

Assessment in practice: achieving joint decisions in oral examination grading conversations

Marit Skarbø Solem, Anne Marie Dalby Landmark, Elizabeth Stokoe & Karianne Skovholt

To cite this article: Marit Skarbø Solem, Anne Marie Dalby Landmark, Elizabeth Stokoe & Karianne Skovholt (12 Sep 2023): Assessment in practice: achieving joint decisions in oral examination grading conversations, Scandinavian Journal of Educational Research, DOI: [10.1080/00313831.2023.2250380](https://doi.org/10.1080/00313831.2023.2250380)

To link to this article: <https://doi.org/10.1080/00313831.2023.2250380>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 12 Sep 2023.



Submit your article to this journal [↗](#)



Article views: 366



View related articles [↗](#)



View Crossmark data [↗](#)

Assessment in practice: achieving joint decisions in oral examination grading conversations

Marit Skarbø Solem^a, Anne Marie Dalby Landmark^a, Elizabeth Stokoe^b and Karianne Skovholt^a

^aUniversity of South-Eastern Norway, Vestfold, Norway; ^bLondon School of Economics, London, UK

ABSTRACT

How do examiners reach joint decisions when they grade oral examinations? While government and policymakers provide general frameworks about grading decisions, we know little about how they are actually accomplished in interaction, particularly when examiners initially disagree. We scrutinized 29 video-recorded grading conversations between secondary school examiners using conversation analysis. Results showed that proposing and deciding grades involved a stepwise calibration through which examiners adjusted their individual positions. While in most cases examiners expressed and sought agreement, we also investigate cases where examiners initially disagreed but eventually reached a joint decision. The paper contributes insights into decision-making in institutional interaction, as well as to our understanding of whether and how guidance is implemented in 'live' assessment situations. Our findings suggest adjustments are needed both to practice and assessment policy. Data are in Norwegian with English translation.

ARTICLE HISTORY

Received 3 January 2023
Accepted 22 June 2023

KEYWORDS

Assessment; summative assessment; oral examination; teachers' joint decision making; conversation analysis

Introduction

One of the most common and routine practices for schoolteachers is to grade their students' work. Research shows that assessment decision-making is complex; it can be less data-driven and more intuitive (or biased), since teachers often know their students and report the need to support them by awarding highest possible grades (Fjørtoft & Morud, 2021; Sandlund & Sundqvist, 2019; Vanlommel et al., 2021). Assessment also occurs in the complex contexts of both national regulation and local and regional assessment criteria, curricula, learning outcomes, and so on (Fjørtoft & Morud, 2021). Assessment can be even more complex when it involves pairs or teams of teachers who must reach *joint* decisions about grades awarded.

Much is already known about formative modes of assessment practice (e.g., Black, 2013; Skovholt et al., 2021). However, summative assessment (involving a grade or a score) has attracted less attention in research (Black, 2013) and, at least in Norway, has mostly focused on written rather than oral examinations despite both modes counting equally on a student's school diploma (e.g., Berge, 2002; Berge et al., 2017; Solheim & Matre, 2014). In most Scandinavian countries, summative assessment occurs during a conversation between two examiners, one internal, normally the student's teacher, and one external, unknown to the student. Given that teachers' grading judgements have been criticised for not meeting standards of reliability, objectivity, and validity (e.g., Allal,

CONTACT Marit Skarbø Solem  marit.s.solem@usn.no  University of South-Eastern Norway, Boks 4, 3199 Borre, Norway

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

2013), there is a need for more knowledge on how oral assessment decision-making is achieved. The present study aims to examine, in detail, how grading conversations are carried out, in order to understand how conversational structures affect assessment practices and even outcomes.

Assessment and grading in school examinations

In Norwegian secondary schools, all students take oral examinations after Year 10 (age 15) and Year 13 (age 16–18) which are summatively assessed. The purpose of oral examinations is to provide information about students' oral competence within a school discipline. Student performance is graded between 1 (lowest) and 6 (highest) based on examiners' interpretations of how well students meet competence aims formulated in curricula. Grades are highly consequential since they are recorded on students' final school diplomas which gives access to post-compulsory education. The 'Regulations to the Educational Act' (2006) in Norway provides a general framework for grading conversations and govern the roles of the internal and external examiners:

For local exams, at least one of the examiners must be external. The school owner may decide that the subject teacher is the internal examiner. In the event of disagreement, the external examiner decides. (Regulations of the Education Act, 2006, § 3-28., our translation)

Accordingly, the external party has primary rights to decide final grades if examiners disagree. However, no further guidelines are provided. Despite the high stakes for students, we know little about how examiners interpret the regulations and how grades are proposed, discussed, (dis)agreed upon, and collaboratively decided in assessment interactions (cf. Nilsberth & Sandlund, 2021). Instead, such conversations happen inside a 'black box' and receive no scrutiny.

Compared to other forms of school-based assessment practice, there is relatively little research that focuses on oral examination practice – in any teaching subject and at any education level (Ministry of Education and Research, 2019). The studies that do exist demonstrate large variation in how oral examinations are executed and assessed. Ambiguous exam constructs are reflected in student performance, with their struggle to identify what is expected and how to interpret and respond to teachers' questions routinely identified (Isager, 2021a, 2021b). In addition, conversation analytic research has shown that examiners' behaviour shapes and impacts student performance (Sikveland et al., 2021; Skovholt et al., 2021; Vonen et al., 2022). For example, variation in the execution of oral exams may impact exam quality and lead to lower validity, reliability and fairness (Bøhn, 2016; Isager, 2021a, 2021b; Kaldahl, 2019; Maugesten, 2011; Skovholt et al., 2021; Vonen et al., 2022). In a Swedish context, research has shown that teachers may agree on *what* to assess when official assessment guidelines are provided. However, there is not necessarily consensus about *how* to assess, based on the guidelines (Byman Frisé et al., 2021, p. 1).

Since all forms of assessment may be subject to error and bias, the teaching profession, school administration, and researchers working with assessment are encouraged to take a 'collective responsibility to look for ways of improving the quality of teachers' judgment' (Allal, 2013, p. 21). Consequently, teacher education should emphasize assessment literacy in line with instructional proficiency (Popham, 2009, p. 5). One important approach to improve the consistency of teacher judgements are activities such as moderation meetings; that is, meetings where teachers discuss, confront, and negotiate grades assigned to student work or other assessment decisions (Allal, 2013, p. 21). Training programs are shown to increase the inter-rater reliability of moderation in the assessment of second/foreign language English oral proficiency (Sundqvist et al., 2020). In Norway, no national grading criteria are provided, and even though local educational authorities are encouraged to publish criteria, it is not mandatory (Norwegian Directorate for Education and Training, 2020). Since grades for oral examinations are reached through a decision-making *conversation*, we argue for the importance of understanding what happens in these interactions. The present study contributes with important empirical knowledge of how grading decisions, that are consequential for students' grades, are achieved.

Assessment and grading as joint decision-making

The practice of awarding grades for oral examinations is the outcome of a decision-making conversation between two examiners who formulate judgements of students' performances. Conversation analytic research has demonstrated that arriving at a joint decision is an interactional achievement that involves the negotiation of particular rights and obligations (Stevanovic, 2012).

First, speakers continuously orient to their own and their co-participant's relative rights to knowledge ('epistemic rights') within a specific domain, and to the relative distribution of rights to decide ('deontic rights') within a specific situation (Stevanovic & Peräkylä, 2012). In teacher-student interaction, the teacher is treated as having primary epistemic and deontic rights (i.e., are more knowledgeable and entitled to decide what happens in classrooms). In grading decision-making, however, the two examiners have more evenly distributed rights to assess a student's performance. But the external examiners are, as mentioned above, given primary rights to decide in cases of disagreement (Norwegian Directorate for Education and Training, 2020).

Second, participants orient to three basic components for reaching a joint decision (Stevanovic, 2012). First, one of the speakers must provide a *proposal* about future action (i.e., a grade proposal) to which all the interactants have *access*. Second, the co-participant needs to express *agreement* with the proposal. Finally, the interlocutors need to display *commitment* to the proposed future action (Stevanovic, 2012, p. 781). Thus, when examiners propose a grade, acceptance of the proposal from the other examiner is crucial for achieving a joint decision. Consequently, withholding agreement or acceptance, using minimal responses, such as 'mm' effectively extend the decision-making process by being treated as *passive resistance* (Stivers, 2006).

In addition to notions of epistemic and deontic rights, the conversation analytic notion of 'preference organization' is crucial for understanding how speakers recognize and manage emerging agreement and disagreement in decision-making and assessment sequences (Pomerantz, 1984). While 'preferred' responses to, say, proposals are typically short and delivered without hesitations or accounts, 'dispreferred' responses are marked by delays, hesitations, mitigations, and accounts (Pomerantz, 1984). Grading decision-making involves the actions of proposing (for which acceptance is preferred) and assessing (for which same or upgraded agreements are preferred). A weak agreement may thus be heard as disagreement with the first assessment. Studies of decision-making in both mundane and institutional settings have documented how participants regularly work to minimize potential disagreement and maximize the chance of agreement (e.g., Costello & Roberts, 2001).

Previous research on assessment in educational contexts has demonstrated the interactional challenge of assessing one's own and others' performances (Mazeland & Berenst, 2008; Skovholt, 2018; Skovholt et al., 2021; Nilsberth & Sandlund, 2021). Disagreement can challenge both social relations and professional identities, where potential conflict is cautiously expressed through extended sequences (Nilsberth & Sandlund, 2021) or through specific language formats, depending on whether they agree with the other assessor or not (Mazeland & Berenst, 2008).

When grading students' work, examiners must therefore handle myriad contingencies, from the outcome for each student to their own professional identity and relationship, to the conversational structures in which the assessment practice is carried out. In this study, we investigate how examiners manage the preference for agreement; the distribution of epistemic and deontic rights, and their professional identities, all while deciding upon a grade which is reified and separated from the context of its production and lent an objective status immediately thereafter. Our research questions are:

- (1) How are decision making achieved in oral examination grading conversations?
- (2) What interactional resources are employed to negotiate grades in order to reach a joint decision?

Data and method",5,7,0,0,3mm,3mm,1mm,0mm>Data and method

Oral skills, along with writing, reading, calculus and digital, are one of five key competences in the Norwegian secondary school curriculum. This implies that each school discipline (e.g., Science, Mathematics, etc.) includes learning outcomes concerning oral competence. For example, students in Year 10 should be able to ‘discuss the form, content and purpose of literature, theatre and films and present interpretative readings and dramatizations’ (Ministry of Education and Research, 2013). Students are examined individually by their subject teacher (the internal examiner) and a teacher from another school in the county (the external examiner). Oral examinations follow a structure comprising two parts: an oral presentation of a given subject, and a conversation in which the examiner poses questions about the curriculum. During the conversation, the participants sit on opposite sides of a desk in a classroom. After the examination, students wait outside the room while their examiners engage in a conversation, the outcome of which should be a unanimous decision. The decision is then communicated to the student by the examiners, usually accompanied by a short explanation.

In Norway, oral examinations are non-standardized so there are no regulations for how the grading conversations should play out. Rather, grading conversations occur in the context of the examiners’ experiences as a ‘community of practice’ (Wenger, 1998). There is a presumed symmetry between examiners: both are expected to have professional expertise in the given school subject and to convey independent professional integrity and agency. However, only internal examiners have knowledge about their students’ prior school achievements, the school’s curriculum, lessons taught and what subjects and topics they have emphasized from the curriculum and syllabus. These factors may impact the oral examination in tangible or tacit ways that are hard to access but may influence their reasoning about students’ competence and ultimately the grades awarded.

The data analysed comprised 29 video-recorded oral examinations in the subject Norwegian language and literature in four secondary schools in Norway, amounting to 18 h of recorded interaction. The participants are 5 internal and 4 external examiners and 29 students aged 16 or 18. The average time of the grading conversations, the data for this study, are 2:38 min. The shortest conversation is 0:38 min and the longest is 9:00 min. The data were recorded with one 360-camera positioned in between the participants (for close-ups) and one camera capturing the activity from 2–3 metres. The data were transcribed using Jefferson’s (2004) conventions for conversation analysis. All participants signed a letter of consent, and the project was approved by the Norwegian Centre of Research Data (NSD).

We used conversation analysis (CA) to investigate video-recorded grading conversations, focusing especially on how grades are initially proposed, negotiated, and decided through a series of actions that constituted the decision-making process. CA is a method for transcribing and analysing social interaction and includes micro analysis of visual and verbal resources used in interaction (Sacks et al., 1974; Sidnell & Stivers, 2012). The focus in the analysis is on the sequential unfolding of actions in interaction that the parties perform successively through their turns at talk. Each turn can be inspected for its treatment of the prior according to the ‘next turn proof procedure’ (Sacks et al., 1974).

Analysis

Analysis of the 29 video-recorded grading conversations revealed that the grading conversations consist of the following activity phases: After an opening, one of the examiners offer a global assessment, indicating the level of the upcoming grade proposal, often followed by more specific assessments before providing a grade proposal. Throughout these phases, the other examiners’ contributions are characterized by minimal or expanded signals of agreement and/or second assessments. In 4 of the 29 conversations, one of the examiners offers a grade proposal upfront, without any preceding accounts or assessments/pre-proposals. In all these cases, the grade proposal is accepted immediately by the other examiner. In three cases no immediate agreement or acceptance occur, instead initiating a negotiation until agreement is reached. Once agreement is achieved, the

Table 1. Overview of the organization of the 29 grading conversations.

	Immediate grade proposal Grade is proposed upfront without pre-accounts	Expanded grade proposal Assessments and accounts precede grade proposal
Immediate agreement Explicit agreement without delay or other markers of dispreference	4	22
Lack of immediate agreement No explicit agreement, but markers of dispreference and disagreement until concessions are given and subsequent grade proposal is agreed upon	0	3

grading conversation moves towards closing and transition to the final part of the examination, where the student is called back into the room for receiving their grade.

Micro analysis of the 29 grading conversations showed that the grading decisions are achieved through three distinct sequential patterns. The decisions are reached via either (1) a grade proposal followed by immediate agreement/acceptance or (2) pre-proposal accounts followed by a grade proposal and immediate agreement/acceptance or (3) pre-proposal accounts and a grade proposal followed by lack of agreement/disagreement and extended negotiation before acceptance. Table 1 summarizes this finding.

In what follows, we will analyse three representative grading conversations that show the three types of sequential patterns found in our data. As the grading conversations with expanded grade proposals and lack of immediate agreement lasts markedly longer than the two other types (9:00 min vs. 1:23/1:26 min), the analysis of the latter type is inevitably markedly longer than the two first types. In addition, we have chosen to also examine the extended type in detail in order to shed light on examiners' management of (dis)agreement in joint inter-professional decision-making, and how the guideline '[i]n the event of disagreement, the external examiner decides' (Regulations of the Education Act, 2006) is managed in practice.

(1) Immediate grade proposal followed by immediate acceptance/agreement

In 4 out of 29 cases, one of the examiners initiate the sequence with a grade proposal, without any form of preceding account, preface or other pre-sequence. In each case, the other examiner agrees immediately, moving the conversation to a close shortly after. Extract 1 is a clear example:

Extract 1:

```
01 EE: .h Du jeg tenker fem jeg.
      .h I am thinking a five.
02 IE: ↑Ja, ((multiple nods))
      Yeah,
03 EE: E:m ja, jeg syns det jeg.=
      E:m yeah, I do think so=
04 IE: =Ja, ((multiple nods))
      =Yeah,
05 EE: forde atte han e:
      because he e: ((short account continues))
```

In line 01, the external examiner opens the grading conversation by immediately proposing the grade five, using a format that invites joint decision-making, formulating the proposal as a thought (Stevanovic, 2013). The internal examiner immediately agrees with a confirming response, delivered with markedly high pitch and combined with multiple nods, thus displaying strong agreement. The reiteration in lines 03–04 confirms their mutual position and agreement. The external examiner moves on with a brief assessment sequence explicating the grounds for the proposed grade (starting in line 05) before the conversation ends (data not shown).

(2) Expanded assessment and grade proposal, followed by immediate acceptance/agreement

The most common sequential organization in our data (22 out of 29 cases) is that the grading conversation starts with a qualitative (that is, non-numerical) assessment before a grade is proposed and immediately agreed upon. Extract 2 illustrates this pattern, where the internal examiner displays strong agreement throughout the external examiner's assessment and grading proposal:

Extract 2:

01 EE: .hh
 02 IE: Tja.
Well.
 03 EE: E:: det er jo uten tvil høy måloppnåelse,
E:: it's without doubt high achievement
 04 EE: [e::] det var jo - han hadde en veldig rød tråd og
[e::] it was (jo)- he had a very red thread and
 05 IE: [Ja,]
[Yeah,]
 06 EE: .h Han hadde jo veldig mye
.h He had (jo) very much
 07 IE: Ja.h
Ye.ah
 08 EE: på hjertet,
on his hart
 09 IE: Ja.h,
Ye.ah,
(21 lines of positive evaluation omitted)
 31 EE: Så jeg syns egentlig ikke det er noen tvil om atte
So I think really that it's no doubt that
 32 det her er en sekser [jeg, rett og slett.]
this is a six [plain and clear.]
 33 IE: [Nei. jeg er heilt-]
[No. I am completely-]
 34 IE: veldig enig med deg i det.
I really agree with you.
((evaluation continues))

In lines 03–08, the external examiner formulates a global assessment where she assesses the student as displaying 'high achievement'. After a series of further positive assessments (data not shown), she proposes the grade as 'a six'. The grade proposal is delivered as a logical consequence, or upshot ('so I think ...') of the preceding assessment (Heritage & Watson, 1979). The lack of hedges and mitigations, together with strong displays of certainty ('no doubt') indicates that the external examiner expects agreement. Looking at the responses from the internal examiner, strong agreement is displayed already in the confirmations throughout the external examiner's initial assessment (lines 05, 07, 09), produced without any delays or other markers of dispreference. Also, the agreement to the actual grade proposal is delivered early, in overlap, and in an upgraded format ('completely', 'very'), signalling strong agreement (lines 33–34) (Pomerantz, 1984). A short assessment sequence follows, before the student is called back in to receive the grade decision.

Extracts 1 and 2 illustrate how grade decisions are reached through immediate agreement characterized by a rapid exchange with no markers of dispreference. Most commonly, a qualitative assessment precedes the numerical grade proposal, as in Extract 2, or occasionally the examiners move straight to the grade proposal, as in Extract 1.

(3) Assessment and grade proposal, followed by lack of immediate acceptance/agreement

In three of the grading conversations, disagreement and rejection of proposed grades lead to a negotiation that expands the sequence substantially until concessions are made and a grade proposal is

agreed upon. In this section, we provide an extended analysis of one of the three cases for illustrating this trajectory in depth. The analysis is divided into Extracts 3a – 3e.

(3a) Opening and initial qualitative assessment

As noted earlier, most grading conversations were characterized by an initial qualitative/global assessment rather than a numerical grade formulated by the external examiner. These qualitative assessments formulate some general level of the students' performance ('that went well') or something specific, as in Extract 3a below.

Extract 3a:

01 EE: nt.h Ja:?
nt.h Yeah:?
 02 IE: Ja?
Yeah?
 03 EE: Hva skal vi si: da?
What shall we say: then?
 04 **(0.4) ((IE changes position in her chair))**
 05 EE: E:m: en grei gjennomgang av e: noen av e: de tingene
E:m: an okaypresentation of e: some of e: the things
 06 **som skiller norrønt fra moderne norsk?=e: hadde e::#**
that separates old norse from modern norwegian? =e: had e::#
 07 **m: e: jeg syns jo flere steder at han kunne sagt litt mer?**
bu-: e: I think (jo) that several places he could have said a
little more?
 08 IE: [Mm,]
 [Mm,]

In the opening (lines 01–04), the examiners negotiate who should go first. The external examiner's 'yes?' in line 1 gives the internal examiner a chance to go first (and potentially give an initial assessment), but the internal examiner echoes the external examiner's turn (line 2) and gives it back, treating the external examiner as responsible for delivering the initial assessment, which they do: 'What should we say, then' (line 03). The utterance takes the form of a pre-assessment that delays the actual initial assessment and indicates a level of uncertainty. This is consolidated in the external examiner's next turn, the initial assessment (line 05–06), sequentially placed as a first assessment (Pomerantz, 1984). The assessment is uttered with hesitation markers ('e:m:'), and the evaluative 'ok presentation' indicates a good grade, though not top score, and the external examiner explicitly states that there are shortcomings in the student's performance, he should have 'said more' 'several times'. Looking at the turn design, we see how the external examiner starts with a (fairly) positive assessment, followed by a negative assessment, highlighting shortcomings. This evaluative turn format, '(fairly) positive x, BUT (fairly) negative y', indicates the latter (negative) part as the most prominent, preparing the grounds for a less favourable grade. The internal examiner does not take an active role in the opening of the grade talk. She does not go first when given the opportunity (line 02), and the 'mm' (line 08) neither express agreement with the external examiner's assessment (Pomerantz, 1984), nor gives a second assessment.

Extract 3a shows how the external examiner's initial assessment implicitly indicates the direction of the grade proposal, and as she points at shortcomings, we know that the student probably is not going to get the top grade. The internal examiner's lack of agreeing responses indicate the emergence of a potential disagreement. By using a general assessment, not moving straight to an explicit grade proposal, the external examiner may be 'testing the waters' (Bergen & Stivers, 2013) through subtle calibration, before proposing a grade that potentially differ with the internal examiner's impression of the student's performance. In the next extract, we see how the external examiner bolsters her initial assessment before proposing a grade.

(3b) Further assessment pointing towards a grade proposal

In Extract 3b, the external examiner offers a series of assessments (lines 09–48), pointing to several weaknesses and downgraded positive elements in the student’s performance. Notably, the internal examiner’s first non-minimal turn in the conversation comes in the form of a disagreement (lines 51–52):

Extract 3b:

09 EE: [e:] for eksempel når han snakker om person og tall, .h e:
[e:] for example when he talks about genus and number, .h e:
 10 i verb, så: kunne han jo snakka om ti(h)d for eksempel også
in verbs, then: he could also have talked about time for example
 11 [(hh)] at] det er noe felles her,=
[(hh) that] there is something in common here,=
 12 IE: [Mm ja,]
[Mm yeah,]
 13 IE: =Mm,=
=Mm,=
 14 EE: =e::m og han var vel inne på litt på slutten atte nt.h e:# at
=e::m and he touched upon a bit at the end that nt.h e:# that
 15 monoftongeringa ikke har slått gjennom på: på [ny]norsk?=
**monophthongization hasn't come through in: in [ny]norsk?=
 16 IE: [mm]**
 17 EE: så det var fint at han fikk med det,=
so it was good that he got that,=
 18 IE: =mm,=
=mm,=
 19 EE: =e::m litt om- men han forklarte om ordstilling og sånn så det
=e::m a little about- but he explained about word order so that
 20 var jo: .hh e: greit og ryddig.
was (jo:) .hh e: nice and neat.
 21 (.)
 22 EE: .hh em:: (0.8) Ha:n kunne ikke helt svare på noen av spørsmålene
.hh em:: (0.8) He: couldn't quite answer some of the questions
 23 som du hadde som oppfølgingsspørsmål?
you had as follow up questions?
 24 ((IE looks in the papers in front of her))
 25 EE: hvordan e:: nt.h det norrøne forsvant?=**han snakka litt om e:**
how e:: nt.h the old norse disappeared?=he talked a bit about e:****
 26 .h svartedauden? og litt om andre::: språk som
.h the Black Death? and a bit about other languages that
 27 påvirker?=**men han va'kke helt sikker på konsekvensene**
influence?=but he wasn't quite sure about the consequences****
 28 av unionen hvordan det (.) spilte inn.
of the union how it (.) played a part.
 29 (.)
 30 IE: Mm,
Mm,
 (17 lines omitted)
 48 e: så der syns jeg også det mangla litt?
e: so there I also think that it lacked a bit?
 49 IE: Mm,
Mm,
 50 EE: e[:m]
e[:m]
 51 IE: [Det] sa han vel egentlig når han snakka om språkhistoria.=
[He]did say that really when he talked about language history.=
 52 EE: =Han sa mer om det [der ja.] det er sant.
=He said more about that [there yes.] that's true.
 53 IE: [ja, mm,]
[yeah, mm,]
 54 IE: Ja,
Yeah,
 55 (.)
 56 IE: (Kanskje han) følte at han hadde sagt det.=
(Maybe he) felt like he had already said that.=
 57 EE: =Ja, det kan [he]nde det,=
=Yeah, that may [be],=
 58 IE: [ja,]
[yeah]
 59 IE: =Ja,
=Yeah,
 60 EE: .hh e: Når han snakka om e:: en middag, så syns jeg han e:
.hh e: When he talked about a dinner, then I thought he e:
 61 snakka seg opp.
talked himself up.
 62 IE: Mm,
Mm,
 63 EE: Den syns jeg han beskrev godt,
That one I think he described well,

In line 09, the external examiner expands on her first assessment (lines 05–07, Extract 3a), exemplifying weaknesses indicated in the initial assessment. After a minimal response from the internal examiner (lines 12–13), the external examiner continues in lines 14–20 with positive assessments, downgraded with hedges (“touched a bit at the end”) and adjectives signalling moderate performance (‘nice and neat’). This pattern continues in lines 22–48, where the external examiner points to additional weaknesses and missing elements (lines 22–23), combined with positive elements that are downgraded with the hedges ‘some’ and ‘a bit’ (lines 25 and 48). Moreover, the turn format observed in extract 3a, ‘x, but y’, with ‘x’ expressing a positive feature, and ‘y’ expressing a negative feature also occur several times here (lines 25–27, and in omitted lines). This extended assessment sequence ends with an upshot in line 48, highlighting lacking elements.

Throughout the external examiner’s extended assessment, the internal examiner only provides minimal responses (lines 16, 18, 49), delayed minimal responses (lines 29–30) or no response at all (line 21). The lack of agreeing responses strongly indicate withheld disagreement (Koenig, 2011; Stivers, 2006). Especially the missing response in line 21, when the external examiner’s turn is prosodically marked as complete, is noticeable, as a response signalling agreement would be expected (Pomerantz, 1984). In line 51, the internal examiner takes her first non-minimal turn in the conversation so far, offering an explicit disagreement: ‘He did say that really when he talked about language history’. The internal examiner objects to a weakness the external examiner has pointed at, suggesting that the student had already covered the topic. Notably, the external examiner immediately agrees (marked with latched turn, recycled wording and explicit confirmation) in line 52. After a short expansion sequence (lines 56–59), the external examiner continues her extended assessment, moving on to another part of the curriculum (line 60). Notably, for the first time in the extended assessment sequence, the external examiner produces a positive assessment without any downgrading features (lines 60–61, 63). The shift in the assessment, from highlighting the negative to highlighting the positive, is framed as related to the student’s performance (the student ‘talked himself up’), not to the ongoing negotiation.

In sum, in this part of the conversation, the external examiner does extensive interactional work, building a case towards a grade proposal that is not top-grade. This prepares the grounds for a grade proposal that may differ from the internal examiner’s view, based on the internal examiner’s lack of agreement so far in the conversation. The external examiner works towards minimizing (potential) disagreement and maximizing the chances of agreement when the grade is finally proposed in the next extract (3c).

(3c) Grade proposal and calibration

In Extract 3c the external examiner proposes a grade, delivered after substantial hesitation markers and pauses (lines 66–69):

Extract 3c:

66 (1.5)
 67 EE: e::m ((turns paper around)) Så-
 e::m ((turns paper around)) So-
 68 (2.0)
 69 han e: m (1.0) mt.h hva skal vi si (1.5) jeg syns at han e:
he e: m (1.0) mt.h what shall we say (1.5) I think that he e:
 70 har noe som er på e: på femmernivå, og noe som er på firernivå.
has something on e: on a five, and something on a four.
 71 IE: Mm,
 Mm,
 72 EE: e::m (1.5) Jeg syns (0.5) nok kanskje at det (.) e: norrøne,
e::m (1.5) I think (0.5) (nok) maybe that it (.) e: old norse
 73 forklaringa på det, språkhistoria, (1.5) # ligger mere på at
the explanation on that, language history, (1.5) # is more that
 74 han i noen grad drøfter,
he to some degree discusses,
 75 IE: [Mm,]
 [Mm,]
 76 EE: [så] han er mere: der?
 [so] he is more: there?
 77 IE: Mm,=
 Mm,=
 78 EE: =e:m Når det gjelder å analysere og tolke på en nyansert måte-
=e:m When it comes to analyse and interpret in a nuanced way_
 79 jeg syns jo han var litt inne på det på den kielland-
I think (jo) that he said something about it on the kielland
 80 teksten, men kanskje på de andre en mer en enkel relevant måte.
text, but maybe on the other more a simple relevant way.
 81 IE: Mm,
 Mm,
 82 EE: em: (1.0) Han e: setter det inn i en sjangermessig historisk
em: (1.0) He e: puts it into a genrelike historic
 83 e:m måte, men det er også noen spørsmål på slutten som han
e:m way, but there are also some questions at the end that he
 84 ikke# svarte helt på, ((looking at paper in her hand))
didn't# answer completely,
 85 de::t (3.5) ((turns paper back again)) jeg føler kanskje at-
i::t (3.5) ((turns paper back again) I feel maybe that-
 86 eller jeg syns kanskje at (0.3) hoveddelen av det han viser
or I think maybe that (0.3) the main part of what he displays
 87 ligger på fire,
is a four,
 88 (.)
 89 IE: Mm,
 Mm,
 90 EE: selv om han er (.) e:: (hh)toucher opp på femmern noen ganger
even if he is (.) e:: (hh) touches the five some times
 91 syns jeg,= men e kanskje ikke hovedinntrykket,=hva (.)
I think,= but e maybe not the main impression,=what (.)
 92 syns du,
do you think,

In line 69, prefacing the grade proposal, the external examiner recycles the ‘what shall we say’-formulation from the initial assessment (see line 03, Extract 1). This meta comment again contributes to marking her global assessment as somewhat uncertain or foreshadowing a grade proposal that is in between or makes relevant two potential grades, as suggested in lines 69–70. Although presented as two equal grade proposals in the wording, subtle prosodic features make it hearable as tilted towards the grade 4. The stressed qualifier ‘something on a five’ indicates *less* of that ‘something’ on level five than ‘something on a four’, without marked stress. Moreover, the substantial delays preceding the grade proposal (lines 66–69), also treats the proposal as dispreferred, orienting to the lack of agreement from the internal examiner until now. In response, the internal examiner only produces a minimal response (line 70), effectively continuing to withhold any signals of agreement. This form of minimal response in a slot where a stronger form of agreement is expected and preferred, displays a subtle, but strong passive resistance to the proposal (Koenig, 2011; Stivers, 2006). In the absence of any uptake from the internal examiner, the external examiner continues specifying several reasons in support of her grade proposal (lines 72–87). The assessments have

the same characteristics as in extracts 3b–3c, with downgraded adjectives (‘to some degree discusses’) and the ‘x, but y’-format, highlighting the weaknesses of the performance as the most prominent feature (lines 82–84).

The extended assessment sequence ends with the external examiner repeating her grade proposal, concluding that the main part of the student’s performance is at level four (lines 86–87). As observed previously, both the assessments and the reissued grade proposal are delivered in a dis-preferred turn format, with hedges, hesitation markers and a self-repair, orienting to the emerging disagreement. The reissued assessments and upgraded grade proposal again makes a second assessment or agreement from the internal examiner the relevant next action. However, the internal examiner only provides yet another minimal and delayed response (lines 88–89), signalling lack of agreement. In response, the external examiner reissues the grade proposal once more (lines 90–91), before explicitly pursuing a response from the internal examiner in a latched turn: ‘what do you think?’, treating the internal examiner’s response to the grade proposal as missing and conditionally relevant for reaching a (joint) decision.

Extract 3c illustrates how the examiners treat agreement or acceptance to a grade proposal as necessary to arrive at a joint decision (Stevanovic, 2012). As this is still not the case at this point, the negotiation may continue.

(3d) Explicit disagreement and reorientation

Extract 3d shows how the internal examiner finally provides an explicit disagreeing assessment, exposing the examiners’ diverging views. In what follows, the external examiner subtly calibrates and reorients the line of argument towards the higher grade five, again preparing the grounds for a joint, but different, decision.

Extract 3d:

133 EE: nt.h Ja syns du det er bedre enn en fire?
nt.h Yeah do you think it is better than a four?

134 (.)

135 EE: at han er e::
that he is e::

136 IE: nt.h .hh Ja æ syns kanskje det.
nt.h .hh Yeah I think maybe that.

137 EE: [Mm,]
[Mm,]

138 IE: [med] tanke på det han si om litteratur,=
[be]cause of what he says about literature,=

139 EE: =Ja,
=Yeah,

140 IE: e:=
e:=

141 EE: =Ja, [her sånn] så var jeg på en måte [helt] klar på
=Yeah, [here so] so I was in a way [completely] clear

142 IE: [()] [ja,]

143 EE: (at det var) e det[te] er bra.
(that it was)= e thi [s] is good.

144 IE: [m,]
[m,]

145 EE: Det er [bå]de:: .h e: klart og tydelig framlagt, og det er e::
It is [bo]th clear and concisely presented, and it is e::=

146 IE: [mm,]
[mm,]

147 EE: [jeg] syns han e: fikk fram essensen i teksten og
[I] think he e: got the essence of the text and the

148 IE: [Mm,]
[Mm,]

150 EE: [hen]sikten med teksten veldi:g godt.
[pur]pose of the text very: well.

151 IE: Mm,
Mm,

152 EE: Så de::t, (1.2) em:: (1.0) og det han sier om e- han fikk jo'ke
So i::t, (1.2) em:: (1.0) and what he says about e- he didn't

153 sagt så mye om om e regn i hiroshima, men e men det er jo
get to say much about Rain in Hiroshima, but e but it is (jo)

154 essensen i det ikke sant,=
the essence of it right,=

155 IE: =Mm,
=Mm,

(26 lines omitted)

180 EE: Du syns at vi skal [gi'n e:n femmer? hhh]
You think we should [give'm a: five? Hhh]

181 IE: [HHhhh]hhh

182 IE: Ei- æ skjønne godt anke:punkta dine.=
Ei- I understand reeally well your conce:rns.=

183 EE: =Mm,
=Mm,

184 IE: E e veldig enig i den e: vurderinga du: (.) gjør,
I'm very on board with the e: assessment you: (.) give,

185 (2.0)

186 IE: Det er jo spørsmål da ka som skal veie tyngst [(hh)]
It is a question about what should count more [hh]

187 EE: [mm,] mm,
[mm,] mm,

The open-ended question ‘what do you think’ in Extract 3c (lines 91–92) did not yield a clear response (data not shown). So, here, in line 133, the external examiner pursues a response with a polar question restricting the response options to a ‘yes’ or ‘no’-response: ‘do you think it’s better than a four?’. The question design indicates that the external examiner expects a confirming response. However, a confirming response will be dispreferred in the sense that it will reveal the so far implicit disagreement between the two examiners, setting up a cross cutting preference (Heritage & Clayman, 2011). In response, the internal examiner’s disagreement finally gets ‘on the record’ with a confirming response in line 136. It is indeed delivered as a dispreferred response, with delay, hesitation markers and downgrading (lines 133, 134, 136), and a subsequent account (line 138). Notably, in response (lines 141, 143), the external examiner provides an upgraded

agreement with the internal examiner's assessment about the of the literary part of the exam (see line 138), indicating strong agreement with the positive assessment. The strong agreement is evident in the upgraded utterances in lines 141, 143 and 145.

In what follows, the external examiner continues with additional assessments (lines 145–161). However, in contrast to the patterns showed in previous extracts, where the external examiner has argued towards the lower grade four, the assessments at this point are designed as positive and upgraded, tilted towards the higher grade five. This is evident in the notable shift in the assessments from consisting of positive, but downgraded elements as described until now, to positive, upgraded assessments. Moreover, the 'x, but y-format' we have seen throughout (extracts 3b–3d), with the negative element appearing last, and thus given most weight, is reversed here: In lines 152–154 the external examiner uses the same format, but with the negative assessment first, followed with the positive assessment, framing the positive part as the most prominent, or working as an excuse for the weakness, both tilting the assessment towards the higher grade five.

After more assessment along the same lines (data not shown), the external examiner for the third time pursues an independent and clear assessment statement from the internal examiner (line 180). The polar declarative formatted question is even stronger geared towards confirming the view that the student should have the grade five. However, in response, the internal examiner refrains from confirming or disconfirming the polar question (lines 180–187). By doing so, she avoids claiming deontic rights to decide, pushing the decision responsibility back to the external examiner. By resisting decisional rights, the internal examiner avoids claiming a unilateral decision in disagreement with the other examiner, while simultaneously pushing towards a decision in line with her own preference (Landmark et al., 2015). Instead, the internal examiner's response works to validate the external examiner's objections (line 182), displaying explicit and upgraded agreement to the external examiner's assessments ('I very much agree', line 184). In this way, the internal examiner simultaneously manages to maintain her opposing view expressed previously, while minimizing the disagreement and preserving the external examiner's deontic right to make the final decision, which we have seen is in accordance with the exam guidelines. In the next extract we will see how a joint decision is finally achieved.

(3e) Joint decision

In all the grading discussions in our data, the examiners treat a joint decision, consisting of a grade proposal and acceptance, as necessary for reaching a grade decision. Extract (3e) shows how the two examiners manage to reach a joint decision, and how they portray the grade decision as not related to the examiners' preceding disagreement but based on general norms of (good) examiner behaviour.

Extract 3e:

211 EE: [Nei jeg] er jo litt sånn e:# tendens til å gå opp=hvis man er
[No I] am (jo) a bit like e:# a tendency to go up=if one is

212 IE: [()]

213 EE: i tvil?=
**in doubt?=
Mm,
Mm,**

214 IE: Mm,
Mm,

215 EE: Man skal liksom komme dem til gode ath
One is supposed to give them benefit that

216 IE: Mm,
Mm,

217 EE: det er en situasjon hvor e (.) du har begrensa tid og det
it is a situation where e (.) you have limited time and there

218 er [splenning i e:
is [ten]sion in e:

219 IE: [Mm,]
[Mm,]

220 IE: Mm,
Mm,

221 EE: Så hvis vi: sier at han e: han gjorde det godt på
So if we: say that he e: he did well on the

222 litteraturen [så] skal det få trekke'n opp,
literature part [then] that shall tilt him up,

223 IE: [Mm,]
[Mm,]

224 IE: Mm,
Mm,

225 EE: Mm,
Mm,

226 IE: Ja,
Yeah,

227 (1.5)

228 IE: .hh Da e e enig i det [hHHHhhhhh] .hh hh
.hh Then I agree with that [HhhHhhHhhHhh] .hh hh

229 EE: [HhhHhhHhhHhh]

230 EE: Da: sier vi det.
The:n we say that.

231 ((E turns papers))

232 EE: Sånn.
Like that.

Extract 3e shows how joint agreement is treated as a necessary step for reaching a joint decision, as agreement is a prerequisite for grading. In this extract a turning point is reached. The calibration process ends, and the decision-making starts. The external examiner provides a general account (lines 211–220) that portrays her decision as based on previous professional experiences as examiner and general grading norms: ‘if I am uncertain, one has a tendency to go up’. She uses first person pronoun ‘I’ in the first part of the account but changes to the indefinite pronoun ‘one’ in the second part. That is, they and/or ‘one’ usually gives students the benefit of the doubt, rooting this particular decision in how they normally perform their job as an external examiner and general norms for examiners, and thus, not as a result of a concession.

The final grade proposal is delivered by the external examiner in lines 221–222. The proposal is designed as a conditional upshot (‘so if...’) inviting agreement. In contrast to previous versions of the grade proposal, it focuses on the parts of the performance that would tilt the grade upwards to a five. Moreover, by using the pronoun ‘we’, the external examiner portrays the decision as a joint accomplishment that they both are accountable for, it is a collaborative achievement. At this point the internal examiner finally gives an explicit agreement, which is confirmed, and the external examiner signals ‘case closed’, turning their papers (line 231).

In sum, this extended grading decision conversation shows how agreement is a prerequisite for reaching a joint decision. The external examiner provides multiple assessments, working

towards reaching a joint decision with explicit agreement from the internal examiner. When that fails, through the internal examiner's withheld agreement, the external examiner gradually reorients their initial grade proposal and line of argument towards the view of the internal examiner. The analysis shows how the internal examiner's lack of explicit acquiescence in terms of silence and minimal responses effectively affects the grading result, changing the external examiner's position in favour of a higher grade. The internal examiner 'gets their way' through passive resistance. However, the external examiner is not just conceding, but accounts for how a five, despite prior arguments pointing towards a four, is rooted in general norms of (good) examiner behaviour.

Discussion and conclusion

In school oral examinations, the purpose of grading conversations is for different examiners to reach a joint decision so as to award a grade to the student being assessed. Our analysis of the 29 grading conversations has shown that they follow a similar trajectory towards a joint decision, regardless of whether the examiners initially agree with each other or not. The analysis showed that the examiners are, in general, oriented towards agreement in the decision-making process. In cases of initial agreement, a joint decision is usually reached after a relatively short time and a few turn exchanges. The difference between cases of initial agreement and disagreement are evident in the expansion of the grading conversation. Examining one extended case in detail, we have shown that disagreement is treated as dispreferred, and something examiners minimize and conceal. It is characterized by prolonged sequences with assessments and accounts calibrating differing views, examiners' reorientation, and revised initial stances. In the specific case analyzed, we have seen how the external examiner pursues acquiescence, while the internal examiner withholds it, both using a variety of subtle interactional resources. Thus, the case is not a deviant case per se, but represents how grading is achieved through interaction.

The oral examination represents situations where professional practitioners are required to reach a joint decision, even though they initially disagree. In the case of the oral exams, the guidelines provide directions on how to deal with potential disagreement: '*In the event of disagreement, the external examiner decides*' (Regulations of the Education Act, 2006, § 3-28, *our translation and emphasis*). However, our study has shown that 'deciding' is not a straightforward, unilateral process, but rather negotiated turn by turn in and through social interaction. This is especially salient when the decision makers have partly diverging stances. In the case analysed in our paper, the external examiner is the one who explicitly proposes the initial and the adjusted grade and hence appear to be the one who 'decides', in accordance with the guideline's instructions. However, close analysis showed that withholding acquiescence may be an effective way of getting the decision maker to reorient and revise their stance. That is, withheld agreement may have direct impact on the decision. Micro details like silence and minimal responses have serious consequences for social practices and outcomes in general, and in our case, for grading. Thus, the study problematizes what it means to decide and shows the importance of micro analysis of institutional practices where decisions are achieved through interaction.

Grading conversations are significant conversations where much is at stake for the participants. The examiners must balance their professional and personal integrity, managing their possible self-presentation as either intransigent or submissive, while also potentially worrying about what disagreements about achievement standards might mean for one's professional competence (e.g., Adie, 2013). The analysis showed that the participants strongly orient to the general conversational norm of preference for agreement (Pomerantz, 1984), and this agreement goes beyond the actual grade proposal. In our 18 h of data, there are overall few traces of explicitly expressed disagreement on how or what to assess of the students' performances, and the examiners engage in a lot of interactional work to avoid or minimize explicit disagreement (cf. Costello & Roberts, 2001; Hudak et al., 2011). However, we have also shown that orientation to these norms can make an external examiner

propose a better grade than they initially would, because of the lack of explicit agreement. The internal examiner gets her way, and ‘the final word’, through passive resistance (Stivers, 2006), only implicitly formulating disagreement.

In our data, the examiners rely on a great deal of tacit examiner norms in the grading conversations. Explicitly formed assessment criteria are rarely referred to in our data, although the examiners often refer to common expectations of what it means to do well in the school subjects examined. For example, when one of the examiners states ‘shall we say eight?’ (the top grade is 6), after a student’s performance, there is no need to discuss the students’ performance any further. The examiners easily agree that the achievement is far above what they could expect from a student on this type of exam, and the student is given the top grade, six, without further discussion. In the case analyzed across Extracts 3a-e, the external examiner rationalizes their changing stance by referral to ‘objective’ assessment norms and how examiners in general assess exams. Throughout our data, the examiners refer to the repertoire they have accumulated through earlier experiences as examiners that they clearly expect the other to share, and they rely on such experiences to make their professional duties manageable, accommodable, and accountable (see also Mazeland & Berenst, 2008, p. 56). As such, the grading conversations are tangible examples of a ‘community of practice’ (Wenger, 1998).

Of course, the grading conversations are significant for students as well, whose educational future may depend on the grade in question. The data revealed that internal examiners especially take this into account, referring repeatedly in the grading conversations to their previous knowledge of the students’ achievements. Perhaps this is why the guidelines explicitly express that an external examiner (i.e., someone who is unknown to the student) must be present during oral exams and has the formal authority to decide the grade. The guidelines therefore highlight the importance of a neutral assessor and thus consider the previous relation between the internal examiner and the student as a potential bias. However, as our study has shown, since institutional policies are realized in and through interaction, the social interactional issues of preference, entitlement, and forms of knowledge – alongside well-documented biases (e.g., of gender, race) – may result in students’ receiving higher (or lower) grades. And, in the case of oral examinations, where criteria are vague and unspecified, such bias is more likely (e.g., Quinn, 2020).

Although our findings are based on a limited number of grading conversations, they show the consequentiality of interaction for assessment outcomes. Our findings have implications for grading as a professional practice. First, our findings challenge the ideal of fairness. The power of even subtle nuances in interaction, points to a lack of transparency in grading conversations. We encourage the use of more specific and explicit assessment criteria, together with evidence-based guidelines on how to implement the criteria in grading conversations. This may reduce the use of implicit assessment norms. Second, when grading is achieved through interaction, the examiners need to be aware of how interactional norms, like ‘preference for agreement’, influence joint decisions. Examiners need to reflect on how to assess (Byman Frisén et al., 2021). We suggest that school leaders need to provide mandatory examiner training or moderation meetings based on empirical knowledge from actual grading discussions like the ones analyzed here, as a way of raising awareness of the consequences of the ‘interactional engine’ (Levinson, 2006) for assessment as professional practice.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was founded by The Research Council of Norway [grant number 273417].

References

- Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, 20(1), 91–106. <https://doi.org/10.1080/0969594X.2011.650150>
- 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>
- Berge, K. L. (2002). Hidden norms in assessment of students' exam essays in Norwegian upper secondary schools. *Written Communication*, 19(4), 458–492. <https://doi.org/10.1177/074108802238011>
- Berge, K. L., Skar, G. B., Matre, S., Solheim, R., Evensen, L. S., Otne, H., & Thygesen, R. (2017). Introducing teachers to new semiotic tools for writing instruction and writing assessment: consequences for students' writing proficiency. *Assessment in Education: Principles, Policy & Practice*, 26(1), 6–25. <https://doi.org/10.1080/0969594X.2017.1330251>
- Bergen, C., & Stivers, T. (2013). Patient disclosure of medical misdeeds. *Journal of Health and Social Behavior*, 54(2), 221–240. <https://doi.org/10.1177/0022146513487379>
- Black, P. (2013). Formative and summative aspects of assessment: Theoretical and research foundations in the context of pedagogy. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 167–178). Sage.
- Byman Frisén, L., Sundqvist, P., & Sandlund, E. (2021). Policy in practice: Teachers' conceptualizations of L2 English oral proficiency as operationalized in high-stakes test assessment. *Languages*, 6(204), 1–23. <https://doi.org/10.3390/languages6040204>
- Bøhn, H. (2016). *What is to be assessed? Teachers' understanding of constructs in an oral English examination in Norway*. University of Oslo.
- Costello, B. A., & Roberts, F. (2001). Medical recommendations as joint social practice. *Health Communication*, 13(3), 241–260. https://doi.org/10.1207/S15327027HC1303_2
- Fjørtoft, H., & Morud, E. B. (2021). Assessment decision making in vocational education and training. *Studia Paedagogica*, 26(4), 119–137. <https://doi.org/10.5817/SP2021-4-6>
- Forskrift til opplæringslova [Regulations of the Education Act]. (2006). <https://lovdata.no/dokument/LTI/forskrift/2009-07-01-964>.
- Heritage, J., & Clayman, S. (2011). *Talk in action: Interactions, identities, and institutions*. John Wiley & Sons.
- Heritage, J., & Watson, D. R. (1979). Formulations as conversational objects. *Everyday language: Studies in ethnology* (pp. 123–162).
- Hudak, P. L., Clark, S. J., & Raymond, G. (2011). How surgeons design treatment recommendations in orthopaedic surgery. *Social Science and Medicine*, 7(73), 1028–1036.
- Isager, J. (2021a). At knække lærerkoden' – en elevperspektivistisk analyse af adressater ved mundtlig eksamen ['Cracking the Teacher's Code' – Students' Perceived Addressees Before Oral Exams]. *Nordic Studies in Education*, 41(4), 295–311. <https://doi.org/10.23865/nse.v41.2692>
- Isager, J. (2021b). Mundtlig eksamen er en kunst' – Danske gymnasieelever til mundtlig eksamen i fagene historie og engelsk ['Oral Exams Are an Art Form' – Danish Upper Secondary Students Attending Oral Exams]. *Nordidactica*, 11(2), 87–112.
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation Analysis: Studies from the first generation* (pp. 13–31). John Benjamins.
- Kaldahl, A. (2019). Assessing Oracy: Chasing the teachers' unspoken oracy construct across disciplines in the landscape between policy and freedom. *L1-Educational Studies in Language and Literature*, 19(3), 1–24.
- Koenig, C. J. (2011). Patient resistance as agency in treatment decisions. *Social Science & Medicine*, 72(7), 1105–1114. <https://doi.org/10.1016/j.socscimed.2011.02.010>
- Landmark, A. M. D., Gulbrandsen, P., & Svennevig, J. (2015). Whose decision? Negotiating epistemic and deontic rights in medical treatment decisions. *Journal of Pragmatics*, 78, 54–69. <https://doi.org/10.1016/j.pragma.2014.11.007>
- Levinson, S. C. (2006). On the human 'interactional engine'. In N. J. Enfield, & S. C. Levinson (Eds.), *Roots of human sociality: Cognition, culture, and interaction* (pp. 39–69). Routledge.
- Maugesten, M. (2011). Mundtlig eksamen. En analyse av åtte studenters forståelse på muntlig eksamen i matematikk. *Norsk Pedagogisk Tidsskrift*, 95(4), 260–272. <https://doi.org/10.18261/ISSN1504-2987-2011-04-03>
- Mazeland, H., & Berenst, J. (2008). Sorting pupils in a report-card meeting: Categorization in a situated activity system. *Text & Talk*, 28(1), 55–78. <https://doi.org/10.1515/TEXT.2008.003>
- Ministry of Education and Research. (2013). *Læreplanverket for Kunnskapsløftet [The Curriculum for Knowledge Promotion in Primary and Secondary Education and Training]*.
- Ministry of Education and Research. (2019). *Kunnskapsgrunnlag for evaluering av eksamensordningen [The foundation of knowledge for evaluating the exam system in Norway]*. <https://www.udir.no/tall-og-forskning/finnforskning/rapporter/Kunnskapsgrunnlag-for-evaluering-av-eksamensordningen/>.

- Nilsberth, M., & Sandlund, E. (2021). On the interactional challenges of revealing summative assessments: Collaborative scoring talk among teachers and students in Swedish national tests. *Linguistics and Education*, 61, 1–23. <https://doi.org/10.1016/j.linged.2020.100899>
- Norwegian Directorate for Education and Training. (2020). Regler for muntlig eksamen [Guidelines for oral examinations]. <https://www.udir.no/eksamen-og-prover/eksamen/muntlig-eksamen/>.
- Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shaped. In J. M. Atkinson, & J. Heritage (Eds.), *Structures of social action* (pp. 57–101). 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>
- Quinn, D. M. (2020). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis*, 42(3), 375–392. <https://doi.org/10.3102/0162373720932188>
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language*, 50(4), 696–735. <https://doi.org/10.1353/lan.1974.0010>
- Sandlund, E., & Sundqvist, P. (2019). Doing versus assessing interactional competence. In M. R. Salaberry, & S. Kunitz (Eds.), *Teaching and testing L2 interactional competence* (pp. 357–396). Routledge.
- Sidnell, J., & Stivers, T. (2012). *The handbook of conversation analysis*. Blackwell Publishing Ltd.
- Sikveland, R. O., Solem, M. S., & Skovholt, K. (2021). How teachers use prosody to guide students towards an adequate answer. *Linguistics and Education*, 61, 1–15. <https://doi.org/10.1016/j.linged.2020.100886>
- Skovholt, K. (2018). Anatomy of a teacher–student feedback encounter. *Teaching and teacher education*, 69, 142–153. <https://doi.org/10.1016/j.tate.2017.09.012>
- Skovholt, K., Solem, M. S., Vonen, M. N., Sikveland, R. O., & Stokoe, E. (2021). Asking more than one question in one turn in oral examinations and its impact on examination quality. *Journal of Pragmatics*, 181, 100–119. <https://doi.org/10.1016/j.pragma.2021.05.020>
- Solheim, R., & Matre, S. (2014). Forventninger om skrivekompetanse. Perspektiver på skriving, skriveopplæring og vurdering i 'Normprosjektet' [Expectations About Writing Competence. Perspectives on Writing, Writing Education and Assessment in the Norm-Project]. *Viden om Læsning*, 15, 76–89.
- Stevanovic, M. (2012). Establishing joint decisions in a dyad. *Discourse Studies*, 14(6), 779–803. <https://doi.org/10.1177/1461445612456654>
- Stevanovic, M. (2013). Constructing a proposal as a thought: A way to manage problems in the initiation of joint decision-making in Finnish workplace interaction. *Pragmatics*, 23(3), 519–544.
- Stevanovic, M., & Peräkylä, A. (2012). Deontic authority in interaction: The right to announce, propose, and decide. *Research on Language & Social Interaction*, 45(3), 297–321. <https://doi.org/10.1080/08351813.2012.699260>
- Stivers, T. (2006). Treatment decisions: Negotiations between doctors and parents in acute care encounters. In J. Heritage, & D. W. Maynard (Eds.), *Communication in medical care: Interaction between primary care physicians and patients* (pp. 279–312). Cambridge University Press.
- Sundqvist, P., Sandlund, E., Skar, G., & Tengberg, M. (2020). Effects of rater training on the assessment of L2 English oral proficiency. *Nordic Journal of Modern Language Methodology*, 8(1), 3–29. <https://doi.org/10.46364/njmlm.v8i1.605>
- Vanlommel, K., Van Gasse, R., Vanhoof, J., & Van Petegem, P. (2017). Teachers' decision-making: Data based or intuition driven? *International Journal of Educational Research*, 83, 75–83. <https://doi.org/10.1016/j.ijer.2017.02.013>
- Vonen, M. N., Solem, M. S., & Skovholt, K. (2022). Managing students' insufficient answers in oral examinations. *Classroom Discourse*. <https://doi.org/10.1080/19463014.2022.2079694>.
- Wenger, E. (1998). *Communities of practice. Learning, meaning and identity*. Cambridge University Press.