

A Case for Deep Learning Through Online Oral Examinations: An Autoethnographical Exploration of a Change of Assessment Method

Jaakko Turunen
Södertörn University, Sweden

The Covid-19 pandemic made digital oral exams a necessity for most higher education courses in Sweden and elsewhere in the world. Yet, more reflection on oral examination could help understand their utility in contemporary higher education. This paper probes the relationship between learning habits and assessment through an autoethnographic exploration into oral examination in an intermediate quantitative methods class for social science students in Sweden. Drawing on theories on constructive alignment and deep learning, this article makes the case for oral examinations as facilitating deep(er) learning. At the same time, the article discusses the oral examination as a critical way to think about the formal assessment culture brought about by ever tightening teaching budgets and external audits.

Keywords: deep-learning, online oral examination, constructive alignment, quantitative methods, methods teaching

INTRODUCTION

One of the consequences of the Covid-19 pandemic and digital teaching in Sweden and elsewhere in the world was the imperative to find out alternatives to sit-in examinations in higher education. A take-home exam was a regular alternative, but what about a digital oral examination? Oral examinations in the humanities and social sciences have not been very popular over the recent decades and there are serious doubts if they are suitable for the task due to risks of subjective assessment and the lack of a written record for auditing purposes (Davis & Karunathilake, 2005; Evans et al., 1966). In this article, I will argue that oral exams, exemplified through digital oral exams, can in fact generate deep(er) learning than take-home exams, and many of the shortcomings outlined in the literature can be avoided or effectively mitigated.

Given the continuing interest in the relationship of how different approaches to learning correlate to academic achievement (Entwistle, 1998; Entwistle & Entwistle, 1991; Herrmann et al., 2017; Marton & Säljö, 1984/1997), this paper contributes to the investigation of this relationship by discussing the advantages of the oral exam in terms of learning achievements and organisational efficiency but also as a way to rethink assessment practices in higher education. Digital oral examinations made an un-planned and often unwanted comeback in the wake of the Covid-19 restrictions and digital education. Yet, oral examinations may be the answer to more questions than just the pandemic. Invoking the languages of constructive alignment

Correspondence: Jaakko Turunen, e-mail: jaakko.turunen@sh.se

Artiklar och reflektioner är kollegialt granskade. Övriga bidragstyper granskas av redaktionen. Se <https://hogreutbildning.se> ISSN 2000-7558

©2023 Jaakko Turunen. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), allowing third parties to share their work (copy, distribute, transmit) and to adapt it, under the condition that the authors are given credit, that the work is not used for commercial purposes, and that in the event of reuse or distribution, the terms of this license are made clear.

Citation: Turunen, J. (2023). «A Case for Deep Learning through Online Oral Examinations: an Autoethnographical Exploration of a Change of Assessment Method», *Högre utbildning*, 13(3), 72–86. <https://doi.org/10.23865/hu.v13.3172>

(Biggs & Tang, 2011) and the Gothenburg school's deep and surface learning (Entwistle, 1997; Marton & Säljö, 1984/1997) this paper provides an autoethnographic case study of three years of oral examinations in a second-year quantitative methods course in the social sciences at a Swedish university.

The purpose of this text is to discuss the oral exam as an underused form of assessment that can address many current challenges in teaching methods courses under "academic austerity," but also as a way of contributing to better, or deeper, learning. I will employ the methods of autoethnography (Adams et al., 2015; Ellis et al., 2011) to do so. Acknowledging that autoethnography and quantitative methods belong to different ontological and epistemological worlds, I try to render the autoethnographic method and the insights it can provide accessible to teachers versed primarily in quantitative methods. This involves making some concessions to often competing scientific paradigms. Even if the oral exam has been discussed a bit more in connection to some disciplines, there is little previous research on oral examination in an online environment (Akimov & Malin, 2020). This paper makes a modest contribution in this respect. Finally, I will also discuss the oral exam as giving rise to an alternative culture of assessment focusing more on the processes of learning than the summative assessment of the product of learning.

I have been teaching this class for over 7 years now. In 2020, I changed the examination from a take-home exam to an oral exam. Although Sweden has returned to campus tuition, my quantitative methods class stayed on Zoom for examinations. Below, I will first discuss the quantitative methods course, then introduce autoethnography as a method for collecting and analysing data making it possible to take the reader with me to follow the change in assessment. I will link my observations from the take-home exam to existing literature, introduce the theories of constructive alignment and deep-learning and discuss my changing way of understanding assessment in the wake of running the oral exams. The conclusion will return to constructive alignment linking the oral exam to other learning activities and discuss their implications for the assessment culture in social sciences.

THE COURSE IN QUANTITATIVE METHODS

The course where I carried out my "experiment" is a second-year course in quantitative research methods. It is expected that the students know the basics of quantitative research, such as what are the variable types (nominal, ordinal, or continuous), what is the basic logic in quantitative research, and how to test an association between two nominal variables (so-called Chi-square test), but it is during this course that they are introduced to a statistical programme (previously SPSS, now Jamovi) and that they are really expected to master the idea that continuous variables (such as income, or temperature), ordinal variables (such as an opinion ranging from 1 to 5), and nominal variables (such as occupation category or gender) must be treated differently in quantitative research, match the right statistical tests with the right kind of variables, and ask questions that feasibly can be answered with the kind of data that is available (i.e. those continuous, ordinal, or nominal variables). The course builds on "applied" quantitative methods and we use a ready database with questions related to the youth, school, free time, and alcohol use. The main learning outcome of the course could be summarised so that the students learn how to test associations between different variables in a dataset and to interpret the results of most basic statistical tests presented in a table. This involves some learning outcomes that require memorising facts, but the most concern the construction of a statistical test and the interpretation of the results where focus is on the way the outcome is produced, what is revealed by the result, and what is disguised, for instance, by different recodings of the variables. The number

of students is normally around 50; the biggest class was just above 70 and the smallest was 35. The previous take-home exam was to be between 1,500 and 1,700 words (seven to eight pages), which produced a pile of around 450 pages to correct.

METHODS AND DATA: AUTOETHNOGRAPHY OF AN “EXPERIMENT IN ORAL EXAMINATION”

The idea to change from a take-home exam to an oral exam was planted in my head in one workshop on alternative examination formats. More and more contextual factors signalled that an oral examination would be worth trying, and finally Covid-19 and the decision to go to digital teaching made the change even look natural. However, the change was only half-planned: I had never carried out oral examinations; I recall only once taking one myself, in Poland in 1999. So not only the oral examination evolved over time, also my own experience in carrying out oral examinations grew. To make sense of all this, I needed a method that could show the evolution of the oral exam in my course and the way how my own stance towards assessment had evolved as a result of carrying out the oral exam.

Autoethnography as a method is quite different from critical realist quantitative research, but it can deliver insights valid across disciplines and research paradigms. Ellis et al. (2011) describe autoethnography as a shift from “physics” to “literature” in social sciences, to literature that would be “meaningful, accessible, and evocative research grounded in personal experience” (Ellis et al., 2011). Autoethnography is sensitive to personal experiences in the research process. The *auto* in autoethnography refers to a selective recollection and narration of past experiences, chosen with the benefit of hindsight. The *ethnography* refers to the cultural in the material such as common values and beliefs as well shared experiences necessary for both the insiders and outsiders to understand the cultural context (Ellis et al., 2011). In this article, “culture” refers to the role of assessment in higher education – and more concretely to the change from summative to more formative assessment in oral examination in my case. The *graphy* further indicates that autoethnography requires an analytical annotation of those experiences that render the culture transparent. In this paper, I will do so by linking my reflections to prior research on oral exams and assessment.

Autoethnography evolved slowly inside ethnographic research from the 1970s onwards, when research practices started to show increasing awareness of researcher’s own influence on the research (Adams et al., 2015). Combined with the rise of identity politics and critical epistemologies, by the 1990s the time was ripe for placing the researcher’s personal narratives in the centre of research. Autoethnography tackles three issues that had become problematic in social scientific research. The first concerned the epistemological ideal of neutrality of the researcher and the objectivity of findings. The second concerned the ethics of research – especially among vulnerable groups. The third concerned the importance and (unavoidable) impact of identities in the processes through which material was obtained. My decision to opt for autoethnography is based on the first concern. First, the material I have is primarily in the form of personal reflection and memory. There are some facts and numbers, which I will roll in, and there are course evaluations. Whilst these count as objective descriptions even in the realist paradigm, I believe they are less interesting without the contextual reflection that embeds them in the practice of oral examination. More useful insights concern that, which *actually* happens during the oral exam (Davis & Karunathilake, 2005). Due to ethical considerations, students are not interviewed for this study making my own reflections about the actual practice of the oral exam the only available material. Second, autoethnography can render the cultural practice transparent.

Quantitative research methods involve constant choices between suboptimal alternatives; the mastery of the method resides more in the logic of decision than in the correct implementation of that decision. Looking back in time, I see that the oral exam changed the way I understood what was at stake in the assessment from summative assessment of a product to a combination of summative and formative assessment of a process of learning.

There is also a third reason for choosing autoethnography as the method of inquiry, which is that the idea of writing this paper emerged only in spring 2021, when I presented my exploration into oral exams at a seminar dedicated to teaching methods. The exam itself had been evolving all the time making any staged comparison between two fixed alternatives impossible. My aim is not to provide a comparison of pros and cons of the oral exam vs. the take-home exam, but to give an accessible description why and how I consider the oral exam to be a beneficial alternative to the take-home exam. In doing so, I will follow the steps – proceeding from personal experienced through sense-making to reflexivity yielding to insights about cultural insider's perspective seeking to produce a description of culture's norms and seek response from audience – described in Adams et al. (2015) with some necessary modifications. The main material is my own reflections on the practice of oral examinations (personal experience). I will describe these reflections primarily with the focus on how I understood i) how the exam is designed, ii) what influence this design has on what is actually taking place during the exam (sense-making) and iii) how I relate to the changes I observed (reflexivity). Together, I hope, this account can show through insider's insights how the oral exam as a culturally embedded practice can both encourage the students to deep learning and reduce the teacher's burden with exam correction that has little benefit for students' learning.

Balancing this study ethically has not been easy. The very idea of interactively constructed identities imply that the students are somehow *present* in my personal account (Edwards, 2021) – yet they have not given their informed consent. The way I have tried to position this study is to eliminate any interactive element in the account thereby perhaps providing not a full ethnographic account of the oral exam – including the students' part in it – but a partial view focusing on my side as an examiner (not as an individual in relationship with the students) and including the students' side only as aggregate information. In other words, the focus is on the culture of examination rather than on the individuals taking part in that culture. In addition, the “students” are treated here as a mass, an aggregate, where the identification of any individual is impossible (which has been often the ethical problem with autoethnographic studies (see Edwards, 2021, for examples). This design has been reviewed by the department where the oral exams have been carried out for ethical considerations.

Growing impatient with take-home exams

The take-home exam is common in Sweden. It can be defined as an open-ended question(s) that requires the students to apply their knowledge to a specific problem within a limited time and with whatever resources available, excluding plagiarism. Organisationally, the take-home exam is flexible, and students who work alongside their studies appreciate the possibility of working on their course papers when it suits them. Pedagogically, the take-home exams fit the needs of the 21st century students and learning (Hall, 2011). Take-home exams reduce student stress and allegedly promote what according to Bloom's taxonomy are “higher-order cognitive skills” (Bengtsson, 2019). Bloom's taxonomy comprises six hierarchically organised levels of learning starting from the root level of remembering and progressing through understanding, applying, analysing, and evaluating to the highest level of creating new original knowledge. Learning

higher-order cognitive skills requires that “students can define problems, predict, hypothesise, experiment, analyse, conclude and are capable of reflective thinking” (Bengtsson, 2019, p. 1). Further, as take-home exams are geared to produce individual answers, they also can reduce chances for cheating (Mohanna & Patel, 2016), whilst allowing for some room for collaborative learning among the students (Johnson et al., 2015). According to Bengtsson’s systematic literature review, there seems to be a consensus that take-home exams suit better for assessing higher-order cognitive skills and should be avoided in introductory classes, where focus is more on memorising and presenting content from textbooks. Examples suitable for take-home exams include applications of knowledge to new areas and synthesising material (Bengtsson, 2019). A methods course is a good example where the take-home exam should bloom and this was very much the culture I inherited and accepted when I took on the course, until it was the time to rethink.

The culture of take-home exams, the summative assessment, and the demands for transparency I inherited were not without problems. My own experience largely concurs with the challenges raised in the literature. First, the main challenge with take-home exams is the possibility of unethical behaviour (Bengtsson, 2019). I have encountered two kinds of plagiarism. In the first case, a student copies other student’s work and is easily caught by electronic plagiarism controls or simply by providing the examiner a “*déjà-vu*” probing to search for the other text. Electronic texts make such a search easy. Given the easy detection of such plagiarism, these cases, though known, are not common nor plenty. The other case is trickier, as now the student has copied parts of the book, perhaps with correct referencing but certainly not indicated that the information is understood, nor applied independently. These do not qualify as plagiarism as such but raise doubts on what grounds a grade can be given (see Bengtsson, 2019 for further references). This concern has only been aggravated by the advances in AI. Second, take-home exams seem to affect study behaviour. Moore and Jensen (2007) show that students who face a take-home exam attended less classes, took part less in non-compulsory learning activities and disengaged from long-term learning behaviour – possible flip sides of less stress. In my statistics class, the take-home exam led the students to choose their preferred statistical method among the many we discuss during the course, and study that. The exam asks the students to carry out a quantitative study but does not require that all the statistical techniques discussed during the class are somehow present in the paper. The exam thus covers only a part of the course content and indirectly indicates that as long the take-home exam passes, the other content of the course is not required. The research community seems most divided over the issue how the take-home exam affects students’ learning patterns, the common problem being that it encourages selective studying (Haynie, 2003). Third, and related, the growing heterogeneity of the student population means that there are students who develop complicated research questions attacking conflicting research results and produce impressive and experimental recordings to test statistical correlations and bring about clarity to existing contradictions. Yet, there are also those who go for the simplest possible test of association between alcohol use and gender or similar. Whilst the first group of students clearly exhibit higher-order cognitive skills like applying and analysing, the latter group hovers between remembering and understanding with minimal creative individual input in the actual methodological work. The take-home exam allows for both kinds of ambition levels without stimulating or encouraging students to aim at higher levels of learning. The fourth problem I encountered was that the take-home exam is laborious to correct, and the feedback may go wasted (Handley et al., 2022). Methods teaching can be at times a pedantic business. I had to fail many take-home exams on small technical mistakes such as the confusion

between “per cent” and “percentage point” or treating the *test value* as a significant or non-significant result. Now, both *are* mistakes, but often it is apparent from the context that the student understands the method right. The *test-value* as such is neither significant nor non-significant but becomes either one when compared to the appropriate critical value. Yet, most statistical programmes automatically flag the test values as significant or not disguising one step in the actual interpretation of the results.

The main reason why I started thinking of alternatives to the take-home exam in quantitative methods was that I was constantly failing between 60 to 70 per cent of the students – and in most cases I was convinced that the student actually did understand, but made a mistake in describing his/her understanding. Yet, the reverse was also true. Among the passed take-home exams there were always those, in which the student did get all the technical terms right, but the formulations were too close to the book to raise concern whether the student understood what he/she had written. Advances in AI will make such problems more likely in the future (e.g. Hern, 2022). Second, despite the promise of promoting higher order cognitive skills, the practice showed the most students used the take-home exam to study selectively – and on the surface. The take-home exam, in the end is a high-stake exam giving rise to strategic choices (Birenbaum et al., 2006). Finally, the unknown faith of the comments I wrote pedantically explaining how the test value acquires its significance, that it first has to be compared to a critical value, which then together yields the p-value, which in fact is the figure you see in the statistical programme made me realise that the idea that a take-home exam could integrate summative and formative assessment started to fail. As a sum, the practice of the take-home exam was not reliable, nor fair, nor did it encourage deeper learning.

Going theoretical: Constructive alignment and deep learning in oral examinations

Once I presented my probes into oral examination in a seminar and was encouraged to develop the ideas, I was “forced” to start thinking in theoretical terms what was happening in the oral exams. I started where I felt the short-comings most acutely: the correction, which I felt was time-consuming and feedback ineffective, and from therein continued to think how to encourage more experimental examinations that would probe something new rather than play the safest card of a simple test of association. I turned to the theory of constructive alignment (Biggs & Tang, 2011) and deep learning Entwistle, 1998; Marton & Säljö, 1984/1997) for inspiration. By constructive alignment I refer to the idea that individual assignments contribute to one another, and that the forms of assessment can be incentives for deep learning. I use the term “deep learning” here to refer to learning activities of higher order cognitive skills through which students internalise the meaning of learnt content in order to be able to test and apply that knowledge (Marton & Säljö, 1997). This can be contrasted to surface learning, which focuses on memorising the content without building relations between the different parts of the content. Yet, by connecting this definition of deep learning to assessment as an incentive my approach to “deep learning” approximates to what Entwistle calls “strategic learning” (1998). Examination is often seen primarily as a summative form of assessment, i.e. that the assessment focuses on what the student knows and can do. However, constructive alignment in teaching emphasises the idea that the examination functions as a motivating force and should be aligned with the learning outcomes of the course. This not only provides for a better overall learning process, but if the examination is also seen as a formative moment, it serves as an additional learning moment that can encourage the students to go further (next time), and it can help universities tackle rising financial constraints and reduced contact education.

Digital oral examinations usually emerged during the Covid-19 pandemic as a “second-choice” option. Sotiriadou et al. (2020) observe that when the oral exam is used in an online environment, it is often seen as a quality control, or a check against academic dishonesty, not as the primary form of assessment. Yet, the argument from constructive alignment is that if designed appropriately, the oral exam should be able to generate deep learning as the processes of learning are reflected in the learning outcomes (Iannone et al., 2020). Surface learning results from attempts to memorise information from the text to get through the assessment. Deep learning, in contrast, is driven by an inner motivation to understand, and it focuses on building links in the material studied. At times, deep-learning is also described as “holistic” in contrast to more “atomistic” surface-learning and connected with explicitly normative assumptions according to which the former is “good” and desirable learning whilst the latter is “bad” and less desirable (Pleijel, 2021). Despite the normative bent in the terminology, the distinction is useful in drawing attention to how deep learning is characterised by building connections between different elements of the material so that together they become more than the sum of the individual (atomistic) elements. Deep learning aims at turning knowledge into an active tool that can be applied in new contexts in a dynamic manner. Entwistle and Entwistle conclude that in deep learning “the answers will depend on what is required by the question, but also on how the individual student understand the topic” (1991, p. 207).

If the mode of learning leads to certain kind of learning outcomes as the Gothenburg school holds, then the examination should be seen as a central formative tool for learning assignments but also as the manifestation of the learning outcomes. Quantitative methods tuition includes a great number of detailed information – such as the difference between “per cent” and “percentage point” – that yields easily to both surface- and deep-learning. And a good way can be gone just by memorising the terminology. The task is thus to design the quantitative methods examination as something that both incites and tests deep-learning. According to Iannone et al., (2020), the more predictable the learning curve appears to be, the less the students have incentives for deep learning, because the learning activity can be tackled as reproduction of information. However, the more unpredictable the learning curve and the requirements for learning are, the more students are inclined to opt for deep understanding of the material.

In the literature, oral exams have some advantages directly connected to constructive alignment and deep learning. Akimov and Malin (2020) point out that the oral exam requires and promotes verbal communication, which is an essential skill in contemporary society. Joughin (2010) has pointed out that the oral exam allows for an easier combination of different modes of assessment making it more adaptable than the take-home exam. Biggs and Tang (2011) as well as Wass et al. (2003) argue that oral exams encourage the students to achieve higher-order cognitive skills as well as make their assessment easier. Finally, I would add two further points. First, previous research has shown that written feedback is not always understood by the student and written feedback has its limits (Cann, 2014); oral exams enable the delivery of feedback in a form of an individual dialogue making sure that the student understands the feedback and makes it versatile enough to discuss, e.g. graphs or charts focusing on their visual and numerical/verbal content (Mulliner & Tucker, 2017). Second, seeing the performance of the student, it is much easier for the teacher to assess what teaching examples have worked giving insightful information about students’ perception of the pedagogical format of the class.

The common problems associated with oral exams concern their reliability and validity (Davis & Karunathilake, 2005; Evans et al., 1996). Reliability is connected to the subjective factor of the examiner. Validity relates to the psychological effects of the exam like stress that can impair

students' performance. To counter some of the negative aspects, Davis and Karunathilake (2005) suggest that the oral exam should take as structured a form as possible, it should be integrated in the curriculum so that students grow familiar with them and more than just one examiner should be involved, and new examiners should be trained for the task. Some of these contradict the advances of oral exams discussed above and could be seen as points to consider in the design of oral examination rather than as arguments against it. I will discuss these points below in more detail.

The practice: Case study of Zoom-facilitated oral examination

I will first discuss the "formal" description of oral examination as it is presented in the study manual available for the students and then go more autoethnographic and discuss my own reflections on the examination and how the oral exam can both incite and render deep learning more explicit to the examiner.

The course where I introduced oral examination is a second-year course in quantitative methods for students in social sciences. The course requires basic knowledge of quantitative methods, but this is the first time the students are introduced to a statistical software (Jamovi), database, and required to carry out independent statistical inquiries using the software and the database. The assessment consists of obligatory seminar attendance, seminar assignments and the final oral examination. In addition, the students have two online quizzes at their disposal. The first covers the content of the preceding introductory course and it should be completed in the beginning of the course; the latter covers the contents of this course, and it can be used as a self-test. Focus is on making sure that students understand what kind of statistical tests can be performed with given variable types thus covering lower-order skills of the course. The purpose of the seminar assignments is to cover "all" content, that is all statistical tests discussed during the course – bi- and multi-variate cross tabulation, T-test, correlation and bi- and multi-variate linear regression – as well as to check that the students can technically implement the required tests in Jamovi. The seminar assignments require varying degrees of work with the database such as recoding and reorganising answer options or computing new variables. The seminar assignments are assessed on a pass/fail basis.

The main high-stake exam is the oral examination in the end of the course. The way I designed the oral examination combines the forms of presentation and interrogation (Joughin, 2010). The presentation part involves an independent quantitative study, where the student is asked to formulate a research question based on existing research and answer it using the available database including research items covering youth alcohol and drug usage, school, free time, and social relations. The instructions for the oral exam are included in the study manual and thus are available from the start of the course. The idea was to follow the basic academic conference presentation model: present and motivate the research question, connect to previous research, explain the data, the results and discuss the findings. The students have 10 to 15 minutes to demonstrate through their presentation that they have achieved the learning outcomes. The presentation is followed by questions concerning the presentation and the whole course content. Five points concerning the presentation are also iterated: i) it must tackle a scientific question that is (loosely/creatively) ii) based on one (pass) or more (pass with distinction) scientific articles, which is answered through iii) logical argument, knowledge in quantitative methods, and the material in the database, and iv) correct statistical test (bi-variate for pass; multivariate for pass with distinction), v) correct interpretation of the results. The whole examination is recorded (sound only) for the purposes of any disputes over the assessment. The grade and motivation are given at the end of the examination, and should the student want a more elaborate

discussion on the assessment a new time has to be booked (as the following student is waiting for their turn). In addition, the research question, the references to used articles, all tables, and a short explication of all possible recodings and/or computed variables must be submitted prior to the examination via email. They were used as a back-up to ensure functioning presentation in the case of possible technical difficulties with Zoom or poorer internet connection.

The interrogative part consists of “follow-up questions” during the remaining five to ten minutes. What these questions exactly would include was left open, requiring the students to prepare in a comprehensive manner. Whilst the presentation part focuses on students’ skills to organise and present information in a logical manner, the questions at the end of the exam assess students’ skills in unrehearsed reasoning and use of knowledge, that is how well they actually master the quantitative methods and concepts (Kehm, 2001). Most questions I asked somehow related to the presentation asking the student to develop it further, or to discuss their choice between alternative but feasible tests and their intricacies, or their choices concerning the recoding of variables. The advantage of relating the questions to their presentation was first to provide the students with a firm ground from which to explore their understanding further, but also for the examiner to skip the stage of planning dozens of questions ahead of the examination. I also asked the students to help me interpret the results of a rather standard cross-tabulation, T-test or linear regression analyses. I shared my screen in Zoom with the student and presented the results so that if the student had used cross-tabulation in his/her presentation I would ask the student to interpret the results of a regression analysis and T-test. These questions provided good insights into how the student understands what is being tested in cross-tabulation, T-test, or linear regression analysis, what can be deciphered from the results, and how familiar the student is with different outputs from Jamovi software. The combination of presentation and interrogation makes the oral examination arguably a better measure of students’ understanding of the subject matter than a written take-home exam, where obscurity can remain and pass.

I have used the university’s study platform to create timeslots for individual examination. Each timeslot is 20 minutes and every now and then I add an extra 5- or 10-minute gap in the schedule to make sure that I am keeping to the allocated times as well as to give me a moment to go through the material that has been submitted via email. I examined on average 20 students a day. Students were asked to book their time and log in 5 to 10 minutes in advance. I also emphasised that there is no time to discuss the grade because the following student is waiting, but that I would be available to discuss the grade further if required. So far no one has required nor made a formal complaint about their grade.

In the beginning of the examinations, I asked the students to repeat the four last digits of their national identification number to control their identity. After this I reminded them to record the examination and start when ready. I had my own Jamovi open and would sometimes repeat the test to check the results if I suspected mistakes. Sometimes the mistake was in recodings or ways in which index variables were computed. Such mistakes would not lead to a fail-grade if the *logic* behind was still correct.

I remained passive during the presentation letting the student continue and made only minimal reactions to any possible quests for support from the students, but I took notes. I noted all possible smaller, technical, mistakes that should be clarified after the presentation. I also made notes about the presentation for myself to make sense of what the student was presenting. I realised I was often drawing arrows between the variables the student picked up during the presentation building my own understanding which variable will be given the position of independent and which will be the dependent variable. This terminology implies

causality – independent effects changes in the dependent variable – but in reality such causality is difficult to prove. The reason for any co-variation between the variables may be due to a third variable, but also that variables affect one another. In a take-home exam any confusion in assigning one variable dependent and the other(s) independent immediately appears as a lack of understanding – because we are assessing a ready product. But because the presentation is more evolutionary in character, I came to realise that the way in which the variables acquire their “statistical” roles also evolved based on knowledge included from previous research and research question. I realised that my arrow-drawing reflected my assessment of the student’s learning *process* more than that of a product. The “product” was assessed *after* the presentation was completed. Thus, the undecidedness what I previously unabashedly deemed lack of understanding emerged in the oral exam as one character of the *process* of learning making the *way* in which variables become dependent and independent a new part of assessment.

After the presentation, I went back to the parts of the presentation I felt further clarification was needed and addressed any smaller, technical, mistakes in the presentation (was the difference really so many per cents or percentage points?). In case there was some confusion, for instance, in the way a cross tabulation was presented (per cent calculated by the wrong variable), I asked the student to re-interpret the table. If they spotted their mistake, it was often easily corrected; however, if they did not understand the mistake, they also started to understand that their performance may not meet the criteria for a pass grade. One unexpected result was that the direct communication during the oral examination increased students’ understanding of their own performance. In case the student had totally misinterpreted the task and the statistical methods in the presentation, I used the remaining time of the examination to discuss the presentation in detail indicating what would be required for the study to work. Implicitly, I was also letting the student know that this examination will not lead to a pass grade and once the grade announced I *very rarely* experienced any anger or dispute. I recall only one time I received a more emotional response and had to tell the student that this discussion, if so requested, will have to continue later. No request, however, was submitted afterwards.

During the interrogative part, I basically asked two types of questions. One type built on the contrast between their chosen method and an alternative method we had discussed. If, for instance, the student had conducted a cross tabulation, I could ask what are the advantages and disadvantages in this particular case of a cross-tabulation over, say T-test? The answers broadly speaking fell into three groups. The first immediately acknowledges the benefits of T-test over their own choice and apologise. If pushed further, they could start to discuss the similarities and differences at the more theoretical level and then perhaps apply to their own case (and realise that the choice of cross-tabulation was the only feasible). Here the formative assessment is achieved through encouraging the student to relay back to the theory and then re-apply the theory to the concrete case (rather than relying on whatever they interpret the authority to imply). The second category of answers would be based on the student’s own interest or “will:” I wanted to do this study, and this is the reason I did not choose any other method. These answers rarely yielded to more fruitful learning outcomes. The last category of answers would go back to discuss the variable types and reason – often correctly – why such variables necessitate cross-tabulation and what would have been required to do a T-test.

The other type of questions I asked concerned ready results of alternative methods. I would present a table with, say results from a multivariate regression analysis and ask the student to help me interpret the results. This type of question showed very quickly whether the student understands *what* is being investigated in a linear regression analysis as well as any detailed

knowledge about the table outputs from Jamovi. Students that immediately turn to the p-value, could be encouraged by asking if something else would also be important. In case they knew what linear regression analysis is all about, they often came back to the right track with this question; those who did not would pick up randomly the figure next to the p-value...

The first type of question is perhaps more conducive to further (deep) learning whilst the latter also helps to assess the summative side of students' learning. Overall, my experience is that being able to link the questions to something the student already is familiar with (their presentation) gives good grounds to engineer the oral examination and its interrogative part as a formative, individually tailored, learning moment.

EFFICIENCY AND DEEP-LEARNING AT THE SAME TIME?

The argument from the theory of constructive alignment and deep learning was that the oral exam should achieve better learning outcomes by inciting deep learning and to do so more effectively. The main efficiency-loss connected with the take-home exams was that many students found it more attractive to play safe than try something more challenging, the correction was laboursome, there were many fails on "technical grounds," and I had a regular feeling that comments were given in vain.

The course in quantitative methods has normally around 50 to 60 students. In take-home exams, the initial fail-rate was around 60 to 70 per cent. After the first round of comments and corrections, a couple of students did not obtain the pass grade. After the introduction of oral exams, the initial fail-rate plunged to 10 to 15 per cent; and after the resit-exams a couple of students did not pass the exam. My interpretation of the big picture is that the oral exam has taken away the fails that were due to technical mistakes in the texts. This saves teacher's and students' time. The second aspect concerns feedback. In oral exams the feedback can be developed through leading questions making the student him/herself realise what the problem is instead of just point it out. The third point – with some variation – is that more students opt for a multivariate analysis aiming at pass with distinction implying deeper learning during the course.

Thinking about deep learning and the examination as *formative* assessment I deem the odds are strongly in favour of oral-examination. First, the possibility of engaging the student to discuss the presentation as well as other statistical methods, asking further questions and using the questions to guide the student to the right logic with quantitative methods makes the oral examination suitable for formative assessment. The versatility of the oral examination allows for an easy combination of more demonstrative and structured parts with more unpredictable and interactive interrogative parts rendering the oral examination broader in its capacity to assess learning both as a process and as a product. Especially this latter aspect is difficult to achieve in a take-home exam because written *text* on paper is understood as completed action, ripe for analysis as it is (not as what it will be) (Ricoeur, 1981). Spoken and written language belong to different kinds of realms of interpretation and therefore of assessment.

The course evaluation gives some insights into what students think about the oral examination as well as their subjective perception of their own learning outcomes. A word of warning is due here. Course evaluations are not obligatory for the students so on average they have a response rate of around 20 to 25 per cent. The figures below should be interpreted against this background. The scale goes from 1 to 5 where 1 is very negative and 5 very positive. On average, during the past three years 80 to 90 percent of the students have positive or very positive (mainly very positive) opinion of the examination. Unfortunately, comparable data from the take-home exam is not available.

I also added course specific questions to the course evaluation that ask about the students' level of knowledge of cross-tabulation, T-test, and linear regression analysis with answer options ranging from "nothing at all," through "understand what it is about," and "carry it out on useable material" to "explain and instruct another student about it." Again about 70 to 90 per cent of replies had chosen "explain and instruct another student" to all three methods. Despite the low response rate, this shows that the skills acquired during the course are understood as *active* skills – in line with the theory of deep learning. This sounds almost too good to be true. My own explanation is that the fact they have explained – orally – their test to me during the examination, answered my questions and defended their approach also ascertained the students that they *can* do this. A written take-home exam, where the feedback comes weeks after, cannot generate similar experience.

LIMITATIONS WITH ORAL EXAMINATION

As with any qualitative inquiry, generalizability of the results is not the priority. I have tried to provide enough contextual information to enable transferability of my reflections to other contexts and hopefully inspire other teachers to take on oral examination into their repertoire of assessment. Autumn 2022 was the first time I had my course on campus after the pandemic. The seminars were still on Zoom as well as the examination. There were no obvious differences to previous years with the oral exam. The reason I had the seminars online was that the assignments require individual, technical skills with Jamovi. Seminars on Zoom are limited in time only when it concerns teacher's time. My experience has been that once I finish the seminar, most students stay on their laptops and continue to exercise with Jamovi. This would not be possible in a classroom. I think the format of the examination and the better utility of the seminars may have contributed to deeper learning.

There is yet another contextual factor that can have influence: previously we used SPSS, which was available on school's computers. Jamovi, however, is an open-source software and can be downloaded on own laptop. Finding an open-source software was necessary for the online tuition under the pandemic, but Jamovi has been commended by students for clarity and the ease of use.

CONCLUDING THOUGHTS: DEEP LEARNING FROM ORAL EXAMINATIONS

Constructive alignment tries to connect the different learning activities to one another so that they together address the learning outcomes of the course (Biggs & Tang, 2011). The main learning outcome of my quantitative methods class was the ability to *apply* statistical techniques, which was the main focus of the oral examination. The introduction of the oral exam effected two broader changes in the course, the first concerning the learning activities the other about the culture of assessment.

The other learning activities involved lectures, individual assignments, quizzes, and group work during the seminars. The oral examination introduced some unpredictability and stress to the course. The other learning activities were designed to use that unpredictability as an encouragement to dive into the world of statistics and variables and try out new things and manage stress. The basic idea I have had was that students needed some experience before the challenges of application of a method could be fruitfully discussed; and to obtain experience, they needed to dare to try out different things. Individual assignments were designed to generate that experience. I repeated multiple times that a genuine attempt – however "wrong" – would still obtain a pass mark from the individual assignment. I believe that the experience of struggling to interpret the mean value of a nominal variable (e.g. the mean of a variable concerning sex coded as 1 for man, 2 for woman, and 3 for other) gives rise to puzzlement that can be very productive for learning

(a nominal variable should be summarised in frequencies, not through any central tendency like the mean value). The seminar questions were discussed in the beginning of the seminar and students were given the opportunity to try again if they wanted, but it was not required. Some assignments that were particularly far off got written feedback pointing out what parts of the literature should be revised again. Lowering the stakes in the seminar assignments encouraged students' own agency in the learning process inciting the mobilisation of higher-order skills without the risk of failure.

In addition, I created two quizzes, one of which covering the contents of the previous quantitative methods course (to ascertain that all knew the basics) and the other testing the contents of the current course. These quizzes served the purposes of lower-order cognitive skills enabling students to test whether they remember and understand the contents of the courses. Locating the "assessment" of lower-order skill to quizzes also served as an indication that the oral examination is qualitatively different from what the quizzes chart.

The oral examination required application of statistical tests but was open for allowing those students who wanted to evaluate their results or creatively combine tests and variables to address more complex problems with the help of the database. In fact, the criteria for pass with distinction requires a multivariate application which builds on the student assessing the combined effect to two variables often necessitating the ability of applying but also evaluating how well the combination of variables address the problem.

The second bigger change concerns the mode of examination itself that is carried out in the medium of *spoken* language instead of *written* language. Ricoeur pointed out that written text is considered as a product up for interpretation; his fellow philosopher of hermeneutics Hans-Georg Gadamer discussed spoken language as a dialogue, a sequence of question and answer that generates the truth that is transformative of the participants to that dialogue (1975). Dialogue is not meant to be subjected to analytical interpretation, but to be experienced as something that changes one's understanding of the world, or quantitative methods in this case. Seeing the oral examination as a Gadamerian dialogue can explain why the students report that they acquired active knowledge in statistical tests, but it also sheds light on my own change from assessing the product to the observation of a process of learning. I have been literally observing hundreds of students applying statistical tests in different ways, with different reasonings and motivations, seen how my questions or comments effect changes in their thinking about quantitative methods and statistical data. This experience has changed the way I understand assessment. At best, oral examination can become a mutual experience of learning.

I started writing this article with a rather pragmatic point in my mind: oral examinations are efficient and economical; once I was faced with the need to think about this theoretically, I also began to see them as rather effective. However, these three e's of New Public Management and audit culture (Power, 1997), in this case become a tool to rethink not only assessment as an instrument in learning, but also as a possible critique of the assessment culture that should univocally yield to transparency, external evaluation and easy reporting. I came to realise assessment as a mutual learning process that can accommodate different kinds of learning processes whilst still abiding to the basic principles of legally certain examination.

ACKNOWLEDGEMENTS

I would like to thank two anonymous reviewers who provided constructive feedback on the earlier draft of this article and the editors of *Högre utbildning* for their deep and insightful comments on a later version of this article.

AUTHOR BIOGRAPHY

Jaakko Turunen has a PhD in political science from Uppsala. He is currently senior lecturer in Public Administration at Södertörn University. His research focuses on information, language, and interaction in processes of organizing and public communication. His teaching encompasses qualitative and quantitative research methods. He is a member of the editorial board of *Finnish Review of East European Studies*.

REFERENCES

- Adams, T. E., Holman, S. J., & Ellis, C. (2015). *Autoethnography. Understanding qualitative research*. Oxford: University Press.
- 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.
- Bengtsson, L. (2019). Take-home exams in higher education: A systematic review. *Education Sciences*, 9, 269. <https://doi.org/10.3390/educsci9040267>
- Biggs, J. & Tang, C. (2011). *Teaching for quality learning at university. What the students does* (4th ed.). McGraw-Hill.
- Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., & Wiesemes, R. (2006). A learning integrated assessment system. *Educational Research Review*, 1, 61–67.
- Cann, A. (2014). Engaging students with audio feedback. *Bioscience Education*, 22(1), 31–41.
- Davis, M. H., & Karunathilake, I. (2005). The place of the oral examination in today's assessment. *Medical Teacher*, 27(4), 294–297.
- Edwards, J. (2021). Ethical autoethnography: Is it possible? *International Journal of Qualitative Methods*, 20, 1–6.
- Ellis, C., Adams, T. E., & Bochner, A. P. (2011). Autoethnography: An overview. *Forum Qualitative Social Research*, 12(1), Art. 10.
- Entwistle, N. J., & Entwistle, A. (1991). Contrasting forms of understanding for degree examinations: The student experience and its implications. *Higher Education*, 22, 205–227.
- Entwistle, N. (1998). Approaches to learning and forms of understanding, In B. Dart, & G. Boulton-Lewis (Eds.), *Teaching and learning in higher education: From theory to practice*. Australian Council for Educational Research.
- Evans, L. R., Ingersoll, R., & Smith, E. J. (1966). The reliability, validity, and taxonomic structure of the oral examination. *Journal of Medical Education*, 41, 651–657.
- Gadamer, H.-G. (1975). *Truth and method*. Continuum.
- Hall, L. (2001). Take-home tests: Educational fast food for the new millennium? *Journal of Management and Organisation*, 7(2), 50–57.
- Handley, K., Price, M., & Millar, J. (2011). Beyond 'doing time': Investigating the concept of student engagement with Feedback. *Oxford Review of Education*, 37(4), 543–560.
- Haynie, W. J. (2003). Effects of take-home tests and study questions on retention Learning in technology education. *Journal of Technology Education*, 14(2), 6–18.
- Hern, A. (2022, 31 December). AI-assisted plagiarism? ChatGPT bots say it has an answer for that. *The Guardian*. <https://www.theguardian.com/technology/2022/dec/31/ai-assisted-plagiarism-chatgpt-bot-says-it-has-an-answer-for-that>
- Herrmann, K. J., McCune, V., & Bager-Elsborg, A. (2017). Approaches to learning as predictors of academic achievement: Results from a large scale, multi-level analysis. *Högre utbildning*, 7(1), 29–42.
- 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, 313–337.
- Johnson, C. M., Green, K. A., Galbraith, B. J., & Anelli, C. M. (2015). Assessing and refining group take home exams as authentic, effective learning experiences. *Journal of College Science Teaching*, 44(5), 61–71.

- Joughin, G. (2010). *A short guide to oral assessment*. Leeds Met Press.
- Kehm, B. (2001). Oral examinations at German universities. *Assessment in Education: Principles, Policy and Practice*, 8(1), 25–31.
- Marton, F., & Säljö, R. (1997). Approaches to learning. In F. Marton, D. J. Hounsell and N. J. Entwistle (Eds.), *The experience of learning* (pp. 39–58). Scottish Academic Press.
- Mohanna, K., & Patel, A. (2016). Overview of open book-open web exam over blackboard under e-learning system. In *Proceedings – 2015 5th International Conference on e-Learning* (pp. 396–402). IEEE. <https://doi.org/10.1109/ECONF.2015.81>.
- Mulliner, E., & Tucker, M. (2017). Feedback on feedback practice: Perceptions of students and academics. *Assessment & Evaluation in Higher Education*, 42(2), 266–288.
- Pleijel, R. (2021). Ytlinläring och djupinläring – en kritisk reflektion kring normativa tolkningar av begrepp i den samtida högskolepedagogiska diskursen. *Högre utbildning*, 11(1), 16–26.
- Power, M. (1997). *The audit society. Rituals of verification*. Oxford University Press.
- Ricoeur, P. (1981). *Hermeneutics of the human sciences*. Cambridge University Press.
- Säljö, R. (2010). Digital tools and challenges to institutional traditions of learning: Technologies, social memory and the performative nature of learning. *Journal of Computer Assisted Learning*, 26, 53–64.
- Sotiriadou, P., Logan, D., Daly, A., & Guest, R. (2020). The role of authentic assessment to preserve academic integrity and promote skill development and employability. *Studies in Higher Education*, 45(11), 2132–2148.
- Wass, V., Wakeford, R., Neighbour, R., & Van Der Vleuten, C. (2003). Achieving acceptable reliability in oral examinations: An analysis of the Royal College of General Practitioners membership examination's oral component. *Medical Education*, 37, 126–131.