What do you assess when you have to do so constantly?

* A teacher inquiry into grading practices within a points-based continuous assessment system

The question of assessment is one of the most difficult for teachers, whether it be designing tests, grading student participation, or conducting an oral exam at the end of a university course with no previous evaluations of student knowledge. Most often, teachers come into a workplace which already has a system of assessment, and are expected to adapt to it – and adopt it wholly. Moreover, often assessment is conducted at limited, separate points throughout the academic year, leaving teachers to rely on their memory and with little room to reflect on what they are actually assessing: a student’s knowledge, effort, attitude, or something else entirely.

The proposed inquiry addresses a specific system of assessment, used at the Department of European Languages at ZZZ, a faculty of a university in Moscow, Russia. It is a points-based system of continuous assessment, wherein instructors assess students’ performance after each class. I believe a teacher inquiry conducted in this context can shed light onto instructors’ teaching and grading philosophies, as well as help refine this particular system.

I will start by explaining the system of assessment used at the department, as it is crucial to this project. A preliminary literature review will follow, and continue into the inquiry questions and procedures. The paper ends with anticipated conclusions and plans for dissemination.
The context of the inquiry

In a school year divided into semesters (September - January and February - June), undergraduate students at the Department of European Languages at ZZZ receive two - three language classes a week (three in their first, second, and fourth years, and two in their third year). In the first three years of undergraduate study, the winter semester is capped off with an oral exam, and the spring semester – with an oral credit pass/fail exam. Instead of assessing students’ work and progress only at the end of the semesters, however, the department also uses a system of continuous assessment, wherein points are assigned to students at the end of each class, and are summed up at the end of the semester.

Students are assigned zero points if they do not come to class, one point if they attend, but are not prepared or do not participate very much, and three points if they attend, are prepared, and actively participate. In addition, tests and presentations (which students are required to do at the end of each textbook unit, so roughly once a month) are judged on a five-point scale and are included in the total tally.

Students can see their points every two months, when they are added up and posted at the department. This is done so that they can get an approximate picture of their work and progress, and can estimate the grade that they are working towards. Grades are produced the following way: points are added up and converted into percentages, which then have corresponding grades. These include: 5 - ‘Excellent’ (85-100%), 4 - ‘Good’ (75-85%), 3 - ‘Satisfactory’ (60-75%), and 2 - ‘Fail’ (<60%). So, a student can achieve 125/150 points in a semester, which is 83%, or a ‘Good’ mark.

There is a particularly interesting and motivating aspect to this assessment system. In the first three years of the undergraduate program, students have the opportunity to get what is called
in Russia an \textit{automatic} grade, meaning that, at the end of a course, they do not have to sit the exam, but can accept the grade that the instructor offers them, based on their work in the semester. Thus, they often finish the course early, even before the exam session starts. The practice of automatic grades is by no means unique to the department, but is widespread in Russian higher education, although most often these are only automatic Excellent or Good marks. Automatic grades are, quite understandably, considered the Holy Grail among students in Russia, and the most successful of them may even sometimes finish a semester without ever sitting an exam in their courses at all.

The department’s policy towards automatic grades is clearly laid out: every student at the end of the semester is offered a grade based on the points that they have accumulated in the course. They can either accept the grade, ranging from Satisfactory to Excellent, or choose to sit the exam. If they choose to sit the exam, their points are discarded and the instructor assesses them only on the basis of their performance on the exam. This is done to level the playing field. For example, a student who missed most of the course and got enough points only to qualify for a Satisfactory mark can sit the exam and receive an Excellent mark. At the same time, it is possible for a student who wanted to improve their automatic Good mark, but did not manage to prepare thoroughly for the exam, to get a Satisfactory mark.

This assessment system has been implemented in order to support both students, who have a clear picture of how regular attendance and work during the semester pays off, and instructors, who are usually able to motivate most of their students into finishing the course with an automatic Excellent or Good mark. In my experience, out of an average group of 10-12 students\textsuperscript{1} usually about two to four would actually sit an exam at the end of a course at the department. Such numbers also allow instructors to allocate more time during the exam to one

\textsuperscript{1} There are usually about 18-20 groups of students in a given year.
student, in order to assess their knowledge and skills more attentively and fairly. As well, the student may decide that they do not have the time or energy to sit the exam this semester, and choose to accept their grade and focus on other, potentially more challenging, exams.

There are, however, distinct issues in this assessment system which have not yet been ironed out. Firstly, the criteria for continuous assessment are not completely clear. Do we judge only participation? Or do we judge knowledge as well? Yet I am not sure we can actually assess knowledge two-three times a week in a language class, when it takes significant time for any language learner to master grammar (e.g. a particular tense) or vocabulary. If we went over the First Conditional on Tuesday, will I be docking points for its incorrect usage on Thursday? How many points do you give to a student who is enthusiastic, but makes mistakes? These are questions that have been raised regularly by myself and my former colleagues.

The second issue directly follows from the first: this system requires instructors to assess students at the end of each class. Moreover, it is a high stakes system, because the whole grade is calculated from it. Does it have enough flexibility to address the complexity of everyday practice? In addition, does it allow for mistakes or setbacks on the part of the instructor? For example, I have yet to reflect on the question of how to assess a student’s participation in a class during which I was not able to successfully facilitate a discussion, or clearly explain an exercise or grammar point.

Thirdly, such nebulous criteria lead to each instructor forming their own set, meaning that the assessment of a student is not systematically cohesive from year to year, which presents difficulties for both students and instructors. A student may study with an instructor who places

---

2 I use the phrase ‘everyday practice’ here, even though most instructors see their students twice a week, because I believe it most closely embodies the intensive academic year at the department. In addition, instructors usually teach several groups of students, so they often teach 4-5 days a week.

3 I have yet to reflect, but I used to give the student 3 points.
emphasis on participation and effort in their assessment, and so receive the highest points, but in the next year they may encounter an instructor who gives importance first and foremost to language accuracy. In my experience, this leads to students sometimes not understanding why their grades (or points) suddenly fall, to instructors not understanding why a student of this language level has been placed in their group, or the student’s previous success, and so forth.

**Preliminary literature review**

The present inquiry is, in its essence, an action research project. As Ferrance (2000) writes, action research is a disciplined inquiry in which the goal is for the teacher to investigate their own practices with the aim of changing these practices for the better. The last part is a crucial element of action research: it is not enough to simply conduct research, it has to inform our practice.

There are several types of action research, as outlined in Ferrance (2000), but the one that corresponds most closely to the present project is collaborative action research, wherein research is conducted by several teachers in several classrooms with a common issue. As has been said above, the assessment system at the department is most certainly a common issue, and the challenges it sometimes presents affect teaching across all of the undergraduate courses. Ferrance (2000, p. 6) writes that one of the side effects of collaborative action research is the development of partnerships. Truthfully, in the present project it may be singled out not just as a potential side effect, but an explicit aim of the inquiry, since the discrepancy between instructors’ assessment criteria presents difficulties for both them and the students.

I am also interested in exploring the potential connection between this inquiry and pedagogical narration. Although pedagogical narration is most closely associated with early childhood education (Atkinson, 2012), several of its principles can be applied to the present
project. Firstly, pedagogical narration aims to make learning visible and to reflect on possibly ordinary moments of everyday practice (Atkinson, 2012). The aim of the present project is to do the same to the practices of continuous assessment among instructors at the department, and their reasoning behind it. Keeping in mind that instructors at the department assess students after each class, the assessment has become an integral part of their everyday practice, and is influencing teaching and learning as well. I anticipate that the “Aha! moments” that come with pedagogical narration (Atkinson, 2012, p. 3) will concern the instructors’ reflections after the class, when they decide on how many points to assign a student, and write out their reasoning behind this.

Pedagogical narration is a long-term study, and inherently collaborative. Teachers are encouraged to share their narrations with their colleagues, students and possibly the latter’s parents (Atkinson, 2012). The collaborative element is crucial to this inquiry, as it aims to stimulate instructors’ collective reflections on the assessment criteria in order to improve the system. There is potential in exploring the value of appreciative inquiry (Moore, 2019) for managing these discussions, since the assessment system is something the department is quite proud of, and working from its strengths may lead to a better discussion.

In addition to basing the project on the literature of action research and pedagogical narration, it would be wise to turn to works on how to teach through discussion (e.g. Brookfield & Preskill, 2005), as well as how to evaluate discussion and participation (e.g. Gillis, 2019; Mainkar, 2008, Norlock, 2016). Mainkar’s (2008) proposed method bears a striking resemblance to the assessment system used at the department in ZZZ, but, whereas at the department it is the teachers who assign points to students, in Mainkar (2008) the students assign points to each other as a form of peer assessment. In addition, the points at the department count for 100% of the
offered grade, whereas Mainkar (2008) writes that students decide the weight that their participation points would carry.

Gillis (2019) addresses the important issue of bias in evaluation, both in that instructors often retrospectively evaluate students’ participation, relying on their own memory, and in the implicit biases such as gender, race, attitude etc. This is important to consider in the present project, and such literature can help single out the biases that instructors will have to reflect on when investigating their assessment criteria.

Finally, it will be necessary to look into the scholarship on assessing language skills and language accuracy (e.g. Green, 2014; Kunnan, 2018) in order to single out the potential criteria that instructors operate from or may want to focus on in their reflections.

The above areas of literature have been selected because of their relevance to the methods of teaching which are encouraged at the department, namely, a discussion-based, communicative approach to language teaching, with little homework and an emphasis on skills development. This approach is mirrored in the primary assessment criteria given to instructors, as can be inferred from the context section of this paper.

**Inquiry question**

The inquiry will focus on instructors’ reflections of how they assess students in the system described above. Specifically, the inquiry questions are:

1. What do instructors assess in a points-based, everyday assessment system?
2. Why do instructors assess this?

By *what*, I mean what is being rewarded and encouraged on a day-to-day basis (e.g. participation, accuracy) and what is not. By *why* – the reasoning behind this. The second
question is given in hopes to stimulate deeper self-reflection on the part of the participating instructors.

**Procedures for the inquiry**

I plan to conduct this inquiry at the Department of European Languages at ZZZ over the course of one academic year. This inquiry will be done in classes of students in their third year of undergraduate studies and their instructors. There are several reasons for this.

Firstly, by their third year, students are completely accustomed to the assessment system at the department, as well as the learning, studying, and teaching culture of the university. In my experience, students in their first year are usually transitioning from school to university and adapting to a new environment, while fourth-year students are focused on writing their Bachelor theses and applying to Masters programs.

Secondly, third-year students receive only two language classes a week, and have one instructor, as opposed to the other years, which have three classes a week and two instructors. It is the reason why it would be preferable to conduct this project with third-year students instead of second-years. The fact that there is only one instructor may lend a consistency to the assessment of students’ performance, as well as a manageable amount of data.

The inquiry itself will be conducted in several stages. In the first stage I will undertake a self-reflexive study to answer the inquiry questions. In the first semester, I will keep a diary in which I will note down the reasoning behind the points I assign to each student. This will be a short text: a couple of sentences, to the point. However, it will aim to answer both inquiry questions. I will write in this diary the evening of each day of class, in order to minimize my reliance on my own memory. This stage of the project will also serve as a pilot study, in order
for me to develop instruments for other instructors to use in their reflections (e.g. a table that they can fill out after each class, or reflection prompts) and to judge the feasibility of the project.

The second stage of the project will start in the second semester. I will ask my colleagues who will be teaching third-year students to follow the procedures I implemented and possibly adjusted in the first semester. Participation in the inquiry will be voluntary. As for myself, I will continue to keep a diary of my assessment practices in order to be able to collaborate and discuss with my colleagues.

Additionally, in the third year there are about three to four times per semester when two groups are combined for one class, e.g. a debate, a presentation competition, or mock United Nations session. In these classes I will ask the participating instructors to evaluate all of the students, in order to have data where there are two evaluations of the same student. I will also ask the instructors not to discuss their evaluations beforehand, as to minimize cross-influence. Ideally, it would be useful to have participating instructors visit each other’s classes, but that is not possible since all the classes are at the same time.

Finally, I hope to be able to recruit participants from among the students themselves. They will maintain a self-assessment diary, much along the lines of the ones the instructors will be maintaining. This will be invaluable information, because it may reveal the students’ and instructors’ perceptions of assessment criteria and performance. However, it will be necessary to consider the issues of power between the instructors and the students. Most likely, only the confidently successful students will want to participate, as they will know that they are working towards a good mark. If my colleagues and I do decide that it is possible to ask the students if they want to participate, this can be done only in the second semester, after rapport and (hopefully) good classroom relations have been established.
The different sources of data that I will try to employ will be an effort to ensure the triangulation of data that is important for action research (Ferrance, 2000).

In order to maintain the collaborative element inherent to both action research and pedagogical narration, but to guard against burn-out, I will organize meetings with participating instructors in the second semester once a month, during which we will share insights from our reflections and discuss these. These meetings will also allow us to potentially introduce changes into the project and the assessment criteria.

A final, collaborative presentation will be shared at a department meeting at the end of the year, in order to put up for discussion our findings and conclusions, and decide together on a way forward.

**Anticipated conclusions and impact**

The proposed inquiry addresses a very specific context in order to improve a specific system of assessment. By confronting teachers with the requirement to justify and reflect on their assessment decisions almost immediately after they make them, and to do so constantly throughout the year, I hope to make explicit the teaching and grading philosophies that have become so automatized through this everyday practice, both in myself and my colleagues. By making this explicit, we will be able to single out the common themes and improve the system, introducing more cohesion and systematization. Hopefully, the students will also agree to participate, since in its essence the system is designed to support them while they study at a highly demanding and challenging faculty of the Russian university.

More generally, the present inquiry will provide invaluable insight into the ways teachers assess students. Often assessment is done only several times a year, so there is a limit to the amount of data one can gather. The environment described in the present paper supports an
intensive, longitudinal inquiry into teachers’ practices, while its collaborative nature will provide the opportunity to see the changes that may occur through discussion, collaboration, and critique. Moreover, I anticipate that constant reflection on assessment will also stimulate reflection on how assessment influences teaching and the classroom.

Although the raw data will only be used inside the Department in order to protect students’ privacy, the general findings concerning assessment in language classrooms and its effects on teaching can be shared in collaborative presentations and articles in language teaching journals, such as the TESOL Journal.

References


