# REPEATED TEST TAKING

## Differences between social groups

Birgitta Törnkvist
Widar Henriksson

EM No 47, 2004


ISSN 1103-2685

## ABSTRACT

The Swedish Scholastic Aptitude Test (SweSAT) is used in the process of selection to higher education in Sweden. The SweSAT is administered twice a year and consists of 122 items. The total raw score is transformed into a normed score, which is used in the selection procedure. The main purpose of this study was to investigate the effects of repeated test taking regarding the SweSAT. The effects are described in relation to the test taker who chooses to retest. Another purpose was to relate the effects of repeated test taking to the individual subtests of the SweSAT and to analyse the self selection process related to repeated test taking. Multivariate linear models for repeated measurements were used to describe the effects of repeated test taking in relation to the test taker. Calibrated scores were used as a basis for relating the effects to subtests. The summarised conclusions were that test takers with high SweSAT score repeated the test more often than those with low scores. Males and test takers from social group I repeated the test more often than females and social group III, respectively. Repeated test taking had its main effect from the first to the second testing for all social groups and both sexes. This main effect of repeated test taking was related to two subtests, WORD and DTM, for all social groups. For the other subtests, social groups II and III were favoured by subtest DS, social group II was favoured by READ and social group III was favoured by subtest ERC. With reference to total normed test score the results indicated that there was a marginal difference between social groups I and II, when controlled for sex and age. Social group I had a higher mean score than the other social groups. Males had higher mean scores than females. The summarised conclusion was that the differences in mean scores between social groups I and III were higher than between males and females.

# TABLE OF CONTENTS

# INTRODUCTION

The selection of students to higher education in Sweden has, since the early 1940s, mainly been based on grade point average (GPA) obtained in upper secondary school. The need for an entrance test for higher education originated in the early 1960s. The main idea at that time was the ambition to broaden the population of students by including new categories. An entrance test provides a solution to the problem of how to select applicants without formal qualifications, and therefore without comparable marks.

To realise this ambition, it was proposed that a Swedish entrance test, of a similar kind to e.g. the SAT in the US, should be developed. Hence applicants without an upper secondary school certificate would be able to qualify by way of an entrance test (Henrysson, 1992).

Also discussed was the way in which an admission test should be used in the selection to higher education, i.e. whether it should be offered to all applicants or restricted to certain categories of students. When the first version of the Swedish Scholastic Aptitude Test (SweSAT) was administered in 1977, the decision was that the test could be taken and used only by applicants who were at least 25 years old and had at least four years of work experience (Henrysson & Wedman, 1975, 1979).

However, this restriction was dropped about 15 years later (1991). Since then, the SweSAT could be used by all applicants to Swedish universities and colleges. One problem discussed at that time was how to reduce the decisive role of grade point average (GPA) from upper secondary school in the selection process. Another idea behind dropping the age restriction was that students from upper secondary school should get a second chance to be admitted to colleges and universities (for further information about the development of the admission system in Sweden, see Wedman, 1992).

## *Social group and the SweSAT*

A number of studies have been conducted with the aim of relating SweSAT scores to different characteristics of the test taker. For example, such characteristics as sex, age and educational level are discussed in Henriksson & Törnkvist (2002) and Törnkvist & Henriksson (2004) (see also Stage 2003, 2004; Cliffordson, 2004).

The social background of the test taker, and its relation to SweSAT score, has also been elucidated in a number of studies. Gustafsson & Westerlund (1994) came to the conclusion that there was a slight tendency that the relation between social background and SweSAT score is lower than the corresponding relation between social background and GPA. Hansen (1997) also concluded that the relation between social group and SweSAT score was about 0.14-0.17.

Stage (2004) examined differences between social groups on item level for the SweSAT. Differential item functioning analyses (DIF) were performed by means of the Mantel-Haenszel method[1] and the summarised conclusion was that there are very few items which function differently for social groups. Cliffordson (2004) included social group in her study of the effects of practice and intellectual growth in order to control for selection effects. Thus, her implicit assumption was that there was a relation between social group, willingness to repeat the SweSAT and SweSAT score.

### *Repeated test taking and the SweSAT*

A number of studies have been conducted with the aim of examining score changes as a function of repeated test taking (Henriksson, 1990; Henriksson & Wedman, 1993; Henriksson & Bränberg, 1994; Henriksson, 1995). With one exception (Henriksson & Bränberg, 1994) these studies have been designed and modelled on the basis of the rules of selection, i.e. the rules regarding the use of SweSAT in the selection procedure. Cliffordson (in press) separated the effects of repeated test taking into two aspects: effects of practice and effects of intellectual growth, over several sessions of the SweSAT. Her conclusion was that there are score gains as a result of practice from the first test administration and that there also are gains associated with intellectual growth. Her results also indicated self selection effects due to age and grade.

[1] For a description of the Mantel-Haenszel method, see for example Holland & Thayer (1988).

### *Purpose*

The main purpose of this study was to investigate the effects of the social group of the test taker and its relation to repeated test taking. Two aspects have been identified; on the one hand whether there is a difference concerning the willingness to repeat the SweSAT, and, on the other hand, whether the observed score changes are related to social group. This latter aspect is related to total score as well as subtest score.

## METHOD

### *Data source*

The total population of this study was based on all test takers at the 00B administration[2] of the SweSAT born in the period 1972 to 1983 who took the test for the first, second, third or fourth time, during a period of two years. The total number of test takers in this study is 20,415.

The test takers who participated in the 00B administration differed with regard to their previous experience in taking the SweSAT test. There were many repeaters. Based on this fact, and with reference to a period of two years (00B, 00A, 99B, 99A), the population, N=20,415, was divided into four subpopulations that were labelled 1, 2, 3, and 4. The designations (1–4) refer to the number of SweSATs taken during the 2-year period. The figure 1 means that 00B was the *first* SweSAT for the test taker, 2 means that 00B was the *second* SweSAT, 3 means that 00B was the *third* SweSAT, and 4 means that 00B was the *fourth* SweSAT.

The total number of items in the SweSAT programme is 122 (Appendix 1), and each item is scored as either 0 or 1. This implies that the range in raw score is 0–122. The raw score is transformed, by equation, into a normed score with the range 0.0–2.0. It is the normed score that is used in the selection procedure, and it was also used in this study. The purpose of the norming procedure is to make scores

[2] The SweSAT is administered twice a year, in the spring (labelled A) and in the autumn (labelled B).

obtained on different testing occasions comparable. The strategy used in the norming procedure is based on different reference populations. These reference populations are chosen by proportional stratified sampling in such a way that sex, age, and educational background are equally distributed at the different test administrations. Two different reference populations, population 1 and population 2, are used as the basis for the norming procedure (Stage & Ögren, 2002).

Further, the population of test takers was divided into subgroups on the basis of sex, age, and social background at the 00B test administration. For the social background the definition given by Statistics Sweden was used with a classification based on parents' education and occupation (Svensson, 1999). The variable social group, i.e., socio-economic background for the test takers was categorized into three socio-economic groups on the basis of the parents' education and vocation: upper middle class (social group I), lower middle class (social group II) and working class (social group III).

The population in the study of four repeated consecutive tests, consisted of 422 test takers; 255 males and 167 females.

### *Instruments*

The SweSAT programme consists of five subtests and 122 items with a total testing time of four hours and ten minutes (for a more detailed description, see Appendix 1). The subtests are:

- WORD, a vocabulary subtest consisting of 40 items
- DS, a data sufficiency subtest measuring mathematical reasoning ability and consisting of 22 items
- READ, a Swedish reading comprehension test consisting of 20 items
- DTM, a subtest measuring the ability to interpret diagrams, tables, and maps, consisting of 20 items
- ERC, an English reading comprehension test consisting of 20 items

The content of the SweSAT has changed over the years from 1977 until today. During the last 15 years, the total number of items has varied between 122 and 148 (Appendix 2).

The subtest STECH (Study techniques) was replaced by the subtest ERC (English reading comprehension) in 1992. At the same time the total number of items was increased from 144 to 148.

Two major changes took place in 1996. Firstly, the total number of items in the test was reduced from 148 to 122, because the subtest GI (General information) was removed from the SweSAT. Secondly, the procedure for field testing newly constructed items was changed by implementing a new pre-test design. Up until 1996, newly constructed items were field-tested in Upper-Secondary schools. In 1996 the pre-test design was changed so that items were pretested together with the actual examination.

## *Statistical analysis and research strategy*

### Self selection

The total population of individuals born in the period 1972 to 1983 included in the VALUTA database comprised 1,348,347 individuals. Information about social group was available for 94% of the population, i.e. for 1,266,598 individuals. The distribution of social group for these individuals was: social group I, 21%; social group II 48%; and social group III, 31% (Svensson, 2002).

The total number of test takers in the present study was 20,415 and this population was divided into four subpopulations labelled 1, 2, 3, and 4. The designations (1–4) refer to the number of consecutive SweSATs taken during the 2-year period. For the total population the distribution regarding social group was social group I, 32%; social group II 48%; and social group III, 20%. In this study the focus was on the distribution of sex, social group and normed score for the four subpopulations. The SPSS package, version 11.0 for windows, was used to calculate mean scores and produce cross tables by sex and social group.

### Effects of repeated test-taking

The repeated measurement module in the SPSS package, version 11.0 for Windows, was used to estimate the effects of repeated test taking. The statistical model was a multivariate linear model with four factors (normed scores and calibrated scores at 99A, 99B, 00A, and 00B) and three confounding variables: age, sex, and social background.

Marginal means of the normed scores for subpopulation 4, conditioned on age, sex, and social background, were used to describe the effects of repeated test taking.

On the subtest level, the description of the effects of repeated test taking is complicated by the fact that the subtests have different maximum scores and are not normed separately. In our study, this problem was handled by using calibrated scores for each subtest (Stage, 2003). The calibration means that for each subtest, and each test administration, the difference between grand mean score over all earlier test administrations since 1996 and the mean score for the subtest is calculated for a reference population 1. This difference is added to each individual score. In a difficult test, the calibrated score becomes higher than the original raw score. In an easy test, the calibrated score becomes lower. A consequence of this calibration technique is that some of the test takers may get calibrated scores higher than the maximum score possible or lower than zero. Here, the calibrated scores are truncated at the maximum. None of the test takers had a calibrated score lower than zero. The highest number of truncations were made on the subtest ERC at the 00A test adminstration, 5% of the test takers had calibrated scores over the maximum. Five per cent of the females and eight per cent of the males in social group I, two per cent of the females and five percent of the males in social group II, and none in social group III had calibrated scores over the maximum. The difference between using calibrated and truncated score is minor. The difference in mean scores for each subgroup (females, males, social groups I, II and III, respectively) is less than 0.01 scores.

## RESULTS

The presentation is organised in the following way: First the self selection process among test takers is studied and second the increase in mean test scores for repeated test takers taking the test four times is studied for normed scores and calibrated scores for subtest WORD, ERC, READ, DS and DTM, respectively .

### *Self selection*

A total of 20,415 test takers born in the period 1972 to 1983 took the 00B test, as their first, second, third or fourth SweSAT. The study only includes repeated test takers that took the tests at consecutive administrations, i.e. for test takers with two tests the first was in 00A, for those with three tests the first test was 99B. For 72% of the test takers 00B was the first SweSAT taken, for 19% it was the second test, for 7% it was the third test and for 2% it was the fourth test taken. The number of test takers who had taken 4 tests was 422 (Table 1).

Table 1. The social group and sex distribution for subpopulation 1 to 4 and for the whole population (%). Number of test takers (N) and proportion of females (%).

| Sub-population | Social group I | Social group II | Social group III | Number of test takers (N) | Females % |
|---|---|---|---|---|---|
| 1 | 30 | 49 | 21 | 14,780 | 55 |
| 2 | 35 | 48 | 17 | 3,874 | 51 |
| 3 | 41 | 44 | 15 | 1,339 | 43 |
| 4 | 43 | 45 | 13 | 422 | 40 |
| N | 6,522 | 9,868 | 4,025 | 20,415 | 10,850 |
| % | 32 | 48 | 20 | 100 | 53 |

The self selection according to social group and sex is evident. Test takers from social groups I and II and males repeated the SweSAT more often than other test takers. Twenty per cent of the test takers at the 00B test administration belonged to social group III and thirty-two per cent to social group I. The proportion of test takers from social group I increases and the proportion of test takers from social group III decreases with the number of SweSAT taken. More than half of the test takers who took the SweSAT test for the first time were females but only forty per cent of the test takers who took the test for the fourth time were females. The self selection to take a first SweSAT test follows the same pattern, i.e. a larger proportion of SweSAT takers are from social group I than social group III. In the total cohort of individuals born in the period 1972 to 1983, 21% belonged to social group I, 48% to social group II and 31% to social group III (Svensson, 2002).

From earlier studies (Henriksson and Törnkvist, 2002; Törnkvist & Henriksson, 2004) we know that test takers with high SweSAT scores repeat the test more often than those with low scores. In this study the mean score at the first test was 0.97 for test takers in subpopulation 4. The corresponding test score for test takers in subpopulation 1 (first timers) was 0.79 (Table 2). The mean score increases with number of SweSAT taken. At the fourth test the mean score was 1.19 for test takers in subpopulation 4.

Table 2. Minimum (Min), maximum (Max), mean normed score (M), and standard deviation (s) at first, second, third and fourth SweSAT for subpopulation 1-4.

| Subpopulation | | N | Min | Max | M | s |
|---|---|---|---|---|---|---|
| 1 | 00B | 14,780 | 0 | 2.0 | 0.792 | 0.436 |
| 2 | 00A | 3,874 | 0 | 1.9 | 0.892 | 0.419 |
| | 00B | 3,874 | 0 | 2.0 | 0.962 | 0.424 |
| 3 | 99B | 1,339 | 0 | 1.9 | 0.955 | 0.407 |
| | 00A | 1,339 | 0 | 1.9 | 1.049 | 0.413 |
| | 00B | 1,339 | 0 | 2.0 | 1.102 | 0.418 |
| 4 | 99A | 422 | 0 | 1.9 | 0.970 | 0.378 |
| | 99B | 422 | 0 | 1.9 | 1.093 | 0.362 |
| | 00A | 422 | 0.1 | 1.9 | 1.156 | 0.377 |
| | 00B | 422 | 0.1 | 1.9 | 1.191 | 0.383 |

There is a relation between sex, social group and score (Stage, 2004). On average, males have higher scores than females and social group I have higher scores than social group III, at the first test. In this study the mean normed score was 0.90 for social group I and 0.67 for social group III and the mean score for males was 0.84 and for females 0.75 for non repeaters (Table 3).

Table 3. Minimum (Min), maximum (Max), mean (M) and standard deviation (s) for normed score in subpopulation 1 by social group and sex.

| Social group | N | Min | Max | M | s |
|---|---|---|---|---|---|
| I | 4,444 | 0 | 2.0 | 0.903 | 0.441 |
| II | 7,236 | 0 | 2.0 | 0.776 | 0.426 |
| III | 3,100 | 0 | 2.0 | 0.672 | 0.413 |
| Sex | | | | | |
| Male | 6,664 | 0 | 2.0 | 0.842 | 0.439 |
| Female | 8,116 | 0 | 2.0 | 0.752 | 0.429 |

The mean normed score at the first SweSAT was 1.05 for social group I and 0.78 for social group III in subpopulation 4. The difference in mean normed score was higher between social groups I and III than it was between males and females, 0.28 scores compared to 0.1 scores (Table 4).

Table 4. Minimum (Min), maximum (Max), mean (M) and standard deviation (s) for normed score at the first SweSAT (99A) for population 4 by social group and age.

| Social group | N | Min | Max | M | s |
|---|---|---|---|---|---|
| I | 180 | 0,1 | 1.9 | 1.052 | 0.364 |
| II | 189 | 0,1 | 1.8 | 0.947 | 0.367 |
| III | 53 | 0,0 | 1.7 | 0.777 | 0.388 |
| Sex | | | | | |
| Male | 255 | 0,1 | 1.9 | 1.010 | 0.357 |
| Female | 167 | 0,0 | 1.8 | 0.910 | 0.401 |

The mean normed scores for the repeaters, at the first SweSAT they took, are higher than for the non repeaters even when controlled for social group and sex, which indicate that social group and sex can only partially explain the self selection procedure. For males in social group I the mean normed score was 0.94 in subpopulation 1 (first timers) and 1.07 in subpopulation 4 for their first attempt at the test (Table 5). The only exception from this trend was for females in social group III. In this group the mean normed score was 0.63 for subpopulation 1 (first timers) compared to 0.60 for subpopulation 4. But, this latter mean score was based on only 25 test takers and the standard deviation was 0.35.

Table 5. Mean normed scores (M) and number of test takers (N) at the first SweSAT for subpopulation 1 and subpopulation 4 by social group and sex.

| Social group | Sex | Subpop. 1 | | Subpop. 4 | |
|---|---|---|---|---|---|
| | | M | N | M | N |
| I | Male | 0.94 | 2,126 | 1.07 | 119 |
| | Female | 0.87 | 2,318 | 1.01 | 61 |
| | Total | 0.90 | 4,444 | 1.05 | 180 |
| II | Male | 0.82 | 3,288 | 0.96 | 108 |
| | Female | 0.74 | 3,948 | 0.93 | 81 |
| | Total | 0.78 | 7,236 | 0.95 | 189 |
| III | Male | 0.73 | 1,250 | 0.94 | 28 |
| | Female | 0.63 | 1,850 | 0.60 | 25 |
| | Total | 0.67 | 3,100 | 0.78 | 53 |
| Total | Male | 0.84 | 6,664 | 1.01 | 255 |
| | Female | 0.75 | 8,116 | 0.91 | 167 |
| | Total | 0.79 | 14,780 | 0.97 | 422 |

### *Effects of repeated test taking*

The analyses of the effect of repeated test taking on the mean scores are based on multivariate models with the variables sex, age and social group controlled for.

The strongest effect of repeated test taking was achieved from the first to the second test for all social groups and for both sexes (Figure 1). This is in accordance with earlier studies of repeated test taking (Henriksson & Törnkvist, 2002; Törnkvist & Henriksson, 2004). In the earlier studies the analyses of changes in mean normed scores were based on multivariate models with the variables sex, age and highest education at the fourth test taken. In this study the education is not controlled for. The mean age in this study was lower (20 years compared to 21 and 22.5, respectively) and as a consequence the education among test takers did not vary much. The highest education for 66% of the test takers was lower secondary school and for 32% upper secondary school. The variable was excluded from the model because it did not contribute to the statistical explanation of the variation of the scores (its contribution to the R-square was very low).

The differences in mean scores between social groups I and II were minor. The standard deviation of scores within these groups was higher than the difference between their mean scores. The changes in mean scores are similar in all compared groups except at the last test administration for social group III. The mean normed score at the fourth test administration was lower than at the third test administration for social group III (Figure 1).

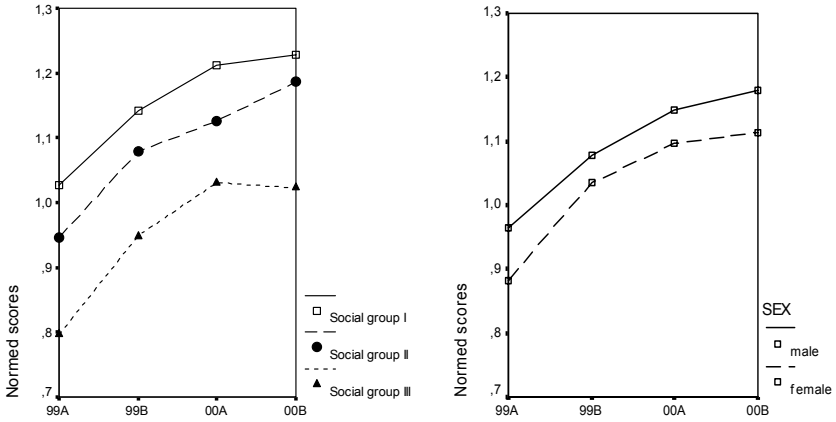



*Figure 1. Mean normed SweSAT score for different social groups, when controlled for sex, and for males and females respectively, when controlled for social background. Evaluated at age 20 years.*

## Subtest

The scores in this section are calibrated scores (cf. p 6) and the statistical models are the same as for the analyses of normed scores.

On the subtest level the differences between the mean scores for the social groups and different sexes are about the same for the subtests DS and DTM, but differ for subtests READ and WORD, compared to the mean normed scores.

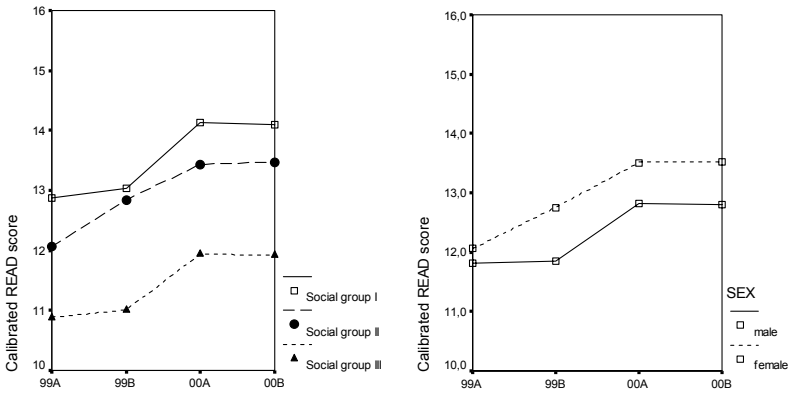
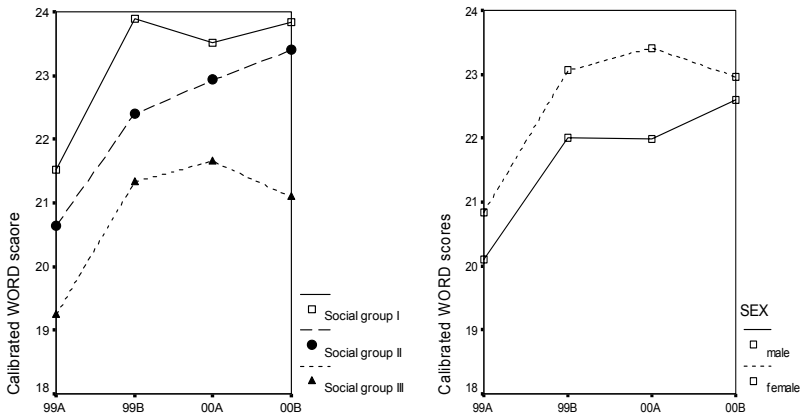



*Figure 2. Mean calibrated score for subtest READ for different social groups, when controlled for sex, and for males and females, respectively, when controlled for social group. Evaluated at age 20 years old.*




*Figur3. Mean calibrated score of subtest WORD for different social groups, when controlled for se, and for males and females respectively, when controlled for social group. Evaluated at age 20 years old.*

The increase in mean scores, from the first to the second test administration, is small for social groups I and III and for males for the subtest READ. Females have higher mean scores than males at subtests READ and WORD. The difference between mean scores for social groups I and III is about the same at each test administration for both subtests and higher than the difference between males and females.

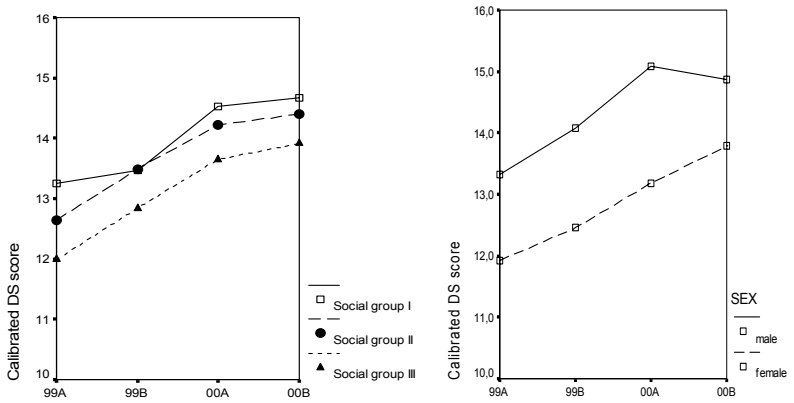



*Figure 4. Mean calibrated score for the sub test DS for different social groups, when controlled for sex, and for males and females, respectively, when controlled for social group. Evaluated at age 20 years old.*

For subtests DS and DTM the difference in mean score between social groups I and III is about the same as the difference between males and females. The DTM subtest has the lowest difference in mean score between social groups I and III of all subtests.

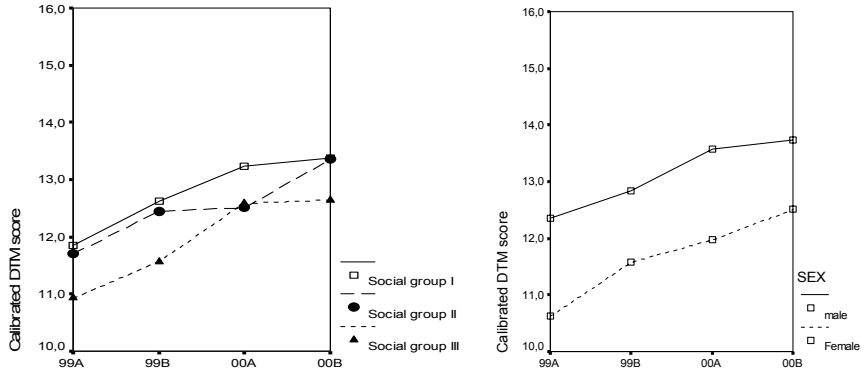



*Figure 5. Mean calibrated scores for the sub test DTM for different social groups, when controlled for sex, and for males and females, respectively, when controlled for social group. Evaluated at age 20 years old.*

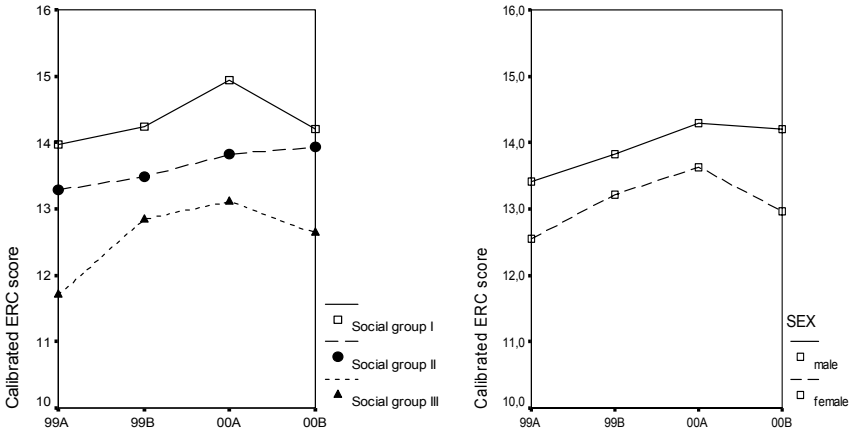



*Figure 6. Mean calibrated scores for the sub test ERC for different social groups, when controlled for sex, and for males and females, respectively, when controlled for social group. Evaluated at age 20 years old.*

For subtest ERC the mean score increases most from the first to the second test administration for social group III, compared to the other social groups and over test administrations. For this test the differences between mean scores are higher between social groups I and III than between males and females.

## DISCUSSION

The purpose of this study was to investigate the effects of social group on mean scores and repeated test taking. Two aspects were identified; on the one hand whether there is a difference concerning the willingness to repeat the SweSAT, and, on the other hand, whether the observed mean score changes are related to social group. This latter aspect is related to total score as well as to subtest score.

With reference to the first aspect, the results indicate that test takers with high SweSAT score repeat the test more often than those with low scores. This result has also been observed in other studies about repeated test taking for the SweSAT (Henriksson & Törnkvist, 2002; Törnkvist & Henriksson, 2004). Another observation was that males, and test takers from social group I, repeat the test more often than females and test takers from social group III, respectively.

But, when comparing with earlier studies, there is a difference between those studies and this study. This difference is that sex and social group are controlled for in this study, but still, the conclusion is valid, i.e., high-scoring test takers repeat the SweSAT more often.

With reference to the second aspect, the main increase in normed score, as a function of repeated test taking, is from the first to the second test occasion for all social groups, when controlled for sex and age.

When comparing this tendency, i.e., highest increase between the first and second test occasion, with the results in earlier studies the finding is that this tendency was also observed in those studies (Henriksson, 1990; Henriksson & Wedman, 1993; Henriksson, 1995; Henriksson & Törnkvist (2002); Törnkvist & Henriksson, 2004). But, in those studies mentioned, the studied characteristics of the test takers were sex, age and educational level, not social group. Thus, the summarised conclusion is that the tendency is exactly the same, irrespective of which individual characteristic is studied.

With reference to normed test scores the results also indicated that there was a marginal difference between social groups I and II, when controlled for sex and age. Social group I had a higher mean score than the other social groups. Males had higher mean scores than fe-

males. However, the summarised conclusion from this study is that the differences in mean scores between social groups I and III were greater than between males and females.

With reference to the second aspect, and subtest score, the finding was that the pattern for the increase in score between the first and second test occasion was the same for all social groups for the subtests WORD and DTM. But, for the other subtests, social groups II and III were favoured by subtest DS, social group II was favoured by READ and social group III was favoured by subtest ERC.

Concerning plausible explanations for the obtained results in this study, as compared to earlier studies by Henriksson and Törnkvist (2002) and Törnkvist & Henriksson (2004), the starting point is the statement that there is a difference in design between these studies. Firstly, social background was the focus for this study. In the other studies (Henriksson & Törnkvist, 2002; Törnkvist & Henriksson (2004) the focus was sex and educational background. Secondly, the population in this study is a cohort of test takers born in the period 1972 to 1983. This cohort was younger, on average, than the test takers in the earlier studies, which meant that the variation in education was very small and did not contribute to the statistical explanation (R-square) of the variation in scores and therefore was dropped from the statistical model.

Thus, the summarised conclusion, based on this and earlier studies is that self selection is operating, i.e., test takers with high SweSAT score repeat the test more often than those with low score. Other summarised conclusions are that males repeat more often than females and social group I more often than social groups II and III. Still another conclusion is that the difference in mean normed scores between the sexes, when controlled for social group and age, is less than the corresponding difference between social groups I and III, when controlled for sex and age.

**Note**

The authors are grateful to the Bank of Sweden Tercentenary Foundation for supporting this research. The study is a part of the project VALUTA.

## REFERENCES


Cliffordson, C. (2004). Effects of practice and intellectual growth on performance in the Swedish Scholastic Aptitude Test (SweSAT). To be published in *European Journal of Psychological Assessment, 20*(3), 192-204.

Gustafsson, J-E., & Westerlund, A. (1994). Socialgruppsskillnader i prestation på Högskoleprovet [Social group differences for SweSAT score]. In Eriksson, R., & Jonsson, J.O. (Red), *Sorteringen i skolan.* Stockholm: Carlssons bokförlag. 264-284.

Hansen, M, (1997). *Social bakgrund och resultat på högskoleprovet* [Social background and SweSAT score] (Rapport Nr 1997:08). Göteborg: Göteborgs universitet, Institutionen för pedagogik.

Henriksson, W. (1990). *Effekter av upprepat provtagande* [The effects of repeated test taking] (PM, Nr 40). Avdelningen för pedagogiska mätningar, Pedagogiska institutionen, Umeå universitet.

Henriksson, W. (1995). *Repeated test taking and the SweSAT* (EM, No 13). Division of Educational Measurement, Department of Education, University of Umeå.

Henriksson, W., & Bränberg, K. (1994). The effects of practice on the Swedish Scholastic Aptitude Test. *Scandinavian Journal of Educational Research, 38*(2)*,* 129–148.

Henriksson, W., & Törnkvist, B. (2002). *The effects of repeated test taking in relation to the test taker and the rules for selection to higher education in Sweden* (EM, No 41). Department of Educational Measurement, University of Umeå.

Henriksson, W., & Wedman, I. (1993). *Effects of repeated test taking on the Swedish Scholastic Aptitude Test (SweSAT)* (EM, No 8). Division of Educational Measurement, Department of Education, University of Umeå.

Henrysson, S. (1992). *Högskoleprovets historia. Några bidrag* [The history of the SweSAT. Some contributions] (PM, Nr 91). Umeå universitet, pedagogiska institutionen, Avdelningen för pedagogiska mätningar.

Henrysson, S., & Wedman, I. (1975). *The content of the Scholastic Aptitude Test* (Spånor från Spint, Nr 3). Umeå university, Department of Education.

Henrysson, S., & Wedman, I. (1979). *Research in Sweden with regard to predictive tests and interviews for admission to higher education* (Spånor från Spint, Nr 13). Umeå university, Department of Education.

Holland, P.W., & Thayer, D.T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H.Wainer & H. Braun (Eds.), *Test validity.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Stage, C. (2003). *Normering, ekvivalering eller kalibrering av delar i högskoleprovet* [Norming, equivaling or calibrating subtests in the SweSAT] (PM, Nr 185). Enheten för pedagogiska mätningar, Umeå universitet.

Stage, C. (2004). *Gruppskillnader i resultat på högskoleprov* [Group differences in SweSAT score] (PM, Nr 192). Enheten för pedagogiska mätningar, Umeå universitet.

Stage, C., & Ögren, G. (2002). *Högskoleprovet våren och hösten 2001*[Score from the SweSAT spring and autumn 2001] (PM, Nr 172). Enheten för pedagogiska mätningar, Umeå universitet.

Svensson, A. (1999). *Socialgruppsbegreppet – sett ur den pedagogiska forskningens synvinkel* [The concept of social group from a perspective of educational research] (IPD-rapporter, Nr 1999:05). Institutionen för pedagogik och didaktik, Göteborgs universitet.

Svensson, A. (2002). Indelningen i socialgrupper [Subdivision of the population into Social groups]. Unpublished manuscript. Valutaprojektet. Institutionen för pedagogik och didaktik, Göteborgs universitet.

Törnkvist, B., & Henriksson, W. (2004). *SweSAT repeat* (EM, No 46). Department of Educational Measurement, University of Umeå.

Wedman, I. (1992). *Selection to higher education in Sweden* (EM, No 1). Department of Education, Division of educational measurement, Umeå university.

# APPENDIX

**APPENDIX 1**

### *The SweSAT Programme*

The SweSAT consists of 122 multiple-choice items distributed over the following five subtests:

Vocabulary (**WORD**) measures the comprehension of words and concepts. It consists of 40 items in which a word or phrase is given, and the task is to identify which of five options has the same, or almost the same, meaning. Words of both Swedish and foreign origin are included in this subtest. The testing time is 15 minutes.

Data Sufficiency (**DS**) aims at measuring mathematical reasoning ability. Each of the 22 items presents a problem and two statements. The task is to decide whether the statements provide enough information for solving the problem. The response format is fixed, i.e. each item presents five identical options. The test does not require advanced mathematical knowledge or skills. The testing time is 50 minutes.

Reading Comprehension (**READ**) measures Swedish reading comprehension in a broad sense. The test consists of five texts with four multiple-choice questions to each text, in total 20 items. The length of each text is roughly one page. Some questions concern details stated in the text, but most of them are designed to test the comprehension of larger parts or the text as a whole. The testing time is 50 minutes.

Interpretation of Diagrams, Tables, and Maps **(DTM)** consists of ten sets of tables, graphs, or maps presenting information about different topics. There are two multiple-choice items to each set, which makes a total of 20 items. The degree of complexity varies from simply reading off a presented graph to problem-solving, i.e. processing information from all the different sources in the material. The testing time is 50 minutes.

English Reading Comprehension (**ERC**) is of the same general type as the subtest READ. However, this subtest is more varied regarding both texts and item format. It consists of eight to ten texts that vary in length. Most of them are followed by one or more questions with four options. The last text in the subtest has sentences where a word or a set of words has been omitted. The test taker is required to choose the

one of four options that best fits the rest of the sentence. The total number of items is 20. The testing time is 35 minutes.

In addition to these five regular subtests, a set of pretest items is included in the testing. This set contains a complete version of DS, READ, DTM, or WORD + ERC. The pre-test items are not included in the test taker's result. The testing time of this set is 50 minutes.

**APPENDIX 2**

***The SweSAT programme 1986, 1991, 1993, 1997 and 2000.***

| Subtest | Type of test | 86B | 91B | 93B | 97B | 00B |
|---|---|---|---|---|---|---|
| **WORD** | **Vocabulary** | 30 | 30 | 30 | 40 | 40 |
| **DS** | **Data sufficiency** | 20 | 20 | 20 | 22 | 22 |
| **READ** | **Reading comprehension** | 24 | 24 | 24 | 20 | 20 |
| **DTM** | **Interpretation of diagrams, tables, and maps** | 20 | 20 | 20 | 20 | 20 |
| **GI** | **General information** | 30 | 30 | 30 | - | - |
| **STECH** | **Study techniques** | 20 | 20 | - | - | - |
| **ERC** | **English reading comprehension** | - | - | 24 | 20 | 20 |
| **Total number of items** | | 144 | 144 | 148 | 122 | 122 |
| **Total number of test takers** | | 3,780 | 56,202 | 54,130 | 47,435 | 20,415 |