LOUIS MOLDEREZ, GAËTAN TEMPERMAN, BRUNO DE LIÈVRE

Analysis of a Learning Process: An Example of the Application of Ethnographic Observation

ABSTRACT

This article proposes a method for evaluating the learning process when using an environment that does not benefit from learning analytics. In the theoretical framework, the reader will find a state of the art of the different methods of evaluation of a computer environment, as well as a presentation of Oppia, the environment used in this study. In the methodology section, the reader will discover the methodology adopted to observe learning in an ethnographic way as well as the research questions we answered thanks to this observation. The results section presents the study of the data to answer these different questions. Finally, the last part outlines the limitations encountered in the use of this ethnographic observation method.

Keywords: human learning computer environment, observation, ethnographic, process, evaluation

1. Introduction

In order to evaluate the process when using an online learning environment (OLE) used in this study, the researcher and/or designer can use the learning analytics offered by the learning environment used. Peraya (2019) highlights the role that these learning analytics (LAs) can play. They provide information about the entire learning process of the learners. With this data, the designer can improve his or her training, determine learner profiles, discover learner errors, determine the use of resources, etc. For example, Boumazguida et al. (2018) use these data, which we can call learning traces, to better understand the students' learning experience during a training course. Harrak (2016) uses them to create learner profiles before learning.

LAs are, therefore, interesting for education, but what if a computer environment does not have them? This is the case with Oppia, the OLE used in this study. It is an environment created by Google in which the designer can program different learning tasks that follow one another. For each task, the designer can propose theoretical content, create a question to question the student and create one or more feedbacks that will adapt to the learner's response. This environment is interesting to use from a pedagogical point of view, but should it be abandoned in view of the lack of LA? How can the teacher visualise the learning process? Are there other methods of gathering information of the same nature as that offered by LAs?

In this article, we propose a method for collecting learning traces from which we will create learner profiles.

2. Theoretical framework

The nature of the learning data and the method we adopt to collect it depends mainly on the evaluation objective. Indeed, the data collected for process analysis is different from that collected for performance analysis. The methodology used to collect this data is also different. What data can be collected and in what ways?

2.1 The different dimensions of an online learning environment evaluation

According to Temperman (2013), an online learning environment (OLE) can be evaluated along three dimensions:

- The product: it is about evaluating individual progress, but also the collaborative progress of the activity. Through this product, the designer can assess progress, equity, transfer or level of mastery.
- The process: this involves evaluating the actions taken during the learning process. In the context of an OLE, this means, for example, analysing and evaluating the interactions between the different members of the groups. Through this process, the designer can evaluate the achievements, uses, (inter)actions and time spent during the learning process.
- Perceptions: this involves understanding how learners perceived the task. In this respect, the designer can assess the usefulness, usability or even the emotions perceived by the learners.

In the early 2000s, research focused on the different ways of evaluating an online learning environment. It recommends that they be evaluated along three dimensions: usefulness, usability and acceptability (Tricot et al., 2003; Nogry et al. 2004; Jamet, 2006).

Evaluating usefulness consists in checking the adequacy between the content of the environment and the learning objective. It involves answering the question if the learner achieves the learning objective with the tasks proposed in the environment. Tricot et al. (2003) distinguish between two ways of assessing usefulness. Empirically, by comparing the performance of groups with each other. For example, a researcher may compare two possible uses of the same OLE, or one use of the OLE with a group not using it. Another way may be to evaluate by inspection. Tricot et al. (2003) propose seven criteria by which the designer can evaluate the environment as useful or not.

Assessing usability consists in checking whether the tasks proposed to learners are adapted to their cognitive abilities. For the designer, it is a question of checking whether the user is able to carry out the proposed tasks (assessing effectiveness), checking whether he/she uses the proposed resources to carry out the tasks (efficiency) and finally determining whether the system is pleasant to use or not (user satisfaction). As with the evaluation of usefulness, usability can be verified in two ways: analytically and empirically.

Evaluating usability analytically consists in studying the interface and the proposed tasks according to certain ergonomic criteria. Nogry et al. (2004) suggest, for example, referring to the typologies of Bastien and Scapin (1993 in Nogry et al., 2004). Another analytical evaluation method consists in imagining the user's behaviour in the interface. Designers can then resort to users from the target audience, or to experts such as teachers or educationalists. This is one of the methods used by Nogry et al. (2007) to evaluate their "Ambre" software. They used pedagogical experts to test and validate the activities proposed in their environment.

Empirical evaluation consists in observing the behaviour of users during their learning. This observation allows us to identify possible technical problems and/or to understand the behaviours that users adopt when using the system.

To observe learners' behaviour empirically, many methods exist (Tricot et al., 2003; Nogry et al., 2004; Jamet, 2006):

- Verbalization collection: this involves asking the learner to say, specify and justify all their actions by speaking out loud throughout the process;

- The ethnographic method: this method consists of observing the process from the student's position. The objective is to observe all the actions that the learner makes throughout the learning process;
- Document collection: some OLEs offer learning analytics to designers. These learning analytics are the learner's data in relation to the task they have just provided. For example, this can represent the number of repetitions of a task, the number of errors, the time spent on each task, etc;
- Recording eye movements: eyes tracking allows the researcher to see where the learner is looking on the interface and with what intensity. The researcher must then interpret this data to understand the difficulties students have in performing the task.

Assessing acceptability consists in checking learners' perceptions of the usefulness and usability of the task. This can be done empirically through the use of questionnaires or interviews, but also by inspection according to specific evaluation criteria (Tricot et al., 2003).

We can summarise the dimensions and methods for evaluating HIEs by comparing those proposed by Tricot et al. (2003), Nogry et al. (2004), and Temperman (2013). Firstly, this comparison allows the designer to benefit from an overview of the dimensions to be evaluated in the use of an HIE. Secondly, to determine which instruments to use to evaluate each of these dimensions. In the Table 1, we propose instruments for evaluating one or more dimensions of HIEs.

In sum, several dimensions of a learning environment can be assessed, each with a specific method. As far as learning analytics are concerned, they can provide a lot of information on the process and usability level. Indeed, they can, for example, allow us to visualise the time it takes the student to solve the tasks, the number of errors he or she makes, the number of times he repeats the task. If an environment does not have them, there are other methods to collect this type of information such as eye-tracking, verbalisation collection or ethnographic observation methods. In this study, we will evaluate the process dimension of the Oppia HIA using ethnographic observation.

2.2. Description of the learning environment

Oppia is an e-learning environment in which the designer can create sequential educational activities.

The designer has to create an exploration in which "cards" are present. Each card is organised in the same way: the designer can insert content (written, audio, visual, audio-visual, etc.), and then propose a task to the learner to solve using the

Table 1. Comparison of evaluation dimensions (Temperman, 2013) and evaluation methods (Tricot et al., 2003; Nogry et al., 2006).

Dimensions Tricot et al. (2003) and Nogry (2004) Dimensions Temperman (2013)		Product	Process	Perceptions
Usefulness	Empirical	Pre- and post-test comparison		
	By inspection	Evaluation grid according to seven criteria (Tricot et al., 2003): • Clarification and presentation of objectives, • Adequacy of content/objectives, • Precision of the didactic scenario, • Suitability of scenario/objectives/content, • Implementation of cognitive and meta-cognitive processes, • Regulation, • Evaluation.		Passing a questionnaire assessing only the usefulness of the OLE
Usability	Empirical		Observation of learner during the learning process: • Eye tracking system, • Ethnographic observation, • Collection of verbalization, • Learnings analytics.	
	Analytical		Use of experts or target audiences to evaluate the system	Passing a questionnaire assessing only the usability of the OLE
Acceptability	Empirical			Questionnaires or interviews assessing acceptability, i.e., both the usefulness and usability of the OLE
	By inspection			Assessing acceptability in terms of suitability for: • Needs or objectives of the institution, • Learners' expectations, • Characteristics of learners. Assess acceptability in terms of compatibility with: • The organisation of time, • The organisation of the premises. Presence of necessary equipment Legible and coherent planning and monitoring Visibility of results

content inserted or not. Then, based on the answers that the learner proposes to this task, the latter receives specific feedback adapted to his or her answer.

This organisation is justifiable by the model of regulated action proposed by D'Hainaut (1980). The learner is led to discover theory (information) and then to check through different activities whether he/she understands/masters this theory (simulation). Then, depending on their response (analysis of the response), the learner receives feedback before carrying out another activity.

According to Dessus (2021), Oppia differs from other learning environments by the finesse of the feedback it offers. Indeed, the designer can create feedback by telling the learner what is wrong with his or her answer, for example, but also by proposing useful tools and clues that will enable the learner to complete the task requested. In this sense, feedback can play a mediating role in learning. Segers (2019) has also studied the use of Oppia situations in a remediation context. She wanted to compare a deferred remediation with Oppia with a classical remediation. The result was that learners who used Oppia situations made more progress than those who did not.

Oppia is therefore an interesting environment to use in a teaching context. Its main flaw is that it does not provide any learning analytics. The aim of this study is therefore to counter this lack of LA by ethnographically observing the learning achieved by learners in Oppia situations.

3. Methodology

In this study, we want to investigate the implementation of an ethnographic observation method, in order to observe learners' learning in the process. We present the objectives, the context and the research questions we answered through this observation. We also explain the methodology adopted and the instruments used. We conclude by demonstrating the results obtained and stating the limitations of this observation method.

3.1. Context

The aim was to create learner profiles from the ethnographic observations made. The sample considered consisted of twelve learners. The experimentation was carried out over a period of three weeks, during which the learners used Oppia situations designed by a teacher. These were based on a theme of the environmental studies course. Through each of these situations, the students were led to discover knowledge and exercise specific know-how and skills in the discipline.

This ethnographic observation allowed us to answer two research questions about the use of Oppia situations:

Question 1: Which student profiles can be distinguished through the use of Oppia situations?

Question 2: What is the link between the emerging student profiles and their performance?

It should be noted that in order to evaluate performance, we did not use the ethnographic observations made, but rather the comparison between a pre-test and a post-test.

3.2. Instrumentation and methodology

The first question concerns the study of the learning process. To arrive at the creation of several profiles, we constructed an observation grid, the construction stages of which we summarise in the Table 2.

3.2.1. Selection of the observation method

First, we selected the observation method best suited to the Oppia environment, in the absence of learning analytics. The method of collecting verbalisations was unsuitable in our case. Indeed, it is impossible for twelve students to detail their learning process at the same time in a classroom. The eye-tracking method requires a lot of resources in terms of equipment. Moreover, it is also difficult to use in a classroom context. We therefore opted for the ethnographic method. As a reminder, this method consists in observing and analysing each action of the learner during the entire learning process. The students were therefore asked to record their screen when they used the Oppia situations. In this way, it was possible to visualise and analyse all the behaviour afterwards.

3.2.2. Creation of an observation grid

Secondly, we created an observation grid to structure our observation of the actions. We created indicators from which it was possible to create learner profiles in relation to the functions offered by Oppia in its system. Each indicator was then translated into an observable ordinal variable in order to structure our observations as suggested by Molinari et al. (2016).

This table lists all the indicators and their respective ordinal observable variables.

Table 2. Observational indicators and their ordinal observable variables.

Indicators	Observable ordinal variables
Number of the attempt to resolve the interaction	
Type of error made (Astolfi, 1997)	1. Errors in understanding work instructions 2. Errors due to school habits or misinterpretation of expectations 3. Errors related to students' alternative conceptions 4. Errors related to the intellectual operations involved 5. Errors relating to the approaches taken 6. Errors due to cognitive overload 7. Errors originating in another discipline (transfer not acquired) 8. Errors caused by the difficulty of the content itself
Importance given to feedback	1. The student does not read the affirmation feedback. 2. The student reads the affirmation feedback 3. The student does not read the error feedback. 4. The student reads the error feedback
Time to solve the task	
Use of documents made available	1. The student does not use the documents provided 2. The pupil makes partial use of the documents provided 3. The student uses all the documents provided
Resolution of the task	1. The student does not solve the task 2. The student has solved part of the task 3. The student has completely solved the task 4. The student has solved the task randomly

3.2.3. The encoding of actions

The observation grid having been created, we then coded each behaviour from this grid. However, we do not have the benefit of all the learners' actions. Indeed, at times, some students were absent, at other times the recording did not work. However, we collected enough data to create the profiles.

3.2.4. Sum of behaviours

We then summed each behaviour, for each observable ordinal variable and for each student.

3.2.5. Creation of cognitive variables

The variables created are said to be behavioural, as they are derived from the observation of the pupil's behaviour. However, behavioural variables do not make it possible to create profiles. This limitation is also pointed out by Molinari et al. (2016). Indeed, these behavioural traces do not give us any information on the cognitive engagement and the strategies used to perform the tasks. We therefore created cognitive variables, born of relationships between two or more behavioural variables. The Table 3 summarises this creation of cognitive variables:

Table 3. Transformation of behavioural variables into cognitive variables.

Cognitive variables	Behavioural variables
Efficiency/inefficiency	Ratio of the total number of tests to the total number of tasks
Rhythm	Ratio of total time to number of tasks completed
Feedback	Ratio of the number of times the student read the error feedback to the number of times error feedback was offered.
The use of documents	The ratio of the number of times the student used the materials provided to the number of tasks where materials were present.
Chance	The ratio of the number of times the student answered randomly to the number of tasks solved.

3.2.6. Transformation into a Z-note

Since our cognitive variables have different measurement scales, we transformed them into a Z-score.

3.2.7. Creation of profiles based on cognitive variables

Finally, based on these cognitive variables and their Z-scores, we created our learner profiles using the automatic hierarchical classification and k-means methods. Indeed, these classification approaches make it possible to group similar data into a single group and thus create profiles.

The second question examined the link between performance and learner profiles. As the profiles were created in the previous question, we summarise below the approaches used to calculate learner performance which we then compared with the emerging profiles. To calculate it, we used a pr e-test and a post-test.

- The relative gain: this makes it possible to determine the educational added value of the system by comparing the results of a pre-test with those of a post-test. According to Gerard et al. (2006) the relative gain is calculated in two ways: if post-test ≥ pre-test, Gain = 100 x $\frac{Post-Pré}{Max-Pré}$; if post-test ≤ pré – test, loss + 100 x $\frac{Post-Pré}{Pré}$.

- Effect size: According to Hattie (2017, p. 359), this allows us to "identify the impact of our teaching over a period of time". It is calculated according to the following formula: *Effect size* = $\frac{Mean\ (Post-test)-Mean\ (Prétest)}{Standart\ deviation}$. According to Hattie (2017), this effect size must be greater than .40 for the learning to be considered different from normal.

3.3. Presentation and analysis of results

Our analysis is structured according to the two research questions previously formulated.

Question 1: Which student profiles can be distinguished through the use of Oppia situations?

As a reminder, this research question was answered through analysing of the learners' learning traces (N = 12) collected through screen recording. These traces were then observed ethnographically and each learner's behaviour was then coded in an observation grid. We then created cognitive variables from which we created the profiles. The table below summarises the Z-scores of each user for each cognitive variable:

Table 4. Z-score of each user for each cognitive variable.

	Variables				
	Efficiency	**Time**	**Feedback**	**Random**	**Use of the docs**
User 1	0.86567246	-1.0629968	-1.3665462	0.21809784	-0.7216728
User 2	-0.1128712	0.12888478	0.22695194	-1.2533239	-0.1416829
User 3	-0.3241201	-0.5527659	-0.3217733	-0.4163301	-0.2233843
User 4	0.61012572	0.56535253	-0.4851265	-0.256166	0.20535144
User 5	-0.116165	-1.3513881	-0.5666549	2.10358561	-1.2380808
User 6	-0.7583943	0.78741645	0.52908653	0.08507263	1.04909687
User 7	-0.966608	2.38420824	-0.3905536	-0.9075001	0.95091558
User 8	-1.4663209	-0.1844217	1.04472956	-0.4163301	0.95091558
User 9	0.0527398	0.20206545	-0.3722492	0.43465051	0.61955374
User 10	0.67337626	0.28676394	0.78093033	0.94434359	0.41442498
User 11	2.26150223	-0.9307301	-1.2312018	0.8475342	-2.2798624
User 12	-0.7189369	-0.2723888	2.15240719	-1.3836343	0.41442498

We then encoded these data in SPSS, then performed a hierarchical classification method and finally a k-means method. This choice to use two classification methods is due to the particularity of the k-means method. Indeed, to establish profiles from this method, the researcher must encode the desired number of profiles before the analysis. In order not to propose a random number of profiles, we first used the hierarchical classification method. This method generates a dendogram that allows us to establish the number of profiles within the same group. By observing this dendogram, we can divide our group of students into three profiles.

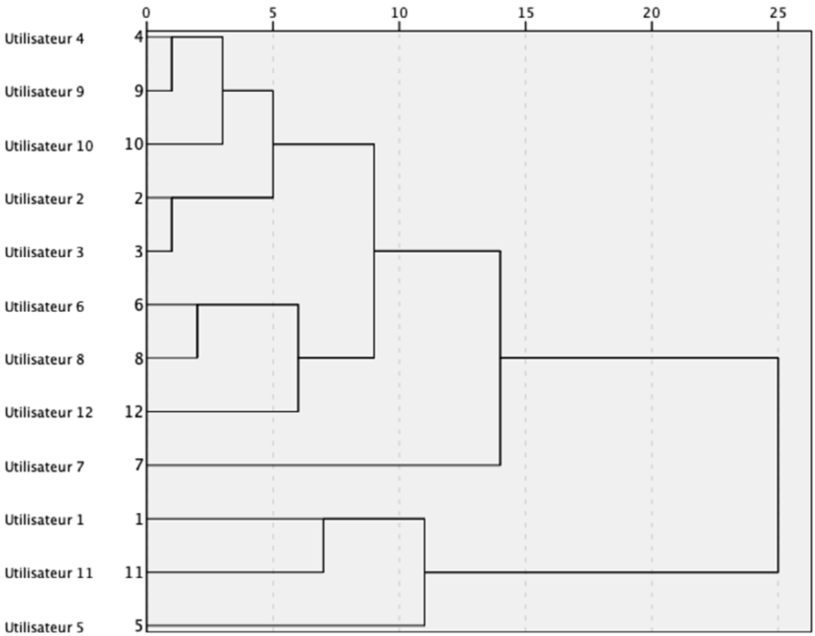

Figure 1. Dendogram obtained from the hierarchical classification.

We then applied the dynamic cloud method to these three profiles. The table below summarises the average Z-scores by variable and for each cluster, as well as the number of staff per cluster.

Table 5. Average Z-score per variable and per cluster.

	Cluster		
	Cluster 1	**Cluster 2**	**Cluster 3**
Efficiency	-0.3897746	-0.2655342	1.00366989
Time	-0.1187855	0.98476067	-1.1150383
Feedback	0.77664914	-0.1797107	-1.054801
Random	-0.5050549	-0.1609857	1.05640588
Use of documents	0.28293967	0.7062294	-1.4132053
N	5	4	3

Presenting the average Z-scores in tabular form is difficult to interpret. Therefore, we created a radar graph to further visualise these differences between clusters. This method was used by Boumazguida et al. (2018).

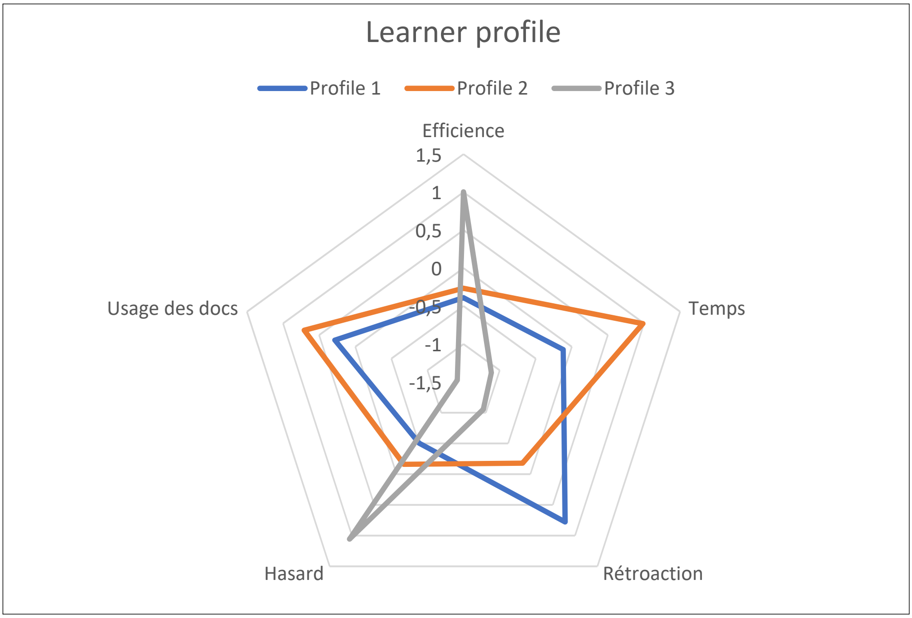

Figure 2. Radar plot showing the profiles and their average Z-scores for each variable.

From this graph we can interpret each of these profiles. A first profile (N = 5) stands out in which the students do not have a high level of efficiency. As a reminder, efficiency is the ratio of the number of trials to the number of tasks performed.

High efficiency means that there is a lot of repetition for the tasks performed, so students regularly fail the questions. In this case, we can talk more about inefficiency. Conversely, low efficiency means that students regularly solve their tasks on the first try. In this profile, students have low efficiency. We also observe a low random response rate, the lowest of the three profiles.

There are two reasons for this low efficiency and random response rate:

- The first concerns the relatively high rate of document use. Indeed, in order to answer the different questions asked of them, students were and should be required to read and understand documents in order to answer the questions correctly. If the students read and understood the documents, they knew how to answer the different questions they were asked.
- The second is the very high rate of reading error feedback. Thus, if students answered the question incorrectly on the first attempt, they read the error feedback which then allowed them to answer the question correctly.

Finally, these students presented a rate close to zero regarding the time taken to solve the tasks. They therefore took the average time to solve the tasks proposed in the situations. From these observations, we can characterise these students as assiduous in relation to the tasks and aids proposed.

A second profile of students (N = 4) stands out. These students also have a low level of efficiency, which is, however, higher than that of profile n°1. This can be explained by two reasons:

- The first concerns the random response rate. This is higher than in profile 1. These students tend to answer the questions more randomly, compared to profile 1 students. They are therefore likely to make more mistakes.
- The second concerns the rate of reading feedback. Indeed, students in this profile do not necessarily read the feedback provided when they make a mistake. The location of the point on the graph (close to zero and therefore to the average) allows us to state that these students read one error feedback out of two. Therefore, if they encounter an error, they do not always read the error feedback offered to them and therefore risk making another mistake.
- Note that even if the feedback rate is high, this does not mean that these students make many mistakes. It means that when they do make one, which does not happen regularly given the low efficiency rate, these students do not necessarily read the feedback offered.

However, despite these two reasons, this level of efficiency remains low, because these students read the proposed documents and understand them. Indeed, we observe a high rate of reading of the documents, more important than that of profile 1.

This behaviour is contradictory: these students read the documents to answer the questions, but do not read all the error feedback that is offered to them.

Finally, concerning the time variable, these students take longer to solve the tasks. This can be explained by the significant amount of time spent reading the various documents to solve the tasks. From these observations, we can characterise these pupils as assiduous with regard to the tasks, but less so with regard to the aids offered.

Finally, the last profile stands out (N = 3), in which these students show little interest in the tasks proposed to them. Indeed, we observe a high rate of efficiency and therefore inefficiency: it is explained by a high rate of random response (the highest of the three profiles), but also by a low rate of feedback reading and document use, also the lowest of the three profiles. In short, these students make a lot of attempts, because they do not use the documents and do not read the feedback provided in case of error. They therefore respond randomly to the different tasks. This is why we observe that they have the lowest time to solve the tasks. From these observations, we can characterise these students as not diligent in relation to the tasks and resources proposed.

Question 2: What is the link between the emerging student profiles and their performance?

As we saw earlier, three student profiles emerged from the analysis of their behaviour during learning. Two of these profiles were characterised as assiduous and one as non-assiduous. Does the student's profile matter in terms of his or her performance at the end of the learning process?

This research question was answered in two stages. First, we determined whether the use of Oppia situations allows students to progress or not by comparing the results of the pre-test with those of the post-test. Secondly, we determined whether the student's profile during the learning process has an influence on his or her progress at the end of the learning process by comparing the average scores on the pre-test and post-test for each cluster.

3.4. Performance analysis

The table below summarises the average pre-test scores, post-test scores, average relative gain/loss and effect size of the group of learners who used the Oppia situations.

Table 6. Pretest Scores, Posttest Scores, Average Relative Gain/Loss and Effect Size – Experimental Group.

	Pre-test scores			Post-test scores			Average relative gain/loss	Effect size
	m (%)	CV	N	m (%)	CV	N		
Total	**52.58**	**22.08**	**12**	**66.42**	**22.67**	**12**	**29.22**	**1.04**

When reading Table 6, we observe that the average score increases from 52.58% (pre-test score) to 66.42% (post-test score). This difference allows us to conclude that there is a learning effect due to the Oppia situations used. Moreover, according to Hattie (2017), an effect size greater than .40 allows us to affirm a progress in learning. In our case, this effect size is well above .40 (1.04 > .40). However, the observation of the average relative gain of 29.22% does not allow us to support this learning effect. Indeed, according to Gérard et al. (2006), there is a learning effect when the average relative gain is greater than 30 or 40%.

In conclusion, we can affirm that the use of Oppia situations allows students to progress, the average post-test score being higher than that of the pre-test, and the effect size being well above .40. The average relative gain obtained is at the limit of the progression threshold.

3.5. Process analysis – performance

As a reminder, the aim of this question is to determine whether the student's profile and therefore the behaviour he or she adopts during learning has an influence on his or her performance.

During the learning process, we could observe that the students adopted different behaviours. Three profiles were distinguished and differentiated in relation to the diligence and seriousness they gave to the task. In a table, we reported the results of the pre-test and post-test, the rate of progression or regression and the cluster with which each user is associated.

Table 7. Descriptive analysis – pre-test and post-test scores, progression or regression rate and cluster for each user – Experimental group

	Scores in Pre-test	Scores in Post-test	Relative gains/ losses %.	Cluster
User 1	8.00	13.75	31.94	3
User 2	12.00	24.25	87.50	1
User 3	13.00	19.00	46.15	1
User 4	14.00	11.00	-21.43	2
User 5	13.00	21.75	67.31	3
User 6	13.50	12.50	-7.41	2
User 7	16.00	20.75	47.50	2
User 8	15.00	17.75	25.00	1
User 9	21.00	19.00	-9.52	2
User 10	13.50	15.25	14.00	1
User 11	13.00	16.50	26.92	3
User 12	12.00	15.75	26.79	1

In a second table, we reported the means and coefficients of variation of the pre-test, post-test scores and the average relative gains/losses for each cluster.

Table 8. Descriptive analysis – mean and coefficient of variation of pre-test, post-test and rate of progression/regression scores for each profile – Experimental group

		Pre-test		Post-test		Average relative gains/ losses	
	N	Average	CV %.	Average	CV	Average	CV
Profile 1	5	50.38	9.5	**70.77**	19.58	**39.89**	**72.76**
Profile 2	4	62.02	21.24	**60.82**	30.26	**2.29**	**1343.61**
Profile 3	3	43.59	25.47	**66.67**	23.45	**42.06**	**52.33**

We compared these profiles to determine whether the student's behaviour during learning has an influence on post-test results and progression/regression rates.

The first profile of students sometimes made use of the materials provided and regularly read the error feedback. These students placed a great deal of importance on the resources provided to complete the various tasks. Table 8 shows that these students made progress with an average relative gain of 39.89%.

However, with a coefficient of variation of 72.76%, we can state that these pupils do not all progress in the same way, some progressing more than others. Indeed, when reading Table 7 we can see that four pupils show a relative gain of less than 50% and only one pupil shows a relative gain of more than 80%, which illustrates this heterogeneity within this group. However, we draw your attention to the fact that the relative gain reflects the progress of the learners. Looking at the post-test score averages, we observe that this profile has the highest post-test average of the other three profiles (70.77%) and that the coefficient of variation is less than 30% (19.58%). In conclusion, these students show a significant progression, which is moreover higher than the minimum threshold of 30% (Gerard et al., 2006). However, their progress was heterogeneous, with one student making more progress than the others. However, we note that these students show an average of 70.77% in the post-test, which is the highest of the three profiles.

The second profile included students who responded more randomly and did not read all the error feedback. However, they regularly read the proposed documents to solve the tasks. These students took the proposed tasks seriously, but paid less attention to the aids provided. We observe for this profile that the average relative gain is 2.29% and is therefore very low. Moreover, we note that the coefficient of variation is 1343%, which reflects a significant difference in relative gains between these students. This difference can be seen when we look at Table 8 Indeed, three students show a relative loss between -1 and -25% and only one student shows a relative gain of more than 40%. Nevertheless, these students obtained a post-test average of 60.82%, which is therefore above average. However, the coefficient of variation is greater than 30%, which reflects a high degree of heterogeneity, although lower than the relative gain, but still significant. In conclusion, these students had an average post-test score of over 50%. This group is very heterogeneous, both in terms of post-test scores (30.26%) and in terms of average relative gain (1343.61%). The latter is very low (2.29%), as three out of the four students in this group show a relative loss.

Finally, the last profile of students who regularly answered randomly and made many mistakes. Given the time they gave to each question, these students did not take learning very seriously. However, when we look at Table 8 we see that these students made progress with an average relative gain of 42%, which is thus higher than the minimum 30% to guarantee a learning effect (Gerard et al., 2006). However, these pupils do not progress in the same way, given the relatively high percentage of the coefficient of variation (52%). We observe in Table 22 that two pupils progress with a relative gain between 25 and 35%, while the last pupil shows a relative gain of over 65%. Furthermore, when observing the average scores on the pre-test and post-test, we see that these students go from an average of 43.59% on the pre-test to an average of 66.67% on the post-test. These students thus show the best progression of the three profiles. However, these results are rather contradictory: in fact, these students did not adopt a "school" behaviour by responding randomly and by not using the proposed resources. However, these pupils managed to progress, and when comparing the averages of relative gains between the profiles, they even progressed more than the pupils in the other profiles.

4. Conclusion and discussion

In conclusion, based on this ethnographic observation, we were able to answer our two research questions.

Firstly, we distinguished three student profiles that differed in their behaviour during learning. Indeed, two of these profiles showed a certain assiduity to learning while one of the profiles was much less assiduous. The assiduous students showed a relatively low rate of efficiency and took a certain amount of time to complete the learning. They did not solve all their tasks by responding randomly. However, they differed from each other in the way they behaved when they had to use the documents provided and/or read the error feedback provided in the environment. Indeed, one of the profiles made more use of the documents, to the detriment of reading the error feedback, while the other profile acted in the opposite way. The last profile showed a less assiduous behaviour. Indeed, these students showed a higher efficiency rate and a higher rate of random response than the other two profiles. Similarly, they showed a low rate of document use, reading error feedback and time to solve tasks.

Secondly, we also compared the average performance of each of these profiles, in order to determine whether there is a link between the behaviour adopted during learning and progression. We cannot say that there is a link between the student profile (and therefore the behaviour during learning) and the rate of progression between the pre-test and the post-test. Indeed, counter-intuitively, it was the students who were the least assiduous and the least receptive to the resources offered during the learning process who performed best. Therefore, we did not use learning analytics. It is therefore quite possible to circumvent this lack of LA in the evaluation of an HIE process. However, the use of this method has some limitations which we mention in the next section.

5. Limits

This method of ethnographic observation has therefore enabled us to collect a set of data with which we have been able to create learner profiles. However, its use has some limitations.

The first limitation is the subjective nature of the observation. As a reminder, Oppia does not have learning analytics, so it is the designer who, on the basis of screen recordings, encoded each behaviour in the table of observable variables. Even if the approach is intended to be as objective as possible, it is logical to denounce its subjective nature. This limitation is also highlighted by Ciccone (2012, p. 70), "the statement of an observation is always already partly an interpretation". Indeed, when encoding the behaviours, the researcher interpreted each of them from the defined observable variables and according to his own claims. If someone else had observed them, he or she would certainly have encoded them in a different way and might have arrived at a completely different result.

This observation is subjective in the understanding of learners' behaviours but is nevertheless interesting in the edumetric evaluation of proposed tasks. As Depover et al. (2012, p. 5) point out, taking learning traces into account provides the designer with "interesting information about the quality of the learning process implemented". For example, from the encoded behaviours for each question, the designer can judge their difficulty rate, and thus modify questions that are too easy or too difficult.

The second limitation is that the observation grid allows us to measure the learner's behavioural engagement and not cognitive. This limitation is also highlighted by Molinari et al. (2016, p. 7) – "While traces can be considered as objective measures of behavioural engagement, they tell us little about cognitive engagement, i.e. the strategies used or the actual degree of intellectual or emotional investment".

Although we have taken steps to create cognitive variables from behavioural variables, none of these variables truly reflect the intellectual or affective engagement of learners.

Moreover, the actions may well be in contradiction with the student's cognitive process, as we have noticed for the less diligent profile. The learner may very well pretend to answer incorrectly, as he or she does not care about the process, but cares more about the final result.

A final limitation is that the evaluation of the learning process through ethnographic observation is resource and time intensive. This would not have been possible with, for example, 30 participants. Nielsen (1993 in Nogry et al., 2006), however, indicates that an observation of at least five participants allows for the identification of problems in the system. An assessment of all the behaviours of each learner is therefore not necessary to edumetrically assess all the tasks of the system.

Ethnographic observation differs from learning analytics in that it visualises the learning traces. Indeed, observation allows us to visualise the whole set of traces, whereas LA only gives us access to the sums of the traces. In the end, the number and nature of the data collected are similar to what LA would offer.

REFERENCES

Astolfi, J.P. (1997). *L'erreur, un outil pour enseigner*. Paris: ESF.

Boumazguida, K., Temperman, G., & De Lièvre, B. (2018). Questioning online activity traces to better understand students' learning experience. *International Journal of Higher Education Pedagogy*, *34*(3), 1–22. **DOI: https://doi.org/10.4000/ripes.1614**

Ciccone, A. (2012). The practice of observation. *Contrast*, *1*(1), 55–77. **DOI: https://doi.org/10.3917/cont.036.0055**

D'Hainaut, L. (1980). *Des fins aux objectifs de l'éducation*. Bruxelles: Labor.

Depover, C., Komis, V., & Karsenti, T. (2012). Quality control: an indispensable tool for establishing the legitimacy of distance learning?. *Formation et profession*, *20*(2), 2–12. **DOI: http://dx.doi.org/10.18162/fp.2012.179**

Dessus, P. (2021, 10 February). *Tutorial – Presentation of the Oppia interactive exploration generator – Course Evaluer les apprentissages, MEEF-PIF*. Université de Grenoble. Retrieved March 23, 2022, from http://espe-rtd-reflexpro.u-ga.fr/docs/scied-cours-qcm/fr/latest/tuto_oppia.html

Gerard, A.F., Braibant, J.M., & Bouvy, T. (2006). *Évaluer l'efficacité pédagogique d'une formation ou d'un cours à l'aide d'un outil d'autoévaluation, 19th International ADMEE-Europe Conference*. Retrieved March 23, 2022, from https://alfresco.uclouvain.be/alfresco/service/guest/streamDownload/workspace/SpacesStore/643ddd81-7da3-11dd-bdb8-b377fd3def91/GerardBraibantBouvy-Admee-2006.pdf?guest=true

Harrak, F. (2016). Analysis of HIE traces for the identification of groups of learners with similar profiles. *6th Young Researcher's Meeting in HIA*, Montpellier, France.

Hattie, J. (2017) *Visible learning for teachers: knowing its impact to maximize student achievement*. Quebec: Presse de l'Université du Québec. (Original work published 2009)

Jamet, E. (2006). A presentation of the main methods of evaluation of HIA in cognitive psychology. *Sticef, 13*, 129–146.

Molinari, G., Poellhuber, B., Heutte, J., Lavoué, E., Widmer, D.S., & Caron, P.A. (2016). Engagement and persistence in e-learning devices: cross-views. *Distances et médiations des savoirs, 13*. **DOI: https://doi.org/10.4000/dms.1332**

Nogry, S., Jean-Daubias, S., & Guin-Duclosson, N. (2007). How to combine evaluation objectives and methods for the iterative design of HIEs? Lessons from the AMBRE-add design. *Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation*, *13*(1), 147186. **DOI: https://doi.org/10.3406/stice.2006.930**

Nogry, S., Jean-Daubias, S., & Ollagnier-Beldame, M. (2004). Evaluation of HIAE: a necessary diversity of methods. *Information and Knowledge Technologies in Higher Education and Industry*, 265271.

Peraya, D. (2019). Learning Analytics: methodological constraints and ethical 'governance' of data. *Distances et médiations des savoirs*, *26*. **DOI: https://doi.org/10.4000/dms.3739**

Segers, N. (2019). *La remédiation différée assistée par ordinateur, analysée sous l'angle des élèves, en vue de remédier aux lacunes en algèbre au CE1D*. Retrieved March 23, 2022 from https://www.researchgate.net/publication/337110801_La_remediation_differee_assistee_par_ordinateur_analysee_sous_l'angle_of_the_students_in_view_of_remedier_to_the_lacunes_in_algebra_at_CE1D

Temperman, G. (2013). *Visualizing the collaborative process and assigning regulatory roles in a distance learning environment*. [Doctoral thesis, University of Mons]. https://tel.archives-ouvertes.fr/tel-01005304

Tricot, A., Plégat-Soutjis, F., Campas, J.F., Amiel, A., Lutz, G., & Morcillo, A. (2003, April 17). *Utility, usability, acceptability: interpreting the relationships between three dimensions of HIA evaluation*. EduTice – Education and Information and Communication Technologies. Retrieved March 23, 2022, from https://edutice.archives-ouvertes.fr/edutice-00000154