SweSAT REPEAT

Birgitta Törnkvist
Widar Henriksson

EM No 46, 2004

ISSN 1103-2685

ABSTRACT

The Swedish Scholastic Aptitude Test (SweSAT) is, together with grade point average (GPA), used in the selection procedure for higher education in Sweden. The SweSAT is administered twice a year, it consists of 122 items and the total raw score is transformed to a normed score which is used in the selection procedure. Certain rules apply to the SweSAT when it is used for selection to higher education and one of these rules is that the test taker can repeat the test. The main purpose of this study was to investigate the effects of repeated test taking for the SweSAT. The effects are described in terms of normed score, partly in relation to the rules for selection, and partly in relation to the test taker who elect to retest. The result indicates that the effects of repeated test taking mainly occur between the first and second testing. This is a tendency that is observed when the effects of repeated test taking are analysed in relation to the rules for selection as well as in relation to the test taker. According to the rules for the SweSAT in the selection system the best score is used if a certain applicant has more than one valid score and the model for description of score gains in this study is based on this fact. Another purpose was to relate the effects of repeated test taking to subtests of the SweSAT. Proportional stratified sampling and multivariate linear models are used when describing the effects of repeated test taking in relation to the test taker, while a reference population and a calibrated raw score is used as a basis when the effects are related to subtests. With regard to subtests, the main effect of repeated test taking is related to three subtests: WORD, DS and ERC. The summarised conclusion that is made in this study is that the effects of repeated test taking between the first and second test occasion, in terms of standard deviation units for total normed score for the SweSAT, is about 0.2 and this finding is in congruence with findings in earlier studies. Plausible hypotheses for the obtained results are discussed.

TABLE OF CONTENT

INTRODUCTION

In Sweden the selection of students to universities and colleges has, since the early 1940s, mainly been based on the grade point average (GPA) obtained in upper secondary school. Since 1977, however, a Swedish test battery, that to some extent resembles the SAT in the US (Henrysson & Wedman, 1975, 1979; Henriksson, Henrysson, Stage & Wedman, 1985), is also used for selection to higher education in Sweden. When the first version of this test battery, the Swedish Scholastic Aptitude Test (SweSAT), was administered in 1977 the decision was also that the test should be taken only by applicants who were 25 years or older and who had at least four years work experience (Wedman, 1994, 1995). However, almost fifteen years later (1991) the restriction with respect to the applicants' ages and work experience was dropped. Since then the SweSAT can be used by all applicants to higher education (Henrysson, 1992).

It is also worth mentioning that the debate in the 1960s and 1970s, that resulted in the development and use of the SweSAT, to some extent had its origin in the ambition to broaden the student population by including new categories of students. An indicator of the efforts to realise that ambition was the decision to include credit points for work experience in the score that was used for ranking this new category of students. It can also be added that in 1977 equal weight was applied to SweSAT score and score for work experience. SweSAT score was transformed to a normed score with a maximum of 2.0 and the quantification of work experience also had a maximum of 2.0 (Henrysson, Kriström & Lexelius, 1985). The weight for work experience has diminished by gradual stages over the years and the last ten years its maximum has been 0.5 (Henriksson & Wolming, 1998). The maximum normed score for the SweSAT has been exactly the same (2.0) over the years.

When the SweSAT is used for selection to higher education in Sweden there are certain rules which are defined and applied. One of the basic rules, which also is the basis for the study reported here, is that the test is optional and that a test taker can repeat the SweSAT as many times as he or she wants. This means that it is up to the individ-

ual whether to take the SweSAT or not, and whether to repeat the SweSAT or not[1].

Effects of repeated test taking from a perspective of the rules for selection

When the SweSAT is used for selection to higher education in Sweden there are certain rules, which are defined and applied. These rules are:

- An obtained SweSAT score is valid for five years
- The best obtained score (normed score) is used in the selection system if a test taker has more than one valid score from the SweSAT
- An applicant is selected on the basis of SweSAT score *or* on the basis of GPA (grade point average)
- If an applicant has a valid SweSAT score *and* a valid GPA, the best result is used in the selection procedure

Effects of repeated test taking in a perspective of the test

Henriksson (1981a) summarised, in his literature review about practice and coaching on test score, the main findings so far by concluding that the effects of repeated test taking (practice) are greater when a test has a speed component than when there is no speed-limit, i.e., the need to respond quickly is susceptible to practice. This means that the gain from repeated test taking on parallel or similar tests is the establishment of a rational time-using strategy. Another finding was that the effects of repeated test taking tend to be more pronounced on non-verbal tests than on verbal tests, such as vocabulary tests and verbal reasoning tests. Still another finding was that the gain from repeated test taking is greater for tests with a complex item format as compared to tests with a simple instruction and a simple format.

This summary, which dates back to the early 1980s, has turned out to be fairly consistent with later reviews in this area (see for example Becker, 1990). It can also be stated that test and test construction has

---

[1] A test taker has to pay a registration fee, 350 SKR (about 50 USD), when taking the SweSAT.

been developed on the basis of knowledge about these findings in order to eliminate irrelevant score gains as a function of repeated test taking (Messick, 1981; Roznowski & Basset, 1992).

Effects of repeated test taking in a perspective of the test taker

In his literature reviews Henriksson (1981a, 1981b) concluded that there is a main distinction between repeated test taking with and without support. Test taking without support was called practice and test taking with support was called instruction or coaching. In a practical situation there is of course no clear-cut border between practice and coaching but in theoretical terms this means that practice implies no support, i.e., no special instruction about strategies for test taking and no teacher support. Based on this distinction one main finding from the literature was that the more able a test taker is, the more gain and benefit from unsupported practice. Another main finding was that the greatest effects of practice occur when the test taker has little or no previous experience of test taking, i.e., when they are completely unfamiliar with test and test situations in general. This latter finding has influenced the test taker and in most cases the consequence is that now test takers are usually fairly familiar with the requirements of each test as well as the whole test situation, i.e., they are ”testwise” in a general and practical sense (Millman, Bishop & Ebel, 1965; Rogers & Bateson, 1992; Roznowski & Basset, 1992). This finding has also influenced those who are responsible for the test, the test administration and use of scores from the test in such a way that different strategies are applied in order to reduce or eliminate the effect of deficiency in this respect.

Scores on retesting are also subject to changes resulting from the test taker’s growth (in this case intellectual growth). If there is a change in the ability, which is measured by a test, this change will be reflected in the test score. This means that the observed test score will be higher on retesting, not as a consequence of repeated test taking per se, but as a consequence of a growth in ability or intellectual capacity (Cliffordson, in press).

Repeated test taking and the SweSAT

As a consequence of the rules for the SweSAT in the selection procedure, and the fact that there is a competition for the available study

places, many applicants repeat the SweSAT. More than one third of the test takers have taken the SweSAT at least once, twice or more before a certain test administration. The fact that many applicants have, or intend to get, two or more valid test scores makes score changes a matter of great concern for all parties involved; test takers, test developers and those who are responsible for using test score for selection purposes.

A number of studies have been conducted with the aim of elucidating 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 for selection, i.e., the rules for the SweSAT when it is used in the selection system.

On the basis of this model for description Henriksson (1990) studied the effect of repeated test taking for the whole population of test takers at the 86B[2] administration of the SweSAT. The result indicated a gain in score between the first and second testing ($M_{d,2}$=0.055[3]). Those test takers who had taken the SweSAT three times obtained a marginally lower score on the third test occasion, as compared to the best of the two earlier scores ($M_{d,3}$=-0.011). The same tendency, i.e., a decrease in score on the last test occasion, was also obtained for those who had taken the SweSAT four times ($M_{d,4}$=-0.022).

To produce evidence in support of the stability of the results for the different categories of repeaters Henriksson & Wedman (1993) used the same methodology for a new sample of test takers (=91B). The results were, on the whole, the same even if the sample was almost ten times larger than in the earlier study. The larger sample was mainly due to the fact that the rules for admission to higher education had been modified. The means in this study were: $M_{d,2}$=0,083, $M_{d,3}$=0,009, $M_{d,4}$=-0,034. The means tended to be somewhat higher than the corre-

---

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

[3] $M_{d,2}$ = mean difference in obtained normed score between the first and the second test occasion for those test takers who have two SweSAT scores.

sponding data in the study by Henriksson (1990), which probably was an effect of growth. Unlike the majority of the test takers in the study mentioned above (Henriksson, 1990) a larger proportion now came from the two highest grades in upper secondary school.

The two studies reported so far were based on exactly the same subtests and the same number of items. In 1992 the SweSAT programme was changed; the subtest STECH was replaced with the subtest ERC and the total number of items were increased from 144 to 148 (Appendix 1). A study of the effects of repeated test taking (Henriksson, 1995) with the same methodology as in the first two studies mentioned, resulted on the whole in the same results as in earlier studies. The means were: $M_{d,2}$=0,111, $M_{d,3}$=0.031, $M_{d,4}$=-0,023. Again the means were somewhat higher as compared to the first study (Henriksson, 1990) and the probable explanation was that it was an effect of growth. The majority of test takers were students from upper secondary school who repeated the SweSAT during their period of schooling in upper secondary school. Another reflection was that the change in the SweSAT programme seemed to have minor effects.

In summary, the studies indicates that the biggest gain in score occurs between the first and second occasion whereas in most cases the score increased only marginally or even decreased on the third and forth testing, in relation to the best test result obtained in any of the previous testings.

Henriksson & Bränberg (1994) focused on the effect of repeated test taking, totally as well as on the level of subtests and groups, and based their study on the fact that the biggest gain in score was obtained between the first and second test occasion. They studied five consecutive populations in order to control for the effect of self selection, i.e., the fact that it is the test taker's own decision whether to repeat the SweSAT or not. The division into subgroups was made on the basis of the variables sex, age and education. On the subtest level, it appeared that the DS, STECH and DTM showed the largest average gain score (see Appendix 3 for a description of the subtests).

Henriksson & Törnkvist (2002) also focused on the effect of repeated test taking, totally as well as on the level of subtests. Their results indicated that the effects of repeated test taking mainly occur between the first and second testing. This is a tendency that is observed when the effects of repeated test taking are analysed in relation to the rules

for selection as well as in relation to the test taker. With regard to subtest, the main effect of repeated test taking was related to two subtests: WORD and DS.

The summarised conclusion that was drawn in the studies reported above was that the effects of repeated test taking, in terms of standard deviation units for total normed score for the SweSAT, is about 0.15-0.25. To make a norm of reference - the magnitude of gain in score between the first and second test occasion for the SAT is about 0.20-0.25 (Messick, 1980; Cole, 1982; Donlon, 1984; Bond, 1989; Powers & Rock (1999). Thus, the findings for the SweSAT are, on the whole, in congruence with findings in international studies.

Purpose

The main purpose of this study was to investigate the effects of repeated test taking with regard to the SweSAT. The effects are described partly in relation to the rules of selection, and partly in relation to the test taker who chooses to retest. Another purpose was to relate the effects of repeated test taking to subtests of the SweSAT and also to compare the results of this study with the results of earlier studies of repeated test taking for the SweSAT.

METHOD

Data source

The total population of this study consisted of all test takers at the 2002 autumn administration of the SweSAT (02B). The total number of test takers at this occasion was 28,589 (N_{tot}=28,589).

The test takers who participated in the 02B administration differed with regard to their experience of taking the SweSAT. There were many repeaters. Based on this fact a period of two years (02B, 02A, 01B, 01A) was selected and the reason for selecting a 2-year period was the ambition to isolate the effect of repeated test taking. A longer time-span than two years increases the probability for other factors than repeated test taking to effect the SweSAT score. This population, which is used to describe the effects of repeated test taking in relation to the rules of selection (N_{1-4}=20,345), was divided into four sub-

populations that were labelled 1, 2, 3, and 4. The designations (1–4) refer to the number of SweSATs taken during the selected 2-year period, i.e., 01A-02B. The figure 1 means that 02B was the *first* SweSAT for the test taker (N_1=14,959), 2 means that 02B was the *second* SweSAT (N_2=4,096), 3 means that 02B was the *third* SweSAT (N_3=1,086), and 4 means that 02B was the *fourth* SweSAT (N_4=204). Thus, those test takers that were labelled 4 had taken all SweSATs administered (02B, 02A, 01B, 01A) during the selected period of observation. It can also be noted that the difference between the total number of test takers at the 02B test occasion (N_{tot}=28,589) and the total number of test takers ($N_{1\text{-}4}$=20,345) in the selected period of observation (02B, 02A, 01B, 01A) depends on SweSATs that were taken before 01A, i.e. 00B or earlier. This means that the designations (1–4) refer to the number of SweSATs taken only during the selected 2-year period, i.e., 01A-02B.

The total number of items in the SweSAT programme is 122 (Appendix 1), and each item is scored 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 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 in this study was divided into subgroups on the basis of sex, age, and educational background at the 02B test administration. As regards age, test takers were divided into four subgroups: under 25 years of age, 25–29, 30–39, and over 39 years of age. The variable educational background was scaled with length of education as the main classifying principle. Seven categories (1–8) were defined. The lower the number is (with the exception of number 8), the lower is the level of education (Appendix 2).

The design in this study was modelled with reference to a two-year period and the reason for selecting a two-year period was that the influence of other factors than the effect of repeated test taking (age,

education etc) would be too plausible if the observed period was longer that two years. The basic idea with the selected design and time period was to describe the effects of repeated test taking for four test occasions. Unfortunately, subpopulation 4 turned out to consist of 204 test takers only, and this is too few observations to make a reliable analysis of repeated test taking for four consecutive test occasions with three independent variables (sex, age and education). Therefore, subpopulation 3 (N_3= 1,086) was used as a basis for describing the effects of repeated test taking for three test occasions. This means that only those who took test 01B, 02A, 02B are studied and this subpopulation is labelled 3* (N_{3*}= 705). This means that we studied repeated test taking for those test takers for which 01B was the first test.

A sample of 504 test takers is selected from this subpopulation (N_{3*}= 705). This sample is used to describe the effects of repeated test taking in relation to subtests and to the test taker. This sample consisted of 233 males and 271 females. The sample was randomly chosen in such a way that the distribution of the normed scores was the same in the sample as in subpopulation 1. The reason for this sampling strategy was to distinguish between the effects of repeated test taking and self-selection. Subpopulation 3 had a higher mean normed score than subpopulation 1, i.e. those who had a rather high score chose to repeat the SweSAT more often than those who had a rather low score. Therefore, it is necessary to control for this selection effect by using a defined sample from subpopulation 3.

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 3). 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 during the years from 1977 until today. Over the last 15 years, the total number of items has varied between 122 and 148 (Appendix 1).

Model of description and statistical analysis

The effects of repeated test taking in relation to the rules for selection

According to the rules regarding the use of SweSAT scores in the selection system, an applicant who has more than one valid score is always allowed to use his or her best result. The model of description of score gains used in this study was based on this fact. This means that if a test taker has, for example, two valid scores and decides to repeat the SweSAT, a score gain is obtained if the score at that third test occasion is higher than *the best* of the two earlier scores.

The effects of repeated test taking in relation to the test taker

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 three factors (normed scores at 01B, 02A, and 02B) and three independent variables: age, sex and maximum education at 02B.

Marginal means of the normed scores for subpopulation 3*, conditioned on age, sex and educational background, were used to describe the effects of repeated test taking. Analogously, estimated marginal means were used in the case of the sample.

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 that they are not being normed separately. In our study, this problem was handled by using raw scores and calibrated scores for each subtest. The subtest scores of reference population 1 (Stage & Ögren, 2002) at the three test administrations; 01B, 02A, and 02B, were used to calibrate the subtest scores. This procedure was used to ensure the comparability of scores from different test administrations. The calibration means that for each subtest, and each test

administration, the difference between grand mean score over all three test administrations and the mean score for the subtest is calculated for reference population 1. This difference is added to each individual score. This means that the calibrated score becomes higher than the original raw score for a difficult subtest (e.g. DTM 02B, Table 10), and analogously, lower for an easy subtest.

RESULTS

The effects of repeated test taking in relation to the rules of selection

Table 1. Test takers who had taken the SweSAT 1, 2, 3 or 4 times grouped according to sex, age and education. Percentage (%).

	Number of SweSATs				
Variable	**1**	**2**	**3**	**4**	**Total**
Sex					
Male	41.1	43.3	48.5	52.9	42.1
Female	58.9	56.7	51.5	47.1	57.9
Age					
–20	51.5	58.4	60.8	62.7	53.5
21–24	17.9	20.1	22.0	20.6	18.6
25–29	9.9	8.9	6.7	7.8	9.5
30–39	15.3	10.1	8.2	6.9	13.8
40–	5.5	2.5	2.3	2.0	4.7
Education					
1	2.4	1.8	1.6	2.0	2.2
2	0.8	0.6	0.4	0.0	0.7
3	13.1	10.3	7.2	5.9	12.2
5	76.4	79.5	79.7	77.0	77.2
6	4.1	6.4	9.1	14.2	4.9
7	1.9	0.8	1.2	0.0	1.6
8	1.3	0.6	0.8	1.0	1.1
Total					
%	73.5	20.1	5.3	1.0	100
N	14,959	4,096	1,086	204	20,345

On the basis of the data in Table 1, two observations can be made. The first one is that the proportion of men increases with the number of SweSATs taken. The second one is that the proportion of test takers in educational category 6, i.e. post secondary education, ≤80 credits (Appendix 2), increases with the number of SweSATs taken.

Table 2. Mean normed score (M) and standard deviation (s) on the 1st, 2nd, 3rd and 4th test occasion for test takers who had taken the SweSAT 1 (n=14,959), 2 (n=4,096), 3 (n=1,086) or 4 times (n=204).

Number of SweSATs	**Test occasion**							
	1		**2**		**3**		**4**	
	M	**S**	**M**	**s**	**M**	**s**	**M**	**s**
1	0.81	0.44						
2	0.84	0.43	0.92	0.42				
3	0.88	0.40	0.99	0.40	1.03	0.40		
4	0.87	0.42	1.00	0.42	1.08	0.42	1.09	0.43

The data in Table 2 show that repeated test taking is a matter of self-selection (cf. Alderman, 1981, Henriksson & Törnkvist, 2002). One indication is that there is a relation between number of SweSATs taken and mean score at the first test occasion. With the exception of those who had taken the SweSAT four times, henceforth called subpopulation 4, the mean at the first test occasion for the other subpopulations (2 and 3) is increasing with the number of test occasions. Another observation is that the mean score for each subpopulation increases with repeated test taking.

The data presented so far are based on background data (Table 1) and average mean score (Table 2) for each subpopulation (1–4) on a certain test occasion. The following Tables (3–8) are based on the model of description defined in relation to the rules of selection. This model allows the description of the relation between the, till then, highest obtained SweSAT score for a certain test taker and the score which is obtained at the subsequent SweSAT administration. A gain in score

means that the score that is obtained at the last test occasion, for those test takers who had taken the SweSAT twice, three times and four times respectively, is higher than the best of the test taker's earlier scores, i.e., the difference in normed score is positive. Analogously, if the score on the last test occasion is lower, the difference is negative.

Table 3. Mean difference and standard deviation for normed score (M_d, s_d) related to sex, age and education for test takers who had taken the SweSAT 2 (n=4,096), 3 (n=1,086), or 4 times (n=204).

Difference in normed score related to variable	**Number of SweSATs**					
	2		**3**		**4**	
	M_d	**s_d**	**M_d**	**s_d**	**M_d**	**s_d**
Sex						
Male	0.090	0.21	0.012	0.18	-0.038	0.18
Female	0.076	0.21	-0.015	0.20	-0.032	0.18
Age						
–20	0.077	0.21	-0.002	0.19	-0.016	0.18
21–24	0.084	0.22	-0.018	0.19	-0.045	0.19
25–29	0.082	0.23	0.016	0.18	-0.156	0.20
30–39	0.096	0.20	0.034	0.17	-0.064	0.12
40–	0.102	0.18	-0.036	0.19	0.050	0.06
Education						
1	0.118	0.20	-0.024	0.13	-0.125	0.10
2	0.048	0.18	-0.125	0.10	0.000	0.00
3	0.108	0.20	0.026	0.18	-0.058	0.16
5	0.080	0.21	0.000	0.19	-0.029	0.19
6	0.054	0.21	-0.024	0.20	-0.024	0.14
7	0.084	0.20	-0.077	0.14	0.000	0.00
8	0.029	0.15	-0.033	0.18	-0.400	0.14
Total	0.082	0.21	-0.002	0.19	-0.035	0.18

The results in Table 3 indicate a gain in score for subpopulation 2 ($M_{d,2}$=0.082) i.e. those who had taken the SweSAT two times. Subpopulation 3 and 4, on the other hand, scored lower compared to the

highest of their earlier valid results ($M_{d,3}$= -0.002; $M_{d,4}$= -0.035). The standard deviation was about the same in each subpopulation (s≈0.20).

There is also a slight tendency towards males gaining more than females, and that test takers in age category 30-39 gained more compared with the other age categories. With reference to educational background, the tendency is that category 3 (upper secondary education, two years) gained more than the other categories.

The model is also used to describe the fact that an applicant who has more than one valid SweSAT score, is allowed to use his or her highest score in the selection procedure, even though the obtained score at the latest SweSAT taken is lower. It is the highest valid score that is used in the selection process and this score is labelled x_{max} for test takers who have more than one valid score. In Table 4 below a distinction is made between test takers who have increased their score and those who have not. The label x_{max} in Table 4 stands for those test takers who did not increase their score at the next test occasion.

Table 4. Mean (M), standard deviation (s), total number (N) and percentage (%) for the best previous obtained result (X_{max}) and the result at the latest test occasion (X_2, X_3, X_4) for test takers who have taken the SweSAT 2, 3 and 4 times respectively.

	Number of SweSATs					
Variable	**2 (n=4,096)**		**3 (n=1,086)**		**4 (n=204)**	
	$\mathbf{X_{max}}$	$\mathbf{X_2}$	$\mathbf{X_{max}}$	$\mathbf{X_3}$	$\mathbf{X_{max}}$	$\mathbf{X_4}$
M	0.94	0.92	1.07	1.03	1.12	1.09
S	0.40	0.42	0.36	0.40	0.40	0.43
N	1,818	4,096	664	1,086	148	204
%	44.4	100	61.1	100	72.6	100

From this table we can see, for example, that 44.4% of the test takers, who had taken the test twice (two valid SweSAT scores), obtained a higher score on the first test occasion, i.e., 55.6% had a higher score (as compared to their first score) at the second test occasion. The mean score for the former group was 0.94 and the mean for the total group at the second test occasion was 0.92.

A more detailed description of the test takers in each category (2, 3 and 4 SweSATs) is presented in Table 5.

Table 5. Total number (N) and percentage (%) for the best obtained results, at a certain test occasion, for test takers who have taken the SweSAT 2, 3 and 4 times respectively.

	Number of SweSATs								
	2 (n=4,096)		**3 (n=1,086)**			**4 (n=204)**			
	Best result		**Best result**			**Best result**			
Variable	**1**	**2**	**1**	**2**	**3**	**1**	**2**	**3**	**4**
N	1,818	2,278	223	441	422	18	44	86	56
%	44.4	55.6	20.5	40.6	38.9	8.8	21.6	42.2	27.4

The relation between Table 4 and Table 5 is that the test takers in each category are divided into subgroups with reference to when they obtained their best result. For test takers in category “2 SweSATs” roughly the same information is presented in Table 4 and Table 5, i.e., 55.6% (n=2,278) obtained a higher score and 44.4% obtained the same or a lower score at the second test occasion. Analogously, test takers in category ”3 SweSATs” are distributed in the following way: 20.5% obtained their best result at the first test occasion, 40.6% obtained their best result at the second test occasion and 38.9% obtained their best result at the third test occasion.

The overall score-gains as a function of repeated test taking can also be described more in detail with reference to increase and decrease between the latest score as compared to the best of earlier scores in terms of tenth of normed score. There are great variations in gains on the individual level which is also indicated by the standard deviations for the differences seen in subpopulation 2, 3 and 4 (Table 3). The more detailed description is presented in Table 6 below.

The lines in Table 6 divide gains and decreases with reference to zero gain (=0). Thus, difference in normed score for test takers who had taken the SweSAT 2, 3, or 4 times was divided into three categories (negative, zero and positive) difference and the percentage was calcu-

lated for each category. This operation was carried out mainly for descriptive reasons.

Table 6. Difference in normed score for test takers who had taken the SweSAT 2, 3, or 4 times. Frequencies and percentage (%).

Difference in	Number of SweSATs						Total
normed score	2		3		4		
-0.7	2		2		1		5
-0.6	1		2		1		4
-0.5	19		9		2		30
-0.4	39	26.5%	31	39.4%	4	46.1%	74
-0.3	148		54		9		211
-0.2	337		117		28		482
-0.1	540		213		49		802
0	732	17.9%	236	21.7%	54	26.5%	1,022
0.1	785		194		33		1,012
0.2	660		135		12		807
0.3	429		66		6		501
0.4	238		21		3		262
0.5	108	55.6%	4	38.9%	1	27.4%	377
0.6	34		1		-		137
0.7	17		1		1		50
0.8	2		-		-		9
0.9	1		-		-		6
1.0	1		-		-		2
1.1	2		-		-		4
1.2	1		-		-		1
Total	4,096		1,086		204		5,386

A closer look at the extremes in Table 6 discloses, for example, that one test taker in subpopulation 2, i.e. a test taker who had taken the SweSAT twice, obtained a normed score gain of 1.2, but also that two

test takers, with exactly the same test experience, lowered their normed score by –0.7.

The overall tendency is, however, that the difference is related to the number of SweSATs taken. The meaning of this tendency can be described by two aspects. One aspect is that the proportion of test takers with zero difference (=0) increases as a function of repeated test taking (17.9%, 21.7%, 26.5%). The other aspect is that the proportions of test takers in the positive difference category (+) decreases, as compared to the corresponding negative category (-) that increases.

The percentages that are described in Table 6 can also be used as a basis for a kind of probability statement for losses and gains in SweSAT score as a function of repeated test taking. For example, the probability of getting a higher score at the second test administration (for those who have taken two SweSATs) is about 56% (p=0.556) and the corresponding probability of getting a lower score is about 27% (p=0.265). Analogously, the probability of getting a higher score at the fourth test administration (for those who have taken four SweSATs) is about 27% (p=0.274) and the corresponding probability of getting a lower score is about 46% (p=0.461). One conclusion from Gustafsson & Benjegård (1996) was also that the effect of repeated test taking gives gains in score and that the gains are diminishing with additional test taking.

In Table 7, a summary of five studies (including this one) is presented: Henriksson, 1990 (86B), Henriksson & Wedman, 1993 (91B), Henriksson, 1995 (93B), Henriksson & Törnkvist, 2002 (97B) and this study (02B). The common basis for all these studies is the model used to describe the effects of repeated test taking, i.e. a model where the effects are described in relation to the rules of selection.

Table 7. Mean difference and standard deviation for normed score (M_d, s_d) for test takers who had taken the SweSAT 2, 3 or 4 times. Summary of five studies (86B, 91B, 93B, 97B and 02B).

Difference in normed score	**Number of SweSATs**					
	2		**3**		**4**	
	M_d	**s_d**	**M_d**	**s_d**	**M_d**	**s_d**
86B	0.055	0.20	-0.011	0.18	-0.036	0.21
91B	0.083	0.20	0.009	0.18	-0.034	0.17
93B	0.111	0.21	0.031	0.19	-0.023	0.18
97B	0.093	0.21	0.010	0.20	-0.028	0.19
02B	0.082	0.21	-0.002	0.19	-0.035	0.18

The last row in Table 7 is identical to the last row in Table 3. The table gives a summary of the effects of repeated test taking during a sixteen-year period, and the summarised conclusion is that the main gain occurs from the first to the second testing.

It should also be noted that the SweSAT programme has changed during this sixteen-year period (Appendix 1). For example, the subtest STECH, measuring study techniques, was included in the studies of 86B and 91B, but not in the studies of 93B, 97B and 02B. The subtest ERC, measuring English reading comprehension, was included in the studies of 93B, 97B and 02B, but not in the studies of 86B and 91B. The subtest GI, measuring general information, was included in 86B, 91B, and 93B, but not in 97B and 02B. Moreover, the total number of items has changed during this period (86B=144, 91B=144, 93B=148, 97B=122 and 02B=122).

Another way of describing the summarised effect of repeated test taking regarding the SweSAT is presented in Table 8. This table is based on exactly the same studies as Table 7.

Table 8. Score gains, in terms of standard deviation units, from the first to the second SweSAT for test takers 86B (n=790), 91B (n=14,509), 93B (n=13,982), 97B (n=12,702) and 02B (n=4,096)

86B	91B	93B	97B	02B
0.121	0.186	0.246	0.207	0.180

Except for the first study (86B), the gain from the first to the second taking of the SweSAT has a range of 0.18–0.25. One difference between the first study and the other ones is that the test takers in the first study were at least twenty-five years old and had at least four years of work experience. This restriction with respect to age and work experience had been eliminated when the other studies were made, which explains the big difference in the number of test takers (cf. Appendix 1).

The effects of repeated test taking in relation to the test taker

Repeated test taking is a matter of self-selection (cf. Table 2). In order to separate the effect of self-selection from the effect of repeated test taking, we selected a sample (n=504) from subpopulation 3* with the same distribution of the normed scores as in subpopulation 1. Subpopulation 4 was used as a basis for sampling in Henriksson & Törnkvist (2002). This subpopulation is not used as a basis for sampling in this study, because of the low number of test takers (n=204).

As an effect of the sampling strategy, Table 9 shows that the normed scores for subpopulation 1 and the sample have approximately the same distribution, mean (M), and standard deviation (s). The mean and the standard deviation for subpopulation 3* (01B, 02A, 02B) on the first test occasion are 0.88 and 0.40 respectively, and the corresponding data of the sample are 0.82 and 0.42.

Table 9. The distribution of normed scores, mean (M) and standard deviation (s), regarding the first SweSAT taken by subpopulation 1 (n=14,959), subpopulation 3* (n=705), and the sample (n=504).

Normed score	**Subpopulation 1 (%)**	**Subpopulation 3* (%)**	**Sample**	
			(%)	**n**
0.00 – 0.49	21.8	15.7	21.8	110
0.50 – 0.99	41.5	42.1	41.5	209
1.00 – 1.49	27.9	33.6	28.0	141
1.50 – 2.00	8.8	8.5	8.7	44
Total (%)	100	100	100	504
M	0.81	0.88		0.82
S	0.44	0.40		0.42

A comparison between the effects of repeated test taking in the sample and in subpopulation 3* (Figure 1) indicates that the tendencies are the same for males and females. The comparisons between males and females and between different test administrations were controlled for age and maximum education in the fall of 2002. Analogously, the comparisons regarding educational background (maximum education) were controlled for age and sex.

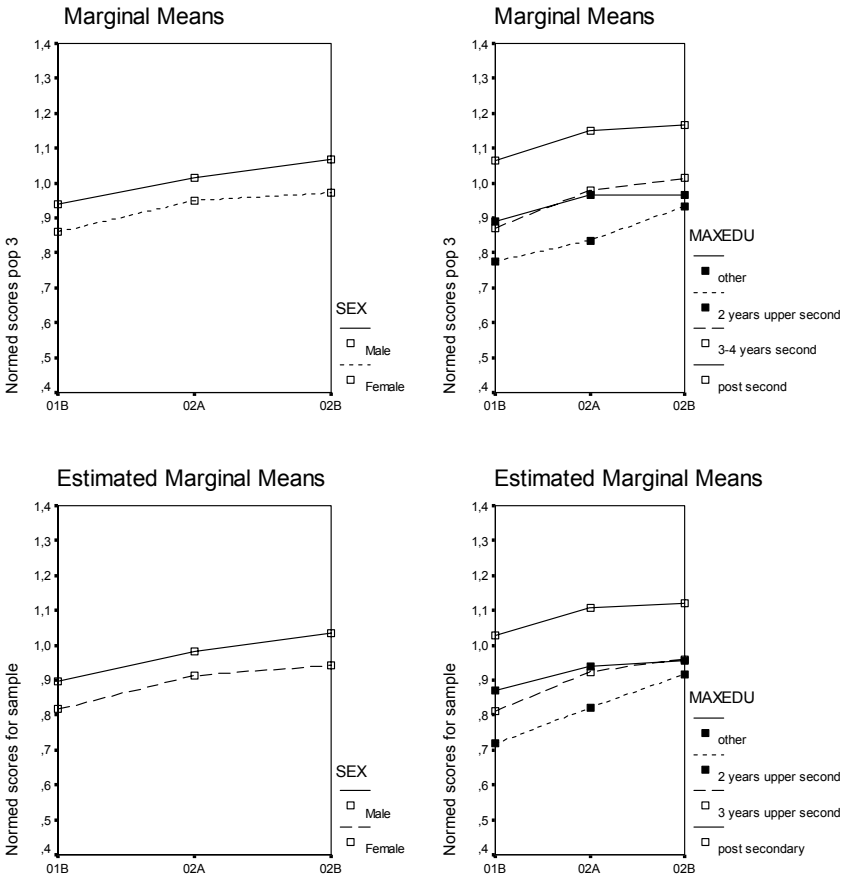

Figure 1. Normed scores (M) for test takers, grouped according to sex and educational background, in subpopulation 3 (n=705) and the sample (n=504). Evaluated at age 22.5.*

Further, a comparison of the effects of repeated test taking within the sample and subpopulation 3* indicates, in both cases, that a higher level of education gives a higher mean score (Figure 1).

For both sexes and all different educational backgrounds there were, except for education category "other" (n=17), significant differences between the mean scores at the first and the second test administration. The mean scores increase with the number of tests taken, but the increase gradually declines. The main benefit of repeated test taking appears from the first to the second taking of the test.

Subtests

The subtests that make up the SweSAT have different maximum scores (Appendix 1), and they are not being normed separately. Our study of the way in which repeated test taking affects the mean scores on different subtests controlled for this fact by using a calibrated score based on reference population 1 (cf. p. 7). The raw scores on each subtest at each test administration are shown in Table 10.

Table 10. Mean raw test score (M) and standard deviation (s) on each subtest for reference population 1 at test administration 01B, 02A, 02B and total mean (M_{tot}) on each subtest for all three test administrations (01B-02B).

Subtest	Test administration						
	01B		02A		02B		01B-02B
	M	s	M	s	M	s	M_{tot}
WORD	24.38	7.75	24.22	7.56	24.05	7.90	24.22
DS	9.98	4.08	9.89	4.20	9.40	3.80	9.76
READ	11.47	3.63	11.90	3.98	12.01	3.76	11.79
DTM	10.27	3.75	10.66	3.72	9.38	3.60	10.10
ERC	10.70	4.27	10.47	4.46	10.67	4.40	10.61
Total raw score	66.80	18.57	67.14	18.96	65.51	18.78	66.48
Total normed score	0.93	0.48	0.93	0.48	0.91	0.48	0.92

The data in Table 10 indicate certain differences in mean total raw score (M_1=66.80, M_2=67.14, M_3=65.51) for the three test administrations. But, as mentioned before, it is the normed score that is used in the selection procedure and the corresponding mean normed scores were M_1=0.93, M_2=0.93, M_3=0.91. There is a little drop on the third test occasion and a closer look at Table 10 reveals that the subtest DTM contributed the most to this drop. The effect of the calibration procedure on the subtests appears in Figure 2 and 3.

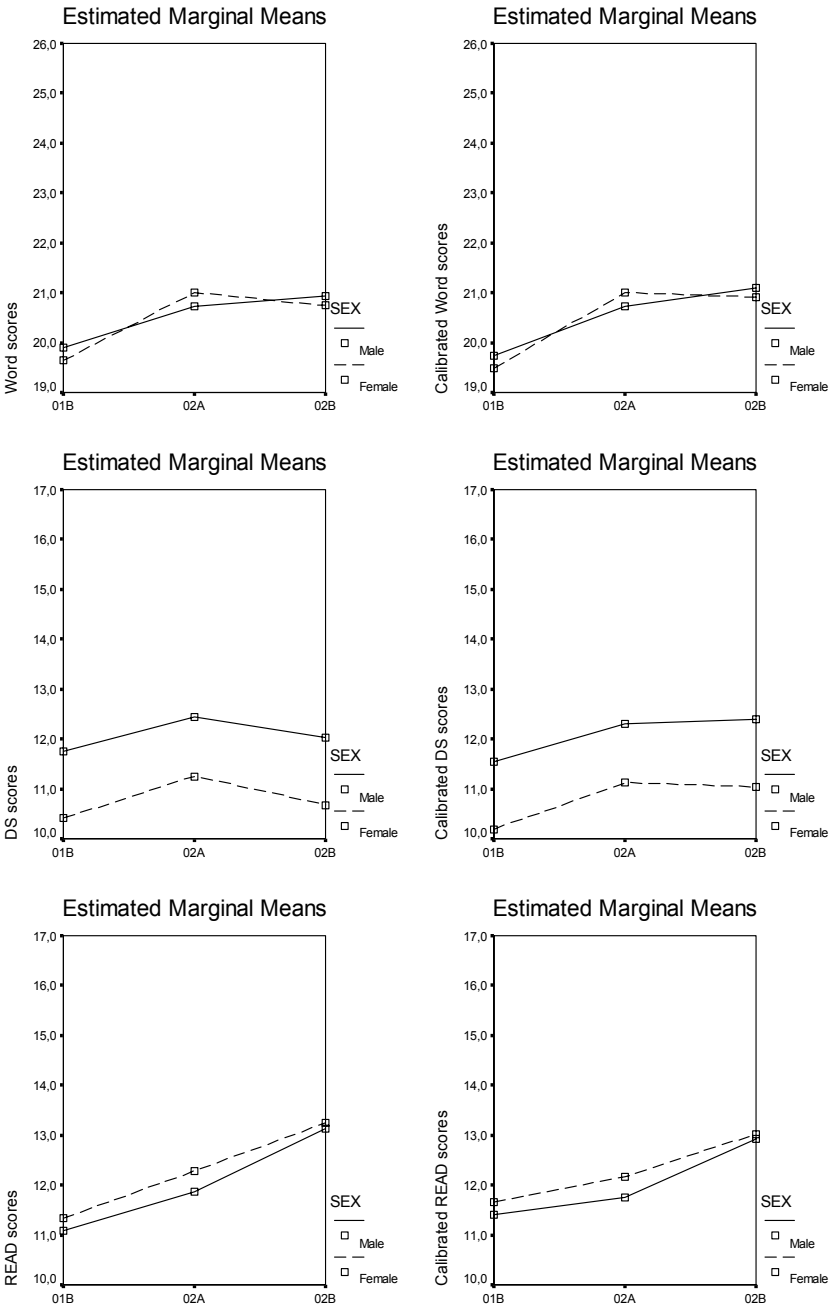

Figure 2. Raw scores and calibrated scores (M) for male and female test takers on the subtests WORD, DS, and READ. Evaluated at age 22.5.

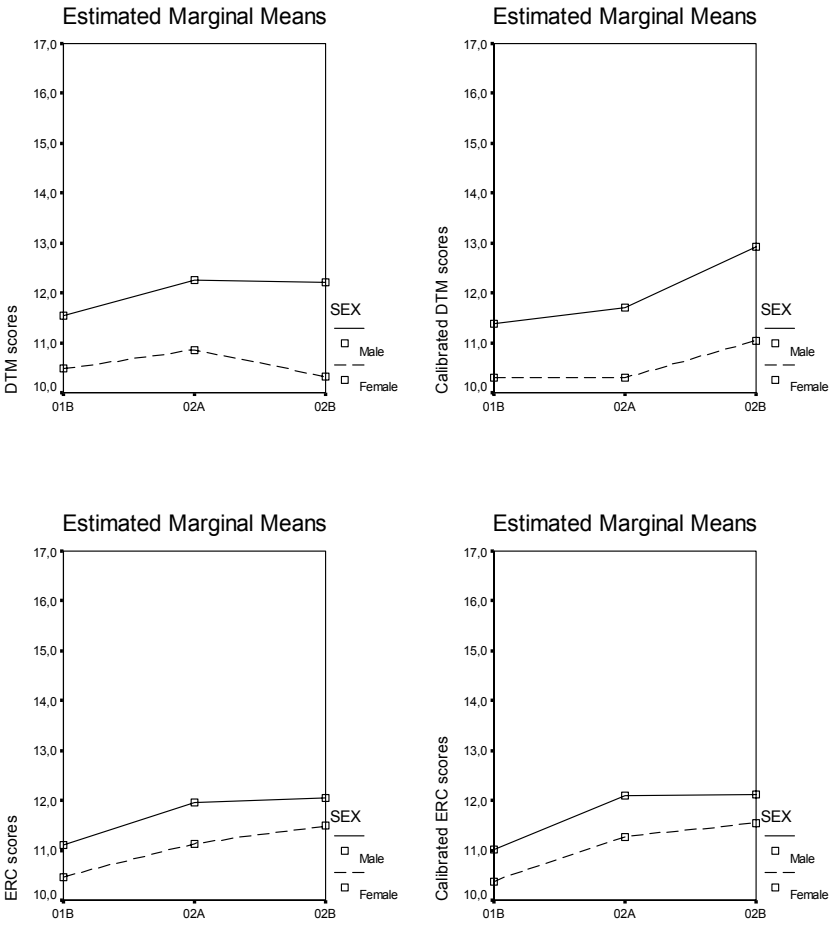

Figure 3. Raw scores and calibrated scores (M) for male and female test takers on the subtests DTM and ERC. Evaluated at age 22.5.

The graphs show that the calibrated WORD, DS and ERC test scores reveal similar trends as the total test scores (Figure 1 compared to Figures 2 and 3). Another observation, which is valid for the total test score as well as for these subtest scores, is that the main benefit of repeated test taking occurs from the first to the second test occasion.

The two remaining subtests (DTM, READ) show a different pattern regarding the effects of repeated test taking, i.e. the gain between the second and the third test occasion is higher than the gain between the first and second test occasion.

When relating the effects of repeated test taking to sex, the main observation is that there were significant differences between males and females on subtests DS and DTM ($p<0.001$). There are no differences between males and females on subtests WORD and READ.

These tendencies can also be described by differences in mean subtest scores between different test occasions. These data are presented in Table 11.

Table 11. Effects of repeated test taking in terms of mean score gain (M) between different test occasions.

Subtest	Score gain (M) $1 \rightarrow 2$	Score gain (M) $2 \rightarrow 3$
WORD	1.3*	0.1
DS	0.9*	0.0
DTM	0.2	1.0*
READ	0.4	1.0*
ERC	1.0*	0.1

* $p<0.05$

DISCUSSION

The purpose of the present study was to describe the effects of repeated test taking in relation to the test taker, to the rules for selection and to subtests. Another purpose was to relate the results of this study to earlier studies of repeated test taking for the SweSAT.

The results in this study indicated that there was a significant increase in normed score between the first and second test occasion and that further gains, as a function of further test repetition, are not as large. The summarised conclusion is that the mean score gain between the first and second test occasion is approximately 0.1 normed score

points. But, on the individual level, there are great variations in increases and decreases.

These findings are, on the whole, in congruence with earlier studies using the same methodology (Henriksson, 1990; Henriksson & Wedman, 1993; Henriksson, 1995; Henriksson & Törnkvist, 2002).

When the gain in score between the first and second test occasion in this study was described in terms of standard deviation units the gain was 0.180. In the perspective of international studies based on the SAT (Messick, 1980; Cole, 1982; Donlon, 1984; Bond, 1989; Becker, 1990; Powers & Rock, 1999) and on the ACT (Andrews & Ziomek, 1998) the magnitude of increase between the first and second test occasion is about 0.20-0.25 standard deviation units.

The conclusion is also that, in the perspective of earlier studies for the SweSAT, the gain in this study is less than the gain that was observed in earlier studies, with the exception of the first study (Henriksson, 1990). Two factors, and the interaction between those factors, are plausible explanations for these differences in gain score for the SweSAT between these studies reported on. One factor is variations in the SweSAT programme and the other factor is the characteristics of the population of test takers.

The SweSAT programme has changed. It was exactly the same subtests in the first two studies (Henriksson, 1990; Henriksson &Wedman, 1993) but in the third study (Henriksson (1995) the sub test STECH (study techniques) was replaced with ERC (English reading comprehension). But, notwithstanding the fact that it was exactly the same subtests in the first two studies referred to, it remains to be explained why the gain was less in the first study.

In the first study (Henriksson, 1990) the test takers were twenty-five years or older and had at least four years of work experience (25-4). A reasonable assumption is that this population was less test-sophisticated than the population of test takers in the 93-study (and the other studies reported). In those latter studies about 80% of the population consisted of students from upper secondary school. But the problem is that this circumstance ought to give the opposite result, i.e., that the 25-4 category of test takers, as compared to the test takers from upper secondary school, should have gained more from repeated test taking. Another, but maybe less plausible hypothesis, is that the

students from upper secondary school were more able and, as a consequence of this fact, gained more from unsupported practice. Still another hypothesis is that the gain score for students from upper secondary school includes a growth-component, i.e., their knowledge and ability has changed between test and retest as a function of schooling and this, in turn, results in a higher retest score. This hypothesis seems reasonable and it is supported by the fact that the SweSAT score is highly related to education (Stage & Ögren, 2002).

Given the fact that the population of test takers in all studies, except the first one, mainly consisted of students from upper secondary school who repeated the SweSAT during their period of schooling, it still remains to explain why the gain between their first and second SweSAT reached a maximum in the third study (Henriksson, 1995, cf. Table 8). One plausible explanation is that the subtest STECH (study techniques) was replaced by the subtest ERC (English reading comprehension). This latter test is, unlike STECH, curriculum-related and, as a consequence, affected by schooling. But then the question is: why is there a decrease in the following two studies? The main reason seems to be that the SweSAT programme has changed. The subtest GI (General Information) was taken away, the total number of items decreased from 148 to 122 and the relation between verbal (WORD+READ+ERC) and numerical items (DS+DTM) was changed. Taken together this implies a main change in structure of the SweSAT and it may also be an explanation for the decrease in gain score between the first and second test. The decrease in range reduces variation and the scenario for changes. The increased importance of verbal ability, which in most cases is changed gradually over a long time-span, also influences the probability for score changes at retest when the time-span is six months.

From the perspective of earlier studies the gain between the first and the second SweSAT taken is also less in this study than the gain that was observed in the study that preceded this study (Henriksson & Törnkvist, 2002). One possible explanation is that the number of test takers is reduced by about 40% if we compare the number of test takers at the fall administration of the SweSAT 1997 and 2002, i.e. that it is a matter of selection.

When the effects of repeated test taking in this study were described in relation to the test taker the main result was that the gain in score is highest between the first and second test occasion. This means that the

two approaches which are used in this study, the rules-for-selection-approach and the test-taker-approach, give approximately the same result. The summarised findings are also that there is no interaction with sex and, on the whole, no interaction with education. This latter finding is not in agreement with findings from other studies (Henriksson, 1990; Henriksson & Wedman, 1993; Henriksson, 1995) indicating that the gain from repeated test taking is related to ability. The finding in this study and the 97B-study (Henriksson & Törnkvist, 2002) gives support to the statement that ability to gain from practice is about the same for all education categories, i.e., that gain is not related to level of education.

One explanation for the conflicting results can be the problem of self selection, i.e., the fact that it is the test taker's own decision whether to repeat the SweSAT or not. In this study, and the study by Henriksson & Törnkvist (2002), the problem of self selection was controlled for by studying the effects of repeated test taking in a randomly selected sample. Nevertheless it should be noted that this conclusion is based on the assumption that the degree of ability and level of education are related to each other.

When observed score gains in this study are related to subtests, expressed in calibrated raw score and described in relation to the test taker, the tendency is that WORD, DS and ERC are most susceptible to the effects of repeated test taking. The finding that DS is effected by repeated test taking is in accordance with findings from other studies (Henriksson, 1981a; Henriksson & Bränberg, 1994; Henriksson & Törnkvist, 2002). The main reason is that DS item format is rather complex and that many test takers are not acquainted with this item type before the first testing.

To explain the score gains for the WORD test by referring to complexity of the item format is less plausible since the WORD subtest has a relatively simple format. It is also an item format that is familiar from school and vocabulary is also highly related to education. Therefore, the most plausible explanation is that the gain is an effect of schooling, i.e., the comprehension of words and concepts increases during the period of schooling and this period coincides with the period of repeated test taking.

Concerning the results and findings in this study it is also worth mentioning that the low number of test takers (n=204, p 12) in subpopula-

tion four did not allow for a description of changes between four test occasions. Instead, the changes over three test occasions were described. Taking into consideration that the results in this study are rather consistent with the results of earlier studies, it is nevertheless important to repeat this study on new data in order to get more reliable estimates of the effects of repeated test taking.

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

Alderman, D.L. (1981). Student self-selection and test repetition. *Educational and Psychological Measurement, 41,* 1073-1081.

Andrews, K.M., & Ziomek, R.L. (1998). *Score Gains on Retesting with the ACT Assessment* (ACT Report No 98-7). Iowa City, IA: ACT.

Becker, B.J. (1990). Coaching for the Scholastic Aptitude Test: Further synthesis and appraisal. *Review of Educational Research, 60*, 373-417.

Bond, L. (1989). The effects of special preparation on measures of scholastic ability. In R. Linn (Ed.), *Educational Measurement.* Third Edition. New York: MacMillan.

Cliffordson, C. (in press). Effects of Practice and Intellectual Growth on Performance on the Swedish Scholastic Aptitude Test (SweSAT). *European Journal of Psychological Assessment.*

Cole, N. (1982). The implication of coaching for ability testing. In A.K. Widgor & W.R.Garner (Eds.), *Ability testing: uses consequencies and controversies.* Washington DC: National Academy Press.

Donlon, T.F. (1984). *The College Board Technical Handbook for the Scholastic Aptitude and Achievement Tests.* New York: College Entrance Examination Board.

Gustafsson, J-E., & Benjegård, M. (1996). Högskoleprovet, social bakgrund och upprepad provtagning [The SweSAT, social background and repeated test taking]. In *Högskoleprovet. Genom elva forskares ögon* [*The SweSAT. Through the eyes of eleven researchers*] (Rapport 1996:22R, 42-49). Stockholm: Högskoleverket.

Henriksson, W. (1981a). *A review of studies dealing with practice, coaching and individual characteristics related to test performance*. Department of Education, Umeå University.

Henriksson, W. (1981b). *Effekter av övning och instruktion på test prestation. Några empiriska studier och analyser avseende övningens och instruktionens betydelse för testprestationen* [The effects of practice and coaching . Some empirical studies about the effects of practice and coaching for the test takers obtained score] (Doctoral dissertation). Umeå universitet.

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., Henrysson, S., Stage, C., & Wedman, I. (1985). *Prov för urval till högskolan* [Test for selection to university] SOU 1985:59.

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.

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. (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., Kriström, M., & Lexelius, A. (1985). *Meritvärdering och studieprognos. Några undersökningar av antagningssystemets effekter* [Evaluation of merits and academic performance. Studies of the effects of selection regulations] (Arbetsrapporter från pedagogiska institutionen, Nr 21). Umeå universitet, pedagogiska institutionen.

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.

Messick, S. (1980). *The effectiveness of coaching for the SAT: a review and analysis of research from the fifties to the FTC* (Research Report No. 80-8). Princeton, New Jersey: Educational Tewsting Service.

Messic, S. (1981). *Issues in effectiveness and equity in the coaching controversy: Implications for educational and testing practice* (Research Report, No 81-19). Princeton, NJ: Educational Testing Service.

Millman, J., Bishop, H., & Ebel, R. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement, 25,* 707-726.

Powers, D.E., & Rock, D.A. (1999). Effects of coaching on the SAT I: Reasoning Test Scores. *Journal of Educational Measurement, 36*(2), 93-118.

Rogers, W.T., & Bateson, D.J. (1992). The influence of test-wiseness on performance of high school seniors on school leaving examinations. *Applied Measurement in Education, 4,* 159-183.

Roznowski, M., & Basset, J. (1992). Training test-wiseness and flawed item types. *Applied Measurement in Education, 5,* 35-48.

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

Wedman, I. (1994). The Swedish Scholastic Aptitude Test: Development, use and research. *Educational Measurement: Issues and Practice, 13*(2), 5-11.

Wedman, I. (1995). Selection to higher education in Sweden. In T. Kallaghan (Ed.). *Admission to higher education: Issues and Practices*. Educational Research Centre, Dublin & International Association for Educational Assessment, Princeton, New Jersey.

APPENDIX

APPENDIX 1

The SweSAT programme 1986, 1991, 1993, 1997 and 2002.

Subtest	Type of test	86B	91B	93B	97B	02B
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	28,589

APPENDIX 2

Classification of education

---

1 = Compulsory school (lower secondary school)

2 = Folk high school

3 = Upper secondary school, 2 years

5 = Upper secondary school, 3–4 years

6 = Post secondary education, ≤80 credits

7 = Post secondary education, >80 credits

8 = Data missing

APPENDIX 3

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.