Processing and Analyzing Assessment Test Logs provided by Digital Pen and Paper

Zaara Barhoumi & Françoise Tort

UMR STEF ENS Cachan, INRP - France

Abstract

In an educational context, the use of new technologies can influence and change teaching practices.

Digital pen and paper, as man-machine interface, appears familiar and do not require any specific training. Analyzing logs recorded by this technology, especially time, seems interesting in order to provide new indicators for evaluation or observation.

As far as we know, time stamped logs recorded by digital pen is barely studied in an educational context by researchers. We explore the use of digital pen and paper solution for paper assessment tests. Especially, we investigate the interest of time stamped answers, considered as traces of the answering process to an assessment test.

As the raw data collected with the digital pen and paper have to be transformed to compute indicators, we have designed and developed a platform named SEGELL. It provides graphical representations, which helps users, as teachers or researchers in education, to analyze students' production.

The paper presents an experiment where we used digital pen and paper for the administration of assessment tests. We detail the analysis of collected data we conducted with the support of our platform. The paper enlightens the type of information we get on students' behavior during the test.

Keywords

Digital pen and paper, time stamped logs, traces analysis, assessment test, education.

Introduction

A digital pen looks and works like an ordinary ballpoint, but it captures handwriting from digital paper forms. In the "Anoto[1]" technology, the digital pen works together with normal paper overprinted with a barely visible pattern of dots. The pen has a tiny camera. When touching the paper, it reads and records strokes in relation to the digital paper's dot pattern. Collected information contains the precise location of each pen stroke from an X,Y coordinates and the date and time of each stroke.

As man-machine interface, the digital pen appears familiar and do not require any specific training. The analysis of the logs recorded by the pen, especially time, seems interesting in order to provide new indicators for evaluation or observation. As far as we know, time stamped logs recorded by digital pen is barely studied in an educational context by researchers.

As we are particularly interested in assessment tests, we decided to explore the use of digital pen and paper solution for assessment tests. Especially, we investigate the interest of time

1 www.anoto.com

stamped answers, considered as traces of the answering process to an assessment test.

As the raw data collected with the digital pen and paper have to be transformed to compute indicators, we have designed and developed a platform named SEGELL. It helps users, like teachers or researchers in education, analyzing students' production. It has 3 main functionalities: reprocessing of data collected by the digital pen, management of data and indicators, interactive graphical representations of indicators.

Section 1 presents an overview of research works concerned with the use of digital pen and paper in educational contexts. Section 2 exposes our research goal and our approach concerning the exploitation of assessment paper test logs. Section 3 describes the main indicators and graphical representations developed for our method and computed by our tool SEGELL. Section 4 presents an experiment in which students used digital pen and paper during assessment tests, in a spreadsheet course at the University. We expose our analysis and the results we get. Finally, we conclude and give the perspectives for this work.

Digital Pen in Education Context: What Research Says?

For few years, digital pen and paper Anoto technology was used in a variety of domains (See Anoto web site). As far as we know, the use of this technology in educational researches is still rare but should become an emerging area. Heidy, Brian and Scott (2006) have used the digital pen as part of a collaborative work for the exchange of ideas on paper expressed spontaneously during a course of computer design. They show that this hybrid technology is as flexible as notebooks or paper to communicate ideas. According to them, it provides the necessary fluidity for a collaborative work in project design.

In an early elementary class, Sugihara et al. (2010) use Anoto technology in order to facilitate the feedback on students' production to teacher. Students' writings are exposed on a screen. The teacher can correct and comment the work for every one in the class. This study concludes that pupils are motivated and enjoy using this system in general.

In mathematics education, Oviatt, Arthur and Cohen (2006) compare the use of this technology in problem solving by college students with other technologies (computer with keyboard, graphics tablet with pen). Authors use times recorded with the digital pen to calculate the total time for solving a problem. Their results show that students using the digital pen are taking less time to solve problems than those who use keyboard or graphics tablet.

These research works are interesting studies of new practices in the classroom. We found no research work that exploit time stamped logs to study precisely students' behavior using a digital pen and paper. However, in human learning environment topic, processing interaction traces are useful to analyze users' behaviors in order to ameliorate the learning process (Dimitracopoulou, 2008).

Use Digital Pen and Paper in Assessment Tests: What for?

A research area, developed at the STEF laboratory, concerns the assessment of students' ICT competencies and skills. Notably, we conducted the DidaTab research project (2005-2008) to get a deeper understanding of spreadsheet knowledge and skills of secondary school students (Bruillard et al., 2008). The method we have adopted was based on the analysis of students' response and productions in cases where they were asked to perform specific tasks involving spreadsheet use. We collected not only the final spreadsheets produced by respondents, but also the video records of all onscreen events during a test. These videos provide useful data on the processes students use to perform tasks, which enriched the evaluation of their production (Tort et al., 2009).

Thus, beyond our findings concerning the French students' spreadsheet knowledge and skills, one of our main findings concerns a methodological issue. That is: assessment of ICT skills may be improved by taking into account the process by which students answer to a test.

In the same research project, we built and administered paper tests. Paper tests are more suitable for large-scale administration, than computer-based tests. Moreover, paper tests are a mean to focus on students' knowledge rather than on how-to-do skills. But the question is how to explore the processes on paper tests?

Digital pen and paper offer a technical solution to capture paper test logs. We want to investigate the interest of time stamped responses, as traces of the answering process to assessment tests. A general question is: what do we learn about the answering process to assessment test? In particular, do we get additional information on students' knowledge and skills under assessment?

From Time Stamped Strokes to Indicators and Graphical Representations

With the Anoto solution, paper tests are fill-in forms. The pen records each stroke with the coordinates of the field where it has been done and with time indication. This raw data have to be transformed to get interesting indicators. While experimenting the use of digital pen and paper on assessment test, we developed an approach inspired by a Trace-Based System concept, described in (Settouti et al., 2009). It is used in processing trace from a user's interaction and navigation through a specific system. Processing data takes place in four major steps: (1) selection and data collection (2) data processing (transforming, indexing data), (3) the application of this treatment to produce appropriate indicators and (4) generating legible and visible representation of these indicators.

We developed a methodology and a tool to support it. It is a platform, named SEGELL. It has three main functionalities, related to our methodology:

(1) Reprocessing of data to calculate time indicators.

(2) Interactive edition and customization of data.

(3) Interactive selection and parameterization of graphical representations.

Time Indicators

Time indicators are the basic automatically calculated by SEGELL. By the mean of the digital pen and paper, we get a couple of first moment of writing and last moment of writing in every field of test forms. With this data, we can compute durations of the answers to questions: time spent to write the answers, but also time spent between writings on two different questions. In order to exploit duration between writings, we used the notion of "latency"

Research works have shown the importance of a response time factor called "latency". Bassili and Fletcher (1991) studied this factor in the analysis of survey questionnaires by telephone. They wanted to understand the behaviors and attitudes of the respondents. More recently, Grant et al. (2000) have coupled telephone surveys in a computerized technique for collecting time. The analysis of latency was used to test the effectiveness and the formulation of questions (Bassili and Scott, 1996). It has been exploited as an indicator of errors (Draisma and Dijkstra, 2004). Finally, Callegaro et al. (2006) made the analysis of a web survey conducted among job applicants on the one hand and employees on the other hand. They used the measure of latency as an indicator of respondents' motivation. The latency is considered an important indicator for assessing the effectiveness and difficulty of the questions, but also the accuracy of responses, attitude and motivation of the respondents.

Works in cognitive psychology, focusing mainly on the writing task, also studied the duration of breaks. Foulin (1995) showed that the durations of pauses were indicators of cognitive processing. These durations are indicative of a rich cognitive task: research knowledge, syntactic structure and grammar of the sentence, sentence design (Olive et al., 2007)

In the case of a questionnaire on the web or on paper, where the respondent must write the answer by him/herself, measuring the latency starts from the moment of reading the question. The cognitive process of answering includes the formulation of syntactic and logical answer. In that case, the pause before writing an answer is part of the total duration of treatment to solve the question.

In our study, we get a couple of first moment of writing and last moment of writing in every field of the forms. With these times we can compute following indicators:

- The duration of writing in one field
- The length of pause between writings in two different fields
- The total duration spent in one field (the pause before plus the duration of writing).
- The order of the writings in the different fields.

Customization of Data

SEGELL offers the possibility to define additional indicators. For instance, it is possible to add marks to each answer.

It is also possible to add classifications over the fields, by defining possible values and attributing a value to each field. For instance, it is possible to classify fields according to the "types of question", with values like "multiple choice", "several answer multiple choice", "short answer". Another instance could be a classification according to the related lessons in the course, or a classification related to a level of difficulty of the questions.

The aim of this functionality is to cross time indicators with these additional indicators or parameters.

Interactive Graphical Representations

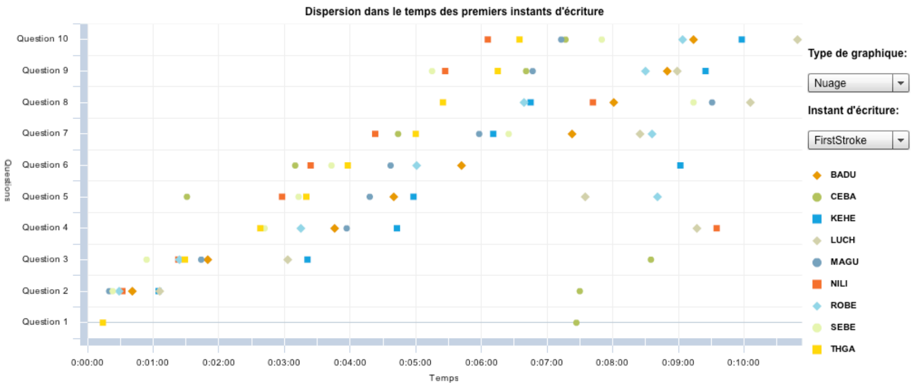
SEGELL offers an interactive interface that emphasizes interesting relations between indicators. This interface allows navigating through different representations and choosing indicators to be represented. All representations display time indicators, and may be customized with additional indicators or parameters.

According to Mitchell (2004), graphical representation of time depends on cognitive representations of time and on native writing and reading directions. SEGELL proposes charts based on left-to-right direction and horizontal time-line, like French reading sentence, to represent time-related data. The timescale does not exceed seconds and minutes. Indeed, we are studying cognitive processes that have very short duration.

We distinguish graphical representation by population of respondents and those by individuals.

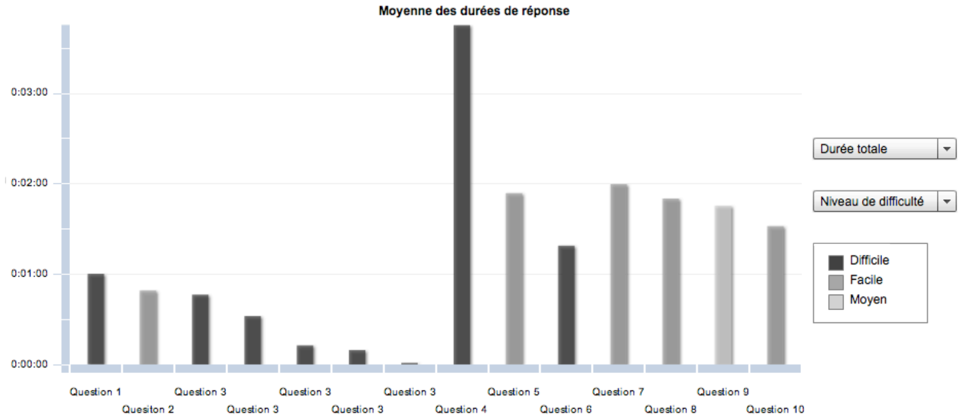
Scatter plot charts display the dispersion in time of first –or last– strokes of respondents to each field (see figure 1). Fields are on the ordinate axis and time is on the abscise axis. The dots in the same colour and shape correspond to the strokes of one respondent. The form of the dot dispersion gives the trend of the answering process of a population of respondents.

Figure 1. Time dispersion of first strokes in fields ('questions') for 9 respondents.



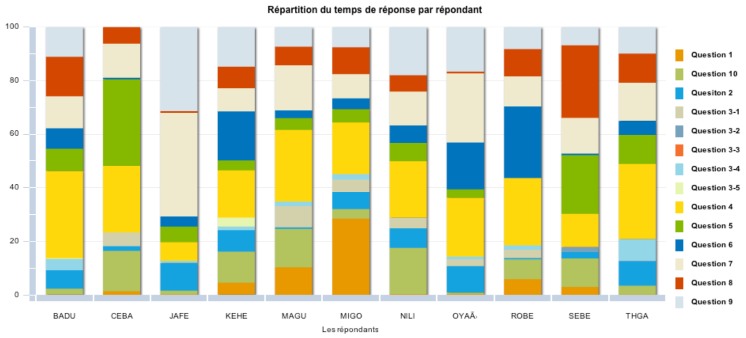
Histogram charts display the average of time spent in fields by all respondents (see figure 2). Three indicators can be displayed: the average latency, the average writing time, or the average total time (latency plus writing). The chart may be customized with user's parameters, by changing the colours of the rectangles.

Figure 2. Average total time on each field (questions). The colour of the rectangles is related to a users' customization, here the level of difficulty of questions: white for easy questions, light grey for medium questions, and dark grey for difficult questions.



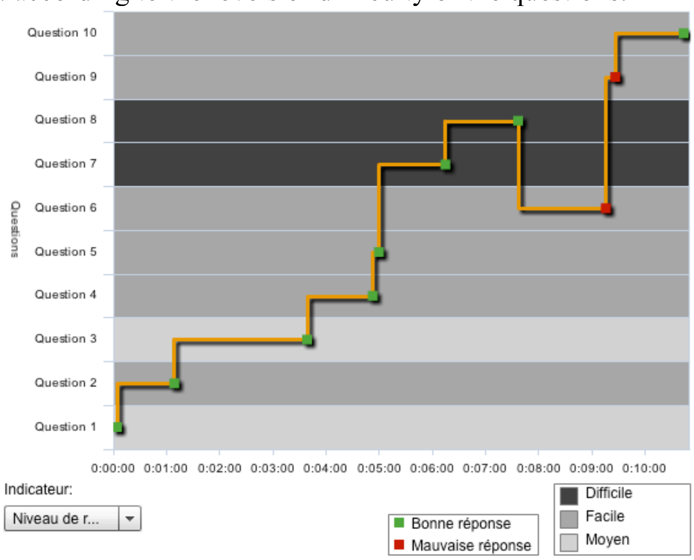
Stacked percentage charts display the percentage of time spent on each field by each respondent among the total time spent on the test (see figure 3). It gives more precise information relative to the time spent on field than the previous chart.

Figure 3. Percentage of time spent on each field (10 questions) for 9 respondents. Each colour is dedicated to a question.



SEGELL computes also line charts that display the answering process of one respondent. We call them 'chronological progressions' (see figure 4). Fields are on the ordinate axis and time is on the abscise axis. The drawn line shows the order in which the respondent filled in the fields, and the time he/she spent on each fields. Coloured dots are added to display marks. Background colours may be customized by user's parameters.

Figure 4. Chronological progression of one respondent to a test that comprises 10 fields (questions). Green dots display good answers and red dots are for bad answers. Background colours are customized according to the levels of difficulty of the questions.



The interface that displays the charts is interactive. In most of the time-related graphs, the user can choose if already defined, parameters that will be illustrated by the colour of bars or of the background lines.

Experiment and Results

Tests and Collected Data

Our experiment took place in a one-semester Bachelor of Science degree course on spreadsheet for management. During this course, we administered 3 paper tests, each spaced by 3 lessons on spreadsheets, to 37 students (35 for test n°2). 9 students, out of the 37, used digital pen and paper during tests.

Each test is composed of the description of a problem with a given screen capture of a spreadsheet, and ten questions. Questions ask students what should be done to reach a given target, or what would happen if some actions were performed. Questions cover all spreadsheet functionalities taught during lessons: cells and sheets editing and formatting, formula writing, chart editing, data table editing and sorting.

There are multiple-choice questions and short-answer questions (see figure 5). Multiple-choice questions accept single answer, except one question. Most of the short-answer questions ask the student to write a formula. In a test, questions are numbered 1 to 10, but they are not ordered according to a given criterion (like spreadsheets functionalities, levels of difficulty, or types of question). Tests are time-constrained according to their levels of difficulty estimated by the teacher: 10 minutes for test n°1; 20 minutes for test n°2 and 15 minutes for test n°3.

We designed the fill-in forms for the tests and instrumented it, by drawing the field and adding the dot pattern. We draw one field for each question, and declare the type of field: multiple choice, or short answer. We used SEGELL to get time indicators, add custom data and edit graphical representations. All screen captures given below are taken from SEGELL.

Figure 5. Test n.1, questions 4 to 6 (English translation).

Question 4
Given that the cell E6 contains the formula =C6*D6. What will be the formula in cell C3 if the contents of cell E6 are copied to that cell?

E6	=

Question 5
The cell E11 displays ######. Choose an explanation:

- O E11 contains an erroneous formula.
- O E11 contains a data that is hidden for confidentiality reasons.
- O E11 contains a number that is too wide for the cell to display it.

Question 6
E11 should display the total purchase price before any taxes. F11 should display the total amount due. Give a formula for E11, which may be copied to F11?

E11	=

Question 7
The cell C2 has no specific format. If one enter 13/12, C2 will display:

- O The text "13/2"
- O The date "13-dec"
- O The fraction "13/12"
- O The decimal number "1.0833333"

Results

We first present the success rate of the 37 students to the 3 tests, and the difficulties they encountered. Note that the marking scheme – one mark for every correct answer– encouraged students to answer all questions, and that's what they did.

Table 1 shows success rate to the 3 tests. In order to get an indicator of students' success rates for each question, we classified questions in 3 categories: "easy" (more than 2/3 of students give a good answer), "medium" (between 1/3 and 2/3) and "difficult" (less than 1/3).

Table 1. Success rates to the tests (37 students).

	Test n.1	**Test n.2**	**Test n.3**
Average number of correct answers	7,1	5	5,9
Number of students who give 5 or less correct answers	5	23	15
Number of "difficult" questions	2	4	3

As expected by the teacher, the test n°1 was easier than the 2 others, and the test n°2 was the most difficult.

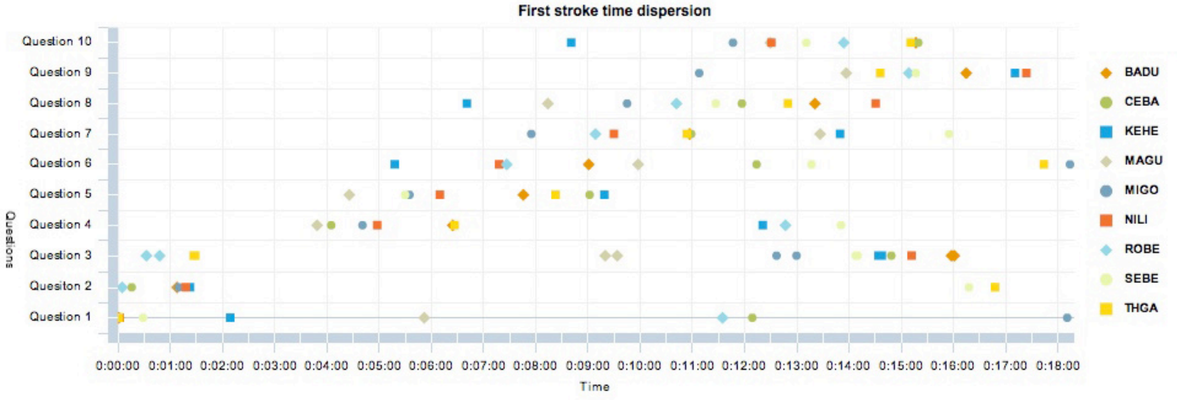
Do time stamped logs give more information on students' performance?

About the order of answering the questions

Digital pens store the time of the first stroke and the last stroke in each field. The scatter plot chart on figure 6 displays, for the test n° 2, the first strokes of each student to the 10 questions. For instance, the student ROBE began to answer to the question 1 after 11:30, after having answered to question 8.

The dots scattered along the diagonal line show that, globally, students answer to questions in the order in which the questions are proposed (question 1 first, question 2 second, etc.). Similar diagonals appear on the scatter plot charts of the two other tests.

Figure 6. Test n.2, first stroke dispersion to each question for the 9 students.



The dots below the diagonal correspond to answers that are postponed. It is likely that the student reads the question, but he/she answers to the question later on, after answering to other questions. For instance, the figure 6 shows that several students have postponed the answers to questions 1, 3, 6 and 9.

The scatter plot graphs of the two other tests show less postponed answers (see test n°1 on figure 1). The test n°2 is more difficult than the others. A hypothesis is that the more the test is difficult, the more students change the order of their answers and postpone questions.

Moreover, considering the three tests, it appears that the most postponed questions are "difficult" or "medium" questions. On the contrary "easy" questions are rarely postponed. An hypothesis is that 1) the students prefer to answer to "easy" questions first, and 2) that students recognize "easy" questions.

Finally, the question 3 of test n°2 is the most postponed question, over the three tests. The question is about error messages in spreadsheets, like two others questions in the other tests, but it is the only multiple-choice-multiple-answer question. It seems that many students when discovering this new type of question decided to postponed it. Moreover, globally, students didn't succeed in answering it, it is a 'difficult' question.

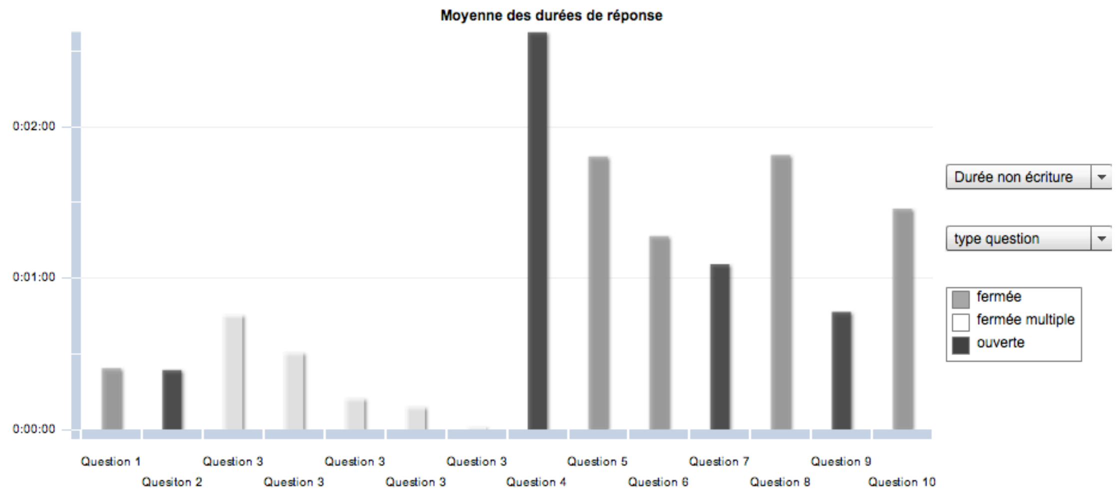
About the time spent on questions

As we already notice, the tests were time-constrained according to the level of difficulty the teacher attributes to them (test 1: 10 minutes, test 2: 20 minutes, test 3: 15 minutes). Digital pens collect time-stamped strokes. SEGELL computes the time spent on each question, adding writing time and time spent before to write (latency).

Considering both time and type of questions, it appears that in average students spend more time (latency plus writing) on short-answer questions than on multiple-choice questions. This result may seem obvious. However considering latency only, the figure 7 shows that some multiple-choice questions of the test n°2 (questions 5, 6 and 8) took more time than some short-answer questions. That is also the case for the two other tests.

Considering both spent time and success rates of questions, no general relation appears between them. It is certainly more interesting to draw up categories of questions: "difficult and time-consuming", "easy but time-consuming", "difficult but quickly answered", "easy and quickly answered", etc. For instance the short-answer question 4, in test n°1 (see figure 5) has a high average latency, but a good success rate (it is classified as 'easy'). On the other hand, the multiple-choice question 7 (see Figure 5) has also a high average latency, but a bad success rate (it is 'difficult'). In other words, some questions are time-consuming but finally successfully completed by students, whereas other questions are time-consuming and are not well completed.

Figure 7. Test n.2, average latency on each question (9 students). Dark grey questions are short-answer questions, and light grey questions are multiple-choice questions. Question 3 is multiple-answer.



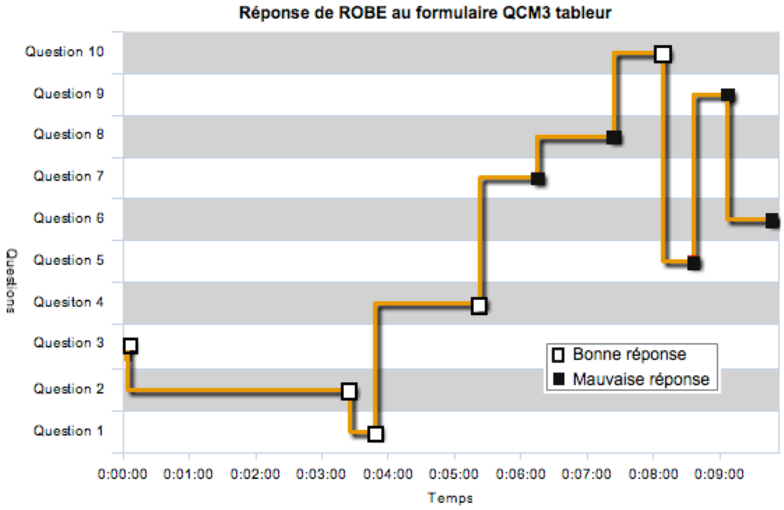
However, the latency ascribed to a question placed after a postponed one may be over-estimated. For instance, on the figure 7, the average latency of question 4 is far higher than others. We observed that 7 students have postponed their answer to the question 3 (the question just before question 4). Part of the latency attributed to question 4 may be in fact a time spent on reading a question 3. This case appears on the chart showing average time, because many students have postponed the question, thus many latencies have been over-estimated.

About individual answering processes

SEGELL provides individual 'chronological progressions'. These charts show how much time one student spent on each question, and how he/she ordered his /her answers to the questions. Figure 8 shows that student ROBE begun by question 3, and then spent more than 3 minutes on question 2. He/she gave 4 correct responses and then failed to answer to questions. It seems that he/she postponed to the end 3 questions (5, 8 and 9) that were difficult for him/her.

Regarding the chronological progressions of students on the three tests, we observed that some students adopt the same way of doing on the three tests. 2 students answered to questions in the order they are proposed, and postponed 1 question or none in each test. On the contrary, 4 students never answer in the proposed order, and postponed from 2 to 6 questions in each test.

Figure 8. Test n.3, chronological progression of the student 'ROBE'. Questions are on the ordinate axis and time is on the abscise axis. The drawn line shows the order in which the student filled in the question, and the time he/she spent on each questions. White dots are added when answers are correct and black dots are for bad answers.



Discussion

The analysis of times, collected with digital pen and paper, give more information on the way students answer to paper assessment tests. How can we exploit this information in an educational perspective?

We have shown that data on time can be achieved on paper tests, using the digital pen and paper. We assume that analysis of times give a characterization of assessment tests and questions that could be used to enhance the assessment and to improve student training. For instance, our results concerning the order of answering the questions show that some questions were not only 'difficult', as there success rates were low, but were also postponed by a lot of students. These questions have hindered students' answering process. The teacher could take it into account for students' assessment. On another hand, we can qualify questions regarding spent time, and especially identify 'time-consuming' questions. A teacher could focus on these questions, discuss them with students in order to determine their difficulties, and train them on similar questions.

We could also take benefits from such data in the area of large-scale assessment tests, like the OECD PISA program. In 2006, some countries have administrated a computer-based assessment of students' scientific literacy in the framework of the OECD PISA program. Collecting *more information, such as time taken per item, keystrokes made and a student's movement through the questions* was among the benefits of the method cited in (OECD, 2010). Unfortunately, such data has not been used in the analysis.

More generally, there is a lack of research on this issue. In order to be able to interpret such data we need references. A solution could be to build calibrated tests, with time references for each question.

Conclusion and Perspective

This first experiment on assessment tests, gives interesting results concerning time on assessment test logs. We could formulate some hypothesis on students' strategy when answering the tests: answering first 'easy' questions, postponing 'difficult' questions. We also observed that some questions are time-consuming: that the case of some 'difficult' questions, but also of some 'easy' questions. We observe that the latency –the time spent on a question before to write the answer– may be significant, regardless the type (multiple-choice or short-answer) of the question.

Although, indicators calculation may be performed with general software tool, as spreadsheets, the platform SEGELL greatly facilitates the work. It computes time indicators automatically, it supports the customization of data by storing classifications of fields, and it edits interesting charts.

Now, we would like to investigate two main directions. One perspective is to adapt our methodology and our tool, SEGELL, to online assessment tests, proposed in e-learning platforms. The idea is to enrich online tests with scripts that record numerous actions and events made by the respondent, like mouse clicks, roll over or focus, etc. We would certainly get richer traces than with digital pen, and could study more precisely the ways students answer to tests, in distance education.

Moreover, until now, we mostly adopted a researcher's point of view, and developed mainly indicators that help to study populations. A teacher, or a student, would certainly prefer to get results in terms of individual performances. For instance, get the mark, the time and the rank, of each answer, and, furthermore, have an appreciation about this time (too long, longer than others), a comparison between the mark, or the performance and time.

The other perspective would be to continue our investigation on the graphical representations of the traces given by time stamped logs. Firstly, we would like to add charts that illustrate a spatial-time representation of traces by drawing the path of the answers on a picture of the form. We can easily do that, by using the x,y coordinates of the strokes.

Bibliography

Bassili, J. N., & Fletcher, J. F., 1991, Response Time Measurement. In Survey Research: A Method for CATI and a New Look at Attitudes. Public Opinion Quarterly 55(3):331-346.

Bruillard, É., Blondel, F.-M., and Tort, F., 2008, DidaTab project main results: implications for education and teacher development. In K. McFerrin, R. Weber, R. Carlsen & D.A. Willis (Eds.), *Proceedings of Society for Information Technology and Teacher Education International Conference, SITE 2008*, pp. 2014-2021, Chesapeake, USA: AACE.

Callegaro, M. et al., 2006. Response latency as an indicator of optimizing: A study comparing job applicants' and job incumbents' response time on a web survey. In C. van Dijkum, J. Blasius & C. Durand (Eds.), Recent developments and applications in social research methodology. Proceedings of the RC 33 Sixth International Conference on Social Science Methodology. Amsterdam 2004 Opanden & Farmington Hill: Barbara Budrich.

Dimitracopoulou, A., 2008. Computer based Interaction Analysis Supporting Self-regulation: Achievements and Prospects of an Emerging Research Direction, In Kinshuk, M.Spector, D.Sampson, P. Isaias (Guest editors). Technology, Instruction, Cognition and Learning

(TICL).

Draisma, S., Dijkstra, W., 2004, Response latency and (para)linguistic expressions as indicators of response error. In Methods for testing and evaluating survey questionnaires ed. Stanley Presser, Jennifer M. Rothgeb, Mick P. Couper, Judith L. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer. New York: Wiley- 2004.

Foulin J. N., 1995, Pauses et débits : les indicateurs temporels de la production écrite. In L'année psychologique. 95(3): 483-504.

Grant, J. T.,et al., 2000, Response Time Methodology for Survey Research. In the Annual Meeting of the Midwest Political Science Association, 27-29 Avril 2000, Chicago, Illinois.

Heidy, M., Brian, L. et Scott, K., 2006, Technology for desing education: a case study. Conference on Human Factors in Computing Systems, Montréal, Québec, Canada 2006 (pp 1067-1072).

Mitchell, M., 2004, The visual representation of time in timelines, graphs, and charts. ANZCA04 conference, Sydney, jyly 2004.

OECD, 2010, PISA Computer-Based Assessment of Student Skills in Science, PISA, OECD Publishing, www.sourceoecd.org/scienceIT/9789264082021

Olive, T., Alves, R. A., & Castro, S. L., 2007, La production de textes : Quels processus sont activés lors des pauses de production et des phases d'écriture ? In J.-M. Hoc & Y. Corson (Eds.), Actes du Congrès National 2007 de la Société Française de Psychologie (pp. 65-72).

Oviatt, S. L., Arthur, A. M., et Cohen, J., 2006, Quiet interfaces that help students think. In Proceedings of UIST '06, Montreaux, Switzerland. ACM: (pp 191-200).

Settouti L. S., Prié Y., Champin P.-A., Marty J.-C., Mille A., 2009, A Trace-Based Systems Framework : Models, Languages and Semantics, inria-00363260 Research Report avalable in http://hal.inria.fr/inria-00363260/fr/.

Sugihara T., Miura T., Miura M., Kunifiji S., 2010, Examining the effects of the simltaneous display of students'responses using a digial pen system on class activity. 10th IEEE International conference on advanced learning tchnologies, pp. 294-296. 2010.

Tort, F., Blondel, F.-M., & Bruillard, É., 2009, From error detection to behaviour observation: first results from screen capture analysis, In D. Ward & G. Croll (Eds.): The Role of Spreadsheets in Organisational Excellence, proc. of the EuSpRIG 2009 Conference, pp 119-132, London: EuSpRIG.