

SONATE, automatic look-over and analysis of tests

Piet van der Zanden

Delft University of Technology, Department of Technical Support

Mekelweg 6, 2628 CD Delft, The Netherlands

E-mail: A.H.W.vanderZanden@DTO.TUdelft.NL

Introduction

Since April 1999 a new system for automatic look-over and quality control of tests is operational in the Delft University of Technology. This system SONATE is developed to replace two old legacy systems. SONATE uses the benefits of those older systems and beats their disadvantages. SONATE is a flexible and user friendly Windows application. With SONATE the teachers themselves can determine the grade and item- and test-analysis so they get a clear insight whether the test meets their expectation, whether it was reliable and valid. This paper describes the history of development, the features of the application, the experiences until now and finally the plans for the future.

History

Until April 1999 two legacy systems were operational for automatic look-over of tests taken on paper. Both systems used optical mark readers. In 1997 those two systems were on the end of their technical lifecycle. The mainframe was to be dismantled and the optical mark readers started to make serious faults. After a desk research and extensive questioning among teachers in 1997 the benefits and disadvantages were gathered.

The system upgrade was done in two steps. In 1998 the optical mark reader was replaced by an image scanner with commercial reading and verifying software. The commercial Teleform software was tuned and tailor-made scripts were written for creating special examination answer-forms. These new answer-forms can be easily manipulated by the teacher - e.g. question numbers and alternatives can be covered up before being multiplied by a copier. The second step was to develop the grade determination and analysis software, which was released in April 1999. In June 1999 SONATE version 1.1 was released and version 1.2 in April 2000.

The features of SONATE

The new system consists of two parts. The central image scanning service of the multiple choice answer forms and the analysis software on the teachers PC. The teacher not only can determine the grade, also information is presented on the quality of the items and the test. Now the teachers themselves have the possibility to improve the quality of the test e.g. by skipping a dubious question.

The scanning of the forms

Various answer-forms are delivered with the software. They are based upon the answer-forms which were already in use. Answer-forms with two, four, five or eight alternatives are available. The four choices answer-forms also are available in a anti-fraud version. Before the teacher multiplies the answer forms he can fill in some data like name of the test, the date and codes. Also questions and alternatives which will not be used can be covered. Multiplying the forms can be done with every copier.

The filled in answer-forms, including a null-form, are scanned by the central scanner and converted to an electronic file. A null-form is the very same form as the batch only filled in by the teacher with the correct answers. After processing the file is sent to the teacher by E-mail.

The grade determination and analysis software

The software is developed according to the Microsoft Windows standard, it is flexible and user friendly. The program operates in four main steps which follow each other in a very intuitive way. When in one screen a value is changed all related parameters in the other screens immediately change as well.

The four steps are:

- 1) **Test Parameters** (see figure 1)

Points	1	1	1	1	1
Alternatives	4	4	4	4	4
A	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
D	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Points	1	1	1	1	1
Alternatives	4	4	4	4	4
A	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
D	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Points	1	1	1	1	1
Alternatives	4	4	4	4	4
A	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
B	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Selected Questions: 30 of 30 Difficulty Index: 0.88 Course Name: Course Code: Course Date: 11-01-2000
Selected Students: 64 of 64 Guessing Probability: 0.31% Example: AB1234
Maximum Allowable Score: 100% Standard Deviation: 4.00 Notes:
Maximum Attained Score: 90% Variance: 16.03 Pre-exam
Minimum Attained Score: 30% Dispersion: 0.13% 1 day after Halloween
Mean Attained Score: 59% KR20 factor: 0.64 It's freezing cold in the hall
Standard Measurement Error: 8%

Figure 1: SONATE; definition of the Test Parameters.

In the first step the file is loaded and the Test Parameters are defined. Data such as number of questions, the score per question, the number of alternatives and the number of versions can be altered afterwards. Immediately after the file is loaded the test-analysis is presented. The test-analysis is shown in every screen. The following parameters are presented:

- Minimum, Mean and Maximum Attained Score.
- Standard Measurement Error.
- Difficulty Index.
- Guessing Probability.
- Standard Deviation.
- Variance.
- Dispersion.
- Reliability KR-20 (reliability-estimator defined by Kuder-Richardson).

2) **Score to Grade Transformation** (see figure 2)

In the second step the teacher can determine the score to grade transformation. There are three possibilities: Piecewise Linear (3 pieces - recommended). Here three points have to be specified: the score for a grade of 1, a grade of 5,5 (Caesura or just-passed) and 10. Piecewise Linear (10 pieces) where for every grade the score is defined and Piecewise Constant where for every grade a score spread is defined.

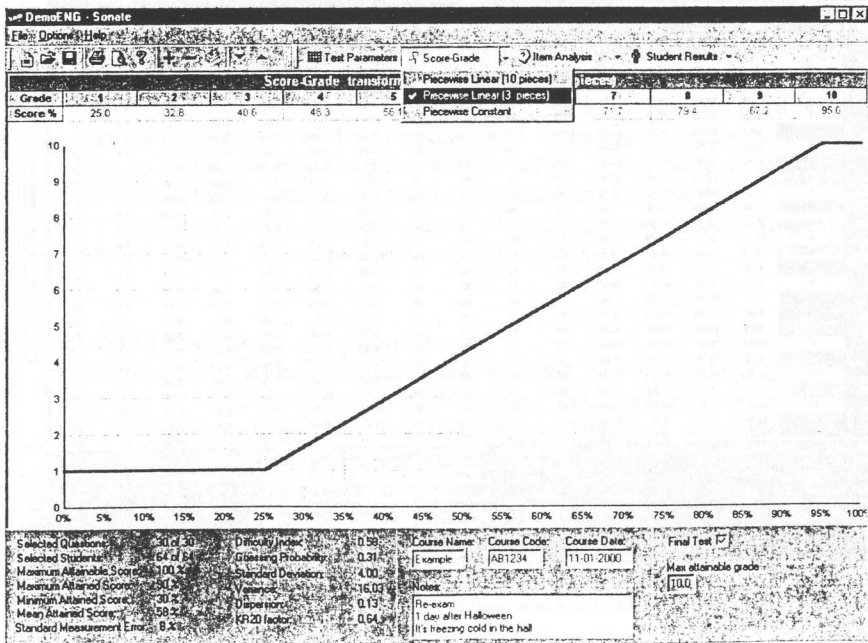


Figure 2: SONATE; the Score to Grade Definition.

3) Item Analysis (see figure 3)

In the third step the item-analysis results are presented. The following parameters are calculated and presented.

- The attraction of the alternatives: Index of Distracter or a-value.
- Mean score of the student per chosen alternative: M-value.
- Item Difficulty Index: p-value for each item or question separately and for the test as a whole.
- Corrected Difficulty Index: p' -value.
- Discrimination Index: Rit-value; how well does this item or question fit in this test (correlation of this item and the test including this item).
- Discrimination Index: Rir value; how well does this item or question fit in this test (correlation of this item and the test excluding this item).
- Guessing Probability, corrected for that item or question and the test as a whole.

Questi	Quality	A	B	C	D	Average score %	P	P'	Rit	Rir	Cp	Points			
1	---	1.56	82.18	6.25	0.00	30	68	58	0	0.92	0.90	0.12	0.06	0.38	1.0
2	---	6.25	1.56	86.63	1.56	44	37	68	30	0.91	0.88	0.43	0.37	0.33	1.0
3	---	28.13	46.31	14.06	12.50	57	64	46	56	0.45	0.27	0.36	0.24	0.29	1.0
4	---	10.94	9.38	7.81	78.31	50	46	50	62	0.70	0.60	0.40	0.29	0.26	1.0
5	---	10.94	10.94	78.31	7.81	47	52	62	50	0.70	0.60	0.42	0.32	0.26	1.0
6	---	9.38	9.38	20.31	68.94	59	44	53	62	0.61	0.48	0.37	0.26	0.29	1.0
7	---	4.69	54.89	9.38	31.26	51	62	51	66	0.31	0.08	-0.14	-0.25	0.38	1.0
8	---	4.69	9.38	10.94	76.08	53	48	53	61	0.75	0.67	0.30	0.19	0.28	1.0
9	---	1.56	6.25	14.06	48.44	47	54	43	68	0.84	0.79	0.29	0.20	0.28	1.0
10	---	23.44	12.50	14.06	48.44	55	60	68	57	0.14	-0.15	0.33	0.25	0.30	1.0
11	---	6.25	66.26	6.25	26.58	48	62	51	56	0.58	0.42	0.31	0.19	0.31	1.0
12	---	1.56	82.81	14.06	1.56	37	68	52	43	0.83	0.77	0.29	0.21	0.39	1.0
13	---	37.50	17.19	37.68	7.81	56	52	54	54	0.38	0.17	0.37	0.26	0.31	1.0
14	---	63.13	6.25	18.75	21.88	64	42	47	58	0.53	0.38	0.48	0.37	0.29	1.0
15	---	14.06	14.06	67.81	14.06	53	54	61	58	0.58	0.44	0.21	0.09	0.25	1.0
16	---	1.56	14.06	48.44	36.94	40	52	56	64	0.36	0.15	0.34	0.23	0.37	1.0
17	---	9.38	9.38	63.13	26.13	52	51	63	54	0.53	0.38	0.38	0.26	0.31	1.0
18	---	3.13	81.26	14.06	1.56	37	68	63	67	0.81	0.75	-0.02	-0.12	0.36	1.0
19	---	1.56	78.68	15.63	3.13	47	61	48	52	0.80	0.73	0.38	0.29	0.37	1.0
20	---	3.13	76.56	10.94	9.38	48	62	46	43	0.77	0.69	0.56	0.48	0.29	1.0
21	---	67.81	9.38	20.31	10.94	63	54	57	41	0.58	0.44	0.38	0.27	0.28	1.0
22	---	6.25	71.88	14.06	7.81	58	68	59	53	0.72	0.63	0.08	-0.03	0.29	1.0
23	---	9.38	18.75	48.44	23.44	57	47	67	51	0.48	0.31	0.61	0.52	0.28	1.0
24	---	21.88	20.31	17.19	49.63	50	62	51	64	0.41	0.21	0.36	0.24	0.26	1.0
25	---	6.25	23.44	66.26	14.06	47	55	62	53	0.56	0.42	0.35	0.23	0.29	1.0
26	---	67.81	7.81	1.56	32.61	68	59	50	55	0.58	0.44	0.17	0.05	0.37	1.0
27	---	34.38	11.25	28.56	7.81	63	63	61	42	0.14	0.13	0.26	0.15	0.30	1.0

Selected Questions: 30 of 30
 Item Difficulty Index: 0.59
 Missing Probability: 0.31
 Standard Deviation: 4.00
 Maximum Allowed Score: 101.13
 Mean Allowed Score: 61.03
 Minimum Allowed Score: 0.00
 Standard Measurement Error: 6.64

Course Name: Course Code: Course Date:
 Example: LAB1234 11-01-2000

Notes:
 Re-exam
 1 day after Halloween
 It's freezing cold in the hall

Figure 3: SONATE; the quality-value and item-analysis.

Based on these parameters a quality indicator is calculated. SONATE marks the items of the test with its own quality-value. This quality-value should be interpreted as a signal or attention caller, it is not an absolute measurement.

The evaluation of an item is based upon a combination of 4 different item-analysis values; the Corrected Difficulty Index (p'-value) (corrected with regard to guessing), the appeal of the alternatives (Index of Distracter a-value.), the Mean Score of the student per chosen item (M-value) and how well the item fits within this test (Rit-value). A perfect item scores 5 points. If a flaw is detected then SONATE assigns a penalty-point. In case an alternative for one item scores more than 90 % of the student-answers then no quality-value is calculated, instead a "---" will be shown. SONATE considers that item as a 'give away'. The quality-value is coloured for better recognition; grey is ok and red is bad. The teacher should carefully review the red(dish) items.

Penalty points are awarded as follows:

- A p'-value (Corrected Difficulty Index) less than 0.25 scores a penalty point.
- An alternative is not a good distracter. Each alternative that scores under a certain level gains a bad-mark. Depending on the p'-value the low-level is calculated as follows:
Is $(0 < p \leq 0.7)$ then the low-level is $\{0.3 / (2 * (\text{number of alternatives} - 1))\}$.
Is $(0.7 < p < 0.9)$ then the low-level is $\{(1 - p) / (2 * (\text{number of alternatives} - 1))\}$.
- The score of the alternatives (wrong answers) is compared to the mean score of the correct answer for that item. Their mean scores should be less than the mean score of that correct answer. In the case more correct answers are possible, the highest mean will be taken. If the score of the wrong answer is higher then the correct score a penalty point is given.
- In case the Rit-value drops under 0.29 then a penalty point is given. If it drops under 0.19 then two penalty points are given.

4) **Student Information** (see figures 4 and 5)

In the fourth step several summaries are presented. The Result-Matrix lists can be sorted by clicking on their column headers. Two graphs can be presented; the Student-Score Graph and the Student-Grade Graph. In the Student-Score Graph the caesura (just-passed indicator) can be changed by dragging it. The consequences of its change are presented directly.

All summaries and graphs can be printed (what you see is what you get). The summaries can be exported to an ASCII file in the same way the are presented.

Student Data										Result Matrix										
Student Index	Student Name	Course	Score	Grade	Correct	Answer	Points	Question	B	C	D	E	F	G	H	I	J	K	L	M
05278330	StudentName01	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
06062803	StudentName02	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
05012303	StudentName03	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
01060303	StudentName04	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
04242803	StudentName05	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
04242012	StudentName06	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
02942802	StudentName07	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
03436103	StudentName08	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
01840203	StudentName09	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
02518703	StudentName10	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
03350604	StudentName11	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
00102003	StudentName12	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
04243203	StudentName13	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
00084803	StudentName14	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
00701004	StudentName15	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
02003503	StudentName16	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
03789003	StudentName17	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
03212004	StudentName18	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
00000203	StudentName19	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
04240003	StudentName20	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
04283003	StudentName21	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
02720003	StudentName22	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
02720003	StudentName23	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M
02720003	StudentName24	ENGL101	85.00	B	10	10	10	1	B	C	D	E	F	G	H	I	J	K	L	M

Figure 2: SONATE; the Student results, the red cells in first column mean that a comment was placed on the answer-form by the student. The red cells with _ mean that no answer was given and the red cells with # mean that more then one alternative was filled in.

Experiences until now

SONATE has been used for over a year now. At this moment there has not been a large scale evaluation. It will be held in this academic year. But already something can be said about SONATE on the basis of individual contacts with users. Generally the responses are very positive. The users of the two legacy systems are pleased with the flexible and clear graphical presentation of the data and the adjustability of SONATE. The new users speak highly of the fast look-over process and the analysis tools for quality improvement. Although there has not been a large scale implementation, the number of users has more than doubled in the last year; from 33 to 84. About 11 Dutch universities have already shown their interest in SONATE.

It seems that almost every user highly appreciates the quality-value feedback. An item with a low quality-value is always reviewed and sometimes adapted. Even sometimes it is passed through to colleagues for reflection. The teachers say that they have skilled their insight in formulating questions.

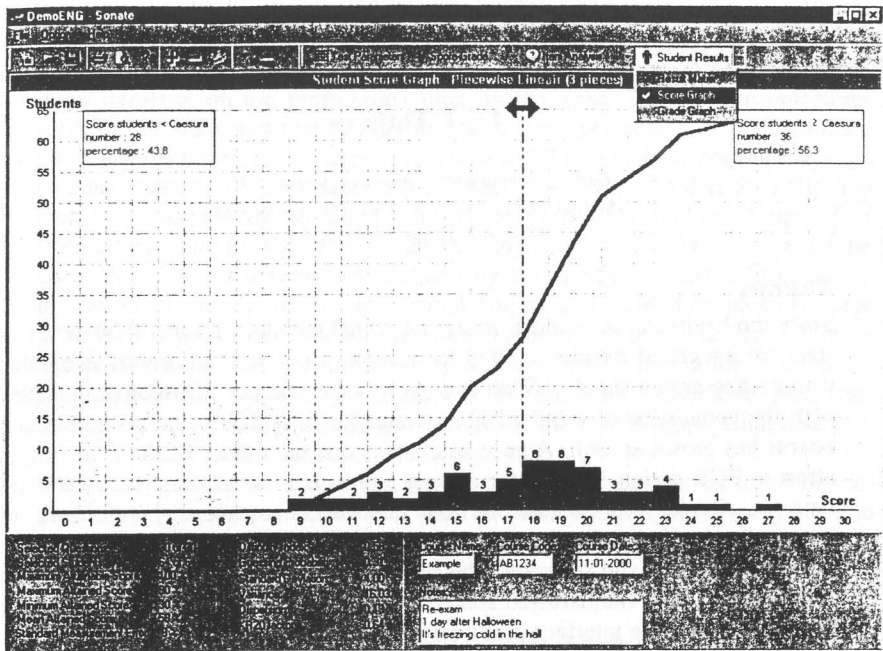


Figure 5: The SONATE screen with the Score Graph. The caesura (dashed line) can be dragged.

Plans for the future

The communication between the teachers and the software developers is good. Their feedback is very important for SONATE. Due to their feedback plans for version 2 of SONATE have been made. The plans include:

- A Test Management functionality for grading and analysing multiple tests at the same time. SONATE version 1 aims at single tests and version 2 at multiple tests.
- Personal student reports: The students want to know their study-results compared to their fellow students, e.g. the percentile scores for every separate test and the total of all tests.

SONATE version 1.3 is to be released by the end of 2000. Version 1.3 will have a more advanced import functionality for it can operate independently from our central scanning service and can import other formats, used by other universities.

References

- [1] Kunst S. and Van der Zanden A.H.W., "Toetsen en ICT in het hoger onderwijs", SURF Educatie (in Dutch), p. 29-35, July 2000.
- [2] Thio, "CITO Toetstechnische begrippenlijst", (in Dutch), 1988.
- [3] Dousma T., Horsten A., Brants J., "Tentamineren", Wolters-Noordhoff (in Dutch), 1997.