respondents; documenters vs. non-documenters; inclusion vs. exclusion from the assessment process.

What becomes clear here is the idea, deeply rooted in European education, of linking the cognitive-practical instruction of students with the assessment of their knowledge. Education and its function of changing people seem conceivable and possible only in this combination of instruction and assessment; the prerequisite for this is examination (oral or written). In other words, teaching and assessment form a referential relationship, because what is covered and worked on in lectures or project seminars is later systematically examined in the exam, and the knowledge that will be tested in the exam is taught, practiced, tested, etc., in the courses in preparation for the exam. In terms of the factual dimension, reference stands for a reciprocal presence: courses are present in the examination, and the examination is present in the teaching. In terms of the temporal dimension, this means replacing the linear view of teaching and assessment as a simple chronology of events with a dialectical relationship, namely as a dialectical practice of temporal references and recursions.

Furthermore, the article analyzed the practices involved in this context of reference, firstly using the example of practical exercises and project-based experimental setups, and secondly using the example of oral examinations. Among other things, it became clear how students are introduced to practical exercises, observed, and corrected; it also became clear how examiners in oral examinations mobilize (communicative) resources and (disappointing) experiences with students to justify their evaluation, and how they attribute successful examinations to the knowledge of the students and less convincing work and answers to the students' lack of willingness and ability, but not to their questions or tasks, hints, or moods. Different, competing judgments indicate that the evaluation of students is a controversial issue. For the participants, these are routine differences over nuances and details. This heterogeneity of judgment is their professional normality and not an exception.[11]

In this context, grades do not only stand for themselves and provide information about the performance of students and their projects, but also about the performance of a course and thus also about teaching. Three levels were identified in these evaluations: first, the evaluators' involvement in the subject of the evaluation (for example, in the project work of student working groups); secondly, strategies of distancing oneself from this involvement, which makes the assessment and evaluation appear objective; thirdly, the self-evaluation of the degree program through a very selective use of the evaluation scale. This means that the actors involved control and coordinate the process via these dimensions, namely professional involvement, consensual distancing, and institutional self-presentation.

Finally, the article provides clues to the development or formation of an engineering perspective on the world. This formation can be attributed both to all the cognitive and practical exercises that show students how they can shape the social world with material objects, among other things, and to their examination and evaluation itself. For the symbolic and performative power of evaluation repeatedly documents the future status of novice engineers: it shows them what they can and cannot do, what they can and cannot become, and what can and cannot be expected of them. This attribution of status is at the core of this institutional form of teaching young

[11] A structural similaritiy can be observed in peer review, see Greiffenhagen (2024), Hirschauer (2015).

people and the classification of their (lack of) success in the form of a ranking on the grading scale.

Author Contributions H.K. wrote most of the manuscript K.E. prepared Figs. 1 and 2, 3. K.E. did the empirical research All authors reviewed the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The ethnographic data that support the findings of this study cannot be made generally accessible in order to protect the participants.

Declarations

Competing Interests The authors declare no competing interests.

References

Atkinson, P., & Coffey, A. (2011). Analysing documentary realities. In D. Silverman (Ed.), *Qualitative research. Theory, method and practice* (pp. 56–75). Sage.

Atkinson, P., & Delamont, S. (1976). Mock-ups and cock-ups. The stage-management of guided discovery instruction. In M. Hammersley, & P. Woods (Eds.), *The process of schooling. A sociological reader* (pp. 56–75). Routledge & Kegan Paul.

Banerjee, M., Feindt, G., Kalthoff, H., & Wihstutz, B. (2025). Leistung im Vergleich. Objektivierungen trotz Differenzierungskontingenz. In S. Hirschauer, P. Hofmann, A. Friedrichs, & G. Schabacher (Eds.), *Humandifferenzierung im Vergleich* (pp. 236–266). Velbrück Wissenschaft.

Bogner, A., Littig, B., & Menz, W. (2009). *Interviewing experts*. Springer.

Boud, D., Dawson, P., Bearman, M., Bennett, S., Joughin, G., & Molloy, E. (2018). Reframing assessment research: Through a practice perspective. *Studies in Higher Education*, *43*(7), 1107–1118.

Bourdieu, P. (1996). *The state nobility. Elite schools in the field of power*. Stanford University Press.

Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record*, *106*(3), 429–458.

Campbell, T. (2015). Stereotyped at seven? Biases in teacher judgement of pupils' ability and attainment. *Journal of Social Policy*, *44*(3), 517–547.

Decker, K., & Hoffmann, C. (2023). Training scenes. Taking science studies to the classroom. *Social Studies of Science*, *53*(3), 402–426.

Eells, W. C. (1930). Reliability of repeated evaluation of essay type examinations. *Journal of Educational Psychology*, *21*(1), 48–52.

Eglin, P. (2009). What do we do wednesday? On beginning the class as university-specific work: A preliminary study. *Canadian Review of Sociology*, *46*(1), 39–57.

Emerson, R., Fretz, R., & Shaw, L. (1995). *Writing ethnographic fieldnotes*. The University of Chicago Press.

Engert, K. (2022). Instruktion und ihre Teilnehmer: Was ist in einem Seminar? In D. Meyer, J. Reuter, & O. Berli (Eds.), *Ethnografie der Hochschule. Zur Erforschung universitärer Praxis* (pp. 83–106). transcript.

Felten, P. (2013). Principles of good practice in SoTL. *Teaching and Learning Inquiry*, *1*(1), 121–125.

Filer, A., & Pollard, A. (2000). *The social world of pupil assessment. Processes and contexts of primary schooling*. Continuum.

Finlayson, D. S. (1951). The reliability of the marking of essays. *British Journal of Educational Psychology*, *21*, 126–134.

Fontana, A., & Frey, J. H. (2000). The interview. From structured questions to negotiated text. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 645–672). Sage.

Forde-Leaves, N., Walton, J., & Tann, K. (2023). A framework for understanding assessment practice in higher education. *Assessment & Evaluation in Higher Education*, *48*(8), 1076–1091.

Fuller, G. (2015). In the garage: Assemblage, opportunity, and techno-aesthetics. *Angelaki*, *20*(1), 125–136.

Gadamer, H. G. (1960). *Wahrheit und Methode. Grundzüge einer philosophischen Hermeneutik*. Mohr.

Greiffenhagen, C. (2014). The materiality of mathematics. Presenting mathematics at the blackboard. *British Journal of Sociology*, *65*(3), 502–528.

Greiffenhagen, C. (2024). Checking correctness in mathematical peer review. *Social Studies of Science*, *54*(2), 184–209.

Greving, S., Lenhard, W., & Richter, T. (2023). The testing effect in university teaching: Using multiple-choice testing to promote retention of highly retrievable information. *Teaching of Psychology*, *50*(4), 332–341.

Hamann, J., & Velarde, K. S. (2023). The emergence of the academic candidate: Evaluation as textual dramaturgy. *Valuation Studies*, *10*(1), 58–89.

Hartog, P., & Rhodes, E. C. (1934). *An examination of examinations*. MacMillan.

Hirschauer, S. (2015). How editors decide. Oral communication in journal peer review. *Human Studies*, *38*(1), 37–55.

Honneth, A. (1995). *The struggle for recognition. The moral grammar of social conflicts*. Polity.

Hopf, C. (2016). *Schriften zu Methodologie und Methoden qualitativer Sozialforschung*. Springer VS.

Ioannidou, M. A. (1997). Testing and life-long learning: Open-book and closed-book examination in a university course. *Studies in Educational Evaluation*, *23*(2), 131–139.

Jack, A. (2016). Harm in asking: Class, acquired cultural capital, and academic engagement at an elite university. *Sociology of Education*, *89*(19), 1–19.

Kain, D. L. (1996). Looking beneath the surface: Teacher collaboration through the lens of grading practices. *Teachers College Record*, *97*(4), 569–587.

Kalthoff, H. (2011). Social studies of teaching and education. Skizze einer sozio-materiellen Bildungsforschung. In D. Sûber, H. Schäfer, & S. Prinz (Eds.), *Pierre Bourdieu und die Kulturwissenschaften. Zur Aktualität eines undisziplinierten Denkens* (pp. 107–133). UVK.

Kalthoff, H. (2013). Practices of grading. An ethnographic study of educational assessment. *Ethnography and Education*, *8*(1), 89–104.

Kalthoff, H. (2019). Organisierte Humanevaluation: Zur Behandlung und Klassifikation von Schulinsassen. In S. Nicolae, M. Endreß, O. Berli, & D. Bischur (Eds.), *(Be)Werten. Beiträge zur sozialen Konstruktion von Wertigkeit* (pp. 221–247). Springer VS.

Kalthoff, H., & Engert, K. (2024). The panopticon of evaluation in educational organizations. In A. K. Krüger, T. Peetz, & H. Schäfer (Eds.), *The Routledge international handbook of valuation and society* (pp. 129–139). Routledge.

Kalthoff, H., & Koelsch, F. (2025). Before (e)valuating: Student testing in history and engineering. *British Journal of Sociology of Education*, *46*(3), 323–340.

Kingston, P. W. (2001). The unfulfilled promise of cultural capital theory. *Sociology of Education*, *74*, 88–99.

Krüger, A. K., & Reinhart, M. (2016). Wert, Werte und (Be)Wertungen. Eine erste begriffs- und prozesstheoretische Sondierung der aktuellen Soziologie der Bewertung. *Berliner Journal für Soziologie*, *26*(3–4), 485–500.

Lennes, N. J. (1923). *The teaching of arithmetic*. Macmillan.

Luhmann, N., & Schorr, K. E. (1988). *Reflexionsprobleme im Erziehungssystem*. Suhrkamp.

Lymer, G., Lindwall, O., & Sellberg, C. (2025). Laughter, failure talk, and the sensitive nature of negative feedback. *Classroom Discourse*, 1–26. https://doi.org/10.1080/19463014.2025.2464564

Lynch, M., & Jordan, K. (1995). Instructed actions in, of and as molecular biology. *Human Studies*, *18*(2), 227–244.

Lynch, M., & Lindwall, O. (2023). *Instructed and instructive actions*. Routledge.
Lynch, M., Livingston, E., & Garfinkel, H. (1983). Temporal order in laboratory work. In K. Knorr-Cetina, & M. Mulkay (Eds.), *Science observed. Perspectives on the social study of science* (pp. 205–238). Sage.
Mannheim, K. (1935). *Mensch und Gesellschaft im Zeitalter des Umbaus.* Sijthoff.
McHoul, A. W. (1978). The organization of turns at formal talk in the classroom. *Language in Society*, *7*, 182–213.
McHoul, A. W. (1990). The organization of repair in classroom talk. *Language in Society*, *19*, 349–377.
Meyer, D., Reuter, J., & Berli, O. (Eds.). (2022). *Ethnografie der Hochschule. Zur Erforschung universitärer Praxis*. transcript.
Nieminen, J. H. (2021). Beyond empowerment: Student self-assessment as a form of resistance. *British Journal of Sociology of Education*, *42*(8), 1246–1264.
Nieminen, J. H., Bearman, M., & Tai, J. (2023). How is theory used in assessment and feedback research? A critical review. *Assessment & Evaluation in Higher Education*, *48*(1), 77–94.
Pryor, J., & Torrance, H. (2000). Questioning the three bears: The social construction of assessment in the classroom. In A. Filer (Ed.), *Assessment. social practice and social product* (pp. 110–128). Routledge.
Ricken, N. (2011). Education and the problems of difference. Reply to Paul Standish. *European Educational Research Journal*, *10*(3), 331–334.
Simmel, G. (1922). *Schulpädagogik. Vorlesungen, gehalten an der Universität Strassburg*. Zickfeldt.
Spradley, J. P. (1979). *The ethnographic interview*. Holt, Rinehart Winston.
Starch, D., & Elliott, E. C. (1912). Reliability of the evaluation of high school work in english. *School Review*, *20*, 442–457.
Stommel, J. (2023). *Undoing the grade. Why we grade, and how to stop*. Hybrid Pedagogy.
Stratton, J. A., & Mannix, L. H. (2005). *Mind and hand: The birth of MIT*. MIT Press.
Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge University Press.
Tsarouha, E. (2019). *Prüfungspraktiken an deutschen Hochschulen. Eine empirische Studie zu systematischen Einflussgrößen auf die Notengebung in Abschlussprüfungen*. Springer VS.
Tyagunova, T. (Ed.). (2019). *Studentische Praxis und universitäre Interaktionskultur. Perspektiven einer praxeologischen Bildungsforschung*. Springer VS.
Tyagunova, T., & Greiffenhagen, C. (2017). Closing seminars and lectures: The work that lecturers and students do. *Discourse Studies*, *19*(3), 314–340.
Ulshöfer, R. (1949). Zur Beurteilung von Reifeprüfungsaufsätzen. *Der Deutschunterricht*, *1*, 84–102.
Verkuyten, M. (2000). School marks and teachers' accountability to colleagues. *Discourse Studies*, *2*(4), 425–472.
von Stumm, S., Hell, B., & Chamorro-Premuzic, T. (2011). The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspectives on Psychological Science*, *6*(6), 574–588.
Waibel, D., Peetz, T., & Meier, F. (2021). Valuation constellations. *Valuation Studies*, *8*(1), 33–66.
Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. Jossey-Bass.
Winston, G. C. (1999). Subsidies, hierarchy and peers: The awkward economics of higher education. *Journal of Economic Perspectives*, *13*(1), 13–36.
Zaborowski, K. U., Meier, M., & Breidenstein, G. (2011). *Leistungsbewertung und Unterricht. Ethnographische Studien zur Bewertungspraxis in Gymnasium und Sekundarschule*. Springer.