

## **On-line Exams, from Intranet to Internet.**

Dick J. Bierman, Bas Roosen & Jeroen Ruwaard  
University of Amsterdam

### **Summary**

*For the last 5 years students, taking the course 'Instrumentation in Psychology', were tested on-line for their final grading. On-line examinations pose stringent criteria to the software used. It is mandatory that the student has the feeling that the on-line aspect does not hinder the expression of the student's knowledge. At any time the student has to be able to see which items have been responded to already. Also, the student should be able to jump from an item to any other item. And when returning to a previously answered item, the student should see which alternative was chosen. Furthermore, special measures should be taken to ensure the integrity of the data even in case of computer failure. Finally computerised examinations should give study recommendations in case the student does not pass. On-line exams are feasible when supervision is available, as has been the case with the on-line testing through the faculties local network. But at this moment there are no ways to guarantee fraud-proof remote examination through Internet. Nevertheless remote examination software may be used to give students an opportunity to assess their own knowledge levels and to get some study advice. The discussion focuses on the transition from the Intranet to the Internet environment.*

### **Introduction**

Traditional Computer Assisted Instruction generally is built around a cycle of recurrent actions. Cognitive diagnosis is followed by remedial exercises and instruction driven by the diagnosed misconception or lack of knowledge (1). This process is repeated until a satisfactory level of knowledge is reached. Computer Assisted Testing on the other hand only covers the diagnostic part of this cycle. It follows therefore that it is often possible to extract from a CAI program a part which would assess the level of knowledge acquired by the student accompanied by a cognitive diagnosis. And similarly it may be a proper strategy to initiate the development of courseware by first developing the diagnostic module. The course 'Instrumentation in Psychology' has used traditional CAI from its early beginning in the late 70's to the present date<sup>1</sup>. Therefore it was quite natural to use the same techniques that were developed to teach the students in the first place, for the final exam of this course. During the last 5 years the exam software that we used over Intranet has gradually changed. The final version was used as the model for implementation of the exam software for Internet. While the former was really used in an examination setting the latter is intended for self-assessment of the actual level of knowledge. We hoped to bring the advantages that were evident from the limited Intranet use to a wider audience on Internet. These advantages were:

- a) the examination software environment proved to be a natural setting for the teacher to improve the items' validity.
- b) the feedback to the student was not only a direct feedback of the grade but more importantly, an advice about what topics needed further study.

### **Computerised testing**

The use of computers for testing purposes is constrained by the limited ability to understand the responses of students. Although AI techniques have been promised to change this picture, it is still not possible to generally apply these techniques in this setting because of the condition that the examination context does not allow for any erroneous interpretation of student responses. The implication is that error-free testing can only be done using limited choice techniques. The items used should be able to tap the knowledge of the student.

---

<sup>1</sup> ... " ... " ... "

Another requirement is that the test should cover all aspects of the knowledge. The general solution to these requirements is to specify a large item pool together with a thematic categorisation of each item. After specification of an examination profile which dictates the number of items for each theme, the software is able to generate a representative exam by random selection of items. Tools should enable the teacher to see the current item's validity and to edit or replace less valid items.

Another aspect of computerised testing is that the cognitive load of the user should be minimal so that all cognitive resources are used for answering the items. This poses some constraints on the user interface. User friendly is not enough. The students screen should be so self evident that even a completely naive user cannot go wrong. For instance, at each moment the student should have an overview about how many items still to go and about which items the student already responded to. Furthermore, the earlier response of the student should be shown if the student decides to return to an item.

Often, the direct feedback of the result of a test is cited as the major advantage of computerised testing. However the immediate analysis of the responses leading to a study advice may be a more important advantage. Furthermore, outliers in response times may be used to assess which items have probably been guessed rather than answered from the student's knowledge. Finally tailored testing may be implemented thereby reducing the testing time.

### Intranet implementation

The Intranet version was implemented using the authoring language Authorware Professional. An overview of the flow of control is given in figure 1.

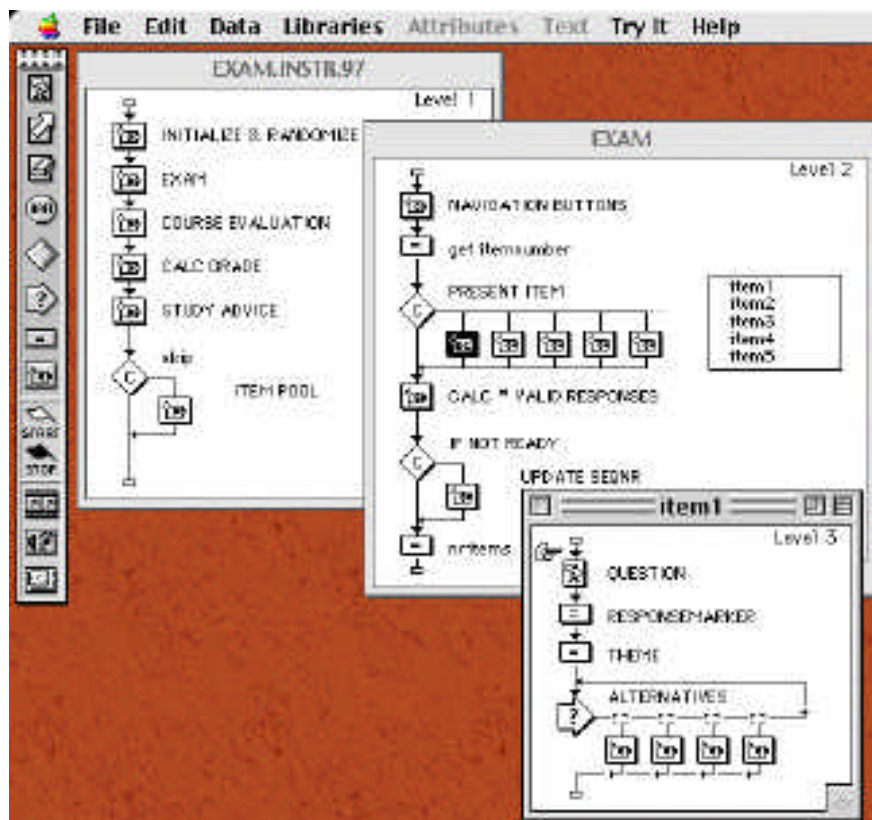


Figure 1

#### Teacher's perspective

The items used are fixed but the teacher can adjust this by moving items from the itempool to the PRESENT ITEM interaction. The items in the pool have all the information, question, alternatives and theme-category embedded. The presentation order of the items is randomised for each user.

Student's perspective

Basically the system is waiting for a student action which can either be a selection of an item or giving a response to an item (see figure 2).

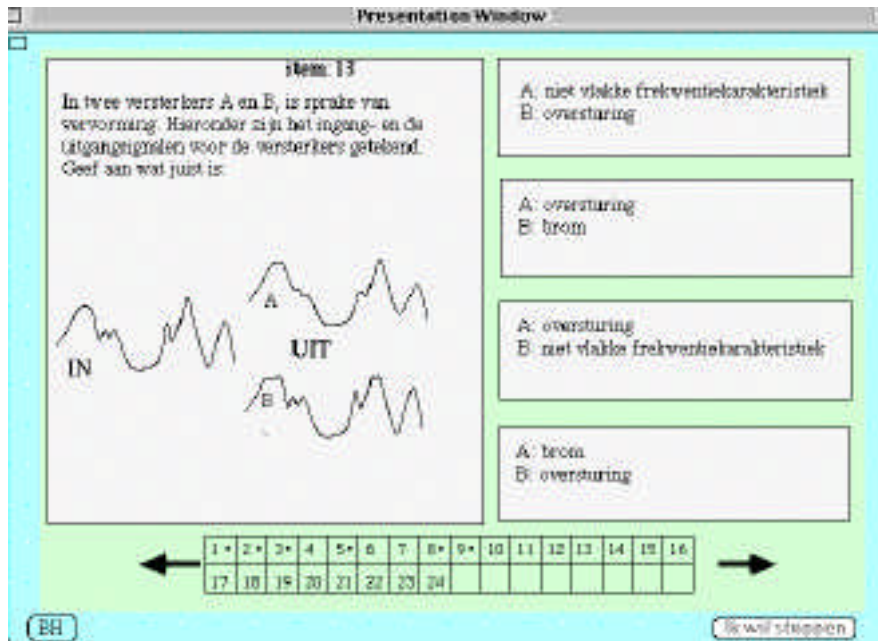


Figure 2

When the student is ready with all items (s)he can leave the exam to answer some evaluation questions about the course. During this time the supervisor still has a secret option to help the student return to the exam proper.

After filling in the questionnaire the grade for this exam is calculated and the student has an option to inspect the list of themes for which (s)he show an inadequate mastery level (figure 3).

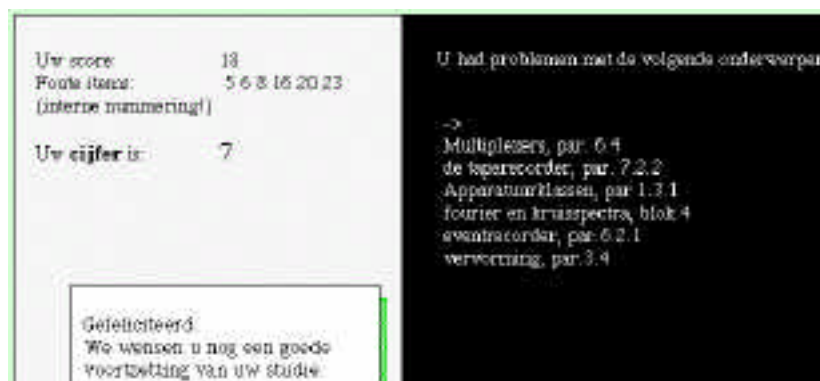


Figure 3

One of the evaluative questions concerns the satisfaction of the student with this way of computerised examination. The figures on a 4 point scale (1=bad.... 4= very good) are given for the last 6 years in Table 1. In 1994 a few computer crashed which gave rise to a global anxiety which is clearly reflected in the score. Since then we have adopted a strategy to have students also log their responses on a paper sheet which could be used to rapidly recover from a computer failure.

Year	Score	sd
------	-------	----

1991	2.7	1.3
1992	3.0	1.0
1993	3.1	0.8
1994	2.4	1.6
1995	2.9	1.1
1996	3.0	1.0

Table 1

**Internet implementation**

The Internet version is implemented using a standard SQL database (Butler) for the storage of items, alternatives, correct alternative, theme-categorisation and responses. The data-model used is given in figure 4.

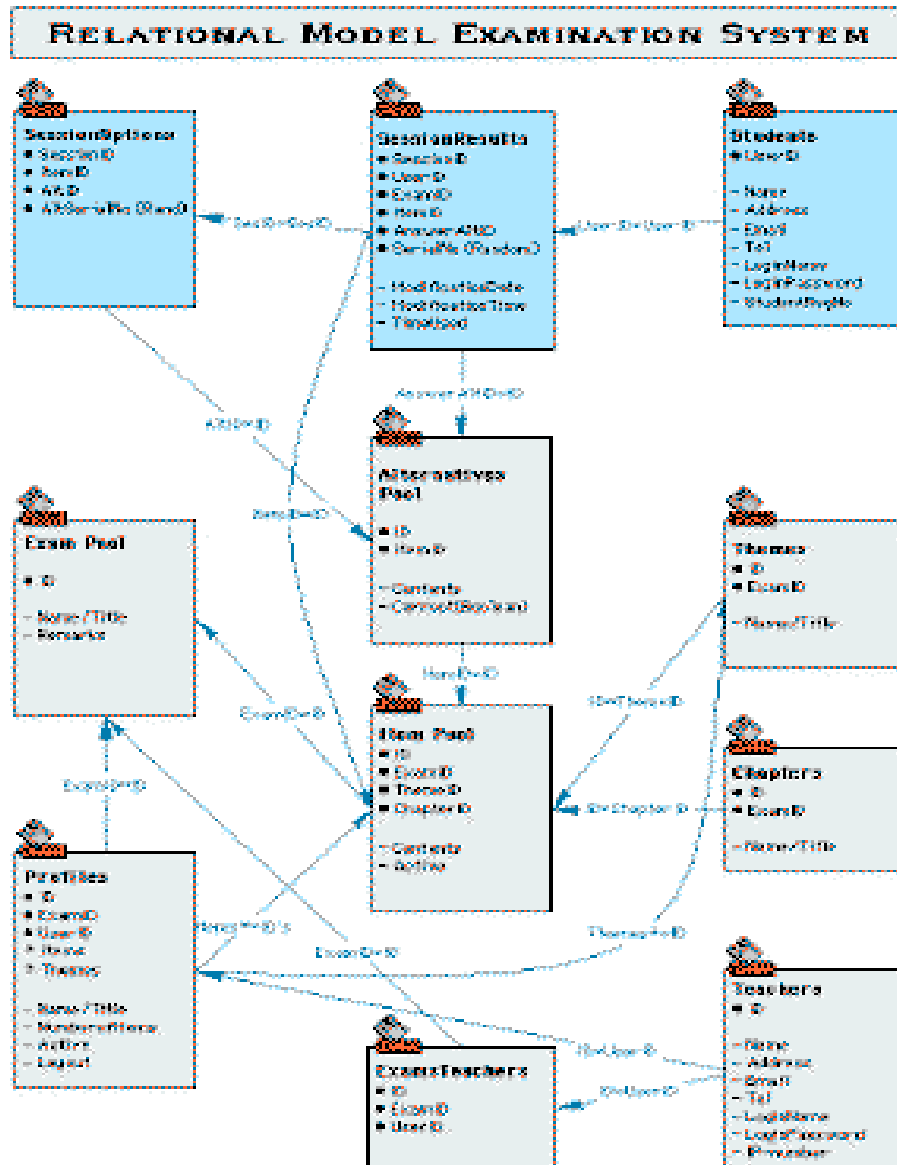


Figure 4

The students pages are dynamically generated from the database. This is done by a "CGI" developed in Weblink Tango. The whole system runs on a WebStar/Tango server combined with a SQL database server. The average time to generate a page is less than 3 seconds.

### The teacher's perspective

The teacher's home page provides options for bulk import of items, review of the statistical properties of items, editing of items and the specification of an examination profile (figures 5 and 6).

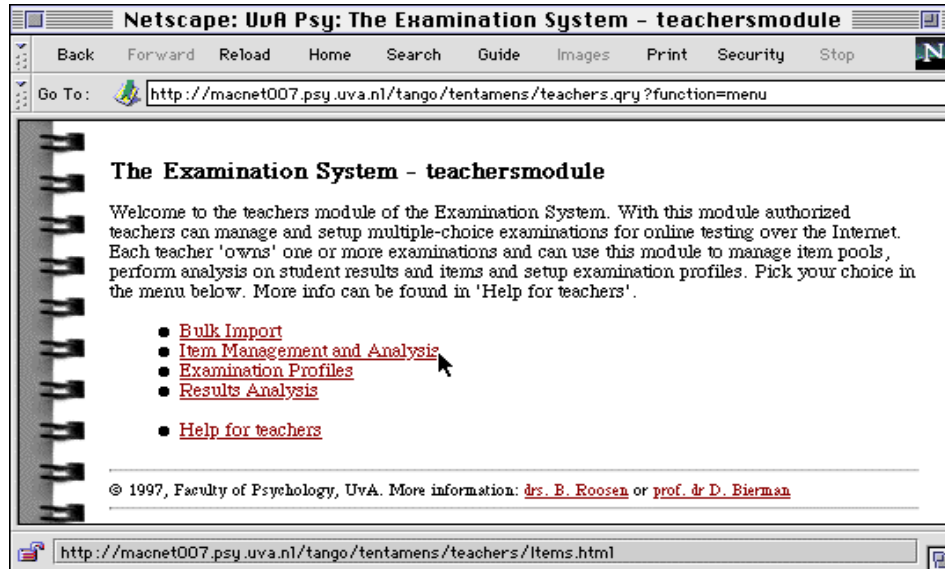


Figure 5

The teacher may allow students to test themes only or may specify an exam profile that exactly replicates the profile of the real examination. The item may be a 3 or 4 alternatives multiple choice item.

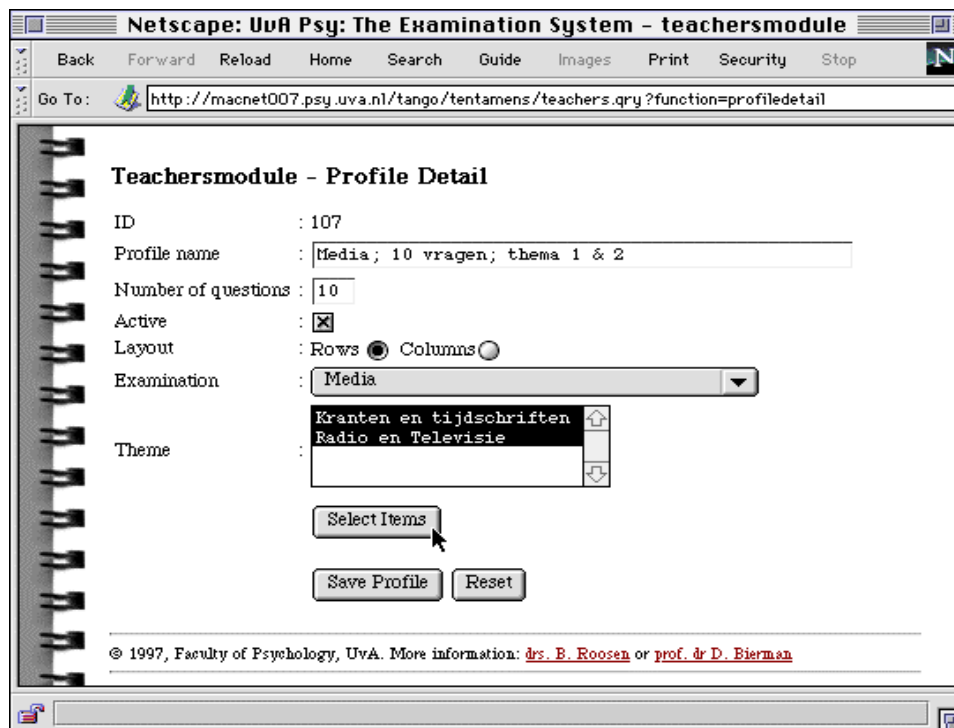


Figure 6

Each item belongs to a thematic category that will be used for the selection of the item as well as for the advice that the system generates at the end of the exam. The presentation format is also specified by the teacher. It is either a 2 column format with the item in the left

Bierman, D.J., Roosen, B. & Ruwaard, J. (1998).  
Procs of the International BITE conf., Uni.of Maastricht pp.218-228.

column and the alternatives in the right column or it is a standard format with the item and alternatives below each other.

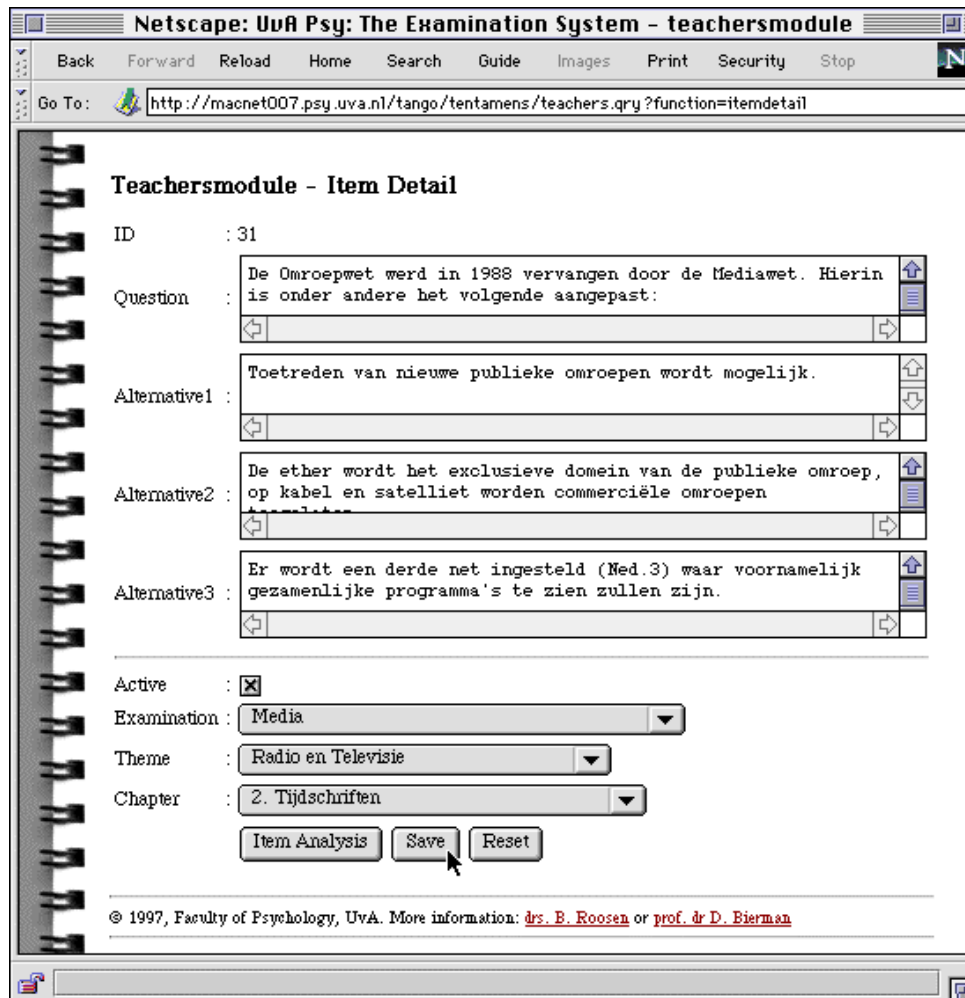


Figure 7

Bierman, D.J., Roosen, B. & Ruwaard, J. (1998).  
Procs of the International BITE conf., Uni.of Maastricht pp.218-228.

### The student's perspective

Students start, depending on their password, on a homepage with one or several examination options to choose from. (figure 8).

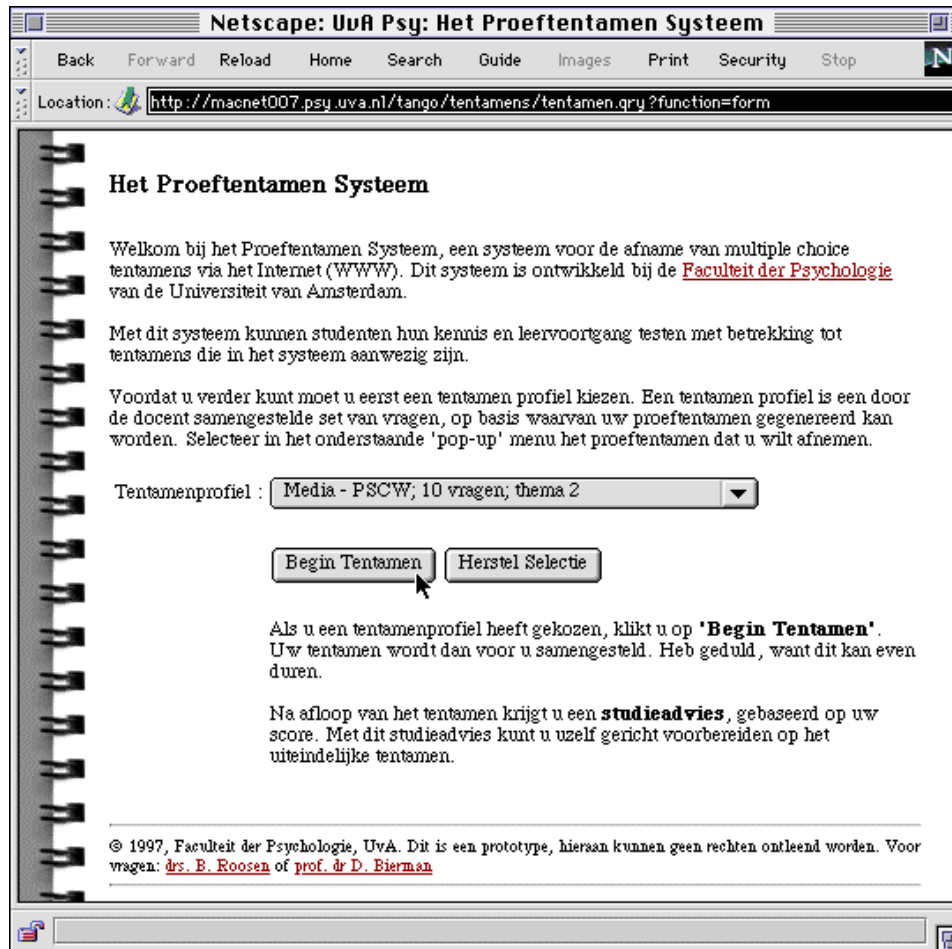


Figure 8

After selection of the examination the software constructs a new examination using the profile. Items and alternatives per item are randomised. The first item is presented in the format specified by the teacher (figure 9). It should be noted that the student may be allowed to have a score calculated before all items have been responded to. In a real test it is not a wise strategy not to respond to items that you are unable to answer. However when the goal is to get the best possible study advice a student should try to avoid guessing because correctly guessed items produce noise with regard to the study advice. When the score calculation is requested the results are displayed (figure 10) in the form of over-all grading and grading per theme. The student is advised to restudy the themes that appear to be insufficient. This advice may be printed.



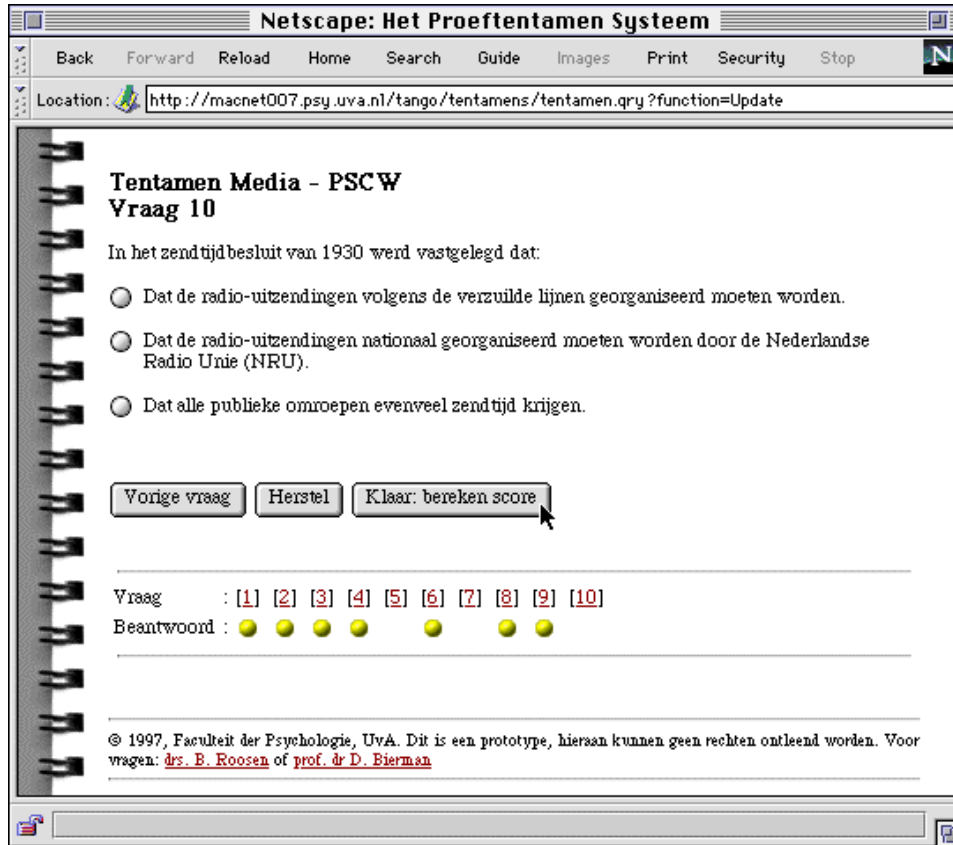


Figure 9

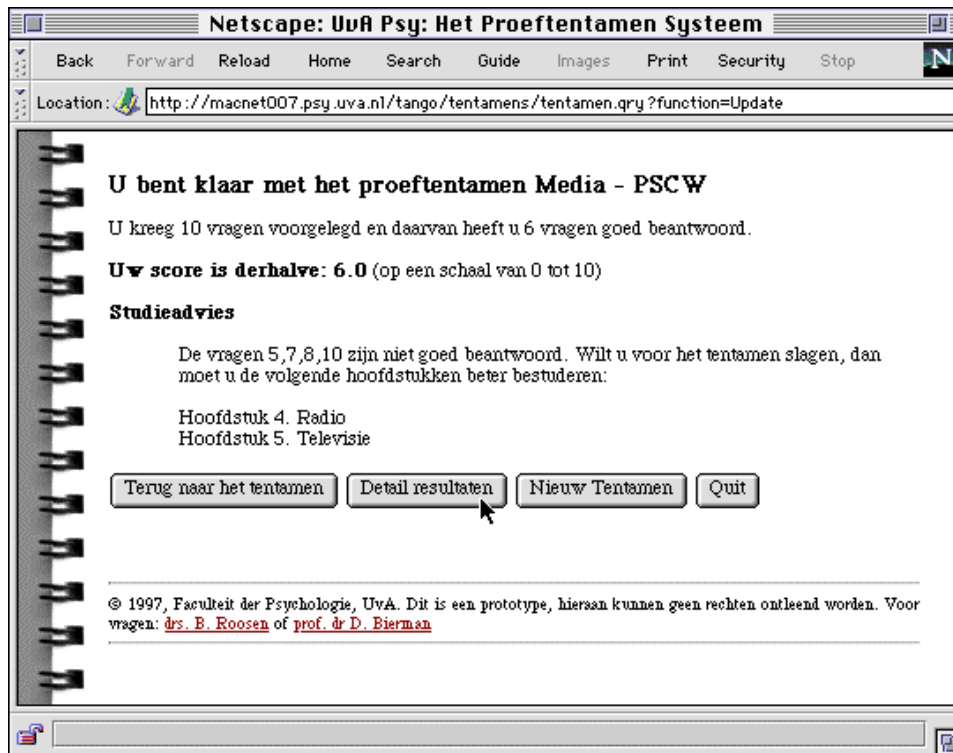


Figure 10

### Future developments

#### Refined diagnoses and advises.

The system is set up in such a way that refined advises can be generated depending on which of the incorrect alternatives has been used. Of course this is only possible if there is a



clear mapping of specific misconceptions and incorrect alternatives. Teachers are encouraged to create items that allow for this kind of refined advises. Further diagnostic improvements may be obtained by gradually implementing AI techniques. Recent developments in essay evaluation may indicate that a valid judgement of this type of questions is feasible in the near future (2). Current trends in AI & Education seem to focus on diagnosis in the framework of simulation environments (3,4). For example in these environments monitor the students behaviour may be monitored in a simulated laboratory and the system may try to diagnose the students misconceptions by comparing predictions of the student with faulty models of the domain. Although this type of diagnosis is not possible for all domains of knowledge it may become a driving force behind the development of better exam validity.

#### Generative items

For a number of domains the large item pool may be replaced by algorithm that generates new items for each exam. Depending on the domain this process may use AI-techniques or may be implemented using standard programming techniques.

#### From Exams to Courseware

In the long run the system may stimulate the development of remedial modules that will be triggered either after the exam or even during the testing. Thus courseware is developed in two phases. This will spread the development costs not only over time but when offered through Internet the costs may spread over space too.

#### **References.**

- (1) Bierman, D.J., & Balk, V.A. (1990), A general purpose architecture for traditional CAI using an advanced authoring system. In: *Proceedings Applica '90* (pp. 1-8). Lille: 1990.
- (2) Noordwijk, K.van & Mulder, R.de (1996), Word use in legal texts: Statistical facts & practical applicability. Procs of JURIX 1996. pp. 91-100.
- (3) Koning, K. de (1997), Model-Based Reasoning about Learner behaviour. *Frontiers in Artificial Intelligence and Applications*, IOS Press, Amsterdam.
- (4) Bierman, D.J., P.A. Kamsteeg, J.A.C. Sandberg (1992), Student-models, scratch-pads and simulations. In: E. Costa, (red.); *New directions for Intelligent Tutoring Systems*, Springer Verlag, Berlijn, 1992, p. 135-145.