Notes from the Ninth International SweSAT Conference

Umeå, June 4 – 6, 2001

The SweSAT has been in existence since 1977, and it has by now become an integrated and generally accepted part of the Swedish educational system. An international scientific advisory board was constituted in 1992, and since then the board has met once a year, every other year in Sweden and the other year in connection with the AERA/NCME annual meeting. The first meeting was held in Umeå in May 1993 (Wedman & Stage, 1994).

This report is a condensed summary of the ninth meeting of the scientific advisory board, which took place in Umeå in June, 2001. A list of participants is enclosed as appendix.

The SweSAT Program since April 2000

Christina Stage

The number of test-takers has continued to decrease. Last fall the number was 38 500 which is 23 % less than in fall 1999. This spring the number was 46 600 which is 24 % less than last spring and only half of the maximum number (in 1992). The decrease depends on the number of 19-years old is especially low at the moment, but the number will rise again in a couple of years according to vital statistics. The number of test-takers is also influenced by the fact that the labor market has improved, so nowadays it is very easy to get a job, while it was difficult a couple of years ago. The number of applicants for higher education has decreased to the same degree. This means that the SweSAT program had a loss of SEK 6 millions year, and negotiations are going on how to finance the activities. Earlier the fees covered all the costs, and the National Agency for Higher Education has applied to the Government to raise the fee from SEK 300 to 350, but nothing has been decided yet.

A discussion, which has been going on during the last year, concerns unfairness against students from poor social background regarding admittance to higher education. There has been no increase in the number of students from lower social classes during the last decade, which is regarded as a big problem. Results published by Jan-Eric Gustafsson (2000) have also shown that the higher social groups are those who take advantage of SweSAT, especially the rule to repeat the test and in that way improve their scores and their chances of getting admitted to higher education. In a proposition from the Ministry of Education, it was suggested that, the share of students admitted on test results should be decreased from today between 30 and 40 per cent to between 10 and 25 per cent. In stead of using the best test result for selection an average should be used, for people with more than one result. This proposition was circulated for comments and I know that at least Umeå University (where Widar Henriksson took part in writing the comments) and The National Agency for Higher Education were against these suggestions. The decision has been postponed due to Sweden´s chairing the EU (no controversial decisions should be taken during this half year), but we are expecting the decision in September.

At the National Agency for Higher Education there is no person responsible for SweSAT at present. The post is vacant, and there are 40 applicants, but the selection has not been made yet.

Regarding our department we have had some staff problems, but now we have two newly employed test developers Jenny Lindberg who is working with DTM and Sandra Scott who is working with WORD. Stig Eriksson has been seriously ill since Easter, he is now improving and out of hospital, but he is on sick leave at least until fall.

The most positive news from the last year is that we, together with Gothenburg, have been granted a big research project on validation of the system for admittance to higher education (VALUTA). This project is financed by The Bank of Sweden Tercentenary Foundation, and the since the project is in many ways connected to the SweSAT program, the last part of this conference will be devoted to it.

Tests for two Disabled Groups – Persons with Defective Vision and Persons with Dyslexia.

Gunilla Ögren

Since fall 1999 there is a special test for persons with defective vision. This group of test-takers can only take the test twice, since there are only two versions of this test, and the tests are not made public after administration. There have been less than 10 test-takers each year. This test consists of 112 items distributed on four of the five sub-tests in SweSAT. The test is transferred to Braille and to cassette. It was not possible to transfer the sub-test DTM, so instead the number of items on the DS subtest was increased from 22 to 30.

For people with dyslexia two different test types have been tried. The first test was used in spring 1996, 1997, and 1998. It was a multimedia test (which means that the test was displayed on a screen at the same time as it was recited). The sub-test DTM was a regular paper test, since it could not be transferred to multimedia. All test-takers also got the other sub-tests in paper versions as well. Analyses of the results on this test showed that the test-takers got lower results on this multimedia test than they got on a regular test. Many of the test takers also had problems with the technology.

Since fall 1999 there is a new design on the test for the dyslexia group. The test is the same as for the regular test takers, but the testing time is 50 per cent longer and there are no pre-test items included. Since the start there has been about 400 test takers each year. To be allowed to take the special test a certificate of the handicap is needed.

Since many of these test takers have taken the regular test as well, it is possible to compare their results on the two test-versions.

Table 1. Differences in normed score between the regular form and the special form for those test-takers who have taken both versions.

Difference	Fall 1999	Spring 2000	Fall 2000	Spring 2001
+0.1	11	15	12	12
+0.2	9	23	12	15
+0.3	13	25	11	12
+0.4	5	22	10	13
+0.5	5	15	11	6
+0.6	-	11	6	3
+0.7	1	2	4	3
+0.8	-	1	1	3
+0.9	-	-	-	-
+1.0	-	1	-	1
N/Total	46/63	115/147	67/82	68/91

The Optical Scanning Procedure

Mats Hamrén

Since 1997 the Swe-SAT answer sheets have been scanned with a new technique, ICR (Intelligent Character Reading), on two high speed color scanners, *ImageTracs,* made in Birmingham, Alabama. These are the fastest scanners in the world that can take full color images. Our scanners, named Falcon 1 and 2, capture both a black and white and a color image with a speed of more than 100 pages (A4 size) per minute and scanner.

The *ImageTrac* scanner has a very useful function, color sense, which makes it possible to automatically send unwanted pages to the reject pocket. By recognizing special colors in certain areas of a page, the scanner can detect pages coming upside down, 180 degrees or an address page (yellow) coming in among the item pages.

We recognize the name and address from handwritten characters and for every marking box in an item we count the number of pixels (black points). For a test administration with 70 000 test-takers there are half a million pages and >13 million item responses, which means > 65 million marking boxes. In every item the "darkest" box is compared with the "next darkest" box and if the difference is big enough the darkest one is automatically chosen. Double marks are shown to the editors for decision by looking at the image. The software system for recognition used is PFP (Premier Forms Processor) made by Mitek in California.

The black/white image is used for recognition and the color image is used for archival on MO (magnet-optical) disks, certified for storage in 30 years. One MO disk (4.6 GB) can store 70 000 color images, which means that all images from the sheets of a spring test administration can be stored on 5-7 MO disks.

After each Swe-SAT administration we scan and edit (4 stations) for 16 hours per day. Thus half a million answer sheets from 70 000 test-takers will be scanned in seven days and the editing will be finished during day eight.

An Item-bank for WORD-items

Sandra Scott, Christina Jonsson

A demonstration was given of the computerized item-bank for the sub-test WORD. The bank is working according to Figure 1.

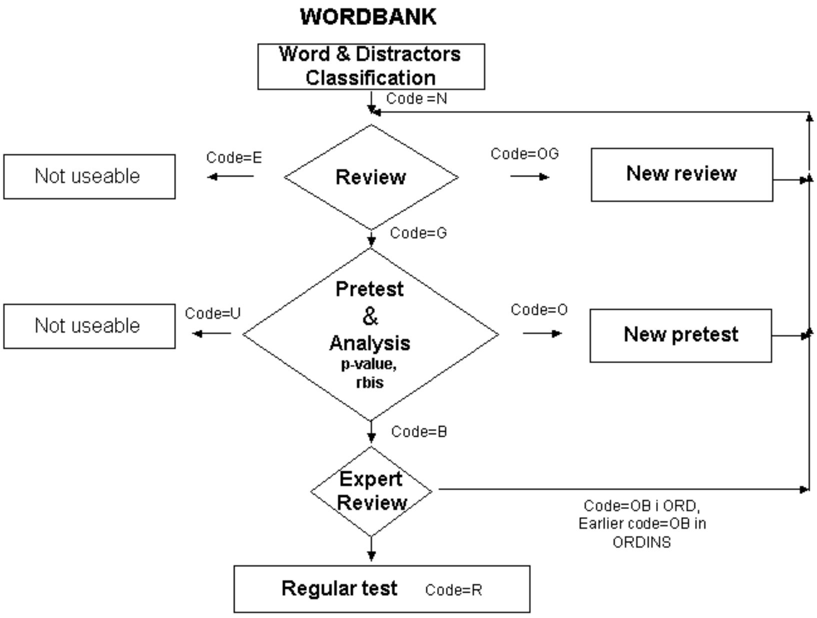

Figure 1. The WORD item-bank

Preliminary Results on a New Sub-test

Anders Lexelius

An analytical reasoning sub-test called VAN, for which no knowledge of formal logic or mathematics is required, has been developed and tried out on a small scale. A total of 160 students (62 males and 98 females) from grades two and three in upper secondary school have answered the items, but the try out will continue in order to get safer results. The results so far are shown in table 1.

Table 1. Results for different grades, programs and gender on the VAN and the DS sub-test.

	VAN		**DS**	
Grade/progr.	Males	Females	Males	Females
2/soc.sc.	8.07	8.35	4.93	4.24
3/nat. sc. tech	10.14	11.38	8.11	7.89
Total	9.24	9.15	6.73	5.12

The results so far seem to be promising regarding gender differences, but it is still too early to draw any conclusions. The work on this sub-test will continue.

Results on SweSAT for Different Social Groups

Christina Stage

During the year we have acquired a data base, which Simon Wolming will describe to you later. In this data base we have for the first time information on social-economic status for at least some of the test-takers and have made some preliminary analyses.

Table 1. Average grades and distribution on social groups for 82 379 students who finished secondary school in 1989

Social group	I	II	III
%	22	50	28
GPA	3.62	3.36	3.11

Of these students 29 959 finished upper secondary school after two years, and their average grades and distribution on social groups is shown in table 2.

Table 2. Average grades after secondary and upper secondary school for the students who chose two years upper secondary school, distributed on social groups.

Social group	I	II	III
%	10	49	41
GPA sec sch.	3.03	2.98	2.89
GPA upper sec.	3.22	3.16	3.05

Of the original students, 46 529 finished upper secondary school after three years, and their average grades and distribution on social groups is shown in table 3.

Table 3. Average grades after secondary and upper secondary school for the students who chose three years upper secondary school, distributed on social groups.

Social group	1	2	3
%	26	51	23
GPA sec. sch.	3.76	3.57	3.29
GPA upper sec.	3.54	3.32	3.15

And the remaining 6000 students left upper secondary school four years after leaving secondary school. Their average grades are shown in table 4.

Table 4. Average grades after secondary and after upper secondary school for the students who left upper secondary school four years after finishing secondary school, distributed on social groups.

Social group	1	2	3
%	36	48	16
GPA sec. sch.	3.74	3.65	3.55
GPA upper sec.	3.53	3.30	3.20

Out of the students who finished secondary school in 1989 a number of 18 892 (23 %) took the SweSAT in spring 1992, and their grades and test results are presented in table 5.

Table 5. Results for different groups on SweSAT spring 1992.

Social group	I	II	III
%	33	51	16
GPA (89)	3.90	3.80	3.73
GPA (92)	3.67	3.51	3.44
Test score	99.30	92.90	88.23
Normed score	1.05	0.89	0.78
WORD	18.75	17.26	16.21
DS	13.21	12.15	11.28
READ	16.00	15.09	14.50
DTM	15.11	14.46	13.79
GI	18.14	17.18	16.51
ERC	18.09	16.75	16.06

The social background of students seem to have a rather important influence on their educational achievement regardless of whether the achievement is measured by grades or by test results. These studies will continue with comparisons on item level.

Repeated Test Taking

Widar Henriksson & Birgitta Törnkvist

When SweSAT is used for selection to higher education there are certain rules which are relevant. These rules are:

- An individual can repeat SweSAT as many times as he/she wants
- An obtained SweSAT score is valid for five years
- If a test-taker has more than one valid score from the SweSAT, the best obtained score (normed score) is used for selection
- An applicant is selected either on the basis of the SweSAT score or on the basis of GPA.

A two-year period is used as a basis for this description - spring 1996 (96A) to fall 1997 (97B). The total population at a certain test administration (97B) is divided into four sub-populations: 0, 1, 2, 3. The number (0 – 3) refers to the number of tests taken before this test administration (97B).

Presentation of the VALUTA Project

Christina Stage, Jan-Eric Gustafsson

The VALUTA project, which started in January this year, is an integrated research program with researchers from Gothenburg and Umeå. VALUTA is an acronym for ”Validation of the System for Admission to Higher Education”, and the project is financed by The Bank of Sweden Tercentenary Foundation.

The aim of the project is to illuminate different aspects of the system for admission to higher education in Sweden. We intend to analyze all components of the admission system and the interaction between them. The frame of reference for the project will be the broadened perspective on validity emanating from Messick (1989).

The components, which have been identified as important are:

1. The selection instruments, i.e. final grades from upper secondary school and SweSAT scores.
2. The criterion, i.e. study success.
3. The selection instruments in relation to the criterion.
4. The interested parties in higher education, students, university institutions, authorities etc.
5. The regulatory system for admission.

The project is divided into sub-projects of which the following are located in Umeå:

Group differences on SweSAT results

Traditional and non-traditional criteria of study success

Examination of two study programs within the Faculty of Social Sciences

Different interested parties and their opinion on the admission system

The effects on selection of repeated test taking

The effects on selection of the possibility of improvement of entrance qualifications

The effects of quota groups on the admission system

And the following sub-projects are loacated in Gothenburg:

The internal structure of the SweSAT

SweSAT in relation to other tests of cognitive ability

Measurement properties of the goal related grades

Properties of some special entrance tests

Examination of study programs within the Engineering sector and the Teachers Training sector.

And the, perhaps most important, sub-project, in which both Gothenburg and Umeå will take part is an integration of the results from the different sub-projects within the framework of Messick´s validity concepts.

Who is a Successful Student? - A Pilot Study

Ewa Andersson

In order to investigate whether the meaning of study success is the same for different interested parties in higher education an interview study is in progress. As teachers indeed may be seen as interested parties to higher education, teachers at the Department of Social Welfare and the Department of Business Administration were chosen.

An interview guide was developed containing different criteria of study success, basically corresponding to cognitive and non-cognitive criteria (10 of each). However, some views of study success seem to include both cognitive and non-cognitive criteria, and therefore the list finally came to comprise 24 criteria of study success.

At the interviews conducted so far the teachers were asked about how they view success in their freshmen as well as their senior students of today and of the future. The teachers were instructed to select freely, among all 24 criteria presented to them, which criteria they found to be relevant for study success. They could select as many or as few criteria they wanted. Furthermore they were asked to select and to rank the five most important criteria among the criteria they had chosen.

The interviews with the teachers also deal with the issue of whether it is possible to measure success according to the criteria they have chosen. The preliminary results show that almost all teachers could provide an example of how to measure the criteria they had chosen. The response rate for almost all criteria chosen is over 80 per cent. Examples of ways to measure a specific criteria is of course written exams but also working papers (pm), written reviews of other students papers, different forms of oral presentations but also seminars.

Furthermore, I asked the teachers whether they consider grades from upper secondary school and SweSAT scores have the ability to select applicants who will be successful in their studies according to each of the chosen criteria. The teachers were also asked what selection instruments they wish for in the future. The data has not yet been analyzed, but my impression from the interviews is that most teachers would like to have special selection procedures as a complement to the selection instruments used at the present.

How do the interested parties of higher education, view the admission system? - A planned study

Ewa Andersson

Since it is vital that the admission system to higher education has legitimacy with its interested parties, it is important to investigate how they view the admission system, especially since the system, from different perspectives, is criticized for not meeting the stated demands. The regulatory system is criticized for being ineffective. The instruments that are used today for selection are charged of not capturing the variety of different competencies and talents of the applicants, which the critics fear might lead to a biased recruitment concerning social background, sex and ethnic minorities.

However, the basic problem is that today there is no assembled knowledge in Sweden of how different interested parties view the issue, i.e. where they differ in their opinions or in what areas they possess a common opinion. Furthermore how different views are related or consistent with the aim(s) of higher education.

The aim of the planned study is consequently to investigate how different interested parties view the admission system of higher education with reference to the regulatory system, the selection instruments and the meaning of study success and how these views are related to the aims of higher education.

Table 1 is a summary of a suggestion to how such a study might be designed concerning who the interested parties to the higher education in Sweden might be, choice of participants and the methods I plan to use.

Table 1. A summary of a planned study of how different interested parties view the admission system to higher education in Sweden

Interested parties identified	**Method**
Applicants to different education programs: *(Admitted students and students who are rejected)*	Questionnaire Interviews
Representatives of the universities and different education programs:*(Vice-chancellors, directors of studies, and teachers)*	Questionnaire Interviews
Representatives of the authorities responsible for higher education: (*National Admissions Office to Higher Education; National Agency for Higher Education)*	Interviews
Representatives for the education policy: *(politicians)*	Interviews
Representatives for the industry, commerce and working life	Interviews

I realize that this is a large study, and that it will be necessary to restrict the study without loosing the perspective of getting an overall picture of the issue. Of importance for how to interpret and value the results is also the delicate question of how the selection of participants is made *within* the different groups of interested parties.

Another question to pose, in this line of work, is who could (or should) be defined as interested parties to the admission system. The concept *interested party* has been chosen because I'd like the study to include all groups that are concerned by higher education. Therefore I don't think that the concepts *consumers* or *beneficiaries* to higher education are sufficient since the aim is to reach a comprehensive view of the issue.

The New Data Base in Umeå

Simon Wolming

Information is obtained from the following databases:

- National student records database (LADOK)
- Tha National Board of Student Aid (CSN)
- The SweSAT database, Department of Educational Measurement
- Primary and Secondary School, The National Agency for Education/Statistics Sweden
- Population and Housing Census, Statistics Sweden
- Total Population Register, Statistics Sweden

Primary school

Graduate class 1988	110 661
Graduate class 1993	97 517
Secondary school:	
Graduate class 1990	93 736
Graduate class 1991	94 019
Graduate class 1992	95 787
All students i secondary school in fall 1994	309 951

Higher education:

All registered students 1993 – 1999	710 852
Swedish students studying abroad	75 182

Available information:

Year of birth

Sex

All grades from primary school

Secondary school programme

Selected grades from secondary school (Math, English, Swedish, Social sc.)

Secondary school GPA, based on all grades

SwSAT results

Socioeconomic background (education, annual income, profession, civil status, citizenship, native country, immigration, housing, employment)

The Paradox of Study Success

Simon Wolming

Instruments used for selection to higher education have traditionally been evaluated on the basis of their predictive validity as the only indicator of validity. Thousands of validity studies have been carried out over the years, mainly focusing on the predictive validity of selection instruments. However, there are several methodological and theoretical problems related to this type of approach. One of them regards the criteria of study success or, more precise, the variation in achievement between students. A methodological requirement is that there must be some variation within the variable used as the criterion of study success. If no variation exists, the result of a correlation will be zero.

It is well known that the variation in study results could be restricted in range by several different mechanisms. This implies that the traditional way of estimating predictive validity leads to a severe underestimation of the predictive validity of the selection instrument. For example, if a group of students all show top results, and these results are used as evaluation criteria for a predictive validity study, the result of the study could be interpreted as if the instrument does not have any predictive validity at all. On the basis of the same top results the instrument may, however, be regarded as perfect—not in terms of its predictive power, but in terms of the students' actual achievement. Therefore, it is somewhat surprising that the predictive validity of a selection instrument often is evaluated on the basis of graduation rate or the rate of the students' throughput. This situation could be called the paradox of study success.

There are several mechanisms that may influence students' achievements. For example, the National Board of Student Aid requires students to complete 75 per cent of the courses during a year in order to receive further financial support. Another aspect is the grading system. Often study programmes divide the students and their achievements between the categories "pass" and "fail" only. Yet another aspect is the fact that students quite often interrupt their studies. These interruptions could have different reasons, e.g. change of study direction, employment, parenthood, studies abroad, etc. It is important to notice that such reasons do not have anything to do with poor achievement. Nevertheless, they are often interpreted in that way. All of the mechanisms mentioned contribute to the problem of range restriction within the criteria. The lack of variation makes it difficult to make valid interpretations of the predictive validity of a selection instrument.

It should also be mentioned that there are different statistical solutions to the problem of range restriction. However, even if the problems with range restriction can be solved and a "true" relation may be shown, the predictions made will always be afflicted with errors. As already mentioned, students' results are affected by mechanisms that cannot be related to the things that the instrument is supposed to measure.

Selection to higher education is not only a matter of prediction. It could also be related to questions of fairness and relevance. Validity investigations must therefore be based upon a much broader perspective. For example, selection instruments must not be biased in any direction so that their use has a negative impact on the grounds of gender or ethnicity. Instruments must also be evaluated in terms of content. In this context, the views among students and within study programmes are crucial.

To sum up, the evaluation of a selection instrument must be based on a variety of approaches. Several mechanisms affect the possibilities to make a valid interpretation of the relation between the instrument and students' subsequent success. Thus, evidence of predictive validity is not alone sufficient as evidence of the value of a selection instrument.

Prognostic Validity under Compensatory Selection Rules

Sven-Eric Reuterberg

The classic problem in determining prognostic validity is that the criterion is known only for those persons who are admitted, and they constitute a selected sample of all applicants. The fact that restriction of range in test scores implies decreased relations with other variables is a well-known phenomenon, and standards procedures are available to solve the problem.

Admittance to higher education in Sweden, however, is a more complicated procedure due to the fact that the applicants are admitted either on their SweSAT scores or on the basis of their leaving certificates from upper-secondary school. This means that the admission procedure is compensatory, and therefore the problem of estimating the correct prognostic validity is not only a matter of restriction of range.

Of course the prognostic validity of a selection test is not a question of its relationship with those who are admitted, but its relationship with the criterion among all the applicants. Thus the problem is that those not admitted are missing information on the criterion variable. However, in admission to higher education all applicants have leaving certificates from upper secondary school and a great majority also have SweSAT scores. By a new statistical method which models incomplete data these new variables make it possible to estimate what criterion scores all applicants should obtain if they had been admitted.

This statistical method has been applied to a simulated data set and new correlation coefficients between SweSAT scores and the criterion have been computed. The actual correlation between the SweSAT scores and the criterion among all applicants amounted to 0.45, so this is the correct prognostic validity coefficient.

Results from a Study of Prognostic Validity of the SweSAT for Civil Engineers

Allan Svensson

The aim of the study was to answer the following questions:

1. What is the prognostic validity of the SweSAT in higher technical education?
2. What differences are there in prognostic validity between SweSAT and school marks (GPA) from upper secondary school?
3. Does the prognostic validity vary between different sub-tests in the SweSAT?
4. Are there any differences between students who started in 1993, 1994, 1995 and 1996?

Table 1. The distribution of students according to year of birth. Per cent.

year	1972	1973	1974	1975	1976	1977	1978	%	N
1993	21	30	48	1				100	3495
1994	11	18	31	39	1			100	4094
1995	6	11	17	30	36	1		100	4736
1996	3	5	9	18	33	31	1	100	4873

Table 2. Marks from upper secondary school. SweSAT scores and number of credits obtained in higher technical education among the students in the investigation.

	school marks		first test		max test		credits	
	M	s	M	s	M	s	M	s
1993	4.07	0.43	1.25	0.36	1.38	0.36	27.8	11.4
1994	4.00	0.46	1.19	0.36	1.34	0.37	26.4	12.9
1995	4.00	0.48	1.17	0.39	1.32	0.40	28.4	11.8
1996	4.03	0.49	1.16	0.40	1.32	0.40	29.2	11.7

Table 3. Marks from upper secondary school and SweSAT scores for all students born between 1972 and 1977 who have completed the science or technical programs.

year of birth	N	school marks		first test[1]		max test	
		M	s	M	s	M	s
72-74	51372	3.51	0.67	1.07	0.40	1.22	0.42
72-75	67604	3.52	0.67	1.06	0.41	1.21	0.43
72-76	82760	3.54	0.67	1.05	0.41	1.20	0.43
72-77	96551	3.56	0.66	1.05	0.42	1.19	0.43

[1] SweSAT scores are available for 75% of all students who have completed the S- or T-program

Table 4. Correlations of school marks and SweSAT scores with credits obtained after the first academic year. Corrected coefficients.

Adm. year	school marks			first test			max test		
	M	F	total	M	F	total	M	F	total
1993	.49	.52	.50	.19	.25	.20	.12	.17	.13
1994	.47	.39	.45	.19	.17	.19	.15	.13	.14
1995	.38	.33	.36	.10	.12	.10	.06	.08	.06
1996	.36	.35	.35	.14	.15	.14	.14	.12	.13

Table 5. Correlations of sub-test scores and credits obtained after one academic year. Corrected scores.

Adm.y	DS	DTM	READ	WORD	GI	ERC
1993	.31	.27	.19	.10	.15	.13
1994	.26	.24	.15	.10	.16	.15
1995	.21	.25	.13	.01	.07	.05
1996	.23	.28	.19	.02	.08	.13

Table 6. Correlations between credits in higher education and results on the quantitative and verbal subtests of the SweSAT. Corrected coefficients.

Adm. year	quantitative sub-tests			verbal sub-tests		
	M	F	total	M	F	total
1993	.33	.36	.33	.13	.20	.16
1994	.30	.27	.28	.15	.13	.15
1995	.28	.27	.26	.07	.08	.07
1996	.28	.30	.27	.11	.11	.11

Separating quantitative and reasoning ability in the SweSAT

Lisbeth Åberg-Bengtsson

As pointed out by some approaches within psychometric research no measurement device such as the SweSAT ever measures just one single ability. On the contrary, performance may be assumed to be influenced by variance from a number of sources at different levels of a hierarchical structure. By using powerful multivariate techniques that have been developed during the last decades, the complex of observed variables in tests of this kind can be decomposed into latent variables and their dimensionality investigated. Such an approach is particularly important when differences in performance between groups of test-takers are investigated.

Previous studies of dimensions of performance in the SweSAT have convincingly argued for two factors, namely one factor which measures general ability corresponding to G or Gf´, a second which measures "knowledge" or Gc´. It has been claimed that the DTM subtest and DS subtest measure, to a great extent, general ability (e.g., Gustafsson, Wedman, & Westerlund, 1992). However, this claim seems somewhat puzzling as these two tests strongly contribute to the gender difference in performance in the SweSAT. Usually, such a difference is not found in tests measuring general analytic ability.

In a previous study (Åberg-Bengtsson, 1999), initially using the test administered in April 1991 (91A), and later replicating on the test from April 1992 (92A) with test takers born in 1972 and 1973 respectively, a "quantitative" factor was identified, in addition to a general DTM factor and "end of test" effect. This quantitative factor was related to items where calculations of one kind or another were to be used, in addition to the reading off of values or finding tendencies etc. in the graphical displays. Scrutinizing the results, it was argued that whereas, in the general DTM factor, the difference between males and females was of relatively limited size and could be explained, for example, by selection effects, the deviation between the sexes in the "quantitative" factor may be regarded as a substantial contributor to the problematic part of the gender difference on the DTM subtest.

In a current follow up performed within the VALUTA project, attempts are being made to separate the "quantitative" factor from the reasoning or analytic ability not only with respect to the DTM subtest but also related to the SweSAT and its dimensionality. At the seminar in Umeå preliminary results from investigations on test version 91A were presented. A structural equation modeling technique is being applied to data organized on item level for the DTM test and organized on a subtest level (sum of scores) for other five subtests of the SweSAT. Three different ways of modeling have been adopted using nested factor models, oblique models, and higher order (HO) models respectively. Subsequently, in all models the sex of the subjects have been included as an independent manifest variable related to identified latent variables or factors.

In the nested factor models a *general* factor, related to all DTM items and all subtests, was identified in addition to a "*quantitative*" factor. The DS subtest as such, as well as those items of the DTM test previously identified as being "quantitative" have loadings on this "quantitative" factor. A third factor related to the four subtests GI (general information), STECH (study technique), WORD, and READ was also found. This latent variable was tentatively labeled "knowledge", although, in accordance with the arguments above, it must be hypothesized to be multi-dimensional and may, for instance, include different verbal abilities. Due to modeling technical considerations the DTM "end of test" factor was also included in the models.

When estimating the oblique model corresponding to the nested factor model described above, the general factor was excluded and successfully replaced by an "analytic" factor related to all "non-quantitative" items of the DTM subtest. Allowing the three latent variables "quantitative", "analytic", and "knowledge" to covary yielded the highest correlation (r=0.91) between the two first factors.

The conventional way of formulating a hierarchical model would be to use so called higher-order modeling. An HO-order model was achieved by reintroducing the general factor, now related to "quantitative", "analytic", and "knowledge" factors respectively, at a higher level (the apex of the hierarchical structure). It was found that the "analytic" factor was loaded by the general factor with an estimated standardized factor loading of 0.98.

When regressing the factors identified on "sex" as an independent manifest variable by far the greatest deviation in favor of males was identified for the "quantitative" factor, related to which standardized factor loadings between 0.47-0.57 were obtained. These values may be compared to loadings of 0.27 on the general and analytic factors. In the latter cases, the deviations are of such a limited magnitude that they might be explained by, for example, selection effects, whereas the more substantial deviation with respect to the "quantitative" factor must have additional explanations.

Finally, it should be emphasized, that even though the "quantitative" factor seems stable and well identified in the results presented, further investigations and replications on later test versions are necessary. In addition to validations that have been begun on tests administered in 1992 through 1995, a most urgent next step is to include grades from the leaving certificate of compulsory school in the models in order to gain a better identification of the "quantitative" factor. Other studies to be considered are investigations of influences from different courses accomplished by the test takers in their upper secondary education that may be hypothesized.

Mixture growth modeling of achievement over time in higher education

Jan-Eric Gustafsson

The study of individual differences in change offers several well-known methodological problems, such as unreliable difference scores and a negative correlation between initial status and change. These problems were for a long time seen as insurmountable and caused many researchers to refrain from studying individual differences in growth. Recently, however, new tools have been developed within the framework of latent variable modeling, which not only solve the old methodological problems but which also contribute new powerful methods for studying correlates, antecedents and consequences of individual differences in change. In these models, latent variables are used to represent parameters such as slope, intercept and curvature of a growth function. Instead of estimating individual values of these parameters, parameters describing the distribution of individual values (i. e., mean and variance) are estimated in a random coefficient modeling approach. Because the parameters of the growth function are represented as latent variables these may be related as dependent or independent variables to other latent or manifest variables in a general structural equations model. This modeling approach may also be extended into mixture growth modeling in which different growth functions are identified for different subsets of the cases. Mixture growth modeling thus segments the sample into different latent classes and fits a separate growth model for each class.

The basic principles of latent variable growth modeling are presented, and the methodology is illustrated in models of data from achievement 3 471 students admitted to the program in 1993 over four years of study in the Civil Engineer program. In the Swedish system of higher education achievement is recorded in a system of credit points (CP), of which the student is expected to earn 40 in an academic year. Not all students achieve 40 CP each year, however, In the paper it is demonstrated how individual differences in achievement over time may be captured by a linear growth model in which an intercept parameter represents individual differences during the first year of study, and a slope parameter represents linear change over the years of study. It is also shown that these parameters to a fairly large extent may be predicted from the selection instruments to higher education (i. e., grades from upper secondary school and the Swedish Scholastic Aptitude (SweSAT) test). In a regression model the intercept parameter thus is predicted with a positive weight by grades and by a reasoning factor derived from the 6 subtests of the SweSAT, while a verbal SweSAT factor had a negative weight in the regression equation. For the slope parameter, however, a reverse pattern was found with negative coefficients for grades and reasoning and a positive coefficient for the verbal factor.

These results show that the Civil Engineer program makes different cognitive demands during the first and the ensuing years. The first year of study makes heavy demands on learning mathematics and basic science, while the following years are more applied and involves more of reading and writing in wider areas. It is, thus, a reasonable interpretation that the different patterns of correlation with the slope and the intercept parameter are due to changing cognitive demands of the Civil Engineer program

In the next step of analysis of the data it is shown that the growth model fitted for the total group of cases may be split into two separate models, for two different groups of cases. One model, which applies to about one third of the cases, has a mean of the intercept parameter close to 40 and a slope parameter close to zero. This is, thus, a consistently high-performing group. The other group has a mean of the intercept parameter close to 20 and with a mean of the slope parameter of 2. In this group the variances of the slope and the intercept parameters are, furthermore, substantial. This group thus evidences a substantial variability both of first-year performance and of later-year performance.

For the consistently high-performing group there are no individual differences, so if we can predict belongingness to this group we achieve all the prediction that is possible. The best predictor of membership to this group proved to be grades from upper secondary school, the group of cases belonging to this latent class having a higher mean grade than the other cases. For the other latent class a similar pattern of predictor relations as was found for the total group of cases was established.

In the discussion of the empirical pattern of results it is emphasized that the changing pattern of relations with achievement over time must be taken into account when methods for selection of students are being discussed. However, most of the research on selection to higher education has concentrated on first-year performance. It is also discussed how the present results may be used to change the Civil Engineer program.

Participants:

Michal Beller, Israel
Linda Cook, USA
Ronald K. Hambleton, USA
Wim van der Linden, The Netherlands

Christina Cliffordsson, VALUTA, Gothenburg
Jan-Eric Gustafsson, SweSAT, VALUTA, Gothenburg
Sven-Eric Reuterberg, SweSAT, VALUTA, Gothenburg
Allan Svensson, SweSAT, VALUTA, Gothenburg
Lisbeth Åberg-Bengtsson, VALUTA, Gothenburg

Ewa Andersson, VALUTA, Umeå
Mats Hamrén, SweSAT, VALUTA, Umeå
Widar Henriksson, SweSAT, VALUTA, Umeå
Christina Jonsson, SweSAT, Umeå
Anders Lexelius, SweSAT, Umeå
Jenny Lind, SweSAT, Umeå
Sandra Scott, SweSAT, Umeå
Christina Stage, SweSAT, VALUTA, Umeå
Birgitta Törnkvist, VALUTA, Umeå
Simon Wolming, VALUTA, Umeå
Gunilla Ögren, SweSAT, Umeå