Linguistics and Education 61 (2021) 100899

Contents lists available at ScienceDirect

Linguistics and Education

journal homepage: www.elsevier.com/locate/linged

# On the interactional challenges of revealing summative assessments: Collaborative scoring talk among teachers and students in Swedish national tests

Marie Nilsberth [a,*], Erica Sandlund [b]

[a] Department of Educational Studies, Karlstad University, 651 88 Karlstad, Sweden
[b] Department of Language, Literature and Intercultural Studies, Karlstad University, 651 88 Karlstad, Sweden

ARTICLE INFO

*Article history:*
Received 6 January 2020
Revised 12 October 2020
Accepted 7 December 2020
Available online 30 December 2020

## 1. Introduction

This study centers on the pedagogical activity of summative grade setting in interaction in two educational contexts: 1) Students in year 6 assessing their own performance in a practice test taken in class as preparation for a high-stakes national test, and 2) teachers assessing year 9 English speaking tests in collaborative assessment moderation meetings in a professional development workshop. In these contexts, both students and teachers use scoring rubrics and are tasked with translating these into a specific grade on a test performance. As standardized tests are increasingly emphasized in education, nationally and internationally, as tools for creating accountability and comparison (c.f. Lingard, Martino & Rezai-Rashti, 2013; Ozga, 2009), such classroom preparations for high-stakes tests have become increasingly common assessment practices in schools – for teachers as well as students. Such practices can also be understood from a perspective of *assessment literacy* development, a concept which Popham (2011) defines as "an individual's understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions" (p. 267). For teachers, the development of assessment literacy is seen as central not only to their professional work of assessing students, but can also serve as a shared knowledge base for professional learning, and as a resource for moderation and equity in assessment (Grainger, Adie, & Weir, 2016; Jönsson & Thornberg, 2014; Popham, 2009). Likewise, for students – from early school years and onwards – the understanding of assessment and grading as such has also become teaching content; in turn expected to enhance learners' abilities to learn and perform well (Smith, Worsfold, Davies, Fisher, & McPhail, 2013; Tanner & Pérez Prieto, 2020). In the Swedish national curriculum, for example, students' ability to evaluate their performance is expressed as one goal of education, where students should be able to "assess their own results and relate these and the assessments of others to their own achievements and circumstances" (Swedish National Agency for Education, 2018, p. 16). Thus, an increased focus on assessment in education has also increased the demands on both teachers and students to develop their assessment literacy. Therefore, a closer look at the practices that evolve from these demands, down to the micro-level of assessment talk, is timely and warranted.

Assessment of student performance in educational contexts encompasses a wide variety of aims (formative vs. summative purposes, e.g., Black & Wiliam, 2009) and activities (classroom assessment, test grading, term grades). At the center of the present study lie the socially situated practices of communicating summative assessments, as we examine assessment conversations between students and between teachers. The conversations in focus are designed to develop shared understandings, or assessment literacy, about how to interpret and apply summative scoring criteria to specific task performances in national tests. While formative assessment aims to support and enhance learning along the way, summative assessments constitute a report *of* learning, which generally involves assigning a score or grade (e.g., Gardner, 2006; Harlen, 2006). Summative assessments, then, are overall judgements of a particular learner's knowledge or abilities at a particular point in time, and are commonly assigned at the end of an instructional unit or educational stage, or in a test performance. While summative assessments in the form of grades are often communi-

* Corresponding authors.
*E-mail addresses:* marie.nilsberth@kau.se (M. Nilsberth), erica.sandlund@kau.se (E. Sandlund).

https://doi.org/10.1016/j.linged.2020.100899

cated in, for example, written report cards, they are also communicated and accounted for in interactions with students, parents, and colleagues.

In the present study, we focus on how the assessment of student performance – of one's own or that of others – becomes an object for social scrutiny and joint decision-making in interaction between peers and colleagues. The analysis is based on empirical data from two different contexts – teachers' collegial talk in collaborative grading of a national test in English as foreign language (EFL) in year nine, and students' self-assessments of performance revealed in peer group talk as part of classroom work in preparation for the national test in science in year six. The pedagogical (and political) justification for such preparation activities is to develop students' and teachers' assessment literacy and to contribute to students' self-regulation of learning. However, the pedagogical activities of practicing assessment, in both contexts, constitute potentially delicate social interactions, which in turn may work against the institutional goals of developing assessment skills. From the ethnomethodological stance of Conversation Analysis (CA) (Gardner, 2019; Sacks, Schegloff & Jefferson, 1974; Waring, 2016), we adopt an emic, participants' perspective to the assessment practices and approach assessments as and in situated social practice, with a focus on the social and interactional accomplishment of assessments of student performance. We specifically investigate how teachers and students, in the context of national testing in compulsory school, manage the institutional activity of revealing to each other their assessment of a particular test task performance. Since assessment discussions involve conversational norms, social concerns, and epistemic constraints, a closer look at the ways in which assessment practice in education play out in social interaction is warranted. In the present study, we demonstrate how assessment practices in both contexts are delicate matters, where participants have to balance social concerns that might come in conflict with the institutional goals about developing assessment literacy. We also highlight some social, interactional, and professional challenges facing students and teachers-as-raters, respectively, in revealing the outcome of an assessment decision. As such, the present study aims to contribute to ongoing discussions on the pros and cons of these types of assessment activities in classrooms, and in teachers' professional arenas for collaborative scoring and moderation.

## 2. Assessment as social practice in education

In this section, we introduce the type of assessment practices in focus in the present study. We detail relevant previous research on students' participation in self-assessment practices (Section 2.1), and continue with a review of teachers' or raters' engagement in collaborative assessment as professional practice or professional development (Section 2.2). Finally, we turn to some conceptual distinctions between the term assessment, as generally understood in educational research and practice, and the conversation analytic treatment of an assessment as a local, situated action-in-interaction (Pomerantz, 1984) in order to clarify the different applications of the term assessment in the present study (Section 2.3). The section ends with a summary and discussion of how previous knowledge on assessment practices has informed the design and focus of the present study.

### 2.1. Students' self-assessment of learning

Students' performance is generally assessed by teachers, but they are also sometimes explicitly asked to make self-assessments, as part of teacher-parent conferences, mentor talks, or in the classroom (e.g.; Hofvendahl, 2006; Waring, 2014). As for students' participation in educational assessment practices, empirical research is relatively limited to date. Research from psychological perspectives on learning has primarily focused on the importance of students' development of meta-cognitive and self-regulatory skills, mostly in connection with pedagogical ideals of formative assessments (e.g., Brown & Harris, 2014; Sadler, 2010). In a research review, Panadero, Brown and Strijbos (2015) identify a lack of consistency as regards to conceptualizations of self-assessment. While the self-regulatory function of students' self-assessments is often presented as positive in terms of the purpose of formative assessment practices, some studies problematize self-regulation as an educational goal, as it promotes students to develop so called "performative identities" (Jeffrey & Troman, 2011, p. 499). The authors point to how coping with performative social practices like test-taking may promote students' ability to 'play the system' rather than expanding their learning (see also (Tanner & Pérez Prieto, 2020). In their study of third-graders taking the national test in Mathematics in Sweden, Sjöberg, Silfver and Bagger (2015) show how assessment trends, such as national tests, affect students' learner identities in how high-stakes testing causes children to position themselves as 'good' or' bad' in Mathematics. They argue that tests constitute " a technology used for disciplining children" (p. 70), and that this disciplining "is often a subtle process that nonetheless produces powerful messages to children about who is "good", "bad", a "winner" or "loser" in the game of the test, and in the future" (p. 71).

Moving toward empirical work on students assessing their own performance, Skovholt, Nordenström and Stokoe (2019) specifically examine how students evaluate their own performance or work in conversations with supervisors or facilitators. They examine interactional sequences where students are asked to evaluate their own performance on a prior task, such as simulation exercises for nursing students, or in individual feedback sessions after an oral presentation in an upper secondary school classroom. The examined examples could mainly be described as formative practices with a prospective learning purpose, and findings show that students' assessments of their own performances bring interactional challenges in relation to eliciting, delivering, and responding to evaluations. The authors demonstrate how students face and display interactional difficulties when asked to offer positive assessments of themselves and their work, which they relate to the well-documented underlying social norm that self-praise should be avoided as to avoid the risk of being viewed as bragging (Pomerantz, 1984; Speer, 2012; Waring, 2014, see also Section 2.3 below). The authors argue that besides interactional norms against self-praise, assessments are also "constrained by who is entitled to perform them" (p. 54), and in educational contexts, the differential institutional roles of student and teacher bear different epistemic rights in terms of assessing performance. While Skovholt et al. (2019) focus on adult learners, their findings have relevance for the younger students' self-assessments in the present study where we also add to this knowledge a focus on students' self-assessments in collaboration with peers, which in terms of social and epistemic positions could be expected to be different.

### 2.2. Teachers' collaborative assessment practices

As for teachers' engagement in assessment practices, it has become increasingly common that policy makers as well as assessment researchers encourage collaborative assessment between raters or teachers as a remedy for challenges with assessment equity, especially in the context of large-scale standardized testing (Erickson, 2009; Swedish Schools Inspectorate, 2013). Moderation (variably referred to as *social moderation* or *consensus moderation*, see Linn, 1993; Sadler, 2013) is defined as "a practice of engagement in which teaching team members develop a shared understanding of assessment requirements, standards

and the evidence that demonstrates differing qualities of performance" (Grainger, Adie, & Weir, 2016, p. 551); an organized practice for the verification of assessment judgements against standards (Bloxham, Hughes & Adie, 2016).

While studies of the long-term effects of moderation are scarce, there appears to be agreement regarding the benefits of moderation activities as professional development. It has been suggested that calibration through moderation activities appears to contribute to a higher degree of consensus on what is to be assessed (Jönsson &Thornberg, 2014, p. 392) and that moderation activities focusing on building communities for teachers or raters have positive effects on "assessors' assessment literacy as well as knowledge of standards" (Bloxham et al., 2016, p. 649). In particular, when raters disagree on the assessment of specific learner performances, quality variations can be discussed and negotiated in relation to standards, leading to opportunities for professional learning (cf. also Adie, Klenowski & Wyatt-Smith, 2012). In line with this also Allal (2013) recommends developing activities where teachers can "discuss, confront, and negotiate grades assigned", thereby improving the coherence and transparency of individual teachers' judgements in summative assessment (p. 21; 33).

In studies of moderation, it has been argued that key to its effectiveness is that the social climate in the group allows for "the representation and exploration of dissensus", which means that the potential embedded in disagreement and challenges is nurtured as an opportunity for learning (Moss & Schutz, 2001, p. 65; see also Sadler, 2013). As part of an interventionist research project on the teaching and assessment of writing in Norwegian primary schools, Matre and Solheim (2016) examined interaction in groups of teachers conducting collaborative, summative assessment of student texts. Their analyses demonstrate that in particular, one of the three teacher groups probed into individual assessments, and that:

> (…) the critical challenging of one another's readings and assessments to a much larger degree open up dialogic spaces with the potential for developing knowledge and understanding. Opposing utterances compel the teachers to argue for their points of view and lead them back into the texts to re-read and make re-assessments, negotiating their judgements. (Matre & Solheim, 2016, p. 200).

The authors conclude that allowing for disagreement in collaborative assessment should be considered a fruitful route in terms of assessment as such, and for professional learning. Research on consensus versus disagreement in professional discourse has shown that professional teams generally do not "welcome explicit threats to the spirit of consensus" in workplace meetings (Marra, 2012, p. 1582; see also Holmes & Marra, 2004), and that participants deploy various interactional measures to avoid the potential threat of conflict. Furthermore, the voicing of deviant opinions may temporarily threaten members' professional identity (cf. Richards, 2006) and when disagreements on, for example, scores, are made public, these divergences are generally treated as socially problematic. Raclaw and Ford (2017) show that in sequences when divergence in preliminary scores delivered by different reviewers in a grant proposal meeting were made public, participants oriented to the displayed divergence as accountable, and as potentially delicate. Laughter was one interactional resource through which participants both indexed and managed the delicacy of moments of disagreement and disaffiliation as it "provides an opportunity for participants to jointly engage in smoothing the potential awkwardness of the moment after divergent scores are first reported" (Raclaw & Ford, 2017, p. 13). Also Pier et al. (2018) noted that while there was low initial agreement among individual reviewers on scores on grant proposals, reviewers tended to conform to majority views during calibration discussions, especially when held accountable for their individual scores. This, in turn, will inevitably also challenge one cornerstone of successful collaborative assessment, namely that different opinions about a particular learner performance create an opportunity for learning and for developing a shared understanding of rubrics (Jönsson & Thornberg, 2014). Like any other collaborative work practice, assessment discussions may, to varying degrees, require participants to reveal the outcome of their individual assessment, and be prepared to have their professional judgements challenged by others. As such, the very act of sharing one's professional judgement also means displaying publicly one's professional competence and/or identity.

### 2.3. Doing assessment in social interaction

Finally, as we focus on assessment practices in social interaction, of relevance to the analyses is the conversation analytic work on assessments that describe a particular type of action in social interaction rather than an educational goal or practice. This framework is central to the present study, as the analyses center on how participants (students and teachers) deliver their individual assessments of a test performance in the form of self-assessments and other-assessments in educational contexts. From a CA perspective, assessments refer to the routine and organized ways in which participants in social interaction both produce and respond to assessments about the social activities in which they participate, or have participated in. In her seminal work on assessments in interaction, Pomerantz (1984) observe how people provide assessments in and about joint activities, when reporting on participation in activities outside the present interaction, and in turns following upon a first assessment. Pomerantz also demonstrated an underlying structural preference for agreeing with a first assessment and demonstrated how disagreements with the first assessment were characterized by delays, mitigations, and hesitations in their production. Disagreeing with a first assessment is thus an accountable, structurally dispreferred action, while agreement (which is structurally invited by first assessor) are preferred actions, generally not showing the production of delays and mitigations (such as *well, uh*, and gaps) as dispreferred ones. Two exceptions to the preference for agreement with a first assessment is in response to self-deprecations, where disagreement is the preferred response, and in the receipt of compliments (cf. Pomerantz, 1978). However, as noted by Lindström and Mondada (2009), Pomerantz' (1984) work centered mainly on second assessments next to first assessments, and how these may include an upgrade, a downgrade, or a same-level assessment as the first. While upgrades do the work of displaying strong agreement, downgrades can generate disagreement sequences, or show weak affiliation with the prior speaker's assessment. More recent work on assessments has also shown that relative knowledge on what is being assessed plays a role in the production of first and second assessments, and inevitably make relevant epistemic issues of in the evaluation of who has the rights to assess persons or objects (Heritage & Raymond, 2005, p. 16). Additionally, institutional contexts in which assessing is the central business of the encounter, are rich contexts for the study of assessments, such as in assessing student performance (Pillet-Shore, 2003; Skovholt et al., 2019), editorial conferences (Clayman, 1998), or performance appraisal interviews (Asmu$\beta$, 2008), to mention a few.

### 2.4. Summing up: assessment practices as social events

Given the faith in collaborative and reflective assessment practices in education today, and the time invested in them, further examination of how they play out in classroom or collegial settings is warranted. Having reviewed work in two distinct assessment contexts – students' self-assessments of performance, and teachers'/raters' collaborative assessment of learner performance,

it can be concluded that participation in assessment practices is no straightforward task, for social/interactional as well as institutional reasons. Firstly, research on the local conditions for students' reflective and self-evaluative practice are scarce, and Skovholt, Nordenström and Stokoe's (2019) represent a strong example of how close examination of such reflective practices can challenge and provide further support for development of fruitful assessment practices. Secondly, a review of studies on the delivery of evaluations in professional collaborative interaction (e.g. Raclaw & Ford, 2017) revealed that public disagreement requires careful interactional work, but may also yield gains in terms of shared understandings of the application of scoring criteria. The present study builds on this previous work and also on conversation analytic work on assessments in social interaction (Pomerantz, 1984) and aims to shed light on the interactional and institutional activity of announcing one's assessments (self- assessments and other-assessments) in interactions with peers and colleagues.

## 3. Data and methodology

### 3.1. Empirical context

The empirical context of this study consists of assessments practices related to the Swedish national testing system. In Sweden, students take national tests in Swedish, English and Mathematics in year 3, 6 and 9. Additionally, during a short period (2013–2015) students in year 6 also took tests in Social Studies and Science, which are no longer mandatory in year 6 but are still so for year 9 students. While not final exams in a traditional sense, the national tests are designed to have "an advisory function in teachers' decision-making regarding students' final grades" (Borger, 2019, p. 155; see also Swedish Ministry of Education & Research, 2017, pp. 22–23), and considered "distinctly high-stakes" (Borger, 2019, p. 155). As the students' own teachers act as the examiners of the national tests, they are also responsible for preparing their classes for the test, for administering the tests, for grading them, and for submitting the results to the authorities. Following the grading system in Sweden, the national tests are graded according to a scale from A – F, where A is the highest grade and E is the lowest passing grade (i.e., meeting the knowledge requirements for a particular level). An F grade means that the knowledge requirements stated have not been fulfilled; thus, an F is a failing grade.

### 3.2. Data and participants

As we are interested in practices surrounding high-stakes assessment – for teachers and students – we draw on two datasets which both center on national testing: the *Testing Talk project* (Swedish Research Council, 2012–4129) and the project *Pupils' stories about grades* (Swedish Research Council, 2013-1668). The student assessment data comes from a larger study on students' experiences of doing national tests and getting grades (see (Löfgren, Löfgren, & Pérez Prieto, 2018; Tanner & Pérez Prieto, 2020). The dataset consists of a total of 13 lessons in two different year 6 classrooms, out of which four lessons had the purpose of practicing self-assessments of students' own performance on written rehearsals, or practice tests, for the high-stakes Biology and Social Studies tests in 6th grade, which were conducted in the classroom. The examples selected for further analysis in this article both come from the Biology lesson, where two different groups of students are recorded as they collaborate. All students in the class were divided into groups by the teacher, and all groups were mixed-gender, but with no other specific criteria. The students produced answers to the old test tasks, and then jointly practiced assessing their own performance in relation to criteria on the grade levels A, C or E. Worth noting here is that although the grading scale spans from A to F, descriptive texts and benchmarks are generally only provided for grades A, C, or E in the assessment materials to teachers. Even if this particular test is no longer in use,[1] preparations using older test assignments is a common practice also in other school subjects that still have the national tests, and therefore they work well as an example in relation to the aims of the present study.

The teacher assessment data consists of four video-recordings of second language (L2) English teachers doing collaborative assessment (CASS) of a paired speaking test in year 9 of compulsory school (the National English Speaking Test, NEST). It is drawn from a collection of four meetings between teachers-as-raters, with a total recording time of 200 min (3,33 h), and the selected recording features three English teachers in one of these four groups. The English test, which is compulsory in 6th and 9th grade, consists of three parts: Speaking (A), Listening and Reading (B) and Writing (C). The data analyzed in the present paper comes from a workshop where four groups of English teachers worked with collaborative assessments of one recorded authentic NEST Speaking test (see Sandlund & Sundqvist, 2019, for a more extensive description of the research design). The teachers were attendees at a professional development day for English teachers organized by a university research center, and had volunteered for a workshop on collaborative assessment of L2 speaking. All ($N = 13$) had a teacher degree in English, and had worked for more than four years, except for one, who had only worked as an English teacher for six months (range 6 months – 25 years). The teachers were provided with one NEST recording featuring two learners and their co-present teacher. The teachers first listened to the selected test recording in full, while taking notes and making initial assessments individually (a similar procedure applied by May, 2011). Raters were then divided into four groups, combined so that they consisted of teachers from different schools. All four CASS moderation meetings were video-recorded in their entirety and the recordings range between 42 and 55 min. Teachers were provided with the authentic assessment materials for the test in question, which consist of a booklet with instructions for assessment, performance descriptors, and written descriptions about characteristics of a performance at the grade levels A, C, and E, were instructed to discuss their views and initial assessments of the two learners, and finally, to agree on a test grade for each learner.

As described above, both datasets thus center on assessing learner performance in a national test; however, there are contextual differences. For one, the datasets have different school subjects in focus and also, different school years, as the classroom data is from year 6 and the teacher data centers on the assessment of a national test in year 9. With these differences in mind, however, the present study centers on interactional practices of revealing summative assessments, and as such, the national testing system with its multifaceted assessment levels, constitutes a relevant overall frame in which the data allows us to shed light upon such practices at different but interrelated arenas. Furthermore, while the classroom data centers on students' self-assessment of their performance according to standards, the teacher data encompasses teachers' professional judgments of learner test performance, but both contexts can shed light on the interactional complexities of revealing to others an evaluation of performance in an educational context.

The combination of data from the two projects is in line with the original scope formulations of both, and the projects have been approved by regional ethics boards. Participants – and, for the stu-

[1] Soon after our data collection, these particular subject tests were withdrawn, but at the time of the study the students took four sub-tests in Biology and in Social studies.

**Table 1**
Overview of data presented.

| Data excerpt | Description | National test subject | School year | Participants |
|---|---|---|---|---|
| 1 | Classroom | Biology | 6 | Students ($N = 4$) |
| 2 | Classroom | Biology | 6 | Students ($N = 4$) |
| 3 | Classroom | Biology | 6 | Students ($N = 4$), biology teacher ($N = 1$) |
| 4 | Teachers' assessment meeting | English | 9 | English teachers ($N = 3$) |

dents – their legal guardians, had given their consent to participate and to be video-recorded. For the sake of protecting participants' identities, pseudonyms are used throughout, and drawings are used for images from the recordings illustrating embodied conduct.

### 3.3. Analytic approach

In order to examine how participants in educational interactions orient to and manage the assessment of performance, we adopt a conversation analytic (CA) approach. CA as an approach takes sequential organization of naturally occurring conversations in and beyond institutional settings as a point of departure (Markee, 2015; Sacks et al., 1974; Schegloff, 2007) and aims to "understand tacit social order through the concrete details of talk-in-interaction" (Malabarba & Nguyen, 2019, p. 3). Grounded in ethnomethodology's members' perspective, CA sets out to uncover the methods that social actors themselves use to accomplish, maintain, and make sense of their social contexts, focusing on actions and the resources (linguistic and embodied) used to build such actions, in sequences of social interaction (cf. Clift, 2016). As such resources are oriented to and treated as meaningful by participants themselves in situated talk, CA transcripts set out to capture the fine-grained details of interaction, including "volume, pitch, pace, intonation, overlap, inbreath, smiley voice, the length of silence, as well as nonverbal conduct" (Waring, 2016, p. 46). The analytic process is inductive, and through close inspection of a piece of recorded data and the accompanying transcript, the analyst's task is to demonstrate the workings of sequences from an emic perspective – i.e. to uncover participants' *own* analyses of a stretch of talk as they interpret prior talk and respond to their co-interactants' contributions.

Data for the present study have been transcribed according to conventions commonly used in conversation analytic research (see Atkinson & Heritage, 1984; Jefferson, 2004: Hepburn & Bolden, 2013). The two datasets were scanned for sequences in which students (Year 6 classroom work on national tests) and teachers (as raters of the Year 9 NEST) announce to their peers an assessment of test performance, and the subsequent trajectories following the assessment revelation. We searched through all recordings in the two datasets and identified all sequences in which students and teachers, in different group constellations, discuss, compare and decide on a grade for a certain student performance on the national tests. As described above, for the classroom data from year 6, two lessons out of 13 recorded were devoted to this particular activity, and for the teacher data, all four recordings contained sequences where the teachers revealed their individual assessments to each other. For the analysis presented in the present paper, we selected for presentation sequences from one of the lessons (the one about Biology) that represent different trajectories identified in the data as a whole. We will show these different aspects of the assessment practices examined in relation to peer-interaction (Excerpts 1 and 2), teacher-student interaction (Excerpt 3) and teacher-teacher interaction (Excerpt 4). The selected video-recorded data presented is illustrated in Table 1.

As the present study is qualitative, each sequence is treated as a single case with its own internal logic, but the excerpts presented were selected because they were considered typical among all identified sequences in terms of how participants orient to assessment revelations.

## 4. Findings

In our analytic presentation, we first examine three segments from a classroom activity where students practice assessing their own performances in a practice test task, in which they are instructed to discuss their assessments in peer groups based on the national test grading criteria. Our analysis of the student sequences show how the act of revealing a judgement of one's own task performance is treated as socially delicate by participants. We focus specifically on demonstrating how the delicacy of revealing positive self-assessments is managed, and on how epistemic asymmetries in classroom interaction impact on the revelation sequences. Subsequently, we examine a sequence from three teachers' collegial assessment discussions regarding student performance in a paired speaking test in L2 English. Our analysis shows how participants treat the revelation of an assessment of student performance as delicate business, and the first professional judgement revealed is treated as an accountable action, where subsequent and divergent judgements are delayed. As such, revealing a professional opinion to colleagues is not a neutral activity, and participants show their management of this delicacy in different ways.

### 4.1. Students revealing self-assessments

The first example comes from a lesson in Biology in Swedish school year 6, with two groups of 12-year-old students preparing for an upcoming national test in Biology, which in Sweden is one out of three Science subjects collectively referred to as NO[2]; the others being Chemistry and Physics. The students have been writing individual answers to a test task, which instructs them to compose advice to a fictional friend mentioned in the task about how a cold might affect participation in a soccer training camp and a soccer cup. The class teacher has talked about the scoring rubrics for this specific task and also highlighted a checklist, which directs the students to give medically and scientifically sound advice to their fictional friend, based on reasonable and appropriate arguments (see Fig. 1).

Excerpt 1 involves a group of the four students Anders, Birgitta, Caj and Doris, sitting at their student desks, facing each other (see Fig. 2).

The sequence is extracted from the discussion in this group. Their task is to discuss and compare their different test answers and evaluate them in relation to the rubrics for grading, and as we enter, there has been a brief silence following a first general discussion of the task. Caj then opens the discussion with an assessment of his own text (line 1):

[2] Swe. Naturorienterande ämnen (NO)

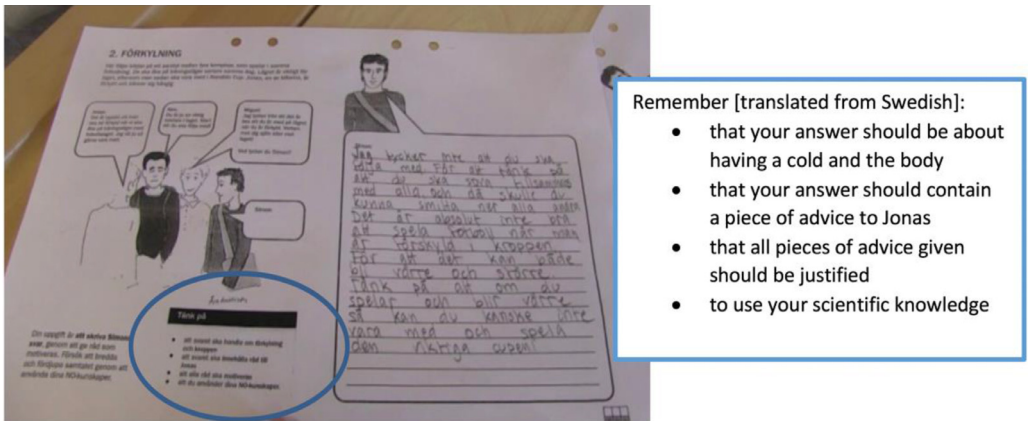


**Fig. 1.** Sample answer to the test task plus task checklist.

| | | |
|---|---|---|
| 1 | Caj: | mitt sög<br>**mine sucked** |
| 2 | | (0.9) |
| 3 | Doris: | nej↑<br>**no↑** |
| 4 | Caj: | jo för jag hade inga bra anledningar >eller jag<br>**yes cause I had no good reasons > or I** |
| 5 | | hade ingen sådär< [ anled- ]<br>**had no such < [ reas- ]** |
| 6 | Birgitta: | [ ja'ra ]<br>**[ me then]** |
| 7 | Doris: | [ jamen ja'ra ]<br>**[well what about but me then]** |
| 8 | Caj: | [ man blir så ] där sjukare i kroppen<br>**[ you get so ] like sicker in the body** |
| 9 | Birgitta: | [och så ska det handla om NO ]<br>**[and then it should be about NO]** |
| 10 | (Anders): | [ (bättre än mina) ]<br>**[ (better than mine) ]** |
| 11 | (Birgitta) | [ (min då) ]<br>**[ (mine then) ]** |
| 12 | Doris: | [ja men jag fa ]ttar inte det står- du ska använda<br>**[but I don't un]derstand it says- you should use** |
| 13 | | dina NO-kunskaper (vi har ju inte lärt)[ eller vad ska]<br>**your NO knowledge °we have not learnt° [or what should]** |
| 14 | ( ) | [ ° nej° ]<br>**[ ° no° ]** |
| 15 | Doris: | =man skriva va [↑det ärå ]<br>**=one write [↑that is then]** |
| 16 | ( ) | [ °ja° ]<br>**[ °yes° ]** |
| 17 | (Birgitta) | [ (nej ja ba-) ]<br>**[ (no I like-) ]** |
| 18 | Caj: | [vi har väl inte lärt] oss sånt om typ förkylning<br>**[ we haven't learned] anything about like colds** |
| 19 | | och nåt<br>**and such right** |

**Excerpt 1.** "Mine sucked".

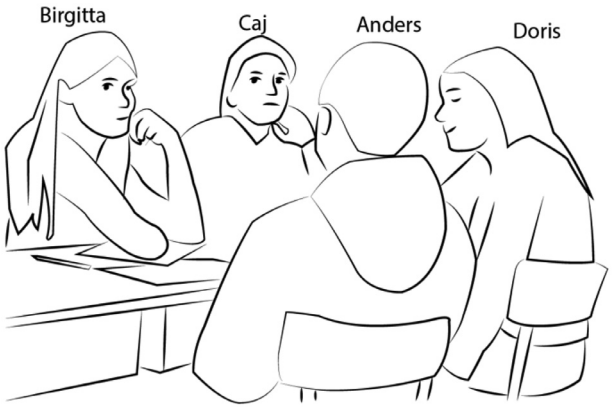


**Fig. 2.** Student configuration at the outset of Excerpt 1.

In excerpt 1, Caj initiates an evaluation of his own test answer with an assessment: *mine sucked* (line 1), leaving the assessable (his test answer) unspecified. This assessment has a sequence-initiating function in that it is the first to attend to the task-at-hand (discussing and assessing their individual performance), and announces to the group his stance towards his own performance. Even if he does not explicitly refer to a specific grade on the *A – F* scale, he evaluates his own performance as substandard. Caj's announcement is not explicitly addressing any specific participant, and it is first followed by a notable silence of 0.9 s (line 2). The delay in responding can be heard as an indication of some kind of interactional trouble – either relating to *how* to answer Caj, or *who* should answer. Subsequently, Doris offers a delayed minimal response with a disagreeing *no* particle (line 3). Making negations is one common way of disagreeing with self-deprecations, such as Caj's, (Pomerantz, 1984, p. 84), and while such preferred answers often come quickly after an initiating turn (Robinson, 2020), Doris' minimal answer is preceded by a pause. As such, her delayed response, while verbally disagreeing with Caj's negative self-assessment, does not offer clear affiliation with Caj, as would an immediate response like "no, I'm sure it didn't". However, Caj appears to treat his own prior assessment, and perhaps in light of co-participants' delayed response, as accountable. He proceeds by accounting for the assessment in lines 4–5: *cause I did not have any good reasons or I did not have any of those reas-* (lines 4–5). His-incomplete production of *reas-* refers to the advice (Swe. *anledningar*, reasons) that are listed in the bullet points in the instructions (see Fig. 1), where the provision of advice supported by reasons is highlighted as an assessment criterion. In response to Caj's account, Birgitta (line 6), in overlap, produces an affiliating self-deprecation by announcing that she, too, had a substandard test answer: *me then,* line 6). Here, the Swedish *ja'ra* is produced as a contraction of the pronoun "jag" (me) and "då" (then), which is a common way of expressing a sentiment like "well you should see mine then" or "well then you should see mine". The first person pronoun, me, is emphasized, and her turn could be heard as regrading Caj's self-deprecation: Birgitta's turn implies that neither she gave any reasons, and that her test answer is just as bad as, or worse than, Caj's. Next, Doris recycles the gist of Birgitta's turn format and adds an assessment of her own test answer: *well what about me then* (line 7), which casts her own performance as equally, or more substandard than preceding speakers. In this context, Doris' and Birgitta's responses could be understood as affiliative in that their self-assessments display solidarity (cf. Clayman, 2002) with Caj's initial self-deprecation by describing their own performances as substandard as well. As such, the production of second self-deprecations could be seen as affiliative actions deployed in service of managing the delicacy of responding to them (see Lindström & Sorjonen, 2012).

In overlap with Doris, Caj (line 8) formulates a candidate test answer to the first bullet point in the checklist (see Fig. 2), which states that the reasons given should be about a cold and the body.

His-*you get like sicker in the body* orients to the bullet point, but its production is hearable as a simultaneous assessment of this type of answer as perhaps too simple: it is delivered with an animated, 'mocking' tone, in a similar way as participants in social interaction use reported speech (cf. Holt, 1996) to both report on and assess some piece of prior talk. In this case, Caj's formulation of a possible test answer seems to simultaneously mock the type of test answer that could be given based on the instructions. Birgitta's next turn, with a turn-initial and marking it as an addition to Caj's turn, also orients to the test instructions: *and then it should be about NO* (line 9). As such, Birgitta adds what she treats as an additional aspect that a test answer should contain – a clear grounding in the school subject NO, which can be heard as Birgitta treating Caj's implied assessment of the instructions as valid and understandable. The discussion is here fast-paced, and it is difficult to hear exactly what is said in lines 10 and 11, but it seems as if two of the students (probably Anders and Birgitta) are still oriented towards Caj's suggestion in line 8, and producing comparisons of that answer to their own. However, in line 12, Doris builds on Caj's and Birgitta's orientations to the instructions, and hearably claims non-understanding of the task: *but I don't get it says you should use your NO skills.* In a softer voice, she produces a complaint-type turn: *we have not learnt* (line 13), which shifts the focus from a lack of understanding on her part to the teaching preceding the test – there is something in what is expected from this task that they haven't learned in class. However, she appears to abandon the turn and restarts with *or what should one write what that is then* (lines 13 and 15). As such, Doris' turn expresses non-comprehension, but also a complaint against the demands of the task. The emphasis on *that* is produced with an intonation rise, and seems to be doing the work of displaying non-comprehension and questioning of the task. The audio quality constrains some of the analysis in lines 14 and 16 as two students (not possible to distinguish who) offer minimal responses that appear to affiliate with her complaint, and in line 17 Birgitta adds on something that we hear as *no I like-* (also difficult to distinguish) which could be heard as agreement with the gist of Doris' complaints – she too found the expectations of the task almost impossible to fulfill. In overlap with Birgitta, Caj shows understanding of Doris' incomplete turn about the teaching, stating that they indeed *haven't learnt anything about like colds and such, right* (lines 18–19). This orients to Doris' previous turn (lines 14 and 15) as complaint-indicative, on which he builds an additional complaint – that medical/physical aspects of colds have not been dealt with in class – which in turn challenges the doability of the task.

In this sequence, the students' task is to come up with summative self-assessments of their own performances on a test task in relation to specific grading rubrics. Aside from displaying that they find the task instructions difficult, and that they also each assess their own attempts at answering the test task as not meeting the requirements of the instructions, the students also face another possible interactional difficulty: they are asked to make qualified judgements that are normally acknowledged to be part of the teachers' expertise (Skovholt et al., 2019; Tainio & Laine, 2015). Assessing one's own performance can therefore be considered risky business, as going either too high or too low may put the speaker at risk of being held accountable, or to impact on the type of student identity that is displayed. As Attenborough and Stokoe (2012) have shown, students, in interaction with other students, have to manage interactional constraints against self-praise (which is dispreferred) as well as institutional expectations to excel in their studies. As a result, being a student means being managing the " 'correctness' of doing publicly observable displays of 'average-ness' not 'brilliance' (Attenborough & Stokoe, 2012, p. 14), which can be seen in attempts at mitigating risks of being seen as too ambitious, knowing too much, or bragging. In our data, these interactional risks are managed through Caj's initial self-deprecation, to which his peers first object but then show solidarity with. Together, they lower the expectations of their individual answers, and in stepwise manner, their self-assessments progress into a complaint sequence where the task and the class teaching are assigned blame for their inability to perform well. Complaint sequences often work as resources to make a personally experienced problem into interpersonal difficulties through being shared with others (Heinemann & Traverso, 2009). This strategy, in turn, works proactively as it pre-empts having to give a positive assessment of one's own performance that in turn could be understood as self-praise and potentially lead to accusations of bragging (cf. Pomerantz, 1978). As the initial self-deprecating assessment is turned into a shared complaint about a third party (the lack of convergence between the test task and NO lessons), the students can affiliate with each other's substandard performances, and join forces against a common enemy. In Lerner's (1993) words, participants collectively position themselves in a conjoined unit against the teaching, which works to mitigate the delicacy of revealing their self-assessments of performance. Consequently, the very act

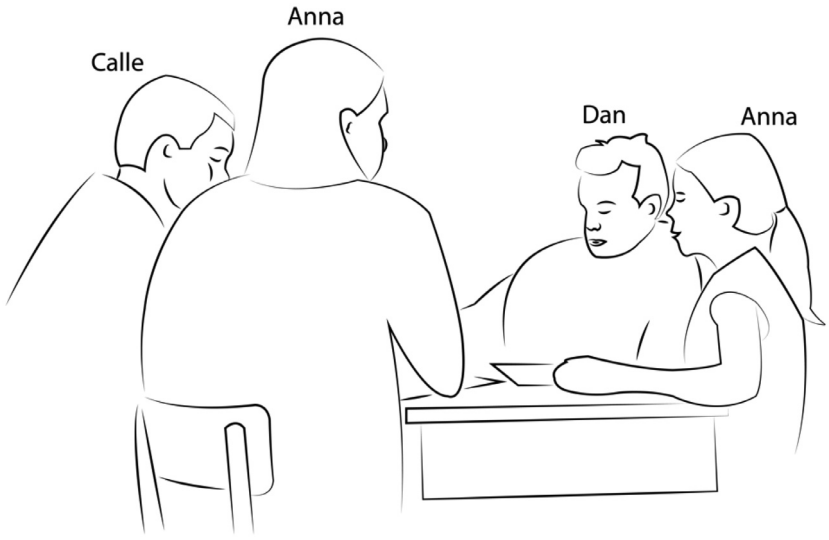


**Fig. 3.** Student configuration at the outset of Excerpt 2.

| | | |
|---|---|---|
| 26 | Britta: | men vi: tänker ju ungefär samma↑(.) att han inte ska |
| | | **but we: do think about the same↑(.) that he should not** |
| 27 | | vara med på träningslägret för å s- se om han kan |
| | | **join the training camp to s- see if he can** |
| 28 | | åka med på cupen |
| | | **go to the cup** |
| 29 | Anna: | ja |
| | | **yes** |
| 30 | Calle: | ja jag skrev väl jag fick ett D |
| | | **yes well I wrote that I got a D** |
| 31 | | **(1.2)** |
| 32 | Britta: | ha? |
| | | **what?** |
| 33 | | **(1.0)** |
| 34 | Dan: | hur vet du det? |
| | | **how do you know that?** |
| 35 | Anna: | **°hma ha ha°** |
| 36 | Britta: | **hm he** |
| 37 | Calle: | för att jag vet det |
| | | **cause I know that** |
| 38 | Britta: | ja ha jag får ett A ja då↑ |
| | | **yes ha I will get an A then↑** |
| 39 | Anna: | mm [jag med↑] |
| | | **mm [ me too↑]** |
| 40 | Calle: | [ näe: ] jag vet för jag kan titta in i framtiden |
| | | **[ no: ] I know cause I can see into the future** |

**Excerpt 2.** How do you know that?

of revealing judgment of one's own task performance is treated as socially delicate, and as such, requires careful interactional management.

Excerpt 2 is taken from a different group during the same activity, involving students Anna, Britta, Calle and Dan (Fig. 3). Prior to the sequence examined, the class teacher has presented the qualitative differences between the grades A, C, and E. Immediately prior to line 26 below, the students in the group have read aloud their different test task answers to each other, and like the students in Excerpt 1, they have specifically noted that their answers should be about colds and the body (lines 1–25, not shown in transcript). As we enter, Britta, who had adopted a leading role in prior talk, offers a candidate summary of the gist of their respective test answers.

Britta's turn identifies, although based on different reasons, a common denominator in their respective answers (*but we think almost the same*) in that they had all provided the same piece of advice to the fictional peer (lines 26–28). Her turn invites confirmation, and Anna responds with an affirmative *yes* (line 29). Calle also responds affirmatively (line 30), *yes,* and, rather abruptly, proceeds with announcing the individual grade he had assigned himself in the task: *well I wrote that I got a D* (line 30). His-announcement comes without any pre-announcement actions, and by using the formulation *I wrote* (as opposed to other alternative formulations such as *I think, I believe*, or even *mine is definitely*…), his selection of *I wrote* puts some distance between himself and the self-assessment – he reports on a past action rather than makes a claim *in situ*. There is a notable silence (line 31) before Britta initiates repair saying *ha?* (Swe) with rising intonation (Swe. "ha" is difficult to translate, but is understood similar to "what?" in English) (line 32). Repair is a mechanism for dealing with problems in production, hearing, or understanding (Hayashi, Raymond & Sidnell, 2013), and Wilkinson and Kitzinger (2006) have shown how displays of surprise are sequentially formed as interactional achievements. Such displays of surprise reveal "norms, values and expectations in a taken-for granted world" (p. 178). In this case, Britta's repair initiation works as to display 'surprise' at Calle's announcement – whether the surprise display is directed at his abrupt announcement or his self-assessment is unclear at this point. Calle does not treat Britta's ha? as a repair initiator, but instead, silence follows (line 33). Dan, then, explicitly holds Calle accountable for his self-evaluation in a question format: *how do you know that* (line 34). Histurn questions Calle's degree of epistemic certainty in the matter (Stivers, Mondada & Steensig, 2011) and requests Calle to provide evidence for his knowing stance. Britta and then Anna both respond to Dan's question with short laughing sounds, with no overlap in their laugh particles. In their responses to Dan, the three other students seem to treat his assessment as unexpectedly high, and as a form of self-praise, where the laughter marks the delicacy of the moment succeeding Dan's question, which is pressing Calle to account for his assessment (lines 35–36). Calle responds immediately, with certainty or even some degree of 'stubbornness' in his voice (for lack of a better technical description), asserting his knowledge in the matter at hand (*cause I know that*, line 37). Note that he does not offer any further accounts for how he came up with the D grade. Next, Britta and Anna seem to align with Dan's questioning stance and disqualify Calle's claim by initiating playful hypothetical assessments of their own answers. Both, in consecutive turns, claim that as a result of Calle's certainty and a lack of a valid account for it, they can be equally sure that they would get even the highest grade A (*yes ha I will get an A then* and *mm me too*, lines 38–39). In overlap with Anna, Calle refuses to give any further account in relation to the grade criteria and claims that his knowledge is based on a talent to look into the future (line 40).

In contrast to Excerpt 1, which was initiated with an overt negative self-evaluation (*mine sucked,* line 1), Calle announces his self-assessment in the form of a D grade, which is one step above the passing limit (i.e., an E). While a D is not automatically a positive assessment, co-participants indeed seem to hear Calle's self-assessment as doing at least some degree of self-praise, as he claims a grade for which he has yet to provide evidence in support. As noted also in relation to Excerpt 1, self-praise is structurally dispreferred when delivered as a first pair part (Pomerantz, 1978; 1984; Schegloff, 2007), since "the subject of the self-praise (the speaker) and the object (the thing being praised) is the same" (Speer, 2012, p. 56; see also Edwards, 2005; Skovholt et al., 2019). As such, self-praise is an interactionally risky matter, and studies have shown an interactional preference for embedding self-praise in relation to a third party attribution (Speer, 2012). A speaker delivering self-praise without support from some other person's judgement or some objective criteria, is often expected by a recipient to provide additional 'evidence', which is what we observed in line 34. The marked silence before Dan's direct question, together with Britta and Anna's laughter and subsequent joking about de-

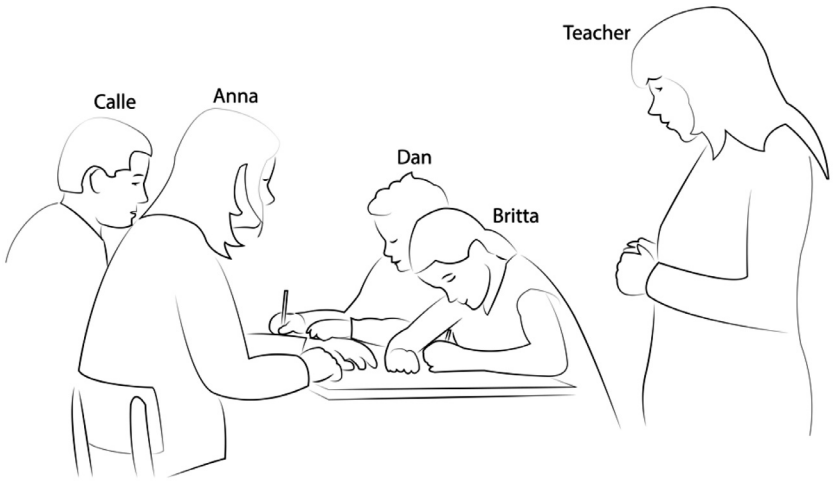


**Fig. 4.** Configuration at the outset of Excerpt 3.

serving an A if Calle deserves a D, also reveal something about the delicacy of responding to self-praise. Britta's and Anna's exaggerated claims appears to function similarly to teasing activities, an interactional resource that tend to occur in sequential environments where a participant has been overdoing complaining, bragging, or elaborating on something (Drew, 1987; Sandlund, 2004).

Overall, Excerpt 2 reveals the social delicacy of making positive self-assessments in peer groups, where the volunteering of a higher assessment than the baseline E grade is risky: It makes the speaker accountable for providing evidence, and vulnerable to joking and teasing from co-participants. In addition, co-participants are put in a difficult situation about how to respond in relation to a (too) high evaluation. As such, our analysis supports those of Skovholt et al. (2019), who conclude that "it is easier to assess oneself negatively than to assess oneself positively" as "students orient towards the constraints against self-praise" (p. 53). Furthermore, we can note that for recipients, it seems easier to respond to self-deprecations (as in Excerpt 1) than to self-praise (as in Excerpt 2).

Later in the same lesson, the teacher moves between the groups to follow up the assessment discussions. In Excerpt 3, the teacher (Fig. 4) approaches the group we saw in Excerpt 2, but a bit later in their interaction.

As the teacher comes up to the student desks, she addresses the students as a group (line 1), asking if they did *end up with something* (line 1). The question targets the task, and bears an assumption that the students have arrived at assessments of their answers. After a 0.9 second gap, Britta and Anna respond, displaying that they orient to the polar question as a request for announcing their assessments (rather than just providing the structurally relevant *yes* or *no*). In overlap, both initiate their turns with *I think* (lines 3 and 4). The use of "I think" here should not automatically be interpreted as an individual's account of an internal mental state of uncertainty, but rather, as a linguistic marker for displaying a particular epistemic stance in relation to co-participants (Kärkkäinen, 2006). Stancetaking in conversation is an intersubjective activity related to asymmetries in terms of authority, responsibility, and access to knowledge in specific contexts (c.f. Kärkkäinen, 2006; Stivers et al., 2011), and here, the students use of "I think" can be understood as positioning themselves as less knowing that the teacher who has the epistemic authority with regards to assessing the accuracy of their assessments of their test answers. As such, the students' "I think" work to downplay claims of accuracy on their upcoming assessments. While Britta moves out of the overlap, Anna presents her suggestion for a grade, *I think I got an E,* and immediately provides support for the assessment, *because I- I didn't give that much like reasons,* (lines 4–5). Her account is referring to the grading criteria as a common ground as they were introduced to whole class prior to the discussions, where the teacher made clear that the passing grade E required one piece of advice accompanied by at least one solid medical reason. "Reasons" is a keyword in the criteria, and in overlap, the teacher and Anna simultaneously complete Anna's turn with the word *reasons* (line 5 and 6). The teacher's *reasons* is followed by a *no,* which is heard as a confirmation that "reasons" is the correct grounds for the evaluation. In overlap with the teacher's *no*, Britta adds her own assessment by aligning with Anna's (*same here,* line 7).

| | | |
|---|---|---|
| 1 | Lär: | kom ni fram till nånting? |
| | | **did you end up at something?** |
| 2 | | (0.9) |
| 3 | Britta: | [jag tro- ] |
| | | **[ I thi- ]** |
| 4 | Anna: | [jag tror] jag fick E för ja- liksom gav inte så |
| | | **[ I think ] I got an E cause I- like did not give** |
| 5 | | mycket typ [motiveringar] |
| | | **that much like [ reasons ]** |
| 6 | Lär: | [motiveringar]↑ [nej] |
| | | **[ reasons ]↑ [no ]** |
| 7 | Britta: | [sam]ma här |
| | | **[sam]e here** |
| 8 | | (1.6) |
| 9 | Dan: | ja tror jag också fick E¿ |
| | | **I think also that I got an E¿** |
| 10 | Lär: | **mm:** |
| 11 | Dan: | för jag [ gav inte så jättemycket. ] |
| | | **cause I [did not give that very much]** |
| 12 | Calle: | [ ja tror jag fick C ] men jag |
| | | **[ I think I got a C ] but I** |
| 13 | | vet inte |
| | | **don't know** |
| 14 | Britta: | nej jag vet inte heller |
| | | **no I don't know either** |
| 15 | Lär: | nej↑ |
| | | **no↑** |
| 16 | | (1.1)**((teacher moves behind Anna and puts herself beside Calle))** |
| 17 | Lär: | men bara man funderar lite grann. |
| | | **but just you think a little.** |
| 18 | Lär: | få se Calle (0.4)((läser)) vila dig? |
| | | **let me see Calle (0.4 )((reads))rest yourself?** |
| 19 | Britta: | hur gjorde du här. ((till Anna)) |
| | | **how did you do this. ((to Anna))** |
| 20 | Dan? | (inaudible) |
| 21 | Britta: | gjorde du så |
| | | **did you do like that** |
| 22 | Calle: | **mm ((teacher looks at Calle's answer))** |
| 23 | | (1.1) |
| 24 | Lär: | <njae:> ett E |
| | | **<we:ll> an E** |

**Excerpt 3.** Did you end up at something?

Two out of four students have now announced their self-assessments, and after a hearable silence of 1.6 s, Dan volunteers a third announcement, which aligns with those of Anna and Britta, *I think I also got an E* (line 9), delivered with an emphasis on *also* and a turn-final intonation rise. The teacher's minimal acknowledgement (line 10) projects that a continuation from Dan is expected, and Dan aligns by offering an account for his E grade *cause I did not give that very much* (line 11). The lexical choice of *give* indexes that he also orients his response to the criteria of "giving" reasons, which provides further support to the solidity of his assessment. Calle, then, overlapping Dan's account, reveals the outcome of his analysis in line 12. Note that this is the same student who was challenged for claiming a D grade in Excerpt 2, and this time, Calle proposes a higher grade for his performance: *I think I got a C but I don't know* (lines 11–12). The epistemic formulation added, *I don't know,* works to hedge the accuracy of his assessment. This time, none of his peers treat his assessment as self-praise, likely because his turn is a second pair part to the teacher's question who is also a primary recipient of his announcement and expected to have epistemic authority to judge. The teacher does not immediately respond, and Calle's proposal is followed by Britta's turn, who also adds an epistemic marker of uncertainty (*no I don't know either,* line 14). Previous work on "I don't know" as an epistemic marker for insufficient knowledge points out how although it can accomplish a variety of different actions, it always indicates some kind of problem with the answerability of the question (Beach & Metzger, 1997; Iversen, 2014; Kärkkäinen, 2006; Keevallik, 2011). In this case, marking their assessments as uncertain show students' orientation to epistemic asymmetries, which is in line with previous work on how claims of "no knowledge" is tied to the sequential contingencies and social ascriptions of knowledge between participants (Keevallik, 2011). The students' use of "I don't know" can also be heard as inviting a second assessment or confirmation/rejection of their assessments from the teacher.

In line 15, the teacher takes the floor with a clearly pronounced *no*, produced with an upward intonation, which seems to confirm that she has understood their uncertainty and expectation of next actions from her. This is followed by a silence, during which the teacher sits down next to Calle so that she can read his answer. Her *but just you think a little* (line 17) invites the students to reflect further. Her use of the generic third person pronoun *man* shows that her turn is not directed to Calle specifically, but functions as a general appeal to the students to step back and reflect on the assessment criteria in relation to their answers. Using Calle as an illustration, she allows herself access to his written test answer (line 17), and after a short silence where she gazes at his piece of paper, she appears to read aloud from his text: *rest yourself* (line 18) (presumably the advice Calle has provided to the fictional peer). There is a parallel interaction between the other students while the teacher is reading Calle's answer (lines 19–21), after which Calle's minimally acknowledges the teacher's formulation (line 22). After marked gap (line 23), the teacher produces her assessment of Calle's answer: *we:ll (*Swe. *njae,* a response particle combining no and yes*) an E* (line 24). Her turn-initial, disagreeing particle is produced slowly, with the shape of a dispreferred turn. As such, having inspected hos written answer, the teacher shows disagreement with Calle's self-assessment and downgrades it two steps – from his suggested C to an E. While the interaction shows participants orienting to the teacher's authority in assessing students' assessments, we can also see that the preference for agreement with first assessments makes downgrading the students' suggestion of a specific grade a delicate matter.

Compared to the first two excerpts that showed student-student interactions, the interaction in this third example is more asymmetric as the responsibility for delivering a second, confirming or downgrading, assessment is assigned to the teacher. Heritage and Raymond (2005, p. 16) point out that second assessments inevitably make relevant questions of relative epistemic rights to evaluate the state of affairs, and in everyday conversation, it is usually the first assessor who claims primary rights to assess. However, in this institutional interaction, it is clearly the teacher who is treated as the one having epistemic primacy (Stivers et al., 2011, p. 9), while students' preceding announcements are delivered with try-marked intonation (Sacks & Schegloff, 1979, p. 18), epistemic markers of uncertainty (such as *I think* or *I don't know*), and notable gaps. In the institutional context of the classroom, teachers are expected to evaluate student contributions, usually in a third turn in so called IRE-sequences (Mehan, 1979), which also involves emotion work and affective stance taking (Tainio & Laine, 2015). This also seems to be the case when the student performance consists of presenting self-assessments. In our data, the teacher's second assessments mostly confirm the student displays, with reference to their shared knowledge of the criteria. But on occasion, as was the case with Calle, the teacher's second assessment downgrades the student's self-assessment. This becomes more interactionally troublesome, evidenced in the teacher's mitigation of the displayed disagreement through a marked silence followed by the partially disconfirming particle "njae"(lines 23–24), which reveals that she treats the downgrades as potentially face-threatening for the students.

In line with what previous research on older students have shown (Attenborough & Stokoe, 2012; Skovholt et al., 2019), assessing own performance is at least equally interactionally demanding for younger students. The three examples from the students' perspective illustrate the social delicacy inherent in publicly making visible self-evaluations, where our analysis focuses on the very act of having to announce and account for a specific grade level in relation to given criteria. In our analysis we have showed how students, in different ways, manage interactional tensions and challenges in the local context of delivering self-assessments. The delivery of self-deprecations seem more unproblematic, such as where a student's low self-assessment yields affiliation, and contributes to an alliance-forming against a third-party (the teaching). In cases where a student makes a claim for a an assessment above the minimal requirement, as in the second example, the assessment is treated as an accountable action, or even as hints at self-praise, and occasions demands for accounting, and subsequent dismissal of the claim through teasing and joking. Finally, our third excerpt showed participants' orientations to the epistemic asymmetry in teacher-student interactions, where the students self-assessments are delivered with disclaimers of uncertainty and where the teacher's downgrade in the teacher's second assessment is produced as a dispreferred action, despite the differing institutional roles of teachers and students.

### 4.2. Teachers-as-raters assessing student performance

We now turn to our second empirical context: teachers revealing their professional judgements of student performance to colleagues. In our analysis of the four video-recorded moderation meetings, we have examined all sequences in which teachers-as-raters, in groups, *reveal to each other their pre-marking grades* of the two learners. These announcement sequences generally take place late in the meeting after the teachers have discussed, in general terms, the scoring rubrics in relation to the two learners. The fact that the revelations of grades takes place late in the rater discussions is not entirely surprising, as each group has been instructed to fill out a joint assessment form with a consensus grading decision before breaking up the meeting, and reaching agreement on a joint grade occasions the announcing of individual grading preferences (see Sandlund & Sundqvist, 2019). For the sake of the present

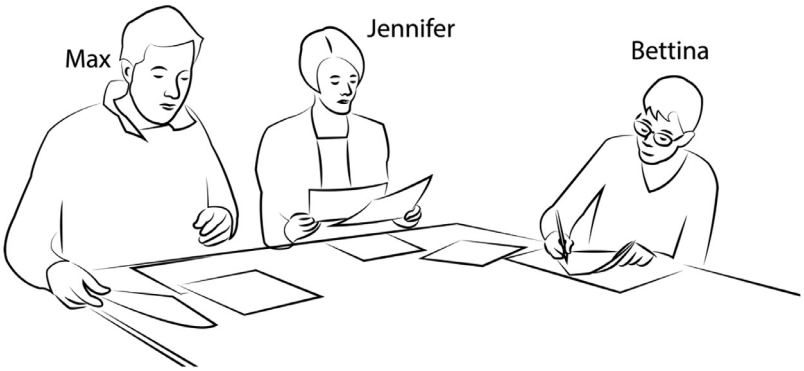


**Fig. 5.** Configuration teachers' moderation meeting.

paper, and because the trajectories of revealing individual assessments are often extended, we have selected one of these four sequences for presentation. In this sequence, we demonstrate how a first grade revelation is treated as accountable by co-participants, and how second and third revelations are stalled, as it gradualy becomes evident that recipients disagree with the first (high) assessment revealed.

The excerpted sequence takes place late in the moderation meeting (Fig. 5), but no candidate letter grades (A to F) have been specified up to this point. Overall, the teachers (in this case, Jennifer [JEN], Max [MAX] and Beth [BET]), have expressed that they are not entirely satisfied with the degree of paired interaction between the two learners, and they have been collaboratively complaining about what in conversation analytic terms could correspond to a lack of post-expansion (cf. Sandlund & Greer, 2020; Sandlund & Sundqvist, 2019) on prior turns: that is, that the students do not engage in or build on each other's responses enough. They have also agreed that the test-takers rely too much on the co-present teacher's prompts, rather than volunteering topical contributions. With this prior talk in mind, it seems unlikely that the raters would expect announcement of the highest grade (A) from their peers. However, as pass grades range from A to E (F as a fail grade), there is still enough of a spectrum of grades to be assigned between B and E to yield a degree of unexpectedness in the revelation of individual assessments. Immediately prior to the sequence in focus (lines 1–16), Beth has initiated talk on procedural matters, where the teachers discuss how to proceed with the collaborative grading and whether they should grade each assessment factor before assigning a holistic grade. As the sequence continues, Max self-selects in moving the procedure forward (line 17 below) with a hearable orientation to the grading criteria.

In line 18, Max specifically references the fact that for the grades E, C, and A, the scoring rubrics provide detailed descriptions of knowledge and abilities that a learner should have at each of the grade levels.[3] Max' turn is produced rather slowly, with an intra-TCU pause and elongated vowels, as if carefully deciding where he is heading with his turn during its production. Beth provides an acknowledgement token (line 19) with an intonation that shows that she is expecting Max to proceed. In line 20, Max initiates an important next step: the indication of a first assessment of the two learners in relation to the grading criteria. With his *anhdu:h (.) I think that it's e: (.) e:h here the question is (.) in both ↑cases th- that e:: one + of the high↑er grades,* Max narrows down the scope of his individual assessment decision as being a grade above the 'middle' – which presumably means no lower than a C for both learners. The entirety of his turn is produced rather slowly, with a tentative, try-marked intonation, as if testing the waters of the others'

| | | |
|---|---|---|
| 17 | MAX | [m: ] nämen så vi har ju kriterierna<br>**[m: ] nowell so we havethe criteria** |
| 18 | | här för (1.1) för e: ce: a:<br>**here for (1.1) for e: ce: a:** |
| 19 | BET | m::, |
| 20 | MAX | oche:h (.) ja tycker ju att det e: (.) e:h<br>**anhdu:h (.) I think that it's e: (.) e:h** |
| 21 | | här e frågan (.) i båda ↑fallen a- att e::<br>**here the question is (.) in both ↑cases th- that e::** |
| 22 | | något +av dom hög↑re betygen.<br>**one + of the high↑er grades.**<br>+gaze up to BET and JEN |
| 23 | BET | m:::,= |
| 24 | JEN | =∘m::∘ |
| 25 | | (0.6) |
| 26 | MAX | d- (.) e::h (.) [( x x )<br>**d- (.) e::h (.) [( x x )** |
| 27 | BET | [hu:r högt då: >tänker du<<br>**[ho:w high then >are you thinkin'<** |
| 28 | MAX | ↑↑ja::: heh heh ↑NU:: .HHH spä+nnande<br>**↑↑ye::ah heh heh ↑NO::W .HHH exc+iting**<br>+gaze down |
| 29 | BET | ∘**hhHHHeh**∘ |
| 30 | JEN | ↑↑**hhh[HEH HEH HEH HEH HEH** |
| 31 | BET | **[mmHEH HEH** |
| 32 | MAX | ja: (.) e:hm (2.5) ja::¿ (1.8) e::h, (0.9)<br>**ye:ah (.) e:hm (2.5) ye::s¿ (1.8) e::h, (0.9)** |

**Excerpt 4a.** "One of the higher grades" [CASS_4].

agreement or disagreement while moving along with the interactional project at hand.

Following upon Max' partial revelation, there is only minimal uptake from Beth and Jennifer (lines 23–24) and a slight pause, which could indicate an upcoming disagreement. Max' possible expansion in line 26 could be the initiation of an account for his prior claim, but it is overlapped by Beth's request for a specification in line 27, which holds Max accountable for revealing a more exact grade in the sweeping continuum provided earlier (*ho:w high then >are you thinkin'*). While prompting Max to be more specific, she withholds her own assessment and/or agreement with Max' initial assessment). Both Jennifer and Beth now have their gaze on Max.

Max, however, does not respond in alignment with Beth's request, but instead produces an assessment, with interspersed laughter, of the current interactional moment: *↑↑ye::ah heh ↑NO::W .HHH exciting*, which occasions laughter (lines 28–31). By

[3] For the grades in between, i.e., B and D, the instructions for assessment state that, e.g. 'most but not all' of the criteria for the grade above should be reached.

formulating the moment as *exciting*, Max keeps the group in suspense while also displaying that what he is being asked to do is interactionally delicate. This is further evidenced in his gaze shift down to the table and away from his co-participants. As research on the organization of laughter in social interaction has shown, invitations to laugh are frequently found in talk about troubles (Jefferson, 1984), in establishing conversational rapport and intimacy (Jefferson, Sacks & Schegloff, 1987), in displaying affiliation/disaffiliation (Drew, 1987; Glenn, 1995) and in dealing with delicate topics (Haakana, 2001). While his co-participants indeed take up the invitation to laugh *with* Max, the momentary tension continues in line 32, as Max continues his attempt to reveal an assessment through consecutive acknowledgement tokens, intra-turn pauses, and hesitation markers.

However, Max's revelation of an exact assessment is further stalled, which is shown in Extract 4b below.

| | | |
|---|---|---|
| 33 | MAX | ja tycker<br>**I think** |
| 34 | | ju:: atte:h (.) båda två e::h (1.5) >båda två:<br>***ju:: thate:h (.) both e::h (1.5) >both*** |
| 35 | | *e* tycker ja uppnår kriterierna som<br>***e* I think reach the criteria in place** |
| 36 | | gäller för betyget c↑e:<br>**for the grade c↑e:** |
| 37 | BET | °m:° |
| 38 | MAX | på på ga: e::h (0.5) på samtliga delar<br>**on on ga: e::h (0.5) on all parts** |
| 39 | | (0.3) >skulle ja<br>**(0.3) >I would <** |
| 40 | | säja (0.7) sedan så (1.1) e:h (.) kan det va<br>**say (0.7) then (1.1) e:h (.) there can be** |
| 41 | | lite (.) som drar ne:r¿ att man<br>**a little (.) that pulls it down so they** |
| 42 | | inte +når upp till a:<br>**don't +reach an a:**<br>+gaze to BET |
| 43 | BET | °mm:° +gaze down |
| 44 | MAX | s'att e::h<br>**so that e::h** |
| 45 | | (4.1) |
| 46 | MAX | ja vi kan väl gå < ig+en↑om> HHHUH HUH [huh ]<br>**yeah why don't we <go +ov↑er> HHHUH HUH [huh ]** |
| 47 | BET | + gaze down [JO:h] |
| 48 | MAX | nu har ja:? sagt va ut(hh)gån(h)gs-(hh)<br>**now I: ? have said wha- the s(hh)tarting point (hh)** |
| 49 | | va(hh) ja(hh) [har- ] va(hh) utgått- .HHH<br>**wh(hh)at I(hh)[have-] what started- from .HHH** |
| 50 | BET | [ja:h ] [m: ]<br>**[ye:ah] [m: ]** |
| 51 | JEN | [m:? ] [m:?] |
| 52 | JEN | ja:<br>**ye:ah** |
| 53 | MAX | lite ja vet ente om ni- håller ni <mä:<br>**a little I dunno if you- do you <agree: with** |
| 54 | | mej> + el[ler<br>**me> + o[r** |
| 55 | JEN | [D-de<br>**[i- it**<br>+gestures |

**Excerpt 4b.** "Now I have said" [CASS_4].

In lines 33–38, Max builds on his prior partial disclosure of "one of the higher grades", and makes a claim that both learners, in his view, reach the criteria descriptors for the grade C "on all parts" (line 38). Continuing his turn, Max then narrows down the scope of his assessment further by accounting (rather unspecifically) for how the learners' talk displays features that disqualify them from the highest grade (A). As a consequence, Max has made public his assessment of each learner as somewhere between a C and a B (see lines 40–42).

Max's talk in the first half of Excerpt 4b is produced with repeated pausing, prolonged vowel sounds, and in a relatively slow pace, as if he is carefully weighing his words in accounting for the scope of his assessment. An abandoned upshot initiation (line 44, *so that e:h*) is followed by an extended silence (line 45), and two visible shifts in gaze from Beth – away from Max and down to the documents in front of her. The silence and the lack of uptake or second assessments from Jen and Beth clearly signal interactional trouble: when first assessments are produced, second assessments are normatively expected (cf. Lindström & Mondada, 2009; Pomerantz, 1984). When no such second assessment is forthcoming, potential disagreement with the first assessment is projected (cf. Pudlinski, 2008). As such, Max now finds himself in a position where he has initiated the interactional project of delivering a first assessments, but where his co-interactants do not display alignment with his project, nor do they affiliate publicly with the stance taken in Max's extended turn.

As neither Beth nor Jen utilize the transition points made available up to line 45, Max shifts the stalled assessment activity into a procedural suggestion in line 46: *Yeah why don't we ⟨go ov↑er⟩ HHHUH HUH [huh.* His-proposal entails not only a procedural next step – to *go over* either the criteria or their respective assessments, but also make the others accountable for participating in the assessment revelation process. As such, Max displays a noticing of the lack of uptake and acknowledgement, and implicitly also invites Jen and Beth to contribute to the task at hand. The turn-final laughter has the characteristics of laughter in the context of troubles-talk (see Jefferson, 1984): it can be heard as Max distancing himself from the delicacy at hand and that "he is managing; he is in good spirits and in a position to take the trouble lightly" (1984, p. 351). Evidence to the fact that this is also how it is heard by Jen and Beth is that they do not reciprocate his laughter, which is the preferred way of handling displayed trouble (cf. Sandlund, 2004; Jefferson, 1984). Max's efforts still do not generate any reciprocal action from co-participants. As a result perhaps, Max's turn in line 48–49 explicitly pinpoints this accountability by stating that he has done his part, so to speak: *now I: ? have said wha- the s(hh)tarting point (hh) wh(hh)at I(hh)[have-] what started- from .HHH.* Despite this second attempt to invite second assessments and actions, Beth and Jen still only provide minimal acknowledgement response particles (lines 50–52), so the interactional trouble remains. Consequently, Max is left trying to determine whether this lack of affiliation indeed signals upcoming disagreement. In line 53, this is exactly what Max addresses. In a turn, explicitly acknowledging that he does not know whether they agree with him, he ends with a turn-final *or,* which relaxes the preference for an agreeing response and opens up for disagreement as well (cf. Drake, 2015). As we move in to the fourth and final part of this 'revelation' sequence, Jen initiates a response in line 55, which continues in line 56 (Excerpt 4c).

Still not revealing her own viewpoint, Jen asks a confirmation question (line 56), which orients to the basis for Max's assessment revelation. Her *what you had put there spontaneously in there* appears to check whether what Max has indicated was his 'spontaneous' assessment of the two learners (*in there* most likely referring to the room the teachers were in when they listened to the

| | | |
|---|---|---|
| 56 | JEN | dä: du hade satt där alltså spontant där inne<br>**what you had put there spontaneously in there** |
| 57 | | men[ar du]<br>**you[mean ]** |
| 58 | MAX | [ja:? ](.) [a:?]<br>**[ye:s?] (.) [a:?]** |
| 59 | JEN | [ja ]<br>**[yes]** |
| 60 | MAX | ja spontant där inne så (.) skulle jag<br>**yeah spontaneously in there (.) I would** |
| 61 | | säja *nn* nånstans, +tils head back and forth<br>**say *nn* somewhere,** |
| 62 | JEN | ja:? (.) ja ↑hade nog:,<br>**ye:ah? (.) I ↑had probably:,** |
| 63 | MAX | ce: be:<br>**cee: bee:** |
| 64 | JEN | ja ja hade- ja hade ö:h (.) de: [de: ] och<br>**yeah well I had- I had e:h (.) dee: [de: ] and** |
| 65 | MAX | [okej]<br>**[okay]** |
| 66 | | be:<br>**bee:** |
| 67 | JEN | ja<br>**yeah** |
| 68 | MAX | ja du >hade de: och be:=du mhm?<<br>**oh you >had dee: an' bee:= you mhm?<** |
| 69 | JEN | m: |
| 70 | BET | m: |
| 71 | MAX | m: |
| 72 | JEN | det va mina spontana ja:<br>**those were my spontaneous ye:ah** |
| 73 | MAX | m: |
| 74 | BET | .hja.hh jag hade också mest de:<br>**.hja.hh I also had mostly dee:** |
| 75 | | [de: och ] ce:<br>**[dee: and ] cee:** |
| 76 | MAX | [ja du hade de.]<br>**[oh you did.]** |

**Excerpt 4c.** "What you had put there spontaneously"[CASS_4].

test recording and filled out a form with their initial ratings and notes). Max confirms Jen's understanding (line 58, 60), also recycling her evaluative description *spontaneously.* The choice of a descriptor here is relevant, we believe: by casting their individual assessment work as spontaneously conducted, the assessments are marked as tentative and intuitive rather than firmly grounded in the scoring rubrics as a spontaneous assessment is open to further revision.

In lines 60–63, Max finally mentions the letter grades previously indicated more indirectly. In confirming Jen's understanding, Max still indicates a range of grades (*I would say *nn* somewhere ce: be:*), but does explicitly state that there are only two grades on the six-grade scale left to consider. In line 62, just prior to Max's mention of the grades C and B, Jen initiates her own assessment revelation, but clearly positions it in the same temporal frame as her confirmation question, using the past tense (*I had*) to mark that she too is drawing on her first, initial assessment prior to the joint rater meeting. She continues her abandoned turn in line 64 and 66 with her own revelation: she had assigned a D and a B, respectively. Max receives this announcement by repeating Jen's assessment embedded in a confirmation-type action, with a prosody indicating that he treats her announcement as news (line 68) (cf. Heritage, 1984; Wilkinson & Kitzinger, 2006). At the end of his turn, the *mhm?* marks receipt of a new understanding. In line 72, Jen produces an upshot of her announcement, again explicitly casting her grades as *spontaneous*. At this point, we have learned that there is discrepancy on one of the learners between Max and Jen where Max had assigned a higher grade for the weaker learner as compared to Jen.

In line 74, Beth orients to the underlying accountability of producing a third announcement. Her *.hja.hh I also had mostly dee:* and *[dee: and] cee:* reveals an assessment of the two learners even one step lower than the two prior announcements – Beth had not assigned a B grade to either learner, and agreed with Jen on the lower D grade for one of the learners. Again, Max receives the announcement by confirming his uptake of the news (*oh you did.*, line 76). At this point in the interaction, all is revealed, and the disagreement projected in the delays in producing second and third grade announcements is evident.

Having examined the grade revelation process in this group of three experienced English teachers, it is obvious that in this particular case, the revealing of a professional opinion to colleagues is a delicate project. As shown in Excerpts 4a-c, Max went first, and also happened to be the one who suggested a more positive assessment of the learners in the test than his colleagues. We now know that Jen and Beth, on a falling scale, had been more severe in their ratings, and that Max's early indication (Excerpt 4a) of "one of the higher grades" projected an upcoming disagreement. As conversation analytic work has demonstrated over decades of empirical work, disagreement is interactionally dispreferred, and when necessary, it is usually postponed, and produced with hesitation markers, dysfluencies, and particles like *well* (cf. Greatbatch, 1992; Pomerantz, 1984). The withholding of second assessments and next announcements on part of co-participants may at first glance seem uncooperative – and places a great deal of the interactional work on Max. However, given the social dispreference pertaining to disagreement, it is possible to see the lack of 'support' from Jen and Beth in a new light – as sensitive orientations to the delicacy inherent in challenging a professional peer on his or her professional accomplishment. As the potential for disagreement becomes obvious to Jen and Beth early on (i.e. that Max reveals that his grade will be from C and higher), this threat to the social smoothness lingers throughout the sequence. Max's management of this interactional trouble begins with explicitly attempting to place the ball in his co-participants' court, and continues with his gradual display of an insight that they may not agree with his professional opinion, which causes him to open up for disagreement as a possible response. Furthermore, the assigning of Max as the 'first revealer' is occasioned by a discussion of procedural matters, and where Max initiates a description of a general assessment. Once this claim has been made public, Beth treats it as incomplete, as Max has not revealed an exact grade. The potential delicacy of the announcement is further underlined in Max's assessment of the current state of affairs as *exciting*, and Max also orients to the lack of reactions and second assessments as a request for accounts for the grade range presented. Consequently, the extended and trouble-laden grade announcement sequence is collaboratively accomplished and not the outcome of the actions of one participant. However, we find the stepwise and gradual treading toward a letter grade announcement does indicate that participants are not only engaged in some neutral professional activity. Instead, in revealing their professional opinions to each other, they also have to manage and negotiate their professional identities, in which revealing a professional assessment that deviates substantially from co-present colleagues is

risky business and poses a threat to the participants' professional identity in interaction.

## 5. Discussion

Even though research on assessment practices covers a broad spectrum of different activities and perspectives, there is a lack of knowledge on how assessments or grades are discussed, negotiated, delivered, and responded to at the micro-level, in situated interactions between students and teachers. In line with previous conversation analytic work on students' self-assessments in feedback sessions (Skovholt et al., 2019), this paper has focused on some interactional challenges that teachers and students, in situations, face when participating in collaborative assessment talk. While the findings from Skovholt et al. mainly focused on students' self-assessments in formative activities, the conversations examined in this paper center on summative assessment, where the purpose of the conversations, broadly, could be described as activities for developing assessment literacy (Popham, 2009). As such, the interactions examined, for students in the classroom, as well as for the teachers-as-raters, centered on developing shared understandings of given criteria in relation to task performances in national tests. Consequently, the findings of the present study offers a window into how summative assessments (in this case, specifically in relation to performances on national tests) are communicated and accounted for in educational assessment talk.

Our analyses of interactions around the announcement of summative assessments revealed that making a summative claim about a performance entails navigating appropriate levels of self-praise or self-deprecation (students), and the gradual accounting for an initial assessment claim which monitors the reactions of co-participants (teachers-as-raters). The detailed CA analysis shows how both teachers and students in these specific institutional contexts rely on shared members' methods (Garfinkel, 1967) for these balancing acts. We saw how disclosing a specific grade indeed is a judgement on performance, which in turn is treated as a sensitive and socially delicate business. As such, the social concerns – such as the locally displayed identities as raters, students, and classmates – are carefully oriented to in all the sequences examined. As CA work has convincingly shown, positive self-assessments puts the speaker at risk of being accused of bragging (Pomerantz, 1978; Skovholt et al., 2019; Speer, 2012), and we saw how students treated grades above the passing limit as accountable actions, requiring 'evidence' from the claimer of such an assessment. We could also observe how initial self-deprecating assessments were relatively unproblematic, and generated affiliative actions from other students. On the other hand, empirical work has shown that there is a general preference for disagreeing with someone else's self-deprecating talk (Pomerantz, 1984) so that a negative self-assessment is usually upgraded or rejected by co-participants. Both types of assessment can create interactionally delicate moments and require careful treading. However, assessments of performance is at the core of education, and as such, high grades, discursively and implication-wise, carry more value than mediocre or low grades. In this respect, for students there is a double burden of managing conflicting expectations between interactional (peer affiliation) and institutional (high-stakes evaluations) norms.

In our examination of summative assessment announcements in a group of teachers-as-raters, we noted how a first grade assessment delivered by one participant turned out to be higher than the subsequent ones, and that co-participants delayed their disagreement, which also demonstrates the principle of preference organization in operation, as disagreements tend to be delayed, mitigated, and produced in dispreferred formats (Pomerantz, 1984). While the teachers fulfill the task they have been asked to perform in the group, their collaborative accomplishment of the assessment revelations show that interactional challenges are continuously monitored and managed. While students are revealing their own assessment of their performance on a test task, the teachers, conversely, are revealing a glimpse of their professional assessment competence as they disclose their evaluations of the two students' test performance. As such, while the issues making the assessment revelations may be delicate for different reasons, all four examples show that revealing summative assessments has its interactional risks.

In all the examples examined, the collaborative discussions clearly center around epistemic aspects tied to the participants' institutional roles as students and teachers, respectively. As demonstrated previously by conversation analysts, assessments in interaction inevitably entail negotiations about epistemic rights (Heritage & Raymond, 2005). However, these epistemic negotiations are played out differently in the different constellations, where the student-to-student interactions as well as the teacher-to-teacher interaction seem to require more negotiations of the relative epistemic rights in the group as compared to the teacher-student interaction. Hence, features of epistemic asymmetry seem to be salient for how collaborative assessment talk is managed in interactions. In the example involving students and their class teacher, the interaction featured an obvious epistemic asymmetry based on the participants' different institutional positions, where the discussion mainly evolved around the teacher's second assessments of the students' proposed grades. This was observable in frequent mitigations and discursive markers of uncertainty in their announcements, showing that participants' treated this practice as troublesome. As a result, it seems that elaborated discussions about different interpretations did not occur. Discussions between participants on an equal level, where the relative epistemic rights between peers (the students) and colleagues (the teachers) is less explicit, were different in the sense that they resulted in different trajectories for students and for teachers-as-raters. For students, we saw the accomplishment of affiliative work against a third party, as well as playful banter and laughter. For the teachers, it became clear how revealing judgements challenges social and structural preferences for agreement and, in line with the findings of Raclaw and Ford (2017), the potential threat of differing views resulted in a lengthy revelation sequence where the lack of displayed agreement yielded accounting, and finally, the revelation of disagreement.

An interesting observation across the two datasets is that for the student groups, going too low rather than too high appears interactionally safer, as a high initial assessment produces a risk of being held accountable for self-praise. For teachers in this particular case, rater leniency (a higher grade in this case) seems to require more accounting than rater severity when the second and third grade announcements are lower than the first (cf. Sandlund & Sundqvist, 2019). As such, it appears as if across our data, high assessments of performance are treated as more interactionally problematic. While this is understandable in light of constraints against self-praise for students, something else is at play in the interactions between teacher professionals. Whether it is the disagreement on evaluations as such, or that rater severity is equated with a higher degree of professionalism (cf. (In press)) remains to be further examined in other contexts. We also saw how participants manage the risks of disagreement, and the displays of professional assessment knowledge displayed in grade disclosures, through interactional preparatory work and the stepwise manner in which opinions are revealed. In line with previous research (c.f. (Marra, 2012; Matre & Solheim, 2016; Pier et al., 2018; Raclaw & Ford, 2017)), we find that disagreement in moderation activities is challenging and that the interactional measures that participants take do the work of mitigating the differences.

To conclude, our findings reveal that the great faith in collaborative assessment practices and in assessment literacy-promoting activities, for both teachers and students, is challenged in terms of their effectiveness by social and interactional norms. Announcing a certain score is not a neutral claim, but for teachers a professional judgement (Allal, 2013), which can put issues of epistemic authority and professional identity at stake. In the student example, making self-assessments involves the risk of either displaying your faults and weaknesses or of being accused of bragging. This finding is in line with Attenborough and Stokoe's (2012) observations on how "students routinely avoid positive assessments of their tasks" (p. 11) in efforts to avoid positioning themselves as either too clever or too much below average. The detailed investigation of the interactional challenges that these kind of sequences invoke shows that social norms seem to override, or at least complicate, institutional objectives of developing assessment literacy through assessment practice activities. These challenges we point at do not rule out that these kinds of collaborative discussions could be important as learning opportunities. What we have demonstrated, however, is that collaborative assessment practices designed to improve assessment literacy and learning may not necessarily generate the intended outcomes as various interactional concerns play a central role in publicly revealing claims about performance. As such, our findings call for a greater awareness of the interactional constraints at play in such collaborative assessment so that the time invested in them is warranted in relation to the purpose and expected outcome – for students as well as for teachers.

## References


Adie, L. E., Klenowski, V., & Wyatt-Smith, C. (2012). Towards an understanding of teacher judgment in the context of social moderation. *Educational Review, 64*(2), 223–240. https://doi.org/10.1080/00131911.2011.598919.

Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and socially situated practice. *Assessment in Education: Principles, Policy & Practice, 20*(1), 20–34. https://doi.org/10.1080/0969594X.2012.736364.

Asmuß, Birte (2008). Performance appraisals: Preference organization in assessment sequences. *Journal of Business Communication, 45*(4), 408–429.

Atkinson, J., & Heritage, J. (1984). *Structures of social action. Studies in emotion and social interaction*. London: Cambridge University Press.

Attenborough, F., & Stokoe, E. (2012). Student life; student identity; student experience: Ethnomethodological methods for pedagogical matters. *Psychology Learning and Teaching, 11*(1), 6–21. https://doi.org/10.2304/plat.2012.11.1.6.

Beach, W., & Metzger, T. (1997). Claiming insufficient knowledge. *Human Communication Research, 23*(4), 562–588.

Black, P., & Wiliam, D. (2009). Developing a theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5.

Bloxham, S., Hughes, C., & Adie, L. (2016). What's the point of moderation? A discussion of the purposes achieved through contemporary moderation practices. *Assessment & Evaluation in Higher Education, 41*(4), 638–653. https://doi.org/10.1080/02602938.2015.1039932.

Borger, L. (2019). Assessing interactional skills in a paired speaking test: Raters' interpretation of the construct. *Apples – Journal of Applied Language Studies, 13*(1), 1511–174.

Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research, 3*, 22–30.

Clayman, S. E. (2002). Sequence and solidarity. In E. J. Lawler, & S. R. Thye (Eds.), *Advances in group processes: Group cohesion, trust, and solidarity* (pp. 229–253). Elsevier Science.

Clayman, S. (1998). Gatekeeping in action: Editorial conferences and assessments of newsworthiness. *American Sociological Review, 63*, 178–199.

Clift, R. (2016). *Conversation analysis*. Cambridge: Cambridge University Press.

Drake, V. (2015). Indexing uncertainty: The case of turn-final 'or'. *Research on Language and Social Interaction, 48*(3), 301–318. https://doi.org/10.1080/08351813.2015.1058606.

Drew, P. (1987). Po-faced receipt of teases. *Linguistics, 25*, 219–253.

Edwards, D. (2005). 'Moaning, whinging and Laughing: The subjective side of complaints. *Discourse Studies, 7*(1), 5–29.

Erickson, G. (2009). Nationella prov i engelska – en studie av bedömarsamstämmighet [national tests in English – a study of inter-rater consistency . Retrieved from http://www.nafs.gu.se/publikationer/.

Gardner, J. (2006). Assessment and learning. Introduction. In J. Gardner (Ed.), *Assessment and learning* (p. 1–8). London: Sage.

Gardner, R. (2019). Classroom interaction research: The state of the art. *Research on Language and Social Interaction, 52*(3), 212–226. https://doi.org/10.1080/08351813.2019.1631037.

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs: Prentice Hall.

Glenn, P. J. (1995). Laughing at and laughing with: Negotiations of participant alignments through conversational laughter. In P. ten Have, & G. Psathas (Eds.), *Situated order. Studies in the social organization of talk and embodied activities* (p. 43–56). Lanham, Maryland: University Press of America.

Grainger, P., Adie, L., & Weir, K. (2016). Quality assurance of assessment and moderation discourses involving sessional staff. *Assessment & Evaluation in Higher Education, 2016*. https://doi.org/10.1080/02602938.2015.1030333.

Greatbatch, D. (1992). On the management of disagreement between news interviewers. In P. Drew, & J. Heritage (Eds.), *Talk at work. Interaction in institutional settings* (p. 268–301). Cambridge: Cambridge University Press.

Haakana, M. (2001). Laughter as a patient's resource: Dealing with delicate aspects of medical interaction. *Text, 21*(1/2), 187–219.

Harlen, W., & Gardner, J. (2006). On the relationship between assessment for formative and summative purposes. In *Assessment and learning* (p. 87–102). London: Sage.

Hayashi, M., Raymond, G., & Sidnell, J. (2013). Conversational repair and human understanding: An introduction. In M. Hayashi, G. Raymond, & J. Sidnell (Eds.), *Conversational repair and human understanding* (pp. 1–40). Cambridge: Cambridge University Press.

Heinemann, T., & Traveso, V. (2009). Editorial. Complaining in interaction. *Journal of Pragmatics, 41*, 2381–2384.

Hepburn, A., & Bolden, G. (2013). The conversation analytic approach to transcription. In J. Sidnell, & T. Stivers (Eds.), *The handbook of conversation analysis* (p. 57–76). Malden, MA: Wiley Blackwell.

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson, & J. Heritage (Eds.), *Structures of social action: Studies in conversational analysis* (p. 299–345). London: Cambridge University Press.

Heritage, J., & Raymond, G. (2005). The terms of agreement: Indexing epistemic authority and subordination in talk-in-interaction. *Social Psychology Quarterly, 68*(1), 15–38.

Hofvendahl, J. (2006). Riskabla samtal – en analys av potentiella faror i skolans kvarts-och utvecklingssamtal *[*Risky conversations – an analysis of potential dangers in mentor/parent conversations in school*]* (p. 338). Linköping: Linköping Studies in Arts and Science.

Holmes, J., & Marra, M. (2004). Relational practice in the workplace: Women's talk or gendered discourse? *Language in Society, 33*, 377–398 10.10170S0047404505043035.

Holt, E. (1996). Reporting on talk: The use of direct reported speech in conversation. *Research on Language and Social Interaction, 29*, 219–245.

Iversen, C. (2014). 'I don't know if I shoud believe him': Knowledge and believiability in interviews with children. *British Journal of Social Psychology, 53*, 367–386.

Jefferson, G. (1984). On the organization of laughter in talk about troubles. In J. M. Atkinson, & J. Heritage (Eds.), *Structures of social action. Studies in conversation analysis* (p. 346–369). Cambridge: Cambridge University Press.

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (p. 13–31). Amsterdam: John Benjamins.

Jefferson, G., Sacks, H., & Schegloff, E. A. (1987). Notes on laughter in the pursuit of intimacy. In G. Button, & J. R. E. Lee (Eds.), *Talk and social organisation* (p. 152–205). Clevedon, PA: Multilingual Matters.

Jeffrey, B., & Troman, G. (2011). The construction of performative identities. *European Educational Research Journal, 10*(4), 484–501.

Jönsson, A., & Thornberg, P. (2014). Samsyn eller samstämmighet? En diskussion om sambedömning som redskap för likvärdig bedömning i skolan. *Pedagogisk forskning i Sverige, 19*(4–5), 386–402.

Kärkkäinen, E. (2006). Stancetaking in conversation: From subjectivity to intersubjectivity. *Text & Talk, 26*(6), 699–731. https://doi.org/10.1515/TEXT.2006.029.

Keevallik, L. (2011). The terms of not knowing. In T. Stivers, L. Mondada, & J. Steensig (Eds.), *The morality of knowledge in conversation* (p. 184–206). Cambridge: Cambridge University Press.

Lerner, G. H. (1993). Collectivities in action: Establishing the relevance of conjoined participation in conversation. *Text, 13*(2), 213–245.

Lindström, A., & Mondada, L. (2009). Assessments in social interaction: Introduction to the special issue. *Research on Language and Social Interaction, 42*(4), 299–308. https://doi.org/10.1080/08351810903296457.

Lindström, A., & Sorjonen, M. (2012). Affiliation in conversation. In J. Sidnell, & T. Stivers (Eds.), *The handbook of conversation analysis* (p. 350–369). Malden, MA: Wiley Blackwell.

Lingard, B., Martino, W., & Rezai-Rashti, G. (2013). Testing regimes, accountabilities and education policy: Commensurate global and national developments. *Journal of Education Policy, 28*(5), 539–556. https://doi.org/10.1080/02680939.2013.820042.

Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education, 6*(1), 83–102.

Löfgren, H., Löfgren, R., & Pérez Prieto, H. (2018). Pupils' enactments of a policy for equivalence: Stories about different conditions when preparing for national tests. *European Educational Research Journal, 17*(5), 676–695. https://doi.org/10.1177/1474904118757.

Malabarba, T., & Nguyen, H. thi (2019). Introduction: Using conversation analysis to understand the realities of English-as-a-foreign-language learning, teaching and testing. In H. Nguyen, & T. Malabarna (Eds.), *Conversation analytic perspectives on*

*English language learning, teaching and testing in global contexts* (p. 1–27). Brisol: Multilingual Matters.
Markee, N. (2015). Introduction. Classroom discourse and interaction research. In N. Markee (Ed.), *The handbook of classroom discourse and interaction* (p. 3–19). Malden, MA: Wiley Blackwell.
Marra, M. (2012). Disagreeing without being disagreeable: Negotiating workplace communities as an outsider. *Journal of Pragmatics, 44*(12), 1580–1590.
Matre, S., & Solheim, R. (2016). Opening dialogic spaces: Teachers' metatalk on writing assessment. *International Journal of Educational Research, 80*, 188–203. https://doi.org/10.1016/j.ijer.2016.07.001.
May, L. (2011). Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly, 8*(2), 127–145. https://doi.org/10.1080/154303.2011.565845.
Mehan, H. (1979). *Learning lessons. Social organization in the classroom*. Cambridge, MA: Harvard University Press.
Moss, P., & Schutz, A. (2001). Educational standards, assessment, and the search for consensus. *American Educational Research Journal, 38*(1), 37–70.
Ozga, J. (2009). Governing education through data in England: From regulation to self-evaluation. *Journal of Education Policy, 24*(2), 149–162. https://doi.org/10.1080/02680930902733121.
Panadero, E., Brown, G. T. L., & Strijbos, J. W. (2015). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review, 28*, 803–830. https://doi.org/10.1007/s10648-015-9350-2.
Pier, E. L., Raclaw, J., Kaatz, A., Brauer, M., Carnes, M., & Nathan, M. J. (2018). "Your comments are meaner than your score": Score calibration talk influences intra- and inter-panel variability during scientific grant peer review meetings. *Research Evaluation, 26*(1), 1–4. https://doi.org/10.1093/reseval/rvw025.
Pillet-Shore, D. (2003). Doing "okay": On the multiple metrics of an assessment. *Research on Language and Social Interaction, 36*, 285–319.
Pomerantz, A. (1978). Compliment responses: Notes on the co-operation of multiple constraints. In J. Schenkein (Ed.), *Studies in the organization of conversational interaction* (p. 79–112). London: Academic Press.
Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In J. M. Atkinson, & J. Heritage (Eds.), *Structures of social action. Studies in conversation analysis* (pp. 57–101). Cambridge: Cambridge University Press.
Popham, W. J. (2009). Assessment Literacy for Teachers: Faddish or Fundamental? *Theory Into Practice, 48*(1), 4–11. https://doi.org/10.1080/00405840802577536.
Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator, 46*(4), 265–273. https://doi.org/10.1080/08878730.2011.605048.
Pudlinski, C. (2008). Encouraging responses to good news on a peer support line. *Discourse Studies, 10*(6), 795–812. https://doi.org/10.1177/1461445608098203.
Raclaw, J., & Ford, C. E. (2017). Laughter and the management of divergent positions in peer review interactions. *Journal of Pragmatics, 113*, 1–15.
Richards, K. (2006). *Language and professional identity. Aspects of collaborative interaction*. Basingstoke: Palgrave Macmillan.
Robinson, J. (2020). Revisiting preference organization in context: A qualitative and quantitative examination of responses to information seeking. *Research on Language and Social Interaction, 53*(2), 197–222.
Sacks, H., & Schegloff, E. A. (1979). Two preferences in the organization of reference to persons in conversation and their interaction. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (p. 15–21). New York: Irvington Publishers.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696–735.
Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535–550.
Sadler, D. R. (2013). Assuring academic achievement standards: From moderation to calibration. *Assessment in Education: Principles, Policy & Practice, 20*(1), 5–19. https://doi.org/10.1080/0969594X.2012.714742.
Sandlund, E. (2004). *Feeling by Doing: The Social Organization of Everyday Emotions in Academic Talk-in-Interaction [Diss.]*. Karlstad: Karlstad University Studies.
Sandlund, E., & Greer, T. (2020). How Do Raters Understand Rubrics for Assessing L2 Interactional Engagement? A comparative study of CA- and non-CA-formulated performance descriptors.. *Papers in Language Testing and Assessment., 9*(1), 128–163.
Sandlund, E., & Sundqvist, P. (2019). Doing versus assessing interaction competence. In R. Salaberry, & S. Kunitz (Eds.), *Teaching and testing L2 interactional competence: Bridging theory and practice.* (pp. 357–396). Oxon and New York: Routledge.
Sandlund, E., & Sundqvist, P. (n.d.). Rating and reflecting: Displaying rater identities in collegial L2 English oral assessment. In Salaberry, M.R., & Burch, A.R. (Eds.), Assessing speaking in context Expanding the construct and its applications. Multilingual Matters. In press.
Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.
Sjöberg, G., Silfver, E., & Bagger, A. (2015). Disciplined by tests. *Nordic Studies in Mathematics Education, 20*(1), 101–121.
Skovholt, K., Nordenström, E., & Stokoe, E. (2019). Evaluative conduct in teacher–student supervision: When students assess their own performance. *Linguistics and Education, 50*, 46–55. https://doi.org/10.1016/j.linged.2019.03.001.
Smith, C. D., Worsfold, K., Davies, L., Fisher, R., & McPhail, R. (2013). Assessment literacy and student learning: The case for explicitly developing students 'assessment literacy'. *Assessment & Evaluation in Higher Education, 38*(1), 44–60. https://doi.org/10.1080/02602938.2011.598636.
Speer, S. (2012). The interactional organization of self-praise: Epistemics, preference organization, and implications for identity research. *Social Psychology Quarterly, 75*(1), 52–79.
Stivers, T., Mondada, L., & Steensig, J. (2011). Knowledge, morality and affiliation in social interaction. In I. T. Sviers, L. Mondada, & J. Steensig (Eds.), *The morality of knowledge in conversation. Studies in interactional sociolinguistics 29*. Cambridge, UK: Cambridge University Press.
Swedish Ministry of Education and Research. (2017). Nationella prov – rättvisa, likvärdiga, digitala. *prop. 2017/18:14 [National tests – fair, equivalent and digital. Government proposition, publication number prop. 2017/18:14]*. Stockholm: Swedish Ministry of Education and Research.
Swedish National Agency for Education. (2018). Curriculum for the compulsory school, preschool class and school-age educare 2011 (revised 2018). Retrieved from https://www.skolverket.se/publikationsserier/styrdokument/2018/curriculum-for-the-compulsory-school-preschool-class-and-school-age-educare-revised-2018?id=3984.
Swedish Schools Inspectorate. (2013). Olikheterna är för stora. Omrättning av nationella prov i grundskolan och gymnasieskolan, 2013 *[The differences are too great.* re-*assessing national tests in compulsory and upper secondary school, 2013]*. Stockholm: Swedish Schools Inspectorate.
Tainio, L., & Laine, A. (2015). Emotion work and affective stance in the mathematics classroom: The case of IRE sequences in Finnish classroom interaction. *Educational Studies in Mathematics, 89*(1), 67–87.
Tanner, M., & Pérez Prieto, H. (2020). "...when it is us the tests are made for". Students' argumentations in a performative education system. "...when it is us the tests are made for". Students' argumentations in a performative education system.. *Scandinavian Journal of Educational Research, 64*(6), 886–900. https://doi.org/10.1080/00313831.2019.1639812.
Waring, H. Z. (2014). Mentor invitations for reflection in post-observation conferences: Some preliminary considerations. *Applied Linguistics Review, 5*(1), 99–123.
Waring, H. Z. (2016). *Theorizing pedagogical interaction. Insights from conversation analysis*. New York: Routledge.
Wilkinson, S., & Kitzinger, C. (2006). Surprise as an interactional achievement: Reaction tokens in conversation. *Social Psychology Quarterly, 69*(2), 150–182.