

Prototype tests @ Leysin American School

Internal report

Luis P. Prieto and Paul Magnuson

24 May 2016

Abstract

This document describes the main results from the Prolearning prototype tests performed at the Leysin American School the 25 April–6 May 2016, involving 6 teachers (plus 5 others as control group). The different quantitative and qualitative data show that participant teachers were able to use Prolearning with remarkably few problems, generating in the process a considerable amount of data about LAS students' experience. These data show the potential of Prolearning to change teacher practice and student experience over time, but also highlight certain limitations of the approach. Although the ideal way of rolling this innovation out at the school level remains unclear, the path forward seems to be through department- or teacher-led customization of certain aspects of the platform.

Introduction

This report is part of a joint effort between [Leysin American School \(LAS\)](#) and the [CHILI Lab of EPFL](#). The main goal of the project is to develop **technologies and approaches to generate data about everyday teaching practice** in a non-threatening way, and to foster **evidence-based teacher reflection**, especially in the context of professional development conversations as encountered in that school.

LAS had already done a first paper prototype pilot with six teachers over two weeks, and later on we performed a second study with nine teachers (also during two weeks), using paper prototypes mimicking different form factors and usage approaches (e.g., student-driven data vs. teacher-driven data), in their everyday practice. This second study showed that it is feasible to obtain a respectable amount of data about both student experience and teacher practice, in a relatively short time. We also found that both teacher- and student-driven data are interesting for teachers (and probably should be used simultaneously). The paper prototype study also hinted to certain actions/experiences to be tracked, and what could be the app's main form factor (the “desktop-like” experience was preferred over the “mobile-like” one).

Using this feedback as input, the CHILI Lab implemented a first prototype of the [Prolearning platform](#), in which teachers could open sessions for gathering student feedback about a number of *learner experience* aspects of a lesson, try to predict the students' responses, and observe the results and their evolution over time.

In this latest trial, we aimed at **assessing the usability and usefulness of the prototype for the teachers, and how its use fits in the teachers' everyday practice**. This would help guide further software development of the application (e.g., for prioritization of features), but also to find the best way of rolling out such an application in a school (and concretely, in the context of LAS's professional development). Also, from a research perspective, we tried to see whether this short usage of the Prolearning platform had any effect on teachers or teacher practice (e.g., more frequent reflection on their own practice, or what is called ‘metacognitive awareness’¹). Thus, the main questions we will respond to in this report are:

Q1. Effects on teachers and teacher practice

- Q1.1 Did the use of the tool affect teachers' metacognitive awareness noticeably?

¹Being aware of one's own knowledge, processes, cognitive and affective states (and their regulation). See Balcikanli, C. (2011). Metacognitive Awareness Inventory for Teachers (MAIT). *Electronic Journal of Research in Educational Psychology*, 9(3), 1309–1332. Available at http://www.investigacion-psicopedagogica.com/revista/articulos/25/english/Art_25_563.pdf

- Q1.2 Did teachers report changes in their practice due to using the tool?
- Q1.3 Did teachers get better at predicting student experience?
- Q1.4 Did student experience reports change over time?

Q2. Usage of the Prolearning platform

- Q2.1 How much did teachers use the platform?
- Q2.2 How many sessions did teachers miss and why?
- Q2.3 What was the overall teacher user experience?
- Q2.4 Was the usage of the tool efficient?
- Q2.5 What was the student user experience?
- Q2.6 Would teachers use it on their own?
- Q2.7 Did teachers use the statistics graphs and how?

Q3. Rolling the platform out in a school

- Q3.1 Were the student experience questions interesting/meaningful for teachers?
- Q3.2 What would be other interesting questions?
- Q3.3 What would be the preferred frequency of use when rolling this?
- Q3.4 What would be the preferred data ownership approach?

Q4. Emergent themes and issues

Method: To answer all these questions, we worked with six teachers, to whom we provided access to the Prolearning platform. To assess any changes in teachers’ reflection/metacognition, a MAIT questionnaire² was administered to the teachers prior and after usage of the platform (as well as to another six teachers who did not use the tool, acting as a control group). Furthermore, semi-structured interviews were conducted with all the teachers that used the tool, to gather their impressions about its usage in everyday practice.

The rest of the document reviews each of the previous questions, providing responses in light of the quantitative and qualitative data gathered in the study.

Research results

From the interviews and meetings with the teachers, we can gather that the teachers come from a wide variety of backgrounds, most of them quite experienced already, having taught internationally for 5-23 years (median value: seven years of experience). Teachers also had different concerns as challenging their teaching practice, from the addressing of diverse students’ needs to classroom management issues.

Q1. Effects on teachers and teacher practice

Q1.1 Did the use of the tool affect teachers’ metacognitive awareness noticeably?

The MAIT questionnaires that the teachers filled in before and after using the Prolearning tool showed an appreciable increase in the average “metacognitive awareness” score after the two weeks of usage (see figure below). However, comparing these results with those of a control group of 5 teachers which did the same test, we can see that their initial values were already higher than those of the control group (hinting that maybe the teachers that volunteered were those with higher metacognitive abilities in the first place), and that control group increase in scores is even higher than that of the experimental group (i.e., maybe

²See note 1 above.

just asking about these issues already makes teachers think more about them, or there may be effects due to control teachers responding less carefully to the second questionnaire). Hence, **it is unclear whether teachers' metacognition was affected due to the use of Prolearning**. This is not surprising given the low number of teachers involved in both groups and the short duration of the intervention. We will have to wait for larger scale tests to see if there is an impact on the (self-reported) metacognitive abilities of teachers, attributable to this new practice.

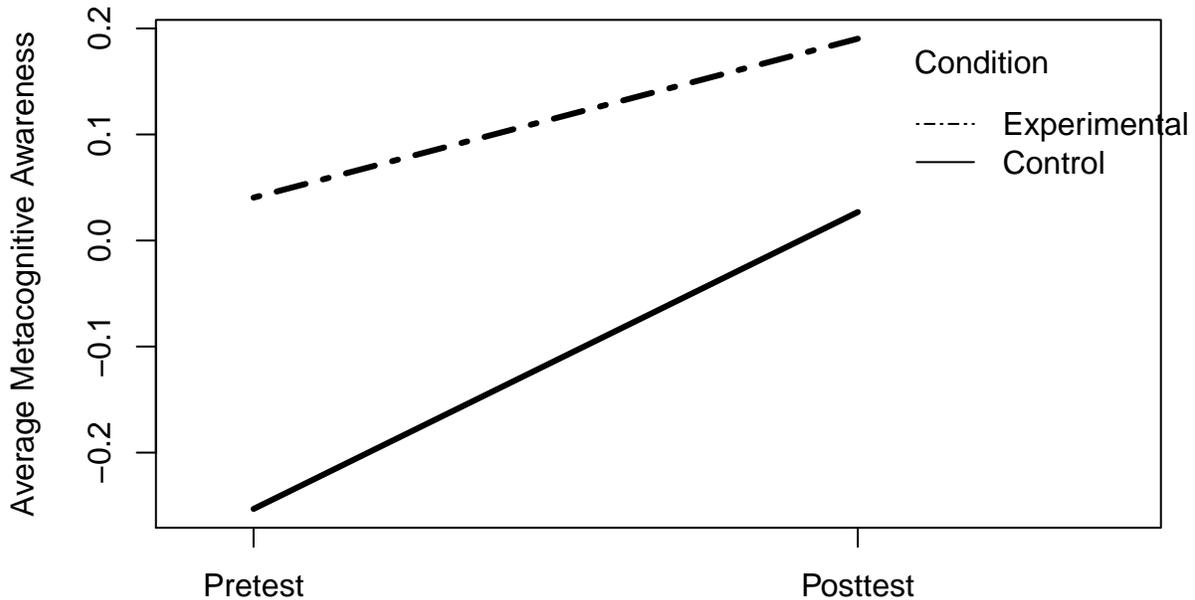


Figure 1: Average MAIT scores of control and experimental groups of teachers, before and after the usage of Prolearning

Q1.2 Did teachers report changes in their practice due to using the tool?

The hypothesis of the Prolearning affecting teachers' metacognition and reflection is somewhat supported by *qualitative* analysis of interview responses, where at least half of the teachers mention (without being asked explicitly) that, over the two weeks of the study, they **thought more about certain student experience aspects** (e.g., *'Did I leave enough time for students to ask their questions?'*), both during the design of the lessons, as well as their execution.

Q1.3 Did teachers get better at predicting student experience?

As we can see in the figure below, although the accuracy of teachers' predictions varied wildly, there was a **significantly decreasing trend in the amount that the teachers were off the student-reported values** (on average, each day the predictions were closer by 0.64 in the 1-100 scale, $p=0.006529$).

Aside from this trend of predictions getting closer to student-reported values, it is also interesting the fact that, in general, **teacher predictions tended to be lower than student responses** (i.e., teachers more often underestimated the amount of students that would report having worked in groups, or having their responses answered): on average, teacher predictions were 13 points below students' actual responses.

In the interviews, some teachers reported actively resisting the urge to adjust their responses to students trends, rather trying to make the prediction they thought was right according to what had happened in the lesson. However, in the teacher predictions we also could observe several teachers answering routinely with a 50 (the default value), which might have been a way for them to avoid making a prediction (because they did not know what to put, or did not dare or bother to make one). Regarding the aforementioned teacher

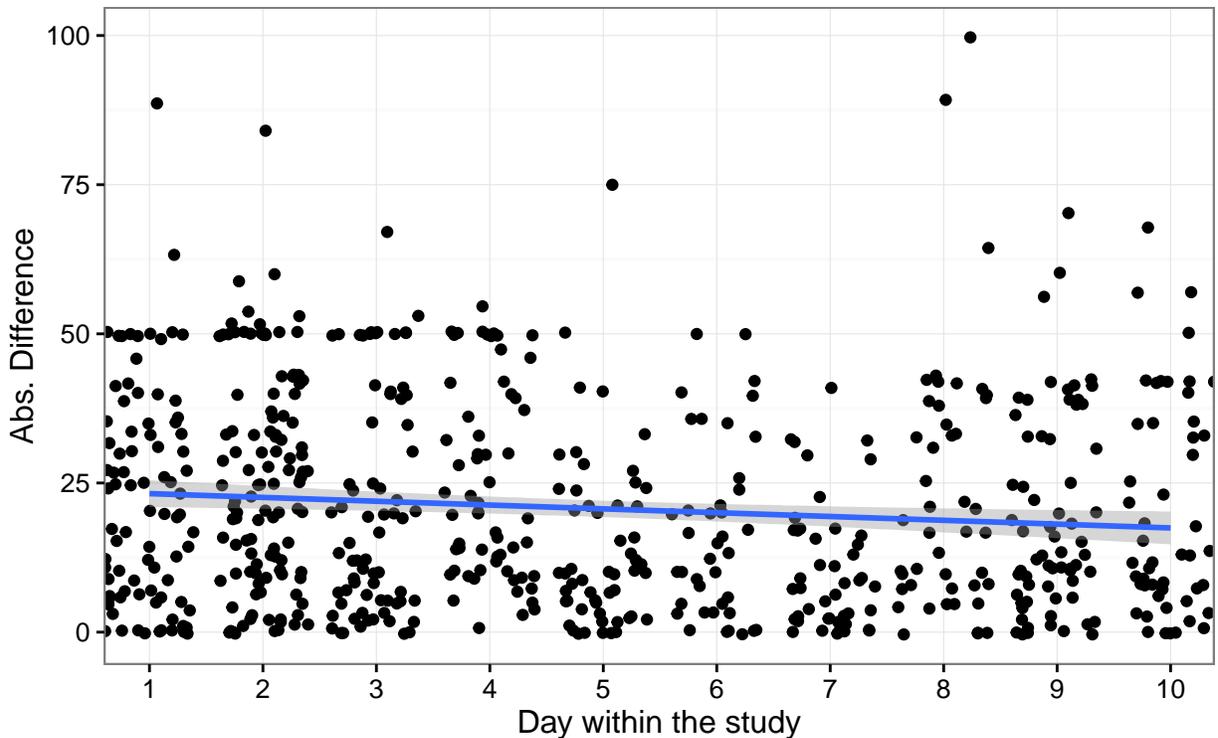


Figure 2: Difference between reported experiences by students and predictions made by teachers, over the 10 days of the study

under-prediction, teachers posited several possible explanations: students misunderstanding the questions, students just clicking away without much thought, students expressing *‘how much they like the teacher’* rather than the actual events of the lesson... , but also teachers being too self-critical (i.e. students actually thinking the lesson had gone better – it being more interesting – than the teacher had thought by looking at their faces and actions).

Q1.4 Did student experience reports change over time?

Another interesting aspect, to check whether the use of the platform had an effect on teacher practice, is to look at the values reported by students in the different kinds of student experiences, and seeing their evolution over time (as a proxy for teachers actually changing their lessons to have more groupwork, more student asking, etc.). Although there is an overall slight upward trend in the student-reported values (an average increase of 0.2297315 points per day), that trend is not significant, and varies a lot from question to question. The figure below shows the trends for the different questions, with significant increases in the Groupwork experience (which was by far the least practiced at the beginning), and slightly less so in the Quiz question, and actually “negative” evolution in the Answers question.

These trends, along with teachers’ remarks in the interviews about how they were more conscious of the aspects mentioned in the questions (and, in some cases, the fact that they tried to address them in each lesson), not only shows the potential of Prolearning in having teachers reflect (before and during the lessons) about certain aspects of teacher practice; they also show how even a relatively short intervention with the tool can have a certain impact on teachers practice and students’ perception of it. The fact that not all questions were affected equally over time can be due to student initial misunderstanding of questions being fixed by teachers over time, the question’s inherent ambiguity, or its relevance for a teachers’ everyday practice, which is bound to vary from teacher to teacher (please read Q3.1 below for more on that topic).

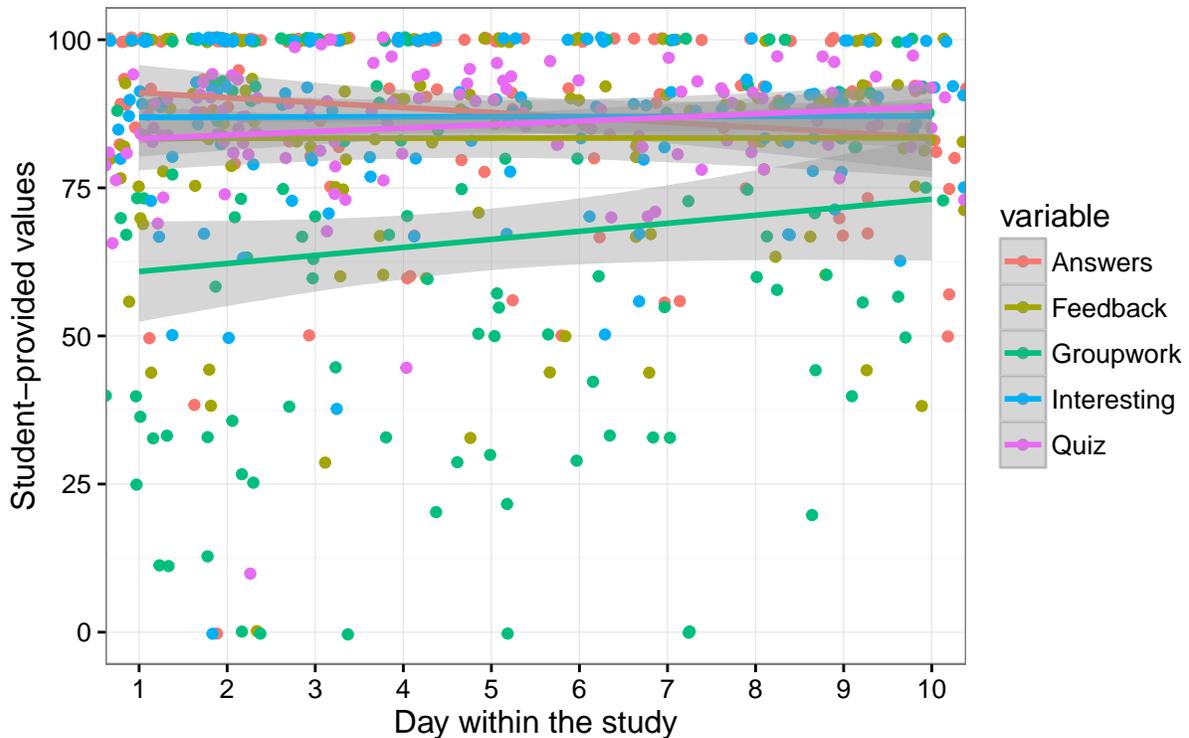


Figure 3: Student experiences reported, for the different kind of experience, over the 10 days of the study

Q2. Usage of the Prolearning platform

Q2.1 How much did teachers use the platform?

In total, the 6 participant teachers involved in the study logged a total of 125 valid data gathering sessions, hence averaging a bit more than **two sessions per day per teacher**. In total, 1147 student responses were gathered (averaging 9.176 students per session). In the graph below we can see how the teachers tended to log more sessions on the first week than on the second (maybe suggesting a waning of the “novelty effect” of using the platform and participating in the study), also showing a quite stable use of the platform after the third day of the study.

Q2.2 How many sessions did teachers miss and why?

Since we do not have accurate information on how many sessions it was *possible* to use the platform (e.g., not counting days the teacher was absent, or the students were on trips or in a lab without their laptops available), we asked teachers for an indication of how many data gathering sessions they had missed. On average, teachers seem to have used the platform for **about 70% of their available sessions**, with extreme values ranging from *no* sessions missed, to about half of the sessions missed.

When asked about the causes for those sessions that were missed, the most common answer was that the teacher **just forgot**, being busy (some mentioned “in the flow”) with the class. In this regard, it looks like teachers often set everything up in their computers, but the absence of a visible cue (which the previous paper versions provided, as something sitting physically on the table) might be the main culprit for teachers not doing the data gathering. Several teachers mentioned that having a student remind them could be a solution (some of them actually did this). In several of the interviews teachers mentioned “not having the habit” of doing such questions at the end (and the need to develop such a habit), as well as how using Prolearning was clashing with some of their **classroom habits** (e.g., take away phones and laptops to avoid distraction) and

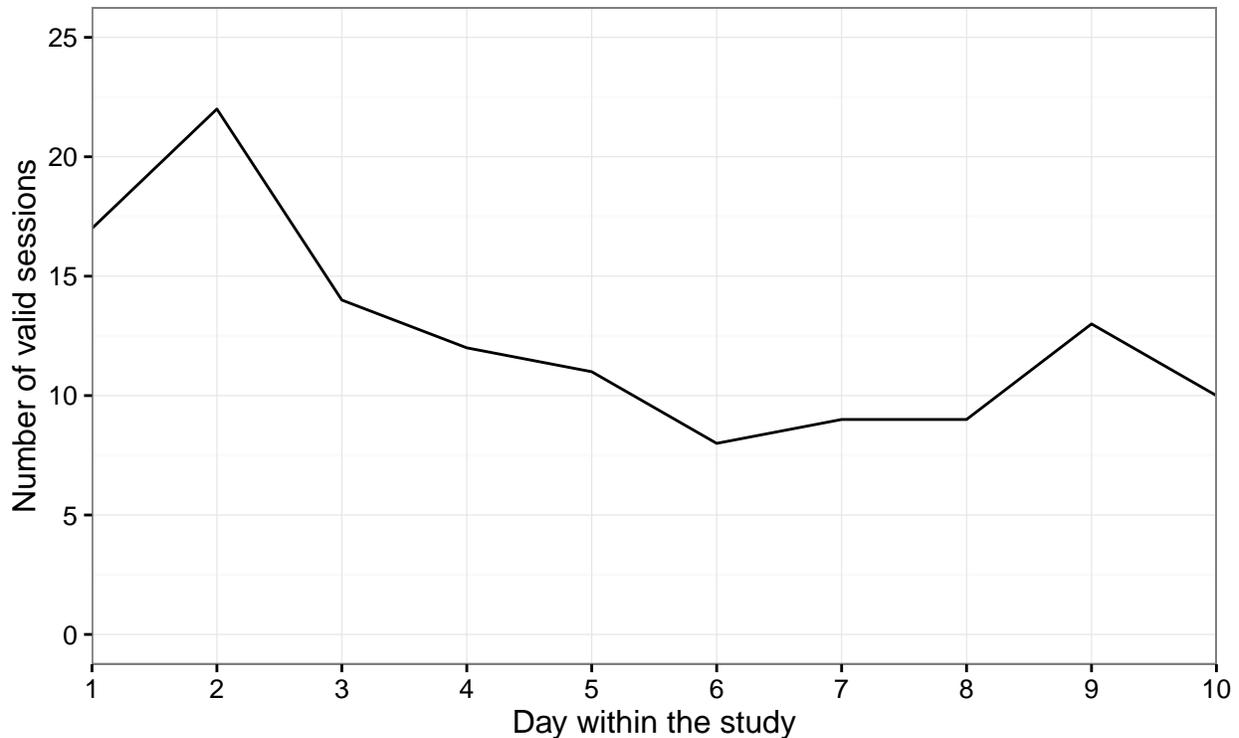


Figure 4: Total number of data gathering sessions over the 10 days of usage

context (e.g., having the lesson in a lab where normally kids do not use the computers). Some other teachers also mentioned the fact that lessons had recently been shortened from 50 to 45 minutes (a perturbation of classroom routine which some teachers still were having trouble with), which made it difficult to fit this ‘*one more thing*’ in an already rushed end of lesson.

Q2.3 What was the overall teacher user experience?

When asked about their overall impression of using the Prolearning tool, responses were wildly **varied**, from ‘*Horrible*’ to ‘*I liked it*’, ‘*good*’ or ‘*painless*’. Although most of the teachers expressed a **positive experience**, all of them also quickly qualified the response: ‘*The first week was very good*’, ‘*The hardest part was remembering to do it*’, or ‘*... every so often*’, or ‘*not all the time*’ (see also Q3.3 below for more on the topic of time and frequency of use).

Q2.4 Was the usage of the tool efficient?

Time efficiency was clearly one of the aspects teachers most positively perceived about the tool (e.g., ‘*very quick*’, or ‘*I can’t see how it can be more efficient*’). Teachers reported that it took them from under a minute to 5 minutes to use the tool in their lessons (with about **2 minutes** being the average usage time reported).

The graph below shows the actual time sessions remained open in the system, and its median value of 151 seconds (about 2 minutes and a half) largely confirms teachers’ perceptions. There is a huge variance around that value, with periods of 15 minutes or more not being uncommon (which hints at teachers sometimes creating sessions in advance, or leaving them open after the class, maybe so that students can fill the questions in on their own time). It is interesting to note that by the second week of the study (in blue), there was less experimentation (i.e., a narrower range of session durations): most sessions lasted for several minutes only (which was the advised way of using the tool).

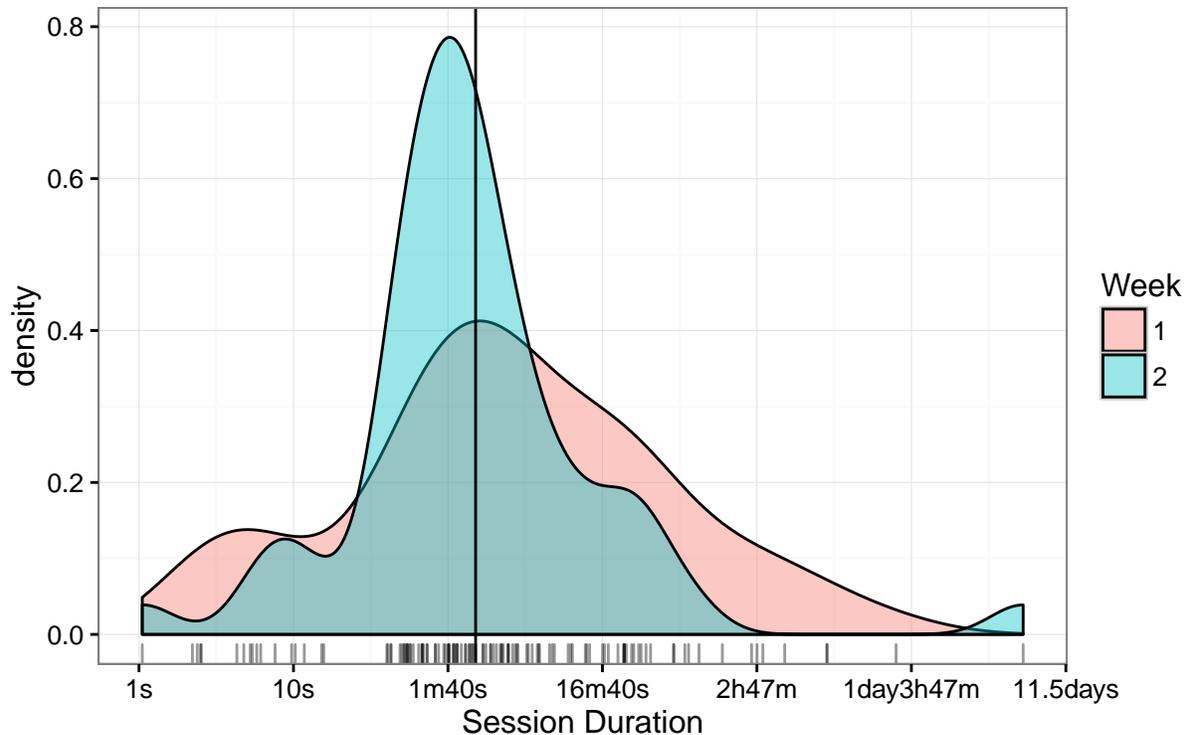


Figure 5: Distribution of times that the sessions were open, in logarithmic scale

Q2.5 What was the student user experience?

An aspect which we did not target explicitly in the design of our study, but which came up quite early on in the interviews with teachers was the **role of the students** and their perception of the tool as very important in order to have reliable indicators of student experience, and for the whole exercise of prediction to have value. When talking about the limitations of the Prolearning approach, the fact (or the possibility) that students were just “clicking away” dedicating little thought to the questions in order to finish and go to the break, very often appeared prominently.

When asked explicitly about what was their perception of student response to the use of Prolearning, teachers mentioned a **variety of attitudes** (and proportions of students sporting them): some teachers mentioned that certain students perceived it as repetitive after a few times, and probably did not perceive the value of doing it (even if teachers asserted having **explained the goal of the exercise**, to different extents), although some teachers also said students *‘loved it’*. Other student issues mentioned by teachers include the students’ age/maturity, the need to develop the habit of doing it and having a fixed time for it (so that it is not perceived as being something *‘between them and the break’*), or the fact that other aspects might be more reflected in the responses than the actual student experience (e.g., how much students like a teacher, or how much they want him to *‘avoid trouble’*). This last issue is another reason against having heavy administrative control of the platform’s data, as it may adulterate the honesty with which both students and teachers use Prolearning (see Q3.4 below about the data ownership approach).

Q2.6 Would teachers use it on their own?

This was probably the most difficult (and most crucial) question of the interview with teachers. In general, teachers avoided an unqualified *‘yes’*, rather providing suggestions on what kind of conditions would be necessary for them to use it. In general, teachers would not use it as a completely personal tool (i.e., only for its own sake, without any external incentive). However, **if presented as part of a larger institutional**

effort for professional development (e.g., as one option to support the current self-appraisal process, or among other options for PD like doing workshops/courses), most teachers said they would prefer Prolearning, mainly due to its quickness and ease of use. Two of the teachers, however, said that even in that case, the **unreliability** of student responses or the fact that it was **taking away time** from already short lessons would probably prevent them from using it.

Q2.7 Did teachers use the statistics graphs and how?

Session stats shown when closing a session (at the end of the lesson). From the interviews with teachers, it appears that most of the teachers **did not spend much time looking at the graph** that appeared after making the predictions. This seems to be confirmed by the fact that very few teachers actually used the “short reflection” field that appeared just below that graph, to record any impressions on seeing the graph (only 32 out of 125 sessions had a comment, 29 of which were from a single teacher, who was the only one to consistently comment on the sessions – to record the main classroom occurrences and rationales behind the predictions made). Upon looking at the graph, teachers reported having felt surprise or relief (as teachers’ under-predictions seemed to hint at lessons that went better than expected, in the students’ opinion), or just making them think about what would be the right mix of student experiences.

Dashboard with overall temporal evolution of each student experience. In the interviews, teachers acknowledged that they did not spend much time (or any time) looking at this dashboard with the temporal evolution of the student experiences and the teacher’s predictions. When asked to look at the dashboard, teachers noticed the general tendency to under-predict the student experiences (see Q1.3 above), and positive feelings about students providing quite high student experience values; others simply wondered why was that so. Interestingly, one of the teachers suggested the idea that sometimes she was making the predictions on the basis of past history with the group of students (e.g., do they tend to find lessons interesting?), rather than the occurrences of the current particular lesson.

Q3. Rolling the platform out in a school

Q3.1 Were the student experience questions interesting/meaningful for teachers?

When asked whether the student experience questions (Answers, Feedback, Groupwork, Interesting, Quiz) were themselves **interesting and relevant**, teachers answered mostly positively, although with some caveats about questions’ ambiguity (e.g., what counts as groupwork?) or difficulty to be responded meaningfully (e.g., in the Feedback question, how can a student “help the teacher know whether she was learning or not?”). Sometimes teachers proposed a slight rewording of the question, either to make it less ambiguous, or to change the time framing of the question (e.g., not necessarily about the present lesson).

Q3.2 What would be other interesting questions?

In this regard, teachers made similar remarks to those of the previous study with paper prototypes: they missed some **content-specific** (or department-agreed) **student experience questions**, or **teacher-customized questions** that had to do with a hypothetical current PD goal of the teacher (e.g., about peer-teaching, or about having enough time to practice speaking in an ESL class). However, it is also worth noting that, despite the fact that teachers were pointed out Prolearning’s ability to add new questions from the “student experience library” to be included in the lessons, **no teacher at all** used it, either because of lack of time, or sometimes consciously to *‘keep it [the questionnaire for students] simple’*.

Q3.3 What would be the preferred frequency of use when rolling out Prolearning in the school?

One of the most common remarks that teachers made about the usage of Prolearning was that they would *not* like to use the tool *every* single day in *every* single lesson (because then it would lose its interest and

become just another chore, or even a source of stress). When asked what would be the right frequency and timing of using the platform, teachers provided a **variety of usage patterns**: once a week, some kind of alternating schema to ask each group of students twice a week, intensively for two weeks at selected points during the year, or asking about a **unit or a week** (instead of every lesson). Some teachers pointed out that regularity (as opposed to selected times in a year) would be important to make the data useful, shorten the feedback loop so that teachers can adapt their teaching immediately, and create the habit to avoid forgetting to do it. A teacher even suggested a schema of asking for **different questions each time** (from a larger set), so that the questionnaires do not become repetitive for students, and at the end data on a variety of aspects is gathered.

Q3.4 What would be the preferred data ownership approach?

In this study we also asked teachers whether the current data ownership scheme (in which the data is personal and only visible to the teacher) was an important aspect. Most teachers responded that **they did not care** about who saw the data, although many were skeptical of whether the school administration would like to see such data or would know what to do with it (rather seeing the value as a **personal tool**, even if it can be used as a source for the reporting/appraisal that is done to the administration). Some teachers alluded at the need for the process of using such data to be transparent, and at removing the ambiguity in the student experience questions, in order to enable comparisons and aggregation between different teachers.

Q4. Emergent themes and issues

Aside from the aforementioned questions that we aimed to answer through the analysis of interviews and the platform usage, there are a number of unexpected insights and issues that have emerged over the course of the analysis. This includes:

- *Teacher cognitive load is always 100%*: An intuition we have been developing over the years in CHILI Lab's eyetracking studies to assess teacher cognitive load during lessons is that "teachers are always on 100% load" during a lesson. The fact that so many teachers *'just forgot'* to use the tool an appreciable number of times during these two weeks (despite the study being a quite salient feature of those two weeks), along with explicit mentions by teachers of seeing themselves *'juggling a hundred things'*, always on the verge of being *'thrown out of whack'*, seem to support this view. This has implications not only for the design of Prolearning but also other classroom technologies that intend to exploit everyday practice or everyday evidence gathering: cues and reminders (either technological, like a small lantern, or social, like asking a kid to remind the teacher) seem necessary for this kind of innovation to be used consistently.
- *Classroom habits and routines*: In relation to the previous one, and since habits and routines are the way teachers cope with the complexity of teaching, it is important to find ways to create the habit in teachers and students to use the platform. Also, it is important to see how the use of the tool may fit into existing classroom habits like putting away the students' computers and devices. In those cases, how can Prolearning still be used successfully?
- *Make the purpose and added value explicit to teachers and students*: In order for the habits and routines to be formed in the classroom, stakeholder buy-in is necessary: the actors involved (mainly, teachers and students) should be made aware of the purpose and value of the platform, and those should be reinforced through other social means (e.g., establishing teacher PD groups that work on the data, showing students the data and how it impacts their lessons, etc.). In this sense, and looking at the teachers that used the tool more successfully, having some kind of "tips and tricks" about how some other teachers managed to use the tool effectively (e.g., asking a student remind the teacher about doing the survey; writing down on the whiteboard the survey along with the other contents to be seen in the lesson; etc.) might also help teachers in adopting Prolearning.
- *Emotional side of things*: Over the interviews there were several mentions to the potential emotional impact of the Prolearning tool, depending on how it is used and within what kind of process: as a source

of stress (both good and bad). Also, certain teacher styles seemed to fit better with this kind of everyday data gathering: teachers that seemed to be more organized and in control, and more systematic in their classroom routines (integrating Prolearning within those routines) were more successful in using the tool and seemed to perceive it more positively. If we are to scale up Prolearning to a wide variety of teachers, some kind of complementary training or materials on how to integrate the innovation with different kinds of teaching styles might be very valuable.

- *Lack of student voices in the study:* Even if the study did not target students as crucial data sources, it is clear from our data that the role of students, their attitudes and buy-in are crucial to the success of Prolearning in its current form. In upcoming studies we should not only explain more carefully the goal and value of the platform to the students (or help teachers do it); we should also interview them to gather their impressions, their perceived value, frequency of usage, etc.
- *Unified school-wide approach:* At different points in the interview it has emerged the fact that both teachers and students are doing assessments, questionnaires, quizzes, etc. which are not perceived as the strict goal of the teaching and learning (e.g., as part of other research projects of different teachers, visiting scholars, other LASER efforts, etc.). While this is the sign of a very active and dynamic school, it can also saturate teachers and students rather quickly, and make them perceive the Prolearning intervention as “just one more thing” to do which is not central to their duties (and hence have a high chance of people falling off the wagon and the habit). We now live in the “age of feedback”, in which we are asked for reviews and feedback about almost every service we enjoy and many other aspects; however, a school rolling this and other innovations out should be careful in selecting such feedback to something that feels coherent and as minimalistic as possible.

Conclusions and next steps

Overall, this second study again has shown promising results, surpassing in some ways the paper user experience of the first study, but also discovering the limits of the new electronic format (as expected). Participant teachers were able to use Prolearning with remarkably few problems, generating in the process a considerable amount of data about LAS students’ experience. Even if the **data gathered are interesting for the school** in and of themselves (e.g., 87% of students considered they had learned something interesting during their lessons), looking at the **temporal evolution** of the data was even more interesting: how teachers seem to get better at predicting student experience (even if they consciously try to avoid being biased by previous responses), or how some of the student experience aspects seem to get “better” over time, or the (non-conclusive) increase in metacognitive awareness after two weeks of tool usage.

Some of the **limitations** of Prolearning found in the study are: a) the lack of perceived added value for students (and some teachers as well), which need to be emphasized to obtain stakeholder buy-in; b) the frequency and length of usage, which can lead to an impression repetitiveness; c) the clash of this attempt at a new classroom routine (asking students every lesson for some kind of feedback) with existing routines and the natural resistance to introduce a new element in an already quite complex teaching practice.

Regarding future steps in this line of work, an obvious one is to **roll Prolearning out** on a real school (e.g., in LAS’s middle school next year). The concrete strategy to be followed in terms of frequency and times of use, or the student experience questions to be included, is not yet determined. Indeed, in this regard we might well find that there is no “silver bullet”, a single configuration that works for the whole school. Rather, school departments or individual teachers may need to **tailor their usage** so as to maximize the perceived added value of the platform.

On the **software development** side, and following the general lines of the previous paragraph, it seems that features in order to **customize usage** (e.g., adding custom experience questions or prompts to the questionnaires) should be given more importance than, say, statistics and graphical elements (which teachers do not seem to use much anyway). The former kinds of features should be completed in the next few weeks, in order to provide LAS with a stable version of the product it has helped develop over the course of this year.