Notes From The Seventh International Swe-SAT Conference

Umeå, June 3 –5, 1999

Christina Stage

The SweSAT has by now been in existence for more than 20 years and has become an integrated and generally accepted part of the Swedish educational system. The International Scientific Advisory Board, which was constituted in 1992, was meeting for the seventh time in Umeå in June 1999.

This report is a summary of the seventh meeting. The main topics for the meeting were: test validity, composition and scoring of SweSAT, and test fairness regarding immigrants. The program for the conference as well as a list of participants are enclosed.

The SweSAT Program since April 1998

Christina Stage

During the last year the number of testtakers has decreased by 7000 in spring and by 5000 in fall. The reason for this decrease is mainly that the number of young people in the proper ages has decreased, but also that the labour market has improved so it is easier for young people to get jobs.

Fall 1998 I was working on a research grant from the Swedish Council for Research in the Humanities and Social Sciences. Gunilla Ögren then took over the project leadership and I am only scientific leader.

The new pretesting procedure which was introduced in 1996 on a two years trial basis, and then prolonged for one year, has now been prolonged for another period of two years.

The cooperation with Twente university has resulted in a master thesis on IRT equating of SweSAT results. We now use the method as a complement to the usual way of equating. What is needed, however, is a better design for composition of the subtests, which takes into account the need for common items which have not been changed between pretest and regular test. We also need more information about how many common items that are needed for the IRT equating.

An issue which has been discussed a lot during the last year is cheating on the test. We still have not solved the problem how to handle testtakers who are suspected cheaters. This year a student has been accepted to medical studies (the most competitive education in Sweden) who is not able to read and write Swedish properly. He had got top results on the test on his third test occasion, but when checked by a

legal graphologist there are with great certainty three different persons who have performed the test in the same name. This is the first case which is going to trial and we are eagerly waiting for the outcome.

Admittance to higher education in Sweden is centralized with the same rules for all universities and colleges. The main rule is that you are qualified for higher education on the basis of grades from upper secondary school. Among those who are qualified selection is made on the basis of average grades OR test results.

It is possible, however, for educations to get permission to do selection by other means. They have to apply to the National Agency for Higher Education and if the arguments for special selection are good enough, permission is granted. So far it is mainly medical and journalistic educations who have their own means of selection. These are both educations with many applicants and good reasources.

The SweSAT, however, has got an increasing importance for admittance to higher education. Since the grading system has been changed from norm related to criterion related, it is less suitable for selection, which in turn means that the importance of the test increases.

This increased importance makes heavy demands on the test. As you already know we have big problems in examining the predictive validity of SweSAT, mainly because of problems with the criterion. This is a topic of great concern since it is claimed that the test is of no use for technical educations since the subtests measuring quantitative abilities are too few and too easy. On the other hand advocates of the faculty of arts claim that the test is unfavourable to their students.

The increased importance of the test as selection instrument has also initiated a discussion on changes/improvements of the composition and the scoring of the test. The National Agency for Higher Education has established an Advisory Board on matters having to do with tests. This Board in turn has appointed a group which should recommend necessary changes of the test. Jan-Eric and I are both members of this group.

Sweden used to be a very homogenous society with one language and the same education for all. This is not true any more, since Sweden has been the recipient of a lot of refugees and immigrants. This group is underrepresented in higher education and the test among other things is probably a hindrance in this regard since it is adapted to people with

Swedish language and Swedish cultural experiences. This is a vey new problem and so far we do not even know how big the problem is, and still less about how to handle it. At present the problem is mainly in upper secondary school, since a big proportion of these groups never finishes upper secondary school with a proper examination. The problem will soon reach us, however, and we would like to be prepared how to handle it.

Various Admission Models and their Pros and Cons.

Michal Beller

Distributing Rare resources:

- Free market
- Equal probablity (e.g., lottery)
- Ranking by seniority (e.g., waiting time)
- Meritocratic approach

Selection to Higher Education: Models

No quota

- Open admissions
- Setting qualitative minimal cut-off points pre studies
- **A pre set quota**
- equal probability - lottery
- merit – setting quantitative cut-off points
  - during studies
  - pre studies

Pre set quota: Merit approach

- **During studies**
  Ranking on the basis of FGPA (fierce competition among students)
- **Pre-studies**
- HSR
- Special achievement tests
- General scholastic aptitude test
- Specific departmental admissions tests
- Previous experience
- Personality inventories

- Interviews
- Combination of the above

The Israeli Admissions Policy

- **Meritocratic Approach**
- **Based mainly on:**
  - HSR (Bagrut)
  - Scholastic aptitude test (PET)
  - Adm = Bagrut + PET (50-50; 40-60; 60-40)
- **Reasons for using a composite score (Adm):**
  - more valid
  - compensatory model
  - gender/group ”equalizer”

Evaluation of the system

- Reliability
- Validity (construct, predictive, face, consequential)
- Culture fairness (group and gender bias)
- Utility – cost-effectiveness

Establishing the Validity of Admission Tests[1]

Ronald Hambleton

The Graduate Record Examination:

What is the Validity Evidence?

Number of Graduate/Professional School Predicitve Validity Studies (in the last 10 years)

1566 + 1 = 1567

Topics

1. Methodological Difficulties

2. GRE Predictive Validity Evidence

3. Additional Validity Evidence

4. Future Research

1. Methodological Difficulties With Predictive Validity (PV) Studies

a. PV Studies Are Rarely Conducted on the Population (or Sample) of Interest – Range Restriction Results.

b. Range Restriction on the Criterion

c. Unreliability of Criterion Scores

d. Inadequacy of the Criterion Variable

e. Use of a Second Predictor Variable or More (Use of a Compensatory Selection Model)

f. Generally Small Samples

[1] Reproduction of trancparancies used for the presentation of a paper by R.K. Hambleton & K. Meara.

Figure 1. Scatterplot Between First-Year Grades and Admission Test Scores, r =.20.

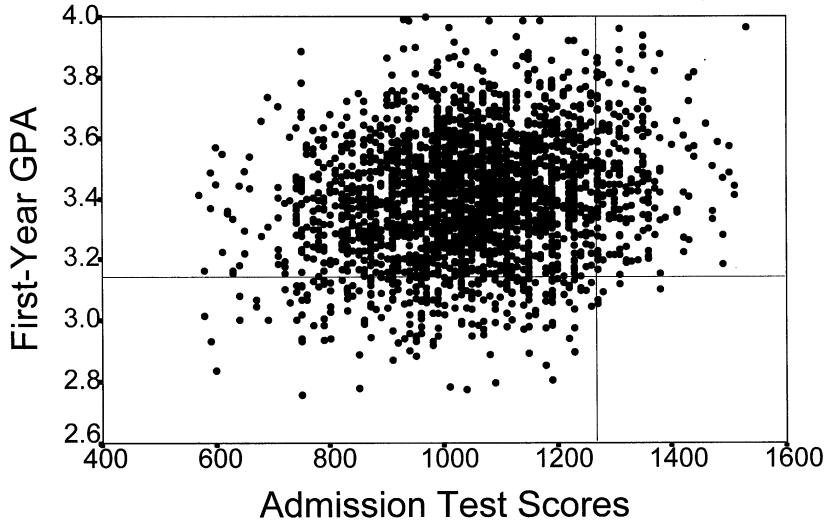

Table 1. *r = .20. Correlations, Standard Deviations, and Positive Hits*

	Percent of Students Selected				
Stat	100%	80%	50%	20%	10%
R	.20	.14	.12	.10	.09
Sd	158	119	94	71	61
Phs	.84	.86	.88	.90	.93

Figure 2. Scatterplot Between First-Year Grades and Admission Test Scores, r =.40.

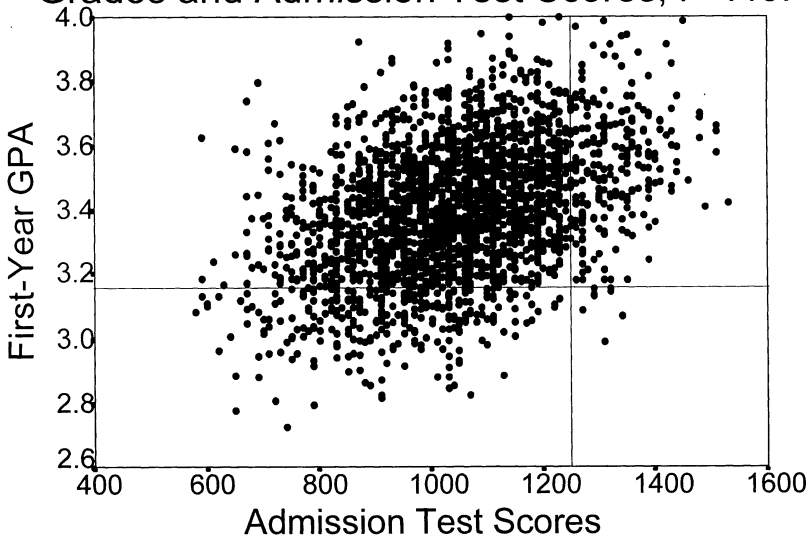

Table 2. *r = .40. Correlations, Standard deviations, and Positive Hits*.

	Percent of Students Selected				
Stat	100%	80%	50%	20%	10%
R	.40	.31	.25	.18	.16
Sd	159	121	95	74	61
Phs	.83	.88	.91	.95	.96

2. Graduate School Predictive Validity Results

Despite the methodological difficulties, what are the predictive validity study results?

Analyzed the recent PVs for the five admission tests to graduate/professional US schools: GRE, GRE Subject, LSAT, GMAT, and MCAT.

- More than 1566 studies.

- Omitted SAT PVs.

Test	Number of PV studies
GRE	1038
GRE Subject	283
LSAT	46
GMAT	145
MCAT	54

GRE Predicitive Validity Studies:

Type of Department	GRE	GRE + G
Natural Science	.31	.44
Engineering	.30	.44
Social Sciences	.37	.48
Human/Arts	.34	.46
Education	.36	.47
Business	.31	.47

3. Additional Validity Evidence

a. Construct Validity Evidence: the GRE measures Verbal, Quantitative, and Analytical Reasoning (Factor analysis)

b. Threats to Validity such as format, speededness, test wiseness have been minimized (if not eliminated).

Future Research Studies

a. Enhanced score reporting for candidates

b. Writing assessments

c. Computer-based formats

Conclusions

1. Predictive validity results are difficult to interpret because of many serious methodological problems that influence the findings.

2. GRE predicitve validity correlations are sizeable (and with UG-GPA), the GRE appears to have considerable utility in the admissions process.

3. The case for the validity of the GRE is seen when other validity evidence is considered – constructs are relevant for success in graduate/professional schools.

4. Several promising directions of research – including writing assessment, advanced mathematical reasoning, and enhanced score reporting (for the benefit of candidates).

Finally – On questions about the validity of the GRE,

- a PV study from a private university with

(1) a very well-qualified applicant pool and a strict admissions process, and

(2) methodological shortcomings, and no corrections for these shortcomings in the interpretation of results, is NOT a study that should influence admissions policy and practices.

There appears to be substantial evidence to support continued use of the GRE.

At the same time, a strong GRE research program – both by GRE and others, should be encouraged.

The Predicitve Validity of SweSAT

Allan Svensson

We are planning a study, the aim of which is to look at the predictive validity of the SweSAT in higher technical education. We have just started the planning which means that we have not got any results and cannot yet present a detailed design. However, we do hope we can get comments and advise of great value from you.

The ETF project

I want to start by giving you some information about a project - a longitudinal project - which is going on in Göteborg. It is called EVALUATION THROUGH FOLLOW-UP (ETF). Its main purpose is:

To make follow-up studies of large and nationally representative samples of students, to ascertain in what way geographic, social and psychological factors affect educational and vocational careers and to discover what changes the Swedish educational reforms have brought about in these respects.

Within this project seven follow-up studies have started. Each sample includes about 10 000 individuals. An overwiev of the project is presented in figure 1. Detailed information may be found in Härnqvist (1998).

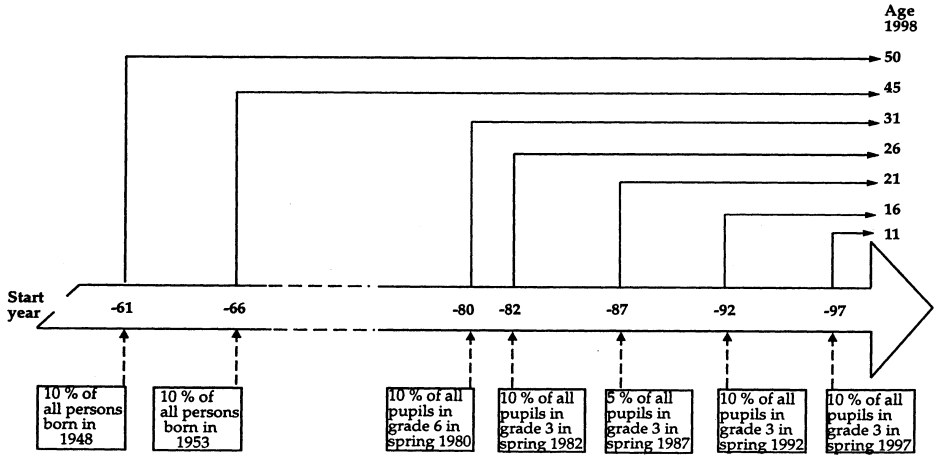

Figure 1. *The cohorts included in the Evaluation Through Follow-up project.*

There are three main categories of information collected:

A. Basic data.

Administrative data from the schools, such as grade and school marks.
Father's and mother's education and occupation.
Scores from intelligence tests and questionnaire responses on school adjustment, interests, educational plans.

B. Follow-up data within the school system.

Administrative data from the schools, similar to the basic data (yearly).
Questionnaire responses from pupils when they are in upper secondary education.

C. Follow-up data from outside the school system.

Excerpts from the records on study finance and from the register of higher education.
Excerpts from the census and income register.

Based on data from the ETF-project a great number of books, articles and reports have been published. Many of them have focused on educational equality, educational reforms and the importance of financial

aid (Svensson, 1980; Reuterberg & Svensson, 1983, 1987, 1994; Härnqvist 1989, 1990).

Data from the ETF-project have also been used for investigations of the SweSAT. For instance socioeconomic and gender differences on the test and differential selection to the test have been studied (Mäkitalo & Reuterberg, 1996; Reuterberg 1997; 1998).

The ETF-R project

Is it possible to use the ETF-data for studying the predictive validity of the SweSAT?

I am afraid that the answer must be "no". These data are not very useful in this field of research. Among other things they suffer from two serious shortcomings:

1. Certainly, the sizes of the samples are quite large - almost 10 000 persons in each - but higher education in Sweden is very differentiated. There are hundreds of different programmes and courses. This means that there will be rather few students in our samples found within the same programme in higher education. In other words the samples in the ETF-project are too small for these kind of investigations.
2. Some of the data we have obtained from Statistics Sweden concern whether the students have completed a degree or not, but we know nothing about how they have succeeded in different courses and how many credits they have earned during each semester. I think it is necessary to have access to this kind of more detailed data for studying the predictive validity of the SweSAT.

What is to be done? One suggestion may be to extend the sizes of the samples - and this has already been done We have created a new and a bigger project called ETF-R, where R stands for "register". This new project includes all Swedes born during the period 1972 - 1977, in all more than 800.000 persons - among them the individuals in our samples born 1972 and 1977.

The ETF-R project does not contain so many different data as the ETF-project. It only contains register data. Nevertheless, I will assure you that this project too, contains a huge amount of data, hundreds of variables for each individual included. Among other things we have got the following data:

- Background variables - sex, social background, ethnic background.
- Information on compulsory schooling.
- Information on secondary education including all study programs.
- All application alternatives to higher education.
- Enrolment in higher education.
- SweSAT scores for all test occasions from 1991 to 1996.

During this spring these data have been supplemented by information on success in higher education. This information includes among other things every course finished.

Some of this information is very detailed, which means that a lot of editing has to be done in order to make the data accessible for analyses. This applies particularly to the application data for upper secondary education, the corresponding data for higher education and the information on courses finished in higher education. This work is in progress at present. When this editing work is finished - hopefully at the end of this summer - it will be possible to use the ETF-R data for predictive studies of the SweSAT.

Why study the predictive validity of the SweSAT in higher technical education?

In the near future there will be a shortage of people with higher technical education in Sweden, and therefore, the provision of education in this domain has been strongly expanded during recent years. However, perhaps not surprisingly this expansion has not solved the problem of shortage because there are:

1. Recruitment problems
2. Graduation problems

The recruitment problems have been thoroughly discussed in a study by Svensson and Reuterberg (1998). Let us now look at the graduation problems. Up to 1975 about 90 per cent of all admitted students in higher technical studies also got their degree, but later on the discrepancy between the numbers of admitted and graduated students began to be great (IVA, 1992). I don't think that there are any causal relations but we can establish that the graduation rate started to fall when the SweSAT was introduced. I am sure that there are lots of causes for the decreasing examination, but nevertheless people may begin to be suspicious of the SweSAT. I hope that our research will eliminate all kinds of such suspicions.

There are recruitment problems to higher technical education, that's true, but these problems are not so large that we don't get enough numbers of applicants (see table 1). There are almost twice as many applicants (about 12 000) as the number of students admitted (about 6 000).

Table 1. *Number of first choice applicants to institutes of technology and number of admitted in 1997.*

	Number of applicants	Number of Admitted	Admitted in relation to applicants
Males	9 566	4 860	2.0
Females	2 926	1 942	1.5
Total	12 492	6 802	1.8

In spite of the rather hard selection process the picture of graduation is not a beautiful one. I just mentioned that the graduation rate has strongly decreased during the last decades and the present situation is shown in table 2. Here you can observe the graduation rate within eight programmes. The rate is reported within five and seven years after the students entered the programmes. Within seven years the graduation rate varies between 44 and 69 percent with an average around 60 percent - somewhat higher for women than for men. When you read these figure, you need to have in mind that the stipulated time for taking a technical degree is four and a half years.

Table 2. *Graduation rate among students in some technical study programmes. Followed through from 1989/90 to 1995/96.*

	Graduation rate within 5 years			Graduation rate within 7 years		
	M	F	Tot	M	F	Tot
Computer science	17	10	17	45	31	44
Electrical engineering	22	25	22	57	62	58
Industrial management	36	34	35	67	63	66
Chemical engineering	23	24	24	61	65	63
Mechanical engineering	30	41	31	68	76	69
Engineering physics	19	13	18	52	60	53
Civil engineering	27	35	29	61	72	64

One reason for the low graduation rate may be that it is not the most able students who have been admitted. One indication of this is the unequal recruitment to the technical programmes (table 3).

Table 3. *Proportions of admitted students in technical study programmes among men and women from different socio-economic groups up to the age of 22. (Individuals born in 1972).*

	Socio-economic group			
	1	2	3	Total
Males	9.2	3.5	1.1	3.6
Females	3.7	1.0	0.3	0.9
All	5.6	2.3	0.7	2.3

There are four times as many male as female students - 3.6 % of all 22 year old men have started a higher technical education as compared to only 0.9 % among the women of the same age. And there are almost eight times as many students from upper middle class homes (group 1) as from working class homes (group 3). Indeed if you compare men from group 1 with women from group 3, there is nearly 30 times as many from the first mentioned group.

The figures presented in table 3 may justify the following question. *Is it possible that the SweSAT selects wrong students?* Oh, what a horrible thought! But if the predictive validity is very low, this condition may be one of the reasons of the low graduation rate. Therefore, I think it is very important to investigate how successful the SweSAT functions in this connection.

Previous research

There is not much previous research in Sweden concerning the predictive validity of different selection instruments for success in higher technical education. However, in an investigation conducted by Henrysson et al. (1985) there are two study programmes in higher technical education included - Mechanical engineering and Electrical engineering. In all there are 200 students in the Mechanical engineering programme and 400 in the Electrical engineering programme. The students started their studies between 1977 and 1979 and their performances were observed for five years.

As predictors Grade Point Average (GPA) from upper secondary school as well as the mean of the grades in mathematics, physics and chemistry were used. Another predictor was additional scores for work experience. As measures of academic success scores on all courses given during the three first years at the technical institutes were used. The results were:

The correlation between GPA and the number of scores awarded in the technical programmes was about 0.40, but the correlation with the mean of the grades in mathematics, physics and chemistry from upper secondary school was somewhat stronger. Additional scores for work experience has a weak negative correlation with success in higher education.

Unfortunately no SweSAT data were included in the analyses, so from this study we cannot get information if the SweSAT is as good as the grades from upper secondary school as a predictor for higher technical studies.

However, there is another study from which we can get some information about this topic (Henriksson & Wolming, 1998). One of the four samples in this investigation consists of 120 students in the engineering physics programme. Their performances were observed for a period of three years.

The students were divided into three groups, according to the grounds on which they had been admitted to the programme. The first group of students had been admitted on the basis of their earlier academic achievement (GPA), the second group on the basis of their SweSAT scores and the third group on the basis of their SweSAT scores with additional scores for work experience. The first group consisted of 80 students and the other two of 20 each.

Academic performance was defined in two ways:

a) The number of credits achieved after each semester
b) The number of dropouts in each group.

The main result was that the students admitted on the basis of their GPA were somewhat more successful than the other two groups. But the differences were rather small. The GPA group had acquired an average of 109 credits after six semesters, the other two groups 101

and 100 respectively. There were no differences at all in drop out rates.

The results of the study give us valuable and important information. They tell us that the SweSAT as a selection instrument is almost as good as the grades from upper secondary school. But there is some information we cannot get. In the study it was not possible to estimate the correlations between the scores on the SweSAT and the criterions because the few students admitted on the basis of their SweSAT scores makes it impossible to get reliable values. Among other things this is what we want to get information on in our planned study.

The planned study

Based on the huge amount of data collected in the ETF-R project, we aim to answer the following questions:

1. What is the predictive validity of the SweSAT in different study programmes in higher technical education? *Success* in higher education will be measured by number of credits achieved after each semester. *Failure* by the number of drop outs.
2. Does the predictive validity vary between different subtests in the SweSAT and from one programme to another?
3. Are there significant differences concerning the predictive validity of the SweSAT between male and female students, between students from various socio-economic groups and between students with different ethnic backgrounds?
4. Are the relationships of equal size at all universities?

The analyses will start at the end of this summer and we hope that we will be able to report the first result during the autumn 1999.

Before I finish I will inform you that this is only the first step in a much bigger investigation. The Bank of Sweden Tercentenary Foundation has promised to give our department - in corporation with the Umeå department - resources to plan a more extensive research programme, which will give much more information about the predictive validity of the SweSAT.

References

Henriksson, W., & Wolming, S. (1998). Academic performance in four study programmes. A comparison of students admitted on the basis of GPA and SweSAT scores, with and without credits for work experience. *Scandinavian Journal of Educational Research, 42*, 135-150.

Henrysson, S., Kriström, M., & Lexelius, A. (1985). Evaluation of merits and academic performance. Studies of the effects of selection regulations. *Arbetsrapporter från Pedagogiska institutionen, Umeå universitet.* Nr 21.

Härnqvist, K. (1989). Comprehensives and social equality. I S. J. Ball,. & S. Larsson, (Red). *The Struggle for Democratic Education*, (18-31). New York: The Falmer Press.

Härnqvist, K. (1990). Long-term effects of education**.** I K. Härnqvist., & N-E, Svensson, (Red). *Swedish Research in a Changing Society*, (323-338). Hedemora: Gidlunds bokförlag.

Härnqvist, K. (1998). A longitudinal program for studying education and career development. *Report from Department of Education and Educational Research, Göteborg University*, 1.

IVA. *Ingenjörer inför 2000-talet.* (1992). Stockholm: Ingenjörsvetenskapsakademin.

Mäkitalo, Å., & Reuterberg, S-E. (1996). Who takes the Swedish Scholastic Aptitude Test? A study of differential selection to the Swe-SAT in relation to gender and ability**.** *Report from Department of Education and Educational Research, Göteborg University*, 3.

Reuterberg, S-E. (1997). Gender differences on the Swedish Scholastic Aptitude Test. *Report from Department of Education and Educational Research, Göteborg University*, 2.

Reuterberg, S-E. (1998). On differential selection in the Swedish Scholastic Aptitude Test. *Scandinavian Journal of Educational Research*, Vol. 42, No. 1,. 81-97.

Reuterberg, S-E,. & Svensson, A. (1983). The importance of financial aid. *Higher education*, 12, 89-100.

Reuterberg, S-E,. & Svensson, A. (1987). Student financial aid and participation in Swedish higher education. *Scandinavian Journal of Educational Research*, 31, 139-162, 987.

Reuterberg, S-E,. & Svensson, A. (1994). Financial aid and recruitment to higher education in Sweden**.** *Studies in Higher Education,* 19, 33-45.

Svensson, A. (1980). On equality and university education in Sweden**.** *Scandinavian Journal of Educational Research*, 24, 79-92.

Svensson, A., & Reuterberg, S-E. (1998). How to get more students to science and technical programmes in higher education. Paper presented at the International Conference on Science, Technology and Society. Tokyo, March 16-22.

Simpson´s Paradox and Culture-Fair Selection

Wim van der Linden

Examples of Simpson´s Paradox:

- Sex bias in graduate admission at Berkely (1975)
- SAT scores decline and grades rise (1998)
- Modern languages in Dutch secondary education (1991)
- Subscription renewals (1979)
- Income tax returns in the US (1982)

Conditional Probabilities

$$P(A \mid B) < P(A \mid \overline{B})$$

$$P(A \mid B \cap C) > P(A \mid \overline{B} \cap C)$$

$$P(A \mid B \cap \overline{C}) > P(A \mid \overline{B} \cap \overline{C})$$

Explanation

The reason of Simpson´s paradox is the existence of a third variable that correlates with both variables.

This additional variable may be unmeasured.

Other manifestations of Simpson´s paradox:

- Correlation analysis: Spurious correlation
- Table analysis: Collapsibility problem
- Experimental design: Unbalanced design or confounded factors
- Multiple regression: Multicollinearity

Spurious correlation

$$r_{xy} > 0$$

$$r_{xy.z} < 0$$

Collapsibility Problem

Under what condition is it possible to collapse a higher dimensional table into a lower dimensional table?

Unbalanced design

	B1	B2	B3
A1	10	15	20
A2	4	6	8

	B1	B2	B3
A1	10	15	20
A2	4	6	10

Multicollinearity

Multicollinearity exists in multiple regression if the predictors correlate.

The correlated predictor may be an unmeasured variable

Multicollinearity has two consequences:

1. Confounding of effects
2. Instability of parameter estimates

Possible Remedies:

Use techniques for experimental control:

- Randomization
- Matching
- Balanced design

Culture fair selection

- Never forget any variable that correlates with group membership variable and selection results (e.g., curriculum attended)
- Analyze conditional rather than marginal probabilities of selection results.

SweSAT and Immigrants – Some Preliminary Results

Sven-Eric Reuterberg

Introduction

During the last decade the number of immigrants in Sweden has increased very much. Many of them have already taken the SweSAT but there are reasons to expect that the number of immigrant SweSAT takers will increase successively. However, little is known about to what extent the immigrants use the SweSAT as a means for admittance into higher education and about their test scores. One reason for that is the fact that no information has been available on these matters. Now, this kind of information is available within the longitudinal data

base called “Evaluation Through Follow-up – Register” ((ETF-R). This data base, which is presented by Svensson in his article, includes more 840.000 persons born during the period 1972 –1979 and for them we know birth country and nationality. The same is also known for their parents. Furthermore, we have data about their SweSAT taking from every test occasion from 1991 through 1996 and these data are on item level which means that we have opportunities for detailed analyses of their test performance when the data based has been completely established. Today, this work is in progress, and therefore, the aim of this article is to present some preliminary results. More precisely, the aims of this article are:

- To study whether there are any differences between Swedes and immigrants, and between different groups of immigrants, in SweSAT taking.
- To study whether there are any differences in SweSAT scores between Swedish and immigrants and between different groups of immigrant test takers.

SweSAT taking

Of the total number of persons in the data base nearly 240.000 have taken the SweSAT at least once, and this number corresponds to more than 25 per cent of the total population. Their distribution on age cohort and test occasion is shown in Table 1. In this table the test occasions are coded as the year (91 – 96) and as the test occasion for each year (A: spring test occasion and B: autumn test occasion).

The table shows some regular trends. There is a tendency of an increasing total number of test takers from 1991 through 1996. The number of test takers is larger in spring as compared to the autumn test occasion and the most common age for taking the test is 18 years.

The table also shows that the number of test takers on each test occasion is so large that it permits analyses for each age cohort over several subsequent test occasions. In the same way the data base makes possible comparisons between different age groups on one and the same test occasion.

Table 1. *Total number of test takers in relation to age cohort and test occasion.*

Birth year	91A	91B	92A	92B	93A	93B	94A	94B	95A	95B	96A	96B
1972	14901	7323	11603	5224	8018	4822	6163	3881	4438	2694	3998	2460
1973	3931	12618	19934	6368	9432	5149	7343	4687	5681	3252	4779	2723
1974	132	2347	8130	12538	17475	7090	9595	5652	7278	4137	6110	3418
1975	2	62	226	1787	4718	12952	18913	7399	9288	4935	7709	4247
1976	1	3	10	41	172	1348	4759	13608	18865	6834	974	4835
1977			1		7	64	246	1158	4796	14223	19317	6338
1978					1	2	8	56	277	1367	7021	16177
1979							1	3	15	91	403	1643
ALL	18967	22353	39904	25958	39824	31427	47029	36444	50638	37533	58411	41841

The categorization of birth country is to a large extent determined by Statistics Sweden which has provided most of the data included in ETF-R. Therefore, we have had only limited possibilities get more detailed information on this variable than that shown in Table 2. In a few isolated cases, however, countries with very few test takers have been put together to one category.

Table 2. *The SweSAT takers' distribution on birth country.*

Birth country	Number of test takers	Percentage of all person in the category
Sweden	228.469	29.5
Finland	689	18.4
Norway, Denmark	551	18.8
Western Europe	547	29.3
Eastern Europe	1.297	23.4
Southern Europe	399	12.2
Great Britain, USA, Canada, Australia	291	30.8
Latin America	1.223	15.1
Africa	392	18.2
Eastern Asia	2.045	23.8
Rest of Asia	1.949	21.8
No information	916	8.9
All	238.768	28.3

Quite naturally, the test takers born in Sweden dominate the number of test takers, but during the period studied more than 10.000 immigrants have taken the SweSAT, and their distribution according to birth country is such that we can expect rather reliable results also for those categories with the smallest number of test takers.

Turning to the percentage of test takers within each category, those born in Sweden are not dominating any more. Now, the Swedes are surpassed by immigrants from the English speaking countries and it has been as common to take the SweSAT among the immigrants from Western Europe as among the Swedes. Immigrants from Southern Europe have taken the test less frequently.

SweSAT taking is a voluntary endeavour and each person is allowed to take test as many times as he/she wants to. It is also a well known fact that repeated test taking tends to imply higher scores (Gustafsson

& Benjegård, 1996). Therefore, it is interesting to know whether repeated test taking is more common within some groups than in others.

In order not to be too extensive the results will be shown for only three age cohorts, those born in 1972, 1975 and 1978, respectively.

Table 3. *Number of test occasions within the different groups. Means among those who have taken the SweSAT.*

Birth country	Born in 1972	Born in 1975	Born in 1978
Sweden	2.03	1.95	1.29
Finland	1.83	1.86	1.31
Norway, Denmark	2.01	2.01	1.41
Western Europé	1.93	1.76	1.33
Eastern Europé	2.01	1.95	1.39
Southern Europé	1.81	2.05	1.41
Great Britain, USA, Canada, Australia	1.87	1.60	1.43
Latin America	1.95	1.75	1.24
Africa	1.99	1.78	1..23
Eastern Asia	2.07	1.95	1.21
Rest of Asia	1.77	1.77	1.23
ALL	2.02	1.94	1.29

First of all, we can see that the number of test occasions decreases from the oldest to the youngest group. This, of course, is due to the fact that the older groups have had more opportunities to take the test.

However, comparing persons with different birth countries, it is obvious that the Swedes are not outstanding. Within each age group there are immigrant groups that have repeated their test takings more times than the Swedes. On the whole, the table shows rather small differences between the different groups.

To sum up we can say that there are some differences between the groups concerning the utilisation of the SweSAT as a means for en-

trance into higher education, but among those who have taken the test there are only small differences concerning repeated test taking.

SweSAT scores

Since SweSAT can be taken more than once and an applicant is given the highest score that he/she has obtained during the last five years, test scores from different test occasions have to be comparable. Therefore, the raw score obtained is transferred to a normed score ranging from 0.0 to 2.0. The results shown in Table 4 are based on the maximum normed score that the test takers have obtained during the total period studied. This score has been multiplied by a constant of 10.

Table 4. *Maximum normed SweSAT score i relation to birth country. The normed scores are multiplied by 10.*

Birth country	Born in 1972		Born in 1975		Born in 1978	
	Mean	*Rank*	Mean	*Rank*	Mean	*Rank*
Sweden	10.16	*3*	9.41	*3*	9.26	*4*
Finland	9.58	*4*	9.71	*2*	9.04	*6*
Norway/Denmark	10.79	*1*	9.85	*1*	12.03	*1*
Western Europe	9.55	*5*	8.81	*5*	9.75	*2*
Eastern Europe	8.45	*7*	8.06	*6*	9.20	*5*
GB/USA/Can./Austr	10.17	*2*	9.15	*4*	9.47	*3*
Southern Europe	6.77	*10*	4.79	*11*	7.08	*8*
Latin America	7.50	*8*	6.44	*9*	5.80	*10*
Africa	7.19	*9*	8.03	*7*	6.00	*9*
East Asia	8.86	*6*	6.68	*8*	7.29	*7*
Rest of Asia	5.26	*11*	4.94	*10*	4.48	*11*
ALL	10.09		9.32		9.18	

The maximum normed scores show a substantial variation. In all age cohorts the test takers from Norway and Denmark have obtained the highest scores. Quite high scores have been obtained also by test takers born in the English speaking countries, Sweden and Finland. Groups that have obtained rather low SweSAT scores are those born in the Rest of Asia, Southern Europe and Latin America.

In this connection it must be said that we know nothing about the reasons for the differences shown in Table 4. It may well be that differential selection effects to the SweSAT play an important role in this connection. There may also be quite large differences as to the social composition of the groups. Another important factor is the fact that I have not been able to take into account the point of time for immigration to Sweden. This information was not available when these analyses were accomplished, but it will be included in the data base within a very near future.

A special problem in this connection is adoption. Within the data base there is a number of persons, who are born abroad but who have been adopted by Swedish parents and they have spent most of their lives in Sweden, often in middle class families.

When the data base has become complete it will contain information which will make us able to control for the factors mentioned above.

Finally, I have made some preliminary analyses of the differences between the Swedish test takers and the immigrant test takers concerning their subtest scores. These differences are shown in Table 5 and they are expressed as standardised beta coefficients in order to make the differences comparable over the various subtests. Since these coefficients are quite abstract, I have supplemented the information by giving the differences expressed in raw scores.

As shown by the table the test takers' age has been kept constant by choosing different test occasions for each age cohort. Throughout the differences refer to the spring test occasion when the test takers were 19 years old.

Table 5. *Differences in subtest scores between Swedes and immigrant test takers. Standardised beta coefficients.*

Subtest	Born 1973 92A	Born 1973 92A	Born 1973 92A	Born 1973 92A	Born 1973 92A	Range in raw scores
ERC	.029	.023	.020	.033	.033	0.57 – 0.80
WORD	.059	.032	.054	.069	.081	0.87 – 2.75
DS	.063	.075	.080	.082	.090	1.34 – 2.32
READ	.056	.058	.061	.082	.090	1.06 – 1.30
GI	.047	.047	.049	.065	---*	0.97 – 1.41
DTM	.092	.094	.091	.091	.097	1.64 – 1.83

* GI was excluded from the SweSAT

All coefficients are positive which means that those born in Sweden have obtained higher scores than the immigrants. However, the differences vary quite substantially between the various subtests. Just as could be expected they are rather small on the English Reading Comprehension test (ERC). In this case the beta coefficients range from 0.20 to 0.30 and this corresponds to raw score differences of 0.60 to 0.80. This result seems quite natural taking into consideration that some immigrants have had English as their first language and that many of them have learnt English during their previous schooling in the same way as the Swedes.

The other subtest differences may be more surprising. In my opinion the most verbal tests, WORD and READ, ought to give greater group differences than the two quantitative subtests Data Sufficency (DS) and Diagram Tables and Maps (DTM), but this was not the case. Instead the largest differences are found for DTM.

I have no firm explanation for that, but I have two tentative hypotheses:

1) Diagram, tables and maps are used more often in Swedish compulsory schools and upper secondary schools than in the corresponding foreign schools and therefore the Swedish SweSAT takers are more accustomed to that kind of material.

2) The material selected for the DTM items is normally taken from Swedish publications and it normally refers to Swedish conditions. Furthermore, the maps, tables and diagrams often contain technical terms that may be more difficult for immigrants to understand as compared to the normal every day language which predominate the items in the other subtests.

Concluding remarks

The results presented here are preliminary, and they should not be used for any far-reaching conclusions. In the same way, the explanations for the results are tentative, since I have not had data enough at hand to make all the controls that ought to be done. However, we are working on making the data more complete, and in a near future the data will reach such a degree of refinement that much more firm conclusions may be drawn. However, I think the problem of immigrants and their SweSAT taking is such an important issue that it must be put forward as soon as possible.

Reference

Gustafsson, J-E. & Benjegård, M. (1996). Högskoleprovet, social bakgrund och upprepad provtagning. In *Högskoleprovet. Genom elva forskares ögon. Stockholm:*

The Swedish Scholastic Aptitude Test

Christina Stage

The SweSAT was introduced in 1977 in connection with a reform of the universities and colleges. It was felt that an admission test would provide a possible solution to two basic problems (1) how to find a method of selection which could be used for applicants without formal qualifications; and (2) how to reduce the decisive role played by marks in the selection process. When the test was first introduced it was, however, only made available for a relatively small group of applicants (those who were at least 25 years old and had at least four years of work experience). Only since 1991 has the test been used for all applicants.

When the use of the SweSAT was restricted to the above-mentioned group the number of persons taking the test was approximately 10 000 each year; 6 000 in the spring and 4 000 in the autumn. Since 1990 the number of examinees has increased dramatically to around 140 000 persons each year; 75 - 80 000 in spring and 55 - 60 000 in autumn.

Since 1996 the test consists of 122 multiple choice questions distributed on five subtests. The test is administered twice a year, once in spring and once in autumn. Students are allowed to take the test as many times as they wish and for those who have several results the best one is used for application. In principle it is optional to take the test; in reality, however, it is compulsory, since only applicants with top marks dare to refrain from taking the test. The content of the test is shown in Table 1.

Table 1. *SweSAT*

Subtest	**Abbreviated**	**No of items**	**time**
Vocabulary	WORD	40	15 min
Data Sufficiency	DS	22	50 min
Swedish Reading Comprehension	READ	20	50 min
Interpretation of Diagrams, Tables and Maps	DTM	20	50 min
English Reading Comprehension	ERC	20	35 min
Total test		122	3h 20min

Vocabulary **(WORD**) measures understanding of words and concepts, and consists of items where the task is to identify which of five presented words has the same meaning as a given word. Both Swedish and foreign words are included in the subtest.

Data Sufficiency **(DS)** aims at measuring numerical reasoning ability. In each item a problem is presented, and the task is to decide whether the information presented is sufficient to allow solution of the problem. The response format is fixed, so each item presents the same five alternatives. The subtest is designed to put as little premium as possible on mathematical knowledge and skills in favour of problem-solving and reasoning.

Reading Comprehension **(READ)** measures Swedish reading comprehension in a wide sense. The examinees are presented with five texts and four multiple choice questions in relation to each text. Each text comprises about one printed page. Some items ask about particular pieces of information but most items are designed to require understanding of larger parts of the text or the text in its entirety.

Interpretation of Diagrams, Tables and Maps **(DTM)** consists of 10 collections of tables, diagrams and/or maps which present information about a topic, with two multiple choice questions in relation to each collection. The degree of complexity of the items varies from simply reading off a presented graph, to some where information from different sources must be combined.

English Reading Comprehension **(ERC)** is of the same general type as the subtest READ. However, in this subtest there is more variability as to both the texts and item formats used. The test consists of 8 to 10 texts of different lengths. Most texts are followed by one or more multiple choice questions with four alternatives. In one of the texts, some words are omitted, and the examinee is supposed to select the omitted word from four alternatives presented alongside the text.

The SweSAT is supposed to measure acquired (developed) abilities and it makes use of the kind of verbal and mathematical skills that develop over the years, both in and out of school. The content of the test does not reflect any specific curriculum, although it is designed to be consistent with school based learning.

The test is designed for selection to different types of university programmes and therefore it is intended to measure the students' general aptitude for studies. Since the test is a selection test it is supposed to rank the applicants as fairly as possible according to their expected academic success. Other requirements on the test are:

- The test should be in line with the aims and content of higher education.
- The test must not have negative effects on the education in upper secondary school.
- It should be possible to score the test fast, cheaply and objectively.
- It should not be possible for an individual to improve his/her test result by means of mechanical exercises or by learning special principles for problem solving.

- The examinees should experience the test as meaningful and suitable.
- The demand for unbiased recruitment should be observed. No group should be discriminated against because of gender or social class.

The test should also be varied and cover many different content areas. It is possible to find the answers to roughly half of the questions in the material provided. In order to answer the remaining questions some background knowledge is necessary.

Test characteristics

In Table 2 the subtests are described by different statistics, which are averages for the tests given from spring 1996 till fall 1998.

Table 2 *Some characteristics of the subtests: means, standard deviations, reliability coefficients, ceiling effects and gender differences.*

Subtest	**Mean**	**Std.dev**	**KR_{20}**	**Ceil eff.**	**d**
WORD	24.4	6.8	.83	0.1 %	-.01
DS	12.9	4.4	.78	0.9 %	.58
READ	12.2	3.5	.67	0.7 %	.08
DTM	11.2	3.8	.73	0.5 %	.59
ERC	12.7	4.2	.80	3.7 %	.24
Total test	72.1	17.5	.92	0.0 %	.00

In Table 3 the intercorrelations between the subtests are given.

Table 3. *Intercorrelations between subtests (reliability coefficients in the diagonal).*

	WORD	DS	READ	DTM	ERC	Tot
WORD	*.83*					
DS	.30	*.78*				
READ	.58	.45	*.67*			
DTM.	.35	.66	.48	*.73*		
ERC	.60	.55	.61	.46	*.80*	
Tot	.57	.55	.69	.59	.69	*.92*

The reliability (KR_{20}) for the subtests DS and DTM together is r = .87, for the combination of WORD, READ and ERC the reliability is r = .90, and for READ and ERC together it is r = .83.

In Table 4 the attenuated intercorrelations between the subtests are given.

Table 4. *Attenuated intercorrelations between subtests*.

	WORD	DS	READ	DTM	ERC	Tot
WORD	*.83*					
DS	.37	*.78*				
READ	.78	.62	*.67*			
DTM	.45	.87	.69	*.73*		
ERC	.74	.70	.83	.60	*.80*	
Tot	.65	.65	.88	.72	.80	*.92*

A factor analysis on subtest level resulted in one factor with an eigenvalue of 2.94, explaining 58.7 percent of the test variance, the second factor had an eigenvalue of 0.94 and the third 0.43.

A factor analysis on item level, however, resulted in one factor with an eigenvalue of 12.44, explaining only 10.2 percent of the variance, a second factor with an eigenvalue of 4.63, a third factor with an eigenvalue of 2 23 and another 33 factors with eigenvalues between 1.6 and 1.0.

Equating of different test versions

All correct answers are given one point, the points are added and the sum is tranferred to a normed scale from 0.0 to 2.0, where 2.0 is the highest result.

The purpose of transferring the raw score to a normed score is to make results from different test administrations comparable. It should always be as easy or as diffucult to get a specific normed score, independent of the difficulty level of the test and independent of the other examinees on a specific test administration.

The procedure used for transferring raw scores to normed scores is equipercentile equating in three different groups: the total group of testees, the testees who are 19 years old and still in upper secondary school, and testees chosen by stratified sampling so that the distributions of sex, age and education are always the same.

We now use IRT as a complement to the usual way of equating. What is needed, however, is a better design for composition of the subtests, which takes into account the need for common items which have not been changed between pretest and regular test. Table 3 shows how the items in the last test were distributed on different pretest occasions.

Table 3. *The items of SweSAT 99:A distributed on pretest occasions*

Pretested	96:A	96:B	97:A	97:B	98:A
WORD	3	1	2	7	8
DS	4	4	3	3	-
READ	-	2	6	1	-
DTM	-	5	3	1	-
ERC	1	-	-	2	-

We also need more information about how many items that are needed to do the IRT equating properly.

Important questions

- What kind of subtests should be added?
- Which subtests should be changed?
- Which subtests should be removed?

How should the scoring be done?

1. As it is done now, supposing there is one underlying ability in the test.
2. One verbal and one quantitative score, supposing there are two underlying abilities
3. Weighting of the subtests in different ways for different educations

Problems and complaints

The WORD subtest takes only 15 minutes (7.5 percent of the total testing time) and gives 35 per cent of the total score.

Technical educations complain that there are too few and too easy quantitative subtests.

Educations of Art (Humanities) complain that the test is unfair against their students.

The test gives rise to gender differences in results and we know that females, when admitted, are as successful as males once they are accepted.

Discussion

For changes of the test, the construct validity should be the starting point; you should start with the criteria and find out what subtests are needed.

In the present test there is too much time per item, which gives too little information. 2.5 min per item is too much 1 – 1.5 min is more appropriate for multiple-choice items. The reliabilities are too low especially on the READ subtest. Very long texts – probably not measuring reading comprehension but skimming skills. One possibility might be to use partial weighting of the distractors. Another might be to collapse the WORD and READ subtests and measure vocabulary in the READ subtest.

Against measuring reasoning ability by means of figures or symbols is the face validity but also that such subtests are highly correlated with quantitative reasoning; an advantage would be (with regard to immigrant testtakers) that such tests are language free.

Verbal production tests has not been very successfull, they are expensive and time consuming and contributes nothing but face validity.

If the subtests can stand for it, i.e. if the reliabilities are high enough it would be prefereable to give the results in two separate scores, weighting, however, puts dubious demands and a lot of validity data.

Since, statistically, two traits are measured by the test two scores would give more information to the testtakers as well as to the schools if the subtests are reliable enough. The schools can do the weighting as they want, by logical or regression procedures.

It is important to distinguish between the admission system and selection test; is the aim to optimize the admission system or to optimize the test. A weigthed composite score of the test and average grades gives a more valid admission decision, it is also compensatory since it works as gender equalizer.

Some comments on the SweSAT from the point of view of research on the structure of intelligence

Jan-Eric Gustafsson

As has been shown in previous research (Gustafsson, Westerlund, & Wedman, 1992; see also Stages contribution at this meeting) the SweSAT is multidimensional. One way to get insights into the measurement properties of the current test thus is to identify the nature of these dimensions. Discussion about further developments of the test also may be stimulated by information about the dimensional structure of the test.

During the last couple of decades some progress has been made in research on models of the structure of human cognitive abilities. This research is primarily based on other kinds of measures than selection tests for higher education, but may still provide useful theoretical insights and suggestions for further developments (see, e. g., Carroll, 1993). One model that has a strong basis in current research, as well as in previous research, is the hierarchical model (see, e. g., Carroll, 1993; Gustafsson & Undheim, 1996). According to this model dimensions of cognitive abilities may be identified which vary in degree of generality, from a general intellectual ability, over a set of broad abilities, down to very specific abilities. Tests such as the

SweSAT may from the point of view of such a model be seen as measuring a mixture of different abilities.

Some studies have been conducted with the purpose of studying the SweSAT from such a perspective (Gustafsson, Westerlund, & Wedman, 1992; Gustafsson, 1994). One main result from these studies is that the SweSAT measures two broad dimensions, along with several specific abilities. One dimension may be identified as "crystallized intelligence" (Gc), and is a dimension of verbal knowledge and skills. The other broad dimension is "fluid intelligence" (Gf), which is an analytic, problem-solving, ability. Among the subtests in the SweSAT WORD, READ and ERC are strongly related to Gc, while DTM and DS are strongly related to Gf. In all subtests there also are specific abilities.

Out of the five subtests in the SweSAT three may thus be classified as measuring primarily Gc, while two primarily measure Gf. One source of potential problems, however, is that both the Gf-tests are numerical/quantitative which implies that the measurement of Gf is confounded with specific abilities of a numerical kind. The broad ability Gf, which also has been shown to be identical with general intellectual ability, can and should, however, be measured with different kinds of content, such as figures, symbols, letters and words. Psychometric research has not shown any gender difference in Gf, at least not in favor of males (Rosén, 1998), but the large performance difference for males on the SweSAT is primarily due to the subtests DTM and DS. This is explained by the finding that the only kind of item content for which Gf-tasks yield gender differences in performance are numerical items (Willingham & Cole, 1997), while other kinds of Gf-items do not give gender differences. From the point of view of gender differences it thus seems unfortunate that the only Gf-tests in the SweSAT are numerical. There may, however, also be other problems associated with the choice of numerical Gf-tests. There is reason to suspect that the numerical content of DTM and DS make test-takers who primarily are focussed upon a "non-numerical" domain, such as law, perform poorly on these tests because they have less experience and/or interest in working with numerical problem solving tasks. If that is the case there is also a risc that the predictive power of the tests is negatively influenced by the numerical content.

All the verbal subtests in the SweSAT are Gc-tests, which measure knowledge and skills which the test-takers have acquired before taking the test. However, it follows from the line of reasoning above that there is good reason to include also verbal reasoning items in the SweSAT. One reason for this is that such an addition would probably improve the SweSAT as a measure of Gf for a large group of test-takers. But there also are reasons to reduce the strong Gc-dominance of the SweSAT. Such tests are thus to a somewhat greater extent influenced by social background than are Gf-tests (Gustafsson & Westerlund, 1994). Another problem with Gc-tests is that performance increases with age, which is not the case for Gf-tests.

One conclusion from this scrutiny of the SweSAT from a psychometric perspective is thus that it should be investigated if it is possible to reduce the current dominance of Gc in the SweSAT, in favor of a larger proportion of Gf-items, and in particular Gf-items with a non-numerical content.

The fact that the SweSAT is dominated by two dimensions also makes it important to consider if the current practice of reporting one total score should not be replaced with a reporting of two separate scores, or, alternatively with a different weighting of the subtests for selection to different educations.

References

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge: Cambridge University Press.

Gustafsson, J.-E. (1994). Dimensions of ability determining performance on the Swedish Scholastic Aptitude Test. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, 1994

Gustafsson, J. E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. Berliner, & R. Calfee (Eds.), *Handbook of Educational Psychology*, New York: Macmillan.

Gustafsson, J.-E., Wedman, I., & Westerlund, A. (1992). The dimensionality of the Swedish Scholastic Aptitude Test. *Scandinavian Journal of Educational Research*, *36*, 21-39.

Gustafsson, J.-E. & Westerlund A. (1994). Socialgruppsskillnader i prestationer på Högskoleprovet. I: R. Erikson & J. O. Jonsson (Red.). *Sorteringen i skolan. Studier av snedrekrytering och utbildningens konsekvenser*. Stockholm: Carlsson Bokförlag.

Rosén, M. (1998b). *Gender differences in patterns of knowledge*. Göteborg: Acta Universitatis Gothoburgensis.

Willingham, W., & Cole, N. (1997). *Gender and fair assessment*. Mahwah, New Jersey: Lawrence Erlbaum.

Educational Measurement in an International Perspective

Professor Ronald Hambleton told us that in the US the interest for measurement has increased during the nineties. In the eighties there was an antitesting movement with less IQ testing and less school-testing, but now the direction has changed and there is a lack of psychometricians. There is also an increased interest in international testing.

The interest is focused on three areas especially, and they are: test accommodation, computer-based testing and standard setting.

Regarding test accommodation there are two aspects. One is accommodation for people with handicaps who demand longer testing time, and the problem is to enhance the validity. Are the test scores comparable when the time is prolonged and is the construct the same? The other aspect is measuring in another language than English. The disabilities act has impact on all issues and the more important the test is the harder they push the language preference (test translation) or alternative assessment. Sometimes the accommodated test is not remotely equivalent to the original test.

As for the second area at present the number of computer-based tests is especially great in the credential field, i.e. GMAT, GRE, TOEFL are all computer-based. The positive thing is the immediate feed-back. The negative thing is the enormous demand on items, since more items are exposed and it is unclear how many items are remembered and passed on to others.

The third area, which is of personal interest is standard setting and score reporting; much more information should be given in connection with score reporting but there is little research done.

There is also an interest in large scale assessment and international comparisons. A national goal is that the US should be number one in the world in math and science in year 2000. NAEP and state to state comparisons is also a great issue. Important political decisions are based on the state ranking and also on ranking of districts within states.

Professor Michal Beller told us that several private colleges have been established in Israel in the last years, offering studies in highly selective areas such as business and law. This causes loss of students for the nonselective university areas.

The selection procedure does not elicit much public discussion at present. PET is still a paper and pencil test. A full version of computer-based test was prepared and circulated among all test registrants as a preparation material. The English section of PET, when administered separately for placement purposes, can be taken in a computerized adaptive manner, using laptop computers provided by NITE. There are also computerized versions available for people who need special test accommodation.

NITE has also developed and administered an advanced mathematics achievement test for engineering schools, and there is a big amd high quality item bank available (which could be used internationally).

The Israeli NAEP has been very successful psychometriaclly but has not had much impact. The only impact was that in some cases the Ministry of Education was forced to rethink the curriculum. At the moment, however, there is no new contract for NAEP.

Professor Wim Van der Linden told us that in the Netherlands there is a big difference between school based testing and credential testing. Regarding school based testing there were no real news. Two years ago a reform had taken place and the new test was now being implemented. The lottery system for admittance to higher education has been replaced by a system where above a certain GPA students have garantueed admission and below that GPA lottery still applies. Also universities will receive the right to select a certain proportion of the candidates themselves.

Regarding credential testing there is an increased interest in testing and psychometricians are in great demand. It is still is difficult to recruit students to the field, however.

CITO became a private organization in January and has undergone a reorganization; at present the top people do not know anything about educational measurements. Still most of test theory research in the netherlands is performed either at CITO or at the Department of Educational Measurement and Data Analysis, University of Twente.

Conference Participants

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

Inger Rydén Bergendahl, The National Agency of Higher Education

Jan-Eric Gustafsson, Gothenburg
Sven-Eric Reuterberg, Gothenburg
Allan Svensson, Gothenburg

Ewa Andersson, Umeå
Kerstin Andersson, Umeå
Stig Eriksson, Umeå
Widar Henriksson, Umeå
Anders Lexelius, Umeå
Christina Stage, Umeå
Gunilla Ögren, Umeå

Program for

The Seventh International SweSAT Conference

Umeå June 3 – 5, 1999

Thursday June 3rd

09.15 Welcome and opening address
The SweSAT program since April 1998
Various Admission Models and their Pros and Cons
Establishing the Validity of Admission Tests
The Predictive Validity of SweSAT. A Planned Study
Simpson´s Paradox and Culture-Fair selection

Coffee 10.00, Lunch 12.00, Coffee 15.00, Dinner 19.00 (Sävargården)

Friday June 4th

09.15 SweSAT and Immigrants. Some preliminary Results
Discussion about Fairness to Testees with Different Cultural and Linguistic Backgrounds.
The SweSAT since 1996. Subtests and Scoring
The Dimensionality of SweSAT
Discussion about Desireable Changes of SweSAT

Coffee 10.00, Lunch 12.00, Coffee 15.00, Dinner 19.00 (First Hotel)

Saturday 5th

09.15 Educational Measurement in an International Perspective
Concluding Discussion and Evaluation

Coffee 10.00, Lunch 12.00