



The Novelty Effect on Assessment Interventions: a Qualitative Replication Study of Oral Performance Assessment in Undergraduate Mathematics

Paola Iannone¹ · Nad'a Vondrová²

Received: 19 September 2022 / Accepted: 12 March 2023 / Published online: 19 April 2023
© The Author(s) 2023

Abstract

This paper reports on a qualitative replication study investigating the impact of the novelty effect on findings from interventions about the assessment of mathematics at university. The replication study used the same data collection tools of a previous study on oral assessment of mathematics, but data were collected in a context where oral assessment is the norm. We aimed to find whether the results of the two studies were comparable and whether there was plausible evidence of an impact of novelty effect on the findings of the original study. The findings of the current study appear to be comparable to those of the original study. Students associate oral assessment with the assessment of conceptual understanding and written assessment with the assessment of procedures; they report being more anxious about the oral assessment, but they perceive oral assessment as a better learning experience than closed book exams. However, in a culture where oral assessment is the norm, we found students engaging with learning also following considerations of the difficulties of other modules taken in the same period of their degree. Finally, in this culture, oral communication of mathematics is also much valued.

Keywords Novelty effect · Oral assessment of mathematics · Students' perceptions · University level

✉ Paola Iannone
p.iannone@lboro.ac.uk

Nad'a Vondrová
nada.vondrova@pedf.cuni.cz

¹ Department of Mathematics Education, Loughborough University, Loughborough, UK

² Faculty of Education, Charles University, Prague, Czech Republic

Introduction

When conducting educational research, we often introduce something new in the setting we are studying and seek to ascertain what the impact of this new activity was on students' learning, perceptions, attitudes, or beliefs according to the focus of the study. In doing so, we incur into something called 'the novelty effect' (Marek & Wu, 2021) by which the outcomes observed may be partly the outcome of the introduction of something new. This may be particularly true when a new assessment method is trialled at university in a discipline which has a very uniform assessment diet such as mathematics (Iannone & Simpson, 2022). The lack of familiarity with the method introduced may cause students to think that the new assessment is more demanding than anything else they have experienced, and they may therefore engage in learning conducive to conceptual understanding because of this perceived difficulty. It may also be that students see the assessment method as so difficult that they disengage completely and revert to procedural ways of learning even if those are inadequate for that assessment method, as in Gijbels and Dochy's (2006) study.

The aim of this paper is to investigate the impact that oral assessment has on mathematics students' engagement with learning and their perceptions of this assessment method in an educational setting where oral assessment of mathematics is the norm. Videnovic and Liljedahl (2018) distinguish between 'non-oral assessment cultures', that is, cultures in which oral exams are non-existent or extremely rare (such as the USA and the UK), and 'oral assessment cultures' where this method is the norm (such as Italy, Germany, and the Czech Republic). Previous studies (e.g. Huxham et al., 2012; Iannone & Simpson, 2015; Iannone et al., 2020) reported encouraging results regarding the affordances of oral assessment for STEM subjects at university—but those were all intervention studies carried out in countries with a non-oral assessment culture, and the novelty effect may have impacted on the findings. The methodology we use for the current study is that of a replication study. We employ the research tools of (Iannone et al., 2020) and deploy them in a country with a strong oral assessment culture—the Czech Republic. We sought to investigate whether the outcomes of our study are in line with those of previous studies and whether the assessment culture of the context impacts the study findings. Therefore, we asked:

RQ1: What are students' perceptions of oral assessment in an 'oral assessment culture'?

RQ2: How do these perceptions compare to those held by students from a 'non-oral assessment culture' who participated in intervention studies such as Iannone et al. (2020)?

Replication Studies and Qualitative Research

Replication of quantitative studies in psychology has gathered much attention after the so-called 'replication crisis' in the early 2000s (Pashler & Wagenmakers, 2012) which led, amongst other initiatives, to the Open Science Collaboration

(2015). In this seminal paper, the authors replicated 100 (quantitative) psychology studies and found that only 36% of the replications had statistically significant results, while 97% of the studies which were replicated did. While the issue of replicability and the importance of replication studies may appear to concern only quantitative methods, Makel et al. (2022) make a strong case for the relevance of a concept akin to replication for qualitative studies and focus on replicability—which implies that researchers must be able to convey their methods with clarity and transparency. Aguilar (2020) offered a classification of replication studies in mathematics education. Replication studies can inhabit a continuum between conceptual replication, where replication aims at reproducing previous results adopting different methods, and direct replication, where replication reproduces the very same experimental procedure of the study it aims to replicate. Aguilar (2020) also proposes a second classification, that is, internal replication, where the same scholar conducted the original study and its replication, and external replication, where the endeavour of replication is taken on by a researcher external to the team that carried out the study that is to be replicated.

The study reported in this paper is a type of direct replication where the study design and the research tools (i.e. questionnaires and interview schedules) follow that of the original study, but it is carried out with a different population in a contrasting culture. The study is also partially an internal replication, and for this reason, it is important to reflect on the positionality of the researchers, as Makel et al. (2022) recommend. Positionality, according to Manohar et al. (2017), requires researchers to position themselves with respect to their research. They may do so in three ways: with respect to the subject of research, to its participants, and to the research context. In the case of the research reported in this paper, the first author designed the study that is replicated (Iannone et al., 2020), and for this study, she was an external observer in that she did not belong to the institution where the study was conducted, although she worked in the same educational context. The first author is now an external observer of the context in which the replication study is conducted, this time external also to the educational context. With respect to the research, the assessment of mathematics at university level is one of the main research interests of the first author. The second author was not involved in the original study, but she is internal to the educational culture in which the replication took place. For the second author, the assessment of mathematics at university level is a new research area. We believe that the positionality of the two authors with respect to the first and second studies allows for a balance of approaches, and the close collaboration that led to this study may help recognising eventual biases.

Even if there are important aspects of replication studies, linked to their potential to better understand some of the phenomena that are found in mathematics education, there is still reluctance from journals to publish them (Aguilar, 2020). To tackle this issue, Star (2018) suggests three criteria for publication of replication studies. These are that the study:

1. Makes a convincing case that the study topic of the replication is of great importance to the field,

2. Makes a convincing case that the field will learn something significant from the replication that is not already known, and
3. Convincingly shows that there is reason to believe that the results of the original study may be flawed (Star, 2018, p. 99).

We believe the study reported in this paper fulfils both the first and the second point in Star's list. The first point—the importance of the topic—is directly related to the uniformity of the assessment diet in mathematics and the need to diversify this diet, as it has been stressed by many in mathematics education and more generally in the wider education literature. Moreover, assessment is one of the key themes for the future of mathematics education emerging from a recent survey of mathematics education researchers (Bakker et al., 2021).

The second of Star's (2018) points is related to the reason for the change of population in the replication study. The original study was an intervention study which introduced oral assessment in a student population that had never experienced such an assessment method. By replicating the same study with a population for whom that assessment method is the norm, we may ascertain which findings of the first study may have emerged because of the novelty effect, and we may learn whether there are other effects that have been hidden because of the novelty of the intervention.

Having mapped the methodological choice of designing a replication study, we now review the literature on oral assessment for STEM subjects from the students' viewpoint. We then report on the methodology of the study and its findings, and we conclude with some final remarks and implications for the assessment of mathematics.

Oral Performance Assessment in University Mathematics

Joughin (2010) distinguishes three types of dialogic assessment:

- *Presentation*, where the student has prepared a presentation of something akin to a small project, often followed by a short Q&A session;
- *Application*, such as the role-play assessment of medicine students when assessing a patient impersonated by an actor;
- *Interrogation*, covering everything from short-form question-and-answer to the viva voce used, in some countries, as final assessment for doctoral students.

This paper concerns the third type, and we maintain the name given to it in Iannone and Simpson (2015) of *oral performance assessment* (OPA). For brevity, we will also use *oral exams* as this is the term that participants use in the interviews. In mathematics, this assessment takes the form of a question-and-answer session with the student writing at the board or on paper while answering the questions that the assessor poses. There are several ways in which OPA can be organised: students may be asked to choose one question amongst a list or may be asked to draw a question from a selection and they may or may not be given some time to

think about the questions. The questions asked may be of a theoretical nature (state definitions, proofs, or theorems, explain passages in proofs) or of a more applied one (working out examples or using algorithms appropriately). Typical of this dialogic form of assessment is the presence of student-tailored contingent questions used by the assessor to probe students' understanding. In what follows, we review studies of OPA in STEM subjects or subjects with a significant component of mathematics (e.g. business and finance).

The literature on OPA at university mostly comprises intervention studies in non-oral assessment cultures and investigates the affordances and drawbacks of this assessment method and the impact that it has on students' experiences of assessment. Students in management and economics who were involved in an intervention where a component of OPA was introduced reported valuing this as a learning experience stating that they learned 'more' for this assessment than for traditional assessment methods (Rawls et al., 2015). The authors' analysis of students' performance in the year before the introduction of the OPA confirmed the students' perceptions by showing they achieved higher marks. The authors hypothesise that this increase in marks could be due to the changed way in which students prepared for the oral exam. In the context of computer science, Ohmann (2019) reports that students commented on the longer time it took to prepare for the oral exam and that they saw this assessment resulting in a better learning experience. These findings are also in agreement with those of Huxham et al. (2012) who compared OPA and traditional written assessment in a biology course. They found that students performed better in the oral exam than in the written one and that this type of assessment may contribute to the development of students' employability skills. They also found that while some students reported higher levels of anxiety associated to oral exams than to written assessment, this could be because they understood this task to require deeper understanding and felt the pressure of having to explain their reasoning to others. Therefore, the authors conclude, this anxiety may have been related to a richer learning experience. Akimov and Malin (2020) showed that OPA can also be successful in an online format, after they introduced this assessment at the end of a course for postgraduate finance students taught online. They conclude that although some students reported having worked harder for the oral exam, they generally valued the conversational nature of this assessment and the feedback that came with it. Similar results were reached by Kamber (2021) in a study with biology students for an online oral examination in a biochemistry course during a distance-learning period. The oral examination was viewed positively by students as a strategy to personalise their learning experience.

Very little literature concerns OPA specifically in mathematics. The studies of Iannone and Simpson (2015) and Iannone et al. (2020) investigated this assessment in two interventions: one in a first-year module where OPA was a small coursework component and the other in two third-year modules in financial mathematics where OPA was the sole assessment component. The first study findings are in line with what other studies reported: students associated OPA to requiring conceptual understanding, although they reported high levels of anxiety and uncertainty regarding fairness (in the sense of Sambell et al., 1997) given the absence of a strict mark scheme, typical of UK closed book exams. The results

of the second study—which is the one replicated in this paper—are of particular interest, and we summarise those in some detail.

In Iannone et al. (2020), OPA was introduced as a high-stake assessment for two third-year modules in a financial mathematics degree in a university in the UK, a non-oral assessment culture. The study was an intervention study—the sole assessment of the two modules was OPA, and it was designed and carried out in collaboration with the lecturers who taught the modules. Because of the novelty of the intervention, it was deemed necessary to introduce a mock OPA assessment during the teaching periods so that students could become familiar—to some extent—with this assessment mode. Data collection followed the data collection of the present paper—although the interviews had two rounds, before and after the OPA. The findings of the study can be summarised as follows: Participants believed that to be able to succeed in OPA, they must engage with conceptual understanding (Skemp, 1976). Because of this perception, many participants reported changing the way in which they prepared from methods associated to procedural understanding (e.g. solving past exam papers) to working with peers and engaging in questions and answers on the topics examined, which is associated to fostering conceptual understanding. Students also stated that they engaged with the lectures more due to the lack of familiarity with the assessment method. However, not all students changed their study habits. Some reverted to old revision methods as they were unsure what the assessment required, only to realise afterwards that the old methods did not prepare them well. Higher than normal levels of anxiety were mentioned by most students, and this was likely linked to the uncertainty of the assessment requirements and the presence of an assessor. However, as in Huxham et al. (2012), the results of the assessment were in line with results in the previous years, indicating that this perceived higher anxiety was unlikely to have impacted on the students' achievement. The relation of OPA to employability was also highlighted positively by many: for these students, the dialogic nature of the OPA was much closer to what they would be asked to do on the workplace, and they much appreciated to be given the chance to practice presenting complex ideas orally.

Methodology

The Local Context

The education of mathematics teachers in the Czech Republic comprises a 3-year bachelor's course and a 2-year master's course, necessary to obtain a teacher qualification. The bachelor and master's degrees are offered as a single-subject (e.g. mathematics) or two-subject course (e.g. mathematics and one more subject).

In Videnovic and Liljedahl's (2018) model of socio-cultural norms pertinent to oral assessment and non-oral assessment cultures, the Czech Republic falls into the former. The following norms are valid for its institutions: students are expected to take written and oral exams in their mathematics modules, to receive up to two marks for the same module: one for the written and one for the oral exam (these are aggregated into one final mark). Students are not shown any assessment rubric

for oral exams, and they must be prepared to explain their mathematical reasoning and their solutions. Mathematics lecturers in this country have relative freedom to decide how to teach and assess students. There are four ways in which university mathematics modules in Czech universities can be assessed, given by the course specifications. Those are:

- Zápočet (accruing credits): The result is pass/fail. It usually comprises tests held during teaching time or seminar works and/or a final written test. The exact combination of coursework and exam is decided by the lecturer and varies between modules;
- Klasifikovaný zápočet (accruing credits and marks): The same as above, but the student is assigned a mark (1, 2, 3, fail). It can also include oral performance assessment;
- Zkouška (examination): The student is assigned a mark (1, 2, 3, fail). The examination can be either written only, oral only, or both (which we will call *combined* from now on). The latter is the most frequent in mathematics modules;
- Zápočet a zkouška (accruing credits and examination): The credit (pass/fail) is a pre-requisite to be allowed to sit for the examination (1, 2, 3, fail).

OPA is traditional in Czech universities. There are no fixed rules for conducting it. It often comprises theoretical questions, but students may also be asked to solve problems during the oral exam. Usually, the student is first assigned a question (either by the instructor or by drawing from a list) and given some time to prepare before starting the OPA.

Methodological Tools

The questionnaire (from Iannone et al., 2020) comprised the Assessment Preferences Inventory adapted to mathematics (API) and the Gibbs and Simpson's questionnaire comparing students' perceptions of assessment between OPA and closed¹ book exam (AEQ, Gibbs & Simpson, 2004). The aim of the questionnaire was to ascertain students' preferences of assessment across a large sample and to investigate the comparative preferences between closed book exams and oral exams. It was translated into Czech by the second author who is a mother-tongue speaker who trained as an English teacher, and the Czech version was piloted with five graduates who did not participate in the main study to check the ease of understanding of the questions in translation.

The sample for the questionnaire was an opportunistic but relevant one. This population was of relatively easy access to the researchers, and their experiences were relevant to the study as these students took university-level mathematics modules. The link to the final version of the questionnaire (Appendix

¹ In this paper, we use the terms 'closed book exam' for a timed exam taken in a lecture hall where students do not have access to supporting material, and 'open book exam' for a timed exam taken in a lecture hall where students have access to supporting materials (e.g. lecture notes, textbooks).

Table 1 Questionnaire responses per year of study

Sample	Invitations sent	Questionnaires completed	
		N	%
Year 2	62	29	47
Year 3	25	17	68
Year 4 + 5 + 6 ^a	34	11	32
Graduated in the past 5 years	69	33	48
TOTAL	190	90	47

^aThe bachelor's lasts 3 years. However, some students study longer—they might have finished all the modules and wait before they apply for the state final examination or the defence of a bachelor's thesis. Or they might repeat failed modules in the subsequent academic year.

1—English version) was sent by email to all the students on the bachelor's degree in Mathematics with Education, except for year 1 students who were deemed not to have enough experience with university assessment at the time of the data collection. For this cohort the return rate was close to 50% (see Table 1). The questionnaire was also sent to students who graduated in the previous 5 years as they were likely to have their university experiences still fresh in their minds. For this cohort, the return rate was 48% (see Table 1). The first items of the questionnaire collected some biographical data, while the last item asked students to leave their email addresses if they agreed to take part in a follow-up interview. The questionnaire was completed anonymously online.

The interviews with the students investigated further the results of the questionnaire. The interview schedule (Appendix 2) was designed merging questions from the first and second rounds of interviews of Iannone et al. (2020). The interviews were conducted by two PhD students trained by the second author of this paper. The pilot interviews with two graduates were audio recorded, transcribed, and discussed during a meeting of the PhD students with the second author so that a shared interview schedule was obtained. The interviews were semi-structured (Flick, 2014) to allow for follow up questions. In total, 21 interviews were conducted, 4 with male and 4 with female students in the bachelor's degree, and 3 with male and 10 with female students in the master's degree.² These were all the students who had replied positively to an invitation to interview in the questionnaire. The interviews lasted between 15 and 30 min and were audio recorded and transcribed.

When asked about their perception of assessment methods, the students were encouraged to think back to their experience of mathematics modules only. Students also took courses in other subjects, and their perception of assessment methods in these may have been different from those of assessment methods in mathematics.

² These are the students who filled in our questionnaire as bachelor graduates and continued with their master study.

Table 2 Examples of code labels and relative definitions

Code label	Label explanation
Combined exam	Capturing the fact that students explicitly mentioned combined exams alongside the written and oral ones
Two-phase preparation	Capturing the fact that some students prepared first for the written exam and when they had passed it, they prepared for the oral exam, each time in a different way
Spontaneous mentioning	Capturing whether a student mentioned a characteristic, e.g. the difference in how they prepare for the written and for the oral exam, spontaneously or after the interviewer's question

Data Analysis

Questionnaire

Table 1 reports the responses from 90 questionnaires returned.

Given the nature of the questionnaire, only descriptive statistics were performed with the data, and those are reported in the 'Findings' section.

The transcripts of the interviews were analysed in an inductive-deductive way. A list of codes was prepared in advance based on the focus of the interviews (e.g. oral exam, types of preparation, materials used, advantages, disadvantages). This list was uploaded to the software ATLAS.ti (<https://atlasti.com>), and three PhD students coded one interview independently as a test. They then met with the second author to negotiate meanings of the codes and add new ones. An example of this process is given in Table 2.

In the third round, the coders coded the remaining interviews in such a way that each coded three interviews, and, when the coding was finished, they checked one another's coding. When the final version was reached, the codes were organised in categories (influence of the type of exam on the student's preparation, oral exam characteristics, written exam characteristics, combined exam characteristics) and quotations reports extracted from ATLAS.ti to prepare a summary text on the categories with illustrative quotations. Once the summary text was ready, the utterance in the categories and sub-categories within it were analysed again by both authors to find emergent themes that cut across the categories provided by the software (second cycle of coding: axial coding, Saldaña, 2015). Indeed, the categories provided by the software were related to the question asked while axial coding reconnected the data that were divided in the process of first coding (Saldaña, 2015). The overarching themes emerging from the analysis process were as follows: impact of the assessment method (on student's preparation), perception of competencies assessed by assessment methods, and students' experience with assessment methods.

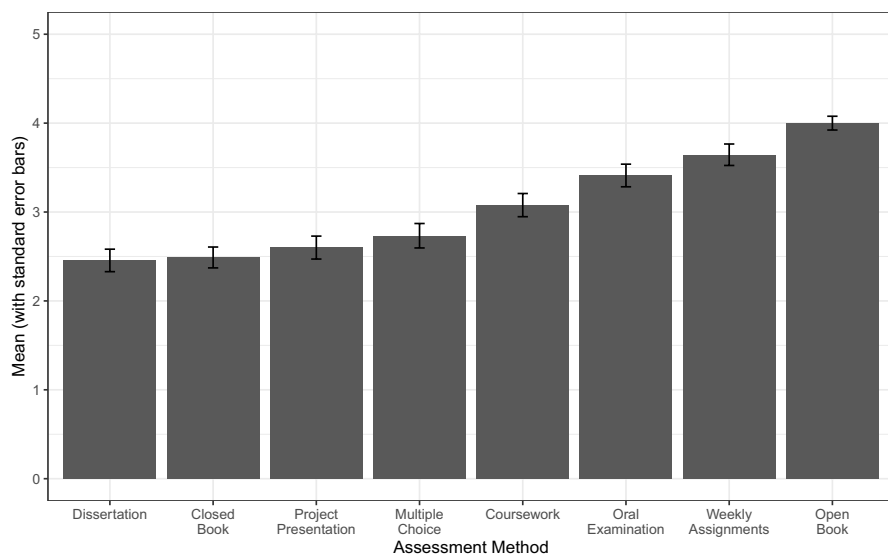


Fig. 1 Mean response (on a 5-point Likert scale) for preference for each of the assessment methods in the API (with standard error bars) for the 90 students in the sample. For the description of the assessment types, see Appendix 1

Findings

Questionnaire

There were 90 questionnaires returned (Fig. 1).

The first part of the questionnaire—the API adapted to mathematics—shows that the most preferred assessment methods for this cohort are open book, weekly assignments, and OPA, and that those are significantly more preferred than coursework. Closed book exams and dissertations are the least preferred methods. Of notice is that, in the Czech assessment diet, open book exams and weekly assignments are rarely used for mathematics, leaving the OPA as the most preferred assessment type of the ones that students have extended experience of.

Figure 2 reports the Gibbs and Simpson's (2004) questionnaire comparing OPA and closed book exams. Students see closed book exams to come with clearer instructions than any other assessment type; they also perceive it not to be challenging when compared to OPA, and think they are more likely to forget what they studied for a closed book exam than for OPA. Apparently contradictory is that participants see that OPA allows them to understand the subject better and to bring their learning together but also to some extent equate preparing for this assessment to memorising. We will elaborate on these findings and relate them to the interview findings during the discussion of this paper.

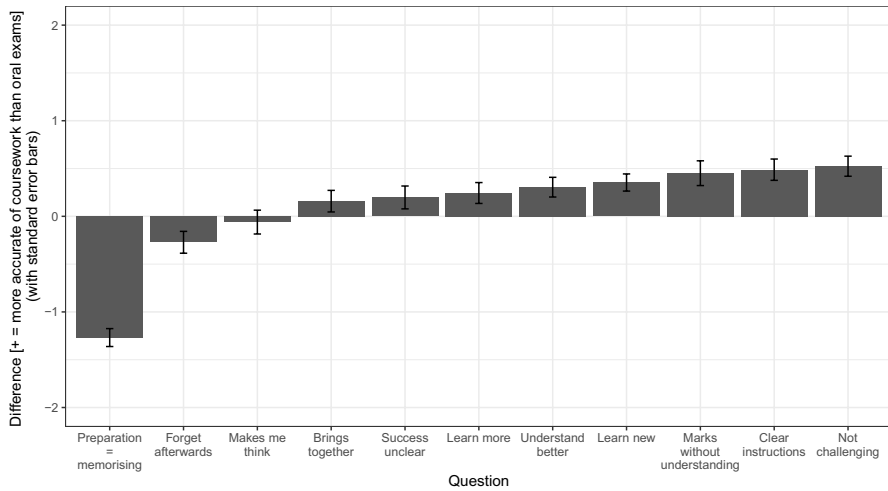


Fig. 2 Mean responses (on a 5-point Likert scale, from +2 [more accurate for written exams] to -2 [more accurate for oral exams]) for 11 statements in the AEQ (with standard error bars) for the 90 students in the sample

Interviews

In what follows, we describe in turn the overarching themes found from the analysis of the interviews.

Impact of the Type of the Assessment Method

In this category, we collected utterances where students describe the impact of assessment methods on their study habits. In describing the impact that the assessment method has on their preparation, students talk about whether they see a difference between how they prepare for oral or written exams, whether their preparation is concentrated at any time during the semester (i.e. cramming), and what materials they use for preparing for each of the assessment methods.

Our participants see a clear difference in the way in which they prepare for the oral and the written exams which is also linked to what they think each method assesses.

So, preparing for written tests basically means trying to solve all possible types of problems, solving different types of tasks. For the oral one, it means focusing on basic theorems and definitions that are used in other parts of the module. (Peter)

This difference manifests itself in the timing of revision, in the materials used, and in the mode of revision, i.e. collaboratively with peers or individually.

The analysis of the utterances related to the timing of students' preparation was difficult as it was often not possible to understand whether the word 'exam' that the

students used meant an oral, written, or combined exam. For those students who stated what exam type they were referring to, it was clear that the main factor that influenced their preparation timing was the perceived difficulty of the module they were preparing for more than the assessment type.

My goal at the beginning of each semester is (laughing): “I will be preparing the whole semester for that [assessment].” Unfortunately, it doesn’t always work out that way. But, for example, for the module of [Mathematical] Analysis, there I think I’ve always worked honestly and been solving those problems the entire semester. (Molly)

Some students described how a cramming strategy for oral exams did not work well and how eventually they changed their revision strategies.

[...] mostly when there was an oral examination, honestly, it was like, one rather took notes and learned it, like, before [the exam]. I began to learn two days, three days before the exam, given how difficult the subject was, but now it has served me better to do it more continually, so that there is not so much [subject matter at once]. (Matt)

Others connect their preparation for the written exam to that for the oral exam for the same module (when there is a combined exam) recognising that preparation during the whole of the teaching period is the best and that rote learning is not going to result in good grades.

When asked what materials students use to revise for exams, lecture notes emerged as the main resource for revisions for oral exams. However, together with the lecture notes, some students report the importance of focusing on the oral comments the lecturers make when explaining something in the lectures.

Well, I prefer to follow the lectures during the semester or at least attend them. I’m definitely not the type to, like, not go to lectures at all. Because I feel like it gives me so much. ... one really hears a lot. (Kitty)

For written exams, students report using problems from the seminars and past exam papers as revision material, but some mention also other resources such as the textbook, internet resources, and videos of problem solving related to the module they are preparing for.

As for those written exams, problems are usually solved. So, I prepare for it by having some problems from the seminars, so I usually go through them. And so far, I probably haven’t come across any situation in which it wouldn’t be enough. It has been enough for me to understand the principle of the problems solved in lectures or seminars. (Kitty)

As for the modes of preparation, students report preparing both collaboratively and on their own for the assessment, but many students (around a quarter of the sample reported this explicitly, while only 2 students explicitly stated that they study on their own) report how revising together with their peers allowed them to see when they had not understood something.

Joint preparation of some sort with my classmates worked best for me. [...] And it has probably been the most valuable, [...], to prepare for any exam or anything I need to learn. That I have an opportunity to consult with someone. (Rocky)

Only two students stated that they saw no difference in the way in which they prepared for oral or written exams—supporting the hypothesis that students do see these assessment methods having different roles and assessing different competencies, as we see in the following section.

Perception of Competencies Assessed

Students were clear when reflecting on the competencies they believe the assessment modes assess. For this study, it is interesting to note that students had extensive experience of combined exams, i.e. exams that comprise a written and an oral part; therefore, they experienced both written and OPA for the same syllabus.

Students recognise the oral exams as suitable to test mathematics theory and conceptual understanding.

[...] the oral part goes in more depth, and it really seems to me that only during the oral exam it becomes transparent whether the student understands or not. Which I think is the main part. (Sally)

The students consistently spoke about assessing theory and assessing procedures within the same topic. Some of them perceive these two aspects to be separate and that it is adequate that one assessment method assesses procedure and the other the theory.

In fact, only the theory is tested in the oral exam. [...] For the written exam, one simply learns how it is calculated, and in most cases, the theory is not needed at all for passing the written exam. But in the oral exam, one needs to know the theory. (Vera)

Others perceive theory and procedures to be linked and that it is only by understanding the theory that a student can master the procedural aspects of the topic.

I know, for example, a lot of the problems are based on theory, so it's also necessary there. But it's not like I'm learning it by heart. For example, in calculus, I just know that there is a theorem that says that the limits are solved so [...] I just know how to use it. But I can't prove it perfectly, explain it and formulate it. [I just learn this aspect] before the oral exam. (Molly)

Despite the way in which the students articulate the relation between competencies in mathematics, they seem to agree that the written exam assesses the ability to solve application-related questions (to which the students often refer as problem solving in the interviews) and that it is possible to do well in the written exam even if the theory has not been fully understood, while for success in the oral exam, the theory needs to be understood and to be explained also using appropriate examples.

Well, I would say that the written exam does not assess if I really understand it, because it's enough to learn the algorithm and one doesn't even have to understand it. While in the oral exam, [the lecturer] can usually see if you understand it. (Dave)

There is another interesting fact emerging from the interviews as well as from the questionnaire: some students perceive the written exam to be more demanding than the oral exam although they see it to be a test of computational proficiency and ease to apply known algorithms while they perceive the oral exam as a test of understanding. Therefore, we hypothesise that the dialogic nature of the oral exam has an impact on the overall experience of assessment, both in a positive and a negative way.

Students' Experience

Most students perceived the lecturer's support during the oral exam as an advantage to them: the lecturer can give the student a hint if the student is stuck or can correct a small mistake when it is made. While in the written exam, it is much more difficult to correct one's own mistakes as there is no dialogic dimension.

The disadvantage [of the written exam] is that if one does not know how to solve it, no one will help him/her, no one will show him/her a direction so that he/she can find the solution. Often it happens that a small numerical mistake is made which in the oral exam, the lecturer would point out and stop the student immediately. (Rita)

The immediacy of feedback was also appreciated by the students who saw oral exams as a learning experience.

I do not know if it is a complete advantage, [but] when a person goes there [to the oral exam] with the feeling that maybe he/she does not understand it well, it often happens that the teacher explains it to him/her. And that he/she learns it during the oral exam. (Molly)

On the other hand, students felt put under pressure by the oral assessment because of the role of contingent questions—with the possibility of dialog students felt there is nowhere to hide. Occasionally, the lecturer can make matters worse by helping which may result in students feeling even more anxious.

The time pressure and the risk of being asked something that the student does not know and would become clear from contingent questioning were mentioned by the students as negative aspects of the OPA. The possibility to ask contingent questions and probe deeper in the oral exam was mentioned by the students both as a positive (the students can show what they can do) and a negative aspect (the student can be found out on what they cannot do).

For example, one can learn only some parts and then somehow glue it together. In the oral exam, it becomes clear when you do not understand it. (Vera)

Finally, the issue of assessment anxiety was raised by most students regarding oral exams, often connected to the perceived pressure of needing to give an answer quickly.

During the oral exam, when this person is sitting in front of me, it's such a pressure as "Hurry! Think of something quickly and say something quickly".
(Maggie)

Again, the presence of the lecturer is perceived as having both positive and negative effects on students' anxiety. Some students perceive the pressure of having the lecturer in front of them waiting for an answer, but others stress how, if there is a good relationship with the lecturer, their presence can help them feel more at ease.

Discussion

This paper reports on a replication of the study by Iannone et al. (2020). It employed the same research tools but was conducted in an educational system where OPA is the norm. Here we discuss the findings of the study (RQ1), and we compare them to those of Iannone et al. (2020) (RQ2).

RQ1 asked about students' perceptions of assessment methods in an oral assessment culture—the Czech Republic—where students are often assessed on the same material in the written exam and in the subsequent OPA. This is an important feature of our sample: unlike the participants in the intervention studies, Czech students can compare their experiences of oral and written exams when employed to assess the same syllabus. The outcome of the questionnaires indicates that students preferred OPA amongst the assessment methods they experienced, and they thought that this assessment method was better at bringing the material together and was also perceived as a learning experience. The perception of OPA as a learning experience supports the findings from other studies (e.g. Huxham et al., 2012), and we hypothesise, with Rawls et al. (2015), that it relates to the presence of immediate feedback during the oral exam. Interestingly in our study, some students thought that OPA is about memorising more than the written exams. This apparent contradiction may be explained by the fact that, in the Czech Republic, OPA can be implemented in many ways and those ways may be perceived differently by the students. Written exams on the other hand are perceived having clearer instructions and marking schemes, which makes it clear what needs to be done to be successful. As for the competencies assessed by these methods, students saw OPA associated to assessing conceptual understanding, while written exams were predominantly perceived to assess computational methods. Therefore, they report different strategies when preparing for the two types of exams: when preparing for written exams, students focus on solving mathematical problems (from the seminars and past exam papers), while when preparing for the oral exams, they mainly focus on theory (theorems and proofs). These findings are noteworthy when considered from the perspective of relational and procedural understanding (Skemp, 1976) as the students in our sample seem to relate success in oral exams to relational understanding and success in written exams to procedural understanding. We recall, however, that for mathematics,

both aspects are fundamental parts of learning: being fluent in applying mathematics algorithms and performing computation is an important part of learning mathematics (Foster, 2018) because by automating certain procedures, the space is created for creative thinking. Although the concept of mathematical fluency in Foster (2018) is richer than what is meant by procedural understanding in Skemp (1976), we argue that being versed in handling complex procedures and algorithms is also an important part of learning mathematics. Finding both these aspects in the students' reflection indicates that they perceive both these competencies to be important for mathematics as these are competencies tested by their lecturers.

Finally, the dialogic nature of the oral exam was seen by participants both as having positive and negative impact on the assessment experience. While most students report higher levels of anxiety for OPA than for written exams, as seen from the questionnaire analysis and the findings in other studies (e.g. Huxham et al., 2012; Iannone et al., 2020), they also appreciate the immediate feedback received during the assessment, again describing this as a learning experience. Some students however highlighted how the relationship to the lecturer is linked to the assessment experience and hinted at how—if this relationship is not perceived positively—the OPA can become very stressful. To sum up, the findings from our study are similar to findings from other studies on OPA, but there are some significant nuances which we report below and are part of the discussion for the second research question.

RQ2 set out a comparison between the result of the current study and that of the study by Iannone et al. (2020) from which it adopted the research tools. The main difference from the quantitative findings of Iannone et al. (2020) is that closed book exams, for students in the current study, are one of the least preferred assessment methods, while this assessment method was amongst the most preferred for students in Iannone et al. (2020). We hypothesise that this preference stems from the familiarity that UK students have with the closed book exam: an assessment method that they experienced extensively and in which they have been successful.³ Comparing the outcome of the second questionnaire with that of previous studies (Iannone & Simpson, 2015; Iannone et al., 2020), we notice that the more complex views of the students in our sample regarding oral assessment may originate from the fact that they can differentiate between implementations of OPA. Therefore, we may infer that the perceptions of OPA in our study are more nuanced than those of students who only experienced one implementation. Research on the impact of different implementations of OPA could help clarify this point. Finally, common to all studies are the high levels of anxiety reported by the students for OPA which seems, however, not to impact significantly on outcomes (see also Rawls et al., 2015). These findings indicate that familiarity does not alleviate anxiety for OPA, and therefore that the high level of anxiety is not likely to be an outcome of the novelty of the intervention.

There are some interesting dimensions to the impact of OPA on study habits that emerge from our study and were not visible in the previous study. For our students, the perceived difficulty of the topic that will be assessed compared to other topics studied in the same period shapes their study habits, as well as the assessment methods used.

³ The university where Iannone et al.'s (2020) study was conducted is amongst the best universities in the UK and admits only students who achieve the highest marks (assessed solely by a closed book exam) in mathematics at the end of their secondary school.

For example, our participants recognised how—for modules perceived as very difficult such as Analysis—they prepare throughout the teaching period by attending the seminars and the lectures. This consideration of the relative difficulty of topics across the degree course did not appear in any of the intervention studies we reviewed. Therefore, there appear to be another dimension contributing to the students' engagement with the subject linked to assessment: namely the perceived difficulty of the module to be assessed in comparison to the other modules in the same semester/degree course.

Another finding worthy of attention which is not present in Iannone et al. (2020) is that students recognise the importance of the oral comments that the lecturer makes during a lecture to be successful in the oral assessment, and report focusing not only on what is written on the board during the lectures but also on what the lecturer says. This is very important for university mathematics—a subject still taught very traditionally in what Artemeva and Fox (2011) call a 'chalk-talk' format (preferred also in the Czech Republic). In a traditional university mathematics lecture, the lecturer writes formal mathematics content on the board and then gives an oral commentary to the students which will often include remarks about the methods used and links to other parts of the syllabus (Weber, 2004). We hypothesise that this attention to the spoken content may come because learning happens in an oral assessment culture where oral exposition is valued and it may link to the students' perception of OPA as a learning experience. Indeed, recent studies carried out in the USA (a non-oral assessment culture) show that while much of the pedagogical reflections of the lecturers is conveyed in lecturers' oral comments (Weber, 2004), students do not pay attention to those comments, do not record them in their notes, and do not remember them after the lecture. They focus instead only on the material that is written on the board (Fukawa-Connelly et al., 2017; Lew et al., 2016), and it is this material only that is recorded in students' notes. This finding is important for mathematics and may be relevant also in the teaching of other STEM subjects: a diverse assessment diet may encourage students to value oral communication.

Finally, in our data, there is no reference to the value of oral assessment for employability skills, as Iannone et al. (2020) found. We hypothesise that this is due to the fact that in the UK, the employability discourse is much present at the institutional level (Chadha & Toner, 2017) in all universities' marketing materials for students, and it is used to justify to students the 'value for money' that a particular university offers. This is not the case in the Czech Republic where most universities (and all universities with teacher education) are still publicly funded. This fact may confirm once more that the educational context plays a role in shaping the students' experience.

Concluding Remarks

In this paper, we reported a study investigating the impact of the novelty effect on findings from intervention studies on oral assessment by way of constructing a replication study to a previous study (Iannone et al., 2020). We found that many of the findings reported in the previous study are still relevant, although the students in our sample had a degree-wide view of their assessment and of the way in which they engaged with OPA, which students in the intervention study could not have. We also found that this degree-wide perception is linked to students' perceptions

of what capabilities are assessed by which assessment method, stressing again the need for variety in assessment. The positive attitude that students have toward OPA and the perception that this assessment tests conceptual understanding are a common thread in OPA studies indicating that this assessment method may indeed encourage students to engage conceptually with the subject and that this is unlikely to be a result of the novelty effect. The main finding of the study, however, is that orality is valued by the students in oral assessment cultures, and this is reflected in the way in which they engage with the teaching indicating once more that assessment conveys messages to students about what the assessors value.

The decision to replicate a qualitative study is also interesting: in the current study we have shown how changing the population of a qualitative study can illustrate some of the ways in which educational context may impact on findings. We have been careful to detail our methods and our research tools so that other researchers may use the same methods in other contexts, taking care of replicability as Makel et al. (2022) recommend.

The limitations of our study concern the sample, only students from one degree course in one university took part, and the positionality of the researchers—of great importance in internal replication studies—may have led to some internal bias. We, however, believe that we have tried to mitigate the impact of bias throughout the research and that the presence of such bias may be unavoidable in qualitative research. For this reason, we have taken great care to make the research process transparent in order to minimise such risk. Despite these limitations, and because of the consistence of findings with previous research, we believe we have illustrated how the novelty effect can impact findings of intervention studies on assessment and how this effect may conceal some important degree-wide outcomes of the assessment studied.

Appendix 1

In this appendix, we include the text of the questionnaire: the first part is the API (Iannone et al., 2020), and the second is the Gibbs and Simpson (2004). For further details on the use of the scale and its uses/reliability and measures, see Gibbs and Simpson (2004).

The questionnaire

About you.

-
- | | |
|---|---|
| 1a. Please indicate your age
I am _____ years old | 1c. Please tell us what your final average grade at
the end was of |
| 1b. Please indicate your gender by ticking the
appropriate box | Year 1 ____ Year 2 ____ |
| ●M ● F ● Prefer not to say | |
-

About Assessment.

1. To what extent would you want your achievements in the course to be assessed by each of the following: (from 1 Hardly at all to 5 Almost exclusively).

		1	2	3	4	5
Multiple choice examination	(e.g. a test taken in an exam room, where for each question you have to select one response from five possible choices)					
Written examination with no support materials	(e.g. a test taken in an exam room, with a separate booklet in which you write solutions, but where you are not allowed to use a calculator, books, or any other support materials)					
Written examination with support materials	(e.g. a test taken in an exam room, with a separate booklet in which you write solutions, but where you are allowed a copy of the standard textbook for the course)					
Weekly examples sheets	(e.g. a test which you complete in your own time over the course of a week, based on the material covered in the course over that week)					
Project coursework	(e.g. a piece of written work submitted in response to a question or problem, undertaken over the course of a number of weeks)					
Project presentation	(e.g. an oral presentation of the results of a project, undertaken in response to a set question or problem, after working on the project for a number of weeks)					
Oral examination	(e.g. working on a mathematical problem on a chalkboard or piece of paper with a tutor present who can provide suggestions or check errors as you work on it)					
Dissertation	(e.g. a substantial piece of written work, on a set topic or problem, undertaken over the course of a long period, such as a term or two)					

About assessment by written exam and assessment by oral exam

Now compare written exams to oral exams. Please rate the statements below where:

- 2: Much more accurate of oral assessment than written exams.
- 1: More accurate of oral assessment than written exams.
- 0: About the same of written exams and oral assessment.
- + 1: More accurate of written exams than oral assessment.
- + 2: Much more accurate of written exams than oral assessment.

	– 2	– 1	0	1	2
Tackling the assessment really makes me think					
I learn more from doing the assessment than from studying the course material					
In completing the assessment, you can get away with not understanding and still get high marks					
The assessment gives very clear instructions about what you are expected to do					
When I tackle an assessment, it is not at all clear what would count as a successful answer					
The assessment is not very challenging					
Preparing for the assessment is mainly a matter of memorising					
Doing the assessment brings things together for me					
I learn new things while preparing for the assessment					
I understand things better as a result of the assessment					
I'll probably forget most of what I learned for the assessment after the assessment is over					

Appendix 2. Students' interview schedule

The initial questions of the interviews were:

- When you prepared for an exam, what information you needed to have in advance? Remember your mathematics modules: which ones ended with a written exam and which with an oral one?
- How did you prepare for these exams?
- What are in your opinions the advantages and disadvantages of a written/oral exam?
- Which one do you prefer for a mathematics module and why?
- Which type of exam you think is more objective?
- Would you welcome another way of being assessed?
- Which characteristics does a student successful in a written/oral exam have, in your opinion?

Acknowledgements We would like to thank Kristýna Nižňanská, David Zenkl, and Pavel Sovič for collecting the interview data and helping with their analysis. We thank Adrian Simpson for his help with the statistical analysis. Finally, we would like to thank the students who took part in the study for giving up their time to engage with our project.

Data Availability The datasets generated during and/or analysed during the current study are not publicly available due to its nature: the data is mostly qualitative, and it is not made publicly available to protect the anonymity of the participants under the European General Data Protection Regulation and the conditions of the ethics agreements signed with the participants.

Declarations

Ethics Approval and Consent to Participate The project was approved by the Committee for Ethics in Research of the Faculty of Education, Charles University, Prague, application number KEV2003. In line with BERA requirements, participants agreed with the points in the informed consent forms which describe the use of the data. Names have been changed to protect anonymity.

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aguilar, M. S. (2020). Replication studies in mathematics education: What kind of questions would be productive to explore? *International Journal of Science and Mathematics Education*, 18(1), 37–50. <https://doi.org/10.1007/s10763-020-10069-7>
- Akimov, A., & Malin, M. (2020). When old becomes new: A case study of oral examination as an online assessment tool. *Assessment & Evaluation in Higher Education*, 45(8), 1205–1221. <https://doi.org/10.1080/02602938.2020.1730301>
- Artemeva, N., & Fox, J. (2011). The writing's on the board: The global and the local in teaching undergraduate mathematics through chalk talk. *Written Communication*, 28(4), 345–379. <https://doi.org/10.1177/0741088311419630>
- Bakker, A., Cai, J., & Zenger, L. (2021). Future themes of mathematics education research: An international survey before and during the pandemic. *Educational Studies in Mathematics*, 107, 1–24. <https://doi.org/10.1007/s10649-021-10049-w>
- Chadha, D., & Toner, J. (2017). Focusing in on employability: Using content analysis to explore the employability discourse in UK and USA universities. *International Journal of Educational Technology in Higher Education*, 14(1), 1–26. <https://doi.org/10.1186/s41239-017-0071-0>
- Flick, U. (2014). *An introduction to qualitative research*. Sage.
- Foster, C. (2018). Developing mathematical fluency: Comparing exercises and rich tasks. *Educational Studies in Mathematics*, 97(2), 121–141. <https://doi.org/10.1007/s10649-017-9788-x>
- Fukawa-Connelly, T., Weber, K., & Mejia-Ramos, J. P. (2017). Informal content and student note-taking in advanced mathematics classes. *Journal for Research in Mathematics Education*, 48(5), 567–579. <https://doi.org/10.5951/jresmetheduc.48.5.0567>

- Gibbs, G., & Simpson, C. (2004). Measuring the response of students to assessment: Assessment experience questionnaire". In C. Rust (Ed.), *Improving student learning: Theory, research and scholarship* (pp. 171–185). The Oxford Centre for Staff and Learning Development.
- Gijbels, D., & Dochy, F. (2006). Students' assessment preferences and approaches to learning: Can formative assessment make a difference? *Educational Studies*, 32(4), 399–409. <https://doi.org/10.1080/03055690600850354>
- Huxham, M., Campbell, F., & Westwood, J. (2012). Oral versus written assessments: A test of student performance and attitudes. *Assessment & Evaluation in Higher Education*, 37(1), 125–136. <https://doi.org/10.1080/02602938.2010.515012>
- Iannone, P., Czichowsky, C., & Ruf, J. (2020). The impact of high stakes oral performance assessment on students' approaches to learning: A case study. *Educational Studies in Mathematics*, 103(3), 313–337. <https://doi.org/10.1007/s10649-020-09937-4>
- Iannone, P., & Simpson, A. (2015). Students' views of oral performance Assessment in mathematics: Straddling 'assessment of' and 'assessment for' learning divide. *Assessment and Evaluation in Higher Education*, 40, 971–987. <https://doi.org/10.1080/02602938.2014.961124>
- Iannone, P., & Simpson, A. (2022). The summative assessment diet: How we assess in mathematics degrees. *Teaching Mathematics and its Applications*, 41(1), 22–31. <https://doi.org/10.1093/teamat/hrab007>
- Johnson, B., & Turner, L. A. (2003). Data collection strategies in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 297–319). Sage.
- Joughin, G. (2010). *A short guide to oral assessment*. Leeds Met Press in association with University of Wollongong.
- Kamber, D. N. (2021). Personalized distance-learning experience through virtual oral examinations in an undergraduate biochemistry course. *Journal of Chemical Education*, 98(2), 395–399. <https://doi.org/10.1021/acs.jchemed.0c01092>
- Lew, K., Fukawa-Connelly, T. P., Mejia-Ramos, J. P., & Weber, K. (2016). Lectures in advanced mathematics: Why students might not understand what the mathematics professor is trying to convey. *Journal for Research in Mathematics Education*, 47(2), 162–198. <https://doi.org/10.5951/jresematheduc.47.2.0162>
- Makel, M. C., Meyer, M. S., Simonsen, M. A., Roberts, M. A., & Plucker, J. A. (2022). Replication is relevant to qualitative research. *Educational Research and Evaluation*, 27(1–2), 215–219. <https://doi.org/10.1080/13803611.2021.2022310>
- Manohar, N., Liamputtong, P., Bhole, S., & Arora, A. (2017). Researcher positionality in cross-cultural and sensitive research. In P. Liamputtong (Ed.), *Handbook of research methods in health social sciences* (pp. 1–15). Springer. https://doi.org/10.1007/978-981-10-2779-6_35-1
- Marek, M. W., & Wu, W. C. V. (2021). Motivational affordances of the novelty effect in TELL. In N. Chen & M. Chang (Eds.), *2021 International Conference on Advanced Learning Technologies (ICALT)* (pp. 268–269). IEEE. <https://doi.org/10.1109/ICALT52272.2021.00086>
- Ohmann, P. (2019). An assessment of oral exams in introductory CS. In E. Hawthorne, M. Pérez-Quinones, S. Heckman, & J. Zhang (Eds.), *SIGCSE '19: Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 613–619). Association for Computing Machinery. <https://doi.org/10.1145/3287324.3287489>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), p. aac4716. <https://doi.org/10.1126/science.aac4716>
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528–530. <https://doi.org/10.1177/1745691612465253>
- Rawls, J., Wilsker, A., & Rawls, R. S. (2015). Are you talking to me? On the use of oral examinations in undergraduate business courses. *Journal of the Academy of Business Education*, 16, 22–33.
- Saldaña, J. (2015). *The coding manual for qualitative researchers*. Sage.
- Sambell, K., McDowell, L., & Brown, S. (1997). "But is it fair?": An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, 23(4), 349–371. [https://doi.org/10.1016/S0191-491X\(97\)86215-3](https://doi.org/10.1016/S0191-491X(97)86215-3)
- Skemp, R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching*, 77, 20–26.
- Star, J. R. (2018). Research commentary: When and why replication studies should be published: Guidelines for mathematics education journals. *Journal for Research in Mathematics Education*, 49(1), 98–103. <https://doi.org/10.5951/jresmetheduc.49.1.0098>

- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education*, 30(4), 325–341. <https://doi.org/10.1080/02602930500099102>
- Videnovic, M., & Liljedahl, P. (2018). Espoused vs enacted model of mathematics assessment in case of oral and non-oral assessment culture. *Open Mathematical Education Notes*, 8(2018), 35–46. <https://doi.org/10.7251/OMEN1802035V>
- Weber, K. (2004). Traditional instruction in advanced mathematics courses: A case study of one professor's lectures and proofs in an introductory real analysis course. *The Journal of Mathematical Behavior*, 23(2), 115–133. <https://doi.org/10.1016/j.jmathb.2004.03.001>