ExamPen: How digital pen technology can support teachers and examiners

Thorsten Karrer
RWTH Aachen University
Ahornstraße 55
52074 Aachen, Germany
karrer@cs.rwth-aachen.de

Moritz Wittenhagen
RWTH Aachen University
Ahornstraße 55
52074 Aachen, Germany
wittenhagen@cs.rwth-aachen.de

Leonhard Lichtschlag
RWTH Aachen University
Ahornstraße 55
52074 Aachen, Germany
lichtschlag@cs.rwth-aachen.de

Jan Borchers
RWTH Aachen University
Ahornstraße 55
52074 Aachen, Germany
borchers@cs.rwth-aachen.de

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
We present *ExamPen*, a digital pen based tool for analysis of written examinations and visualization of student performance. While written exams play an important role in higher education, there is little tool support to ensure that exams are fair, well-designed, and transparently graded. *ExamPen* enables teachers to better understand how their questions are solved by the students over the course of a written test. We believe that this can help teachers to identify possible flaws in the exam design or grading process. *ExamPen* has been implemented as a prototype and its applicability tested in a proof-of-concept study.

Introduction
Obtaining a degree in higher education certifies fulfillment of certain standards in skill, knowledge, or intellectual capacity. To verify that these standards are met, the issuing authorities such as universities often rely on written examinations. It is paramount for the quality of education that the way teachers derive the grades from the performance of the students during such exams is transparent, reproducible, and fair. This comprises all steps from the preparation of exam tasks to the review of the deliverables. In the following we focus especially on the creation and grading of written exams as they are common in university-level studies.

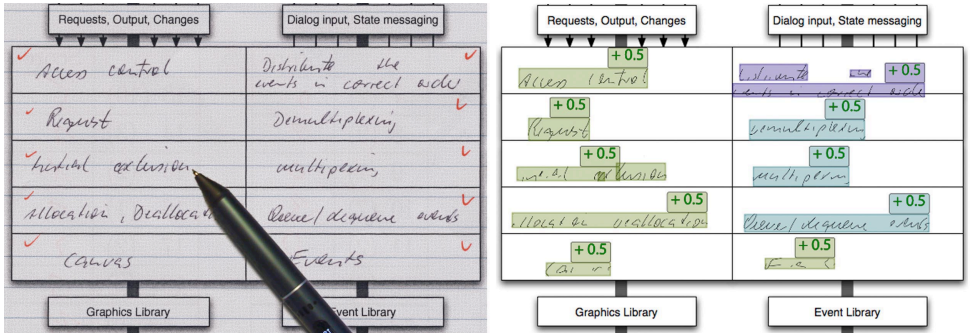

Figure 1. Left: original exam sheet with the students answers and the corrector's annotations. Right: corresponding screenshot from our GUI. *ExamPen* colors *strokes* according to timestamps, and the teacher annotates them with the achieved marks.

Often, there are no strict guidelines for the design of these exams. The process relies only on the experience of the teacher who is asked to come up with questions and tasks that are suitable to clearly indicate and differentiate the distinct levels of students' performance. At the same time, these questions have to be easily understandable, must not be ambiguous, and should clearly communicate the amount of work a student ought to put in the solution during the exam. They should, of course, also yield an appropriate amount of marks to reflect their demands in knowledge and work time in the context of the exam. Devising fair exams is, in a word, difficult and time consuming.

Teachers often reuse exam questions from previous years which leads, having gathered a repository of students' answers in the past, to a refined a posteriori estimate of these questions' difficulty. This is one way to help tuning the rating of the questions or spotting misleading question texts. What is not regarded by methods like this, however, is that the rating of an exam question not only has to reflect the difficulty of the question but may also depend on the amount of time the teacher expects students to spend on the task during the exam. Apart from test-driving an exam with TAs and stop watches, there are little ways to effectively determine if this aspect of the exam is sufficiently calibrated to allow realistic and fair grading of each question. All in all, there is a lack of tools that allow teachers to critically assess the quality of their own exam designs.

We are interested in how we can leverage digital pen technology (in contrast to using tablets like [1]) to support teachers in the design of written tests and the tutoring of students. Writing exams using digital pens gives, in addition to the final 'snapshot' represented by the ink on the students' hand-ins, the timing information for each pen stroke. Teachers can then not only analyze the content of the answers (as it is done, e.g., by [2]) but also interactively explore the temporal evolution of their students' papers over the course of the exam. Data on how long it took to answer a certain question, how often an answer was revised, or the order in which a student chose to work on the questions becomes accessible. Individual strategies of students on how they approached the exam are revealed and can help teachers to provide better personalized tutoring on areas students have difficulties with. The goal of our research is to investigate how this kind of support can increase the quality of teaching and the fairness of written exams.

ExamPen - a Prototype

Our *ExamPen* prototype consists of a set of Anoto[1] digital pens which are used by the students to write their exam papers with and a software that is used by the teacher to review, grade, and analyze the exam afterwards. This is somewhat similar to [3] but extends the focus beyond grading support. The exam handouts

[1] www.anoto.com

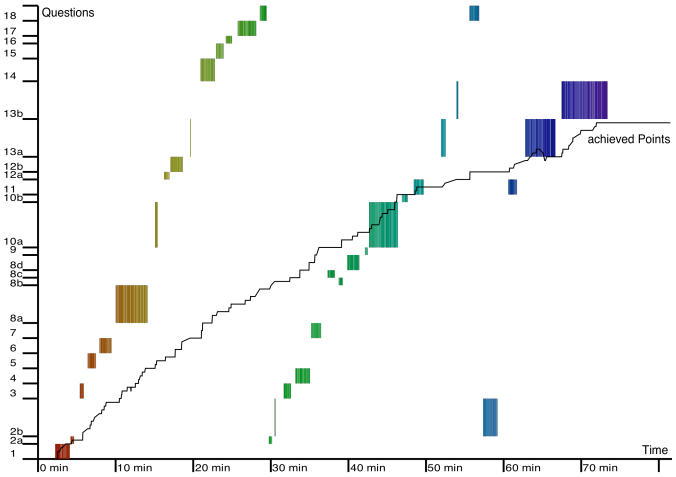

Figure 3. The diagram shows how a student progresses through the exam in three passes: the second and third pass hardly revisit the questions solved in the first, instead the student concentrates on more difficult questions. The horizontal position of the blocks show when and in which question she made pen strokes; the height of each block shows how many marks can be achieved for each question.

are printed on Anoto micro-dot pattern and all exam questions are interleaved with enough space to allow students to fill in their answers directly onto the handout sheets. The digitals pens record every stroke each student draws together with a timestamp. Each stroke is automatically assigned to a single exam task or question based on the position on the exam paper. When the teacher reviews the exam in our software color coding indicates the time when a stroke was written by the student during the exam (Figure 1 right). This already gives a quick and intuitive overview on how long a student took for a task and how often she had to revisit it for annotations or corrections. The *ExamPen* allows the teacher to classify students' answers as being correct, wrong, or neutral (like margin notes or emphasis marks). Similar to the normal grading workflow of many teachers, the software also allows marks to be given for correct answers or deducted for wrong answers (Figure 1 right).

This mixture of automatically and manually collected data *ExamPen* can be interpreted in two scopes: on the one hand, we can analyze individual students' performance and strategies over the course of the exam, on the other hand we can present statistics on individual exam questions across all students.

Figure 3 is an example for a diagram that shows a student's distribution of strokes for each question over the time of the exam. A teacher can determine the student's strategy at a glance—in this case the student went through the whole of the exam in three cover-to-cover passes—and if and how often each question has been revisited. The strokes can also be cumulated by question to determine the total time spent on each question.

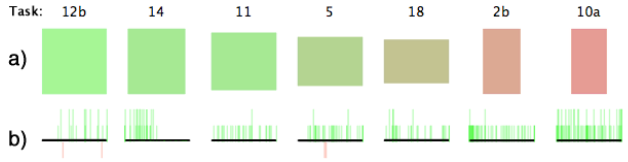

Figure 2. Question footprints (a) and marks over time distribution (b). A quick glance reveals questions with wrongly estimated durations (18,2b,10a), or which are graded unbalanced towards the end (14).

Assuming the questions are designed in a way that a correct answer should approximately take a fixed amount of time, the ratio of minutes spent on a question in average to this expected duration indicates if the task's complexity was estimated correctly by the teacher during the design of the exam. This ratio is visualized as a *footprint* for each question (Figure 2a): width corresponds to average time spent and height corresponds to the expected time; a square is a desirable result while an unbalanced rectangle may indicate a poorly designed question. An *average marks*

over time distribution for each task (Figure 2b) reveals at which point in time marks are gained or lost relative to the time the students worked on that question. This way, a teacher can assess if progress prior to finishing the task is already rewarded properly or if students write more than what was asked for.

Current State and Future Work

Our aim is to explore the potential of digital pen technology to support teachers in assessing the quality of their exams and to help them tutoring their students. We developed *ExamPen*, a prototype of such a system, which we used to conduct an informal proof-of-concept study analyzing an undergraduate level exam on HCI. Our experience was that teachers could not only uncover flaws in the exam design but that the collected data provided insight into students' exam-solving strategies and their individual performance that would have been difficult to gather otherwise.

Having only started to investigate what can be done with the timing information from digital pens in written exams, more questions have yet to be answered:

- Can teachers provide more detailed individual mentoring to students based on the additional insight the tool provides?
- Does providing the students themselves with the results of the analysis for self-tutoring help them approach their next exams more efficiently?
- Will clustering the annotated strokes for each question to yield tag-clouds of correct and wrong stroke chains enable teachers to keep grading that is done by different persons more consistent?
- Can and should this technology be used to detect plagiarism during exams?

We are looking forward to discuss these ideas with other participants of the "Next Generation of HCI and Education" workshop and identify the most important research questions in that area.

Acknowledgements

This work was partially funded by the German B-IT Foundation and by the German Government through its UMIC Excellence Cluster for Ultra-High Speed Mobile Information and Communication and its HumTec Human Technology Center at RWTH Aachen University.

References

[1] R. de Silva, D. Bischel, W. Lee, E. Petersen, R. Calfee, and T.Stahovich. Kirchhoff's Pen: A Pen-based Circuit Analysis Tutor. In *Proc. of the 4th Eurographics workshop on sketch-based interfaces and modeling,* 75-82, 2007.

[2] S. Friedler, Y. Tan, N. Peer, and B. Shneiderman. Enabling Teachers to Explore Grade Patterns to Identify Individual Needs and Promote Fairer Student Assessment. *Computers & Education,* 51(4):1467-1485, 2008.

[3] M. Nakagawa, N. Lozano, and H. Oda. Paper Architecture and an Exam Scoring Application. *Proc. 1st pen-based learning technology,* 113-118, 2007.