The Swedish national aptitude test:
A 25-YEAR TESTING PROGRAM

current status and future development

Högskoleverket 2002

Högskoleverket (the National Agency for Higher Education) •
Birger Jarlsgatan 43, Box 7851 • SE-103 99 Stockholm •
phone: +46 8 563 085 00 • fax: +46 8 563 085 50 • e-mail: hsv@hsv.se •
www.hsv.se

The Swedish national aptitude test: a 25-year testing program
Current status and future development

Produced by the National Agency for Higher Education in July 2002
National Agency for Higher Education Report Series 2002:22 R
ISSN 1400-948X
Content: National Agency for Higher Education, Department of Research and Policy Advice, **Nils Olsson**.
Grahic design National Agency for Higher Education, Information Department
Printed by the National Agency for Higher Education, Stockholm, August 2002

TRYCKT PÅ MILJÖMÄRKT PAPPER

Contents

Summary

The Swedish national university aptitude test, SweSAT, has been used to select applicants for higher education programmes since 1977. From 1977 until 1990 the test was only used for the selection of individuals who lacked comparable school-leaving qualifications but had attained the age of 25 and had at least four years of vocational experience. In 1991, however, all applicants were given the opportunity of taking the test. Since 1991, therefore, applicants have been able to use both results from the test and their school-leaving grades in their applications for higher education programmes.

In connection with its 25th anniversary in 2002, it was decided to conduct an expert appraisal of the test. The aim was to allow leading international experts in the field of testing to study the national university aptitude test and assess whether it could still be regarded as an effective instrument for selection and also make concrete recommendations about its future.

The panel of experts considers that on the whole the quality of the national university aptitude test and the research linked to it is high. They also consider that it is very important to continue research about the test. At the same time the panel points out that even though the quality of the current test is excellent, the purposes it serves have changed since 1991. It is no longer being used today, as it was prior to 1991, only for applicants who lack school-leaving grades but is now open to all qualified applicants as well. The panel considers that this widening of its use is a good reason for beginning to investigate alternative forms of test construction.

The panel gives the following recommendations for the future use of the SweSAT. At the moment applicants for enrolment in higher education in Sweden can compete either on the basis of their school-leaving grades or their results from the national university aptitude. The panel considers that analysis should be made of what consequences a method involving combination of the two sets of results to provide a total sum of points would have on enrolment.

The panel also comments on the current use of school-leaving grades, in which all the grades are totalled to produce a median value. This value is then used uniformly, or in other words in the same way, for basically all forms of programme. The panel considers that consideration should be given to the possibility of weighting grades in various subjects differently for different programmes. The reason given is that this would make it possible to enrol more appropriate students for the various programmes.

The group also considers that it should be possible to weight the results of the national university aptitude test in the same way, for instance by dividing the total score into a verbal section and a mathematical section.

The results of the national university aptitude test could be used to inform testees of their strengths and weaknesses to a greater extent than is the case today.

In this context it should also be made clear what ability or abilities were being measured in each of the subtests.

It should be possible to use the ability profiles obtained on the basis of school-leaving grades and test results in attempts to even the gender balance and other disparities (in terms of social group, for instance) in programmes.

It would be desirable to investigate the possibilities of using more achievement based tests for enrolment. Tests of this kind should, as in the current national university aptitude test, comprise tasks that involve analytical thinking but should be linked to a greater extent than today to specific disciplines, such as mathematics.

The panel refer to some advantages of computer-based testing. The use of computer technology simplifies the administration and marking of tests that evaluate written production. In addition, modern technology enables "adaptive testing", which means that tasks adapted to the capabilities of the testee can be selected from a battery of items. A third advantage of using computers is that it enables the design of tests for individuals with functional disabilities for whom paper-and-pen tests present practical difficulties.

Finally the panel considers that the test should be studied from points of view such as cost-effectiveness and its impact on the educational system as a whole. This means, for instance, that the cost of basing selection on the test should be compared with the use of other instruments.

The assessors

John Fremer is education director of the Global Institute and a senior development leader in International Projects at ETS in the U.S. His primary areas of expertise are national testing program development, testing standards, trends in testing, and communicating about testing to various audiences. E-mail: jfremer@ets.org

David F. Lohman is professor of education at the University of Iowa, U.S. His research interests are mainly individual differences in abilities and achievements. He also teaches courses in educational psychology and human intelligence. E-mail: david-lohman@uiowa.edu

Werner W. Wittmann is professor of psychology at the university of Mannheim, Germany. His research interests include evaluation research and assessment of intelligence and human performance. E-mail: www@tnt.psychlogie.uni-mannheim.de

Project manager and editor

Nils Olsson, National Agency for Higher Education, Sweden. E-mail: nils.olsson@hsv.se

Introduction

Nils Olsson

The present evaluation project started in November 2001 when we contacted leading experts in the field of educational testing. Several criteria applied to the choice of expert assessors. We wanted independent, foreign experts. They had to be familiar with the normal methods of evaluating selection tests, or in other words there had to be documentary evidence of their experience of assessing tests on the basis of standard criteria such as the accuracy of measurement and prognostic capacity. In addition we wanted the assessors to be familiar with what could be referred to as the macroperspective on selection tests, by which we mean the capacity to evaluate a test on the basis of the way in which it is used in the educational system as a whole, taking into account factors such as its economic benefits and its impact on the various sectors of the educational system involved.

The assessors were supplied with a translated version of the national university aptitude test that was currently being used. They also received about twenty research reports on the test and a summary of the most recent government bill on higher education "Open Higher Education". The panel also visited Stockholm and were given an oral presentation of the Swedish system of higher education by the National Agency for Higher Education.

The assessors received no explicit guidelines for their work, such as indication of which aspects of the test to concentrate on. Instead, we chose to allow them to focus on what they themselves considered important. The advantages of avoiding any direction of the assessors' attention to certain problem areas were considered to outweigh any possible risk of some aspects being disregarded. The experts' opinions on the test have been compiled in this report. We have decided to publish both the individual opinions of each expert and their joint comments. What follows is a review of the expert's conclusions. A few reflections on the evaluation follow the review.

Individual reviews

The assessors consider that the national university aptitude test and the research and development that have gone into it are of a very high standard from an international point of view. However, the assessors consider that enrolment procedures in Sweden should be differentiated to a greater extent than is possible using the current aptitude test. Enrolment would become more effective, in their opinion, if the aptitude test were supplemented by, or in some cases even replaced by, several sub-sections which could be weighted differently for enrolment to different kinds of programs.

Changes in the Swedish enrolment system that fit in with the assessors' proposals

have admittedly been introduced recently into the Higher Education Ordinance and are in the process of being incorporated to an even greater extent in the latest government bill "Open Higher Education" (2001/02:15). The most important consideration here is that the current regulations on enrolment allow "selection to be based on specific tests other than the national university aptitude test…" (Higher Education Ordinance 7:15). However, tests of this kind may only be used in connection with "programs offering qualifications for professional areas requiring certain personal characteristics or special competence, or programs with an artistic specialisation". In addition in "Open Higher Education" it is stated that "A provision allowing alternative selection to ten per cent of beginners' places should be introduced." (p. 65). Even though these amendments concur with the recommendations of the assessors, they still consider that it is important to go further and explore the possibilities of differentiated enrolment procedures in more detail. It should, however, be noted that none of the assessors explicitly recommends the use of interviews as a method of selection. Instead their ideas revolve around various types of ability tests and achievement tests.

The following section contains a brief review of the opinions of each of the assessors.

John Fremer considers that from an international point of view the standard of the national university aptitude test is very high. However, he proposes a number of potential developments. Fremer begins by pointing out that in the United States counterparts to the national university aptitude test are used as one element in a major overall assessment of each applicant. There are, however, great variations between universities, some combine test results and school-leaving grades, which Fremer believes should be tested in Sweden as well, other also use essays, letters of recommendation and interviews, in procedures that increasingly resemble normal staff recruitment methods. Fremer's proposed methods of extending the test include, for instance, testing written skills. Current technology would enable use of the Internet to distribute essays for grading, and in addition software exists that can grade essays directly. Like David Lohman writes below, Fremer considers that a supplementary achievement test would be suitable for some programs. Fremer also refers to the use of the test for diagnostic purposes, with feedback being provided to students about their strengths and weaknesses.

Further research should, in Fremer's opinion, also study the links between scores on the subtests and success in various programs and courses, which could reveal variations in the prognostic value of different subtests for different programs. Fremer also considers it important to investigate whether the national university aptitude test offers a real possibility for students with various forms of functional handicap to do themselves justice while taking it.

David Lohman considers that the national university aptitude test is one of the best tests of its kind in the world at the moment. Lohman does, however, suggest a few changes. Today the test measures general study skills and Lohman would like to add specific achievement tests, such as a mathematics test for those apply-

ing for natural science programs. Lohman points out that even though these tests are expensive to produce and that achievement tests give rise to greater variation between different social groups they also have advantages: the prognostic value of the test increases and the risk of categorising an individual as deficient in study skills is avoided. This last observation means that individuals can find it more crushing if a test indicates that they lack the general aptitude for study rather than that their knowledge of mathematics is inadequate, a situation that can be rectified.

Another idea that Lohman proposes is that of "weighting" the current test in terms of the programme applied for, for instance ascribing more significance to its verbal elements for applicants for the humanities. However, for technical reasons, a weighting system would require the development of additional subtests. Lohman is also of the opinion that school-leaving grades should be weighted as well. In addition, he would like the current test, if the this type of test of general study skills is retained, to be extended to include a test of spatial and perceptual abilities. This leads Lohman to advocate a test with three dimensions—verbal, quantitative and spatial—instead of the two dimensions the current test is normally considered to measure, the verbal and the quantitative. Items with at least three different formats should be used to test each dimension. On the subject of further research about the test, Lohman considers that it is particularly important to study other criteria of success than the number of credit points attained during the first semesters in higher education, citing, for instance, important vocational abilities such as the capacity to solve problems and creativity.

Werner Wittmann considers that the elements of the test that to the largest part deal with non-verbal abilities, such as the items that involve interpreting maps, tables and diagrams, are particularly commendable. These items correspond more than well to the tasks individuals are faced with when they enter the labour market, in his opinion. Wittmann agrees with Lohman in believing that in any future development of the test it would be desirable to pay more attention to the ensuring that the national university aptitude test (like school-leaving grades) measures not only one dimension, such as verbal abilities, but other dimensions as well, for instance the ability to solve problems. Differentiating test scores (or school-leaving grades) could not only improve their prognostic value for selection but also enable them to be used in student counselling, and teachers in higher education should also be informed of students' results on the various subtests. Specified test results would, for example, enable students and teachers to find out about individual strengths and weaknesses. Wittmann has more faith in this way of recruiting students from backgrounds with little contact with higher education—i.e. the use of a test that indicates ability profiles—to engineering programs, for instance, than methods that involve setting quotas for under-represented groups. One consequence could also lead to individuals from these backgrounds being inspired by others, from their own social group for instance, who have been successful in their studies.

On the whole Wittmann is positive to selection tests of the national university aptitude test type for yet another reason. The aptitude test and school-leaving

grades are often said to have higher prognostic value for successful study than other methods of selection, but Wittmann also points out that this greater prognostic value—which can moreover be improved even further by differentiating test results—results in considerable economic gains in the form of a reduction of the number of drop-outs. In this context it can be pointed out that the relatively low fee charged for taking the test (compared for instance to what a system based on interviews would cost) should also appeal to Wittmann. In Appendix 1 Wittmann also presents the findings of his own research about the national university aptitude test, which show that even though women do less well than men as a group in the quantitative subtests nevertheless a much smaller proportion of the women who do perform well in these subtests elect to study engineering programs. The low proportion of women in the engineering professions can therefore be attributed to a great extent to interest and stereotypes . In Appendix II Wittmann demonstrates the other potential uses of the national university aptitude test referred to above, for student counselling for instance. In Appendix III Wittmann studies David Lohman's proposed addition of a subtest on spatial perceptual abilities. Wittmann demonstrates that adding a subtest of this kind would not increase gender-related variation in the results, which is somewhat surprising.

A few reflections on the evaluation

Evaluating a test like the national university aptitude test with a relatively long history and which has been and is being taken by large numbers of testees is a very complex task, as is evident from the experts' assessment, which covers a number of different aspects of the test. This is based on the fact that the impact of a test which is as comprehensive as the national university aptitude test will vary in many ways on those that can be described as its potential users, or in other words the groups of students, public agencies, professional groups in the labour market, etc. who are affected in some way by its outcomes. Some of these effects may appear to be easy to identify, such as the way in which results vary for the men and women who take the test, but others are more indirect and difficult to discern, such as, for instance, what effect the very existence of the test has had on social bias in enrolment to higher education. The test and enrolment have been the topic of constant debate. However, it is far from easy to trace any simple connection. In the case of social bias, the problem is that it is difficult to isolate the impact of the test in view of the multiplicity of other factors that affect an individual's readiness to apply for higher education. The opportunity offered by the test in the form of a second chance for those who cannot compete on the basis of their school-leaving grades is often one of the arguments advanced for the existence of the national university aptitude test. Even though this second chance has at the same time definitely played a role in reducing social bias, it is more than probable that financial considerations, both during studies and afterwards, are much stronger motives for most students enrolling on programs. Perhaps, therefore, it would be more reasonable to go no further than to demand

that the administration of a test like the national university aptitude test should not discriminate against any specific group of individuals. If the aim is to extend recruitment, more powerful instruments are needed than those provided by the national university aptitude test.

Having said this about the impact of the test on recruitment, some mention the yardsticks used in evaluating testing activities may be of benefit to the reader. There are a number of different yardsticks or criteria that can be justly invoked in evaluating testing activities but which ones to emphasise, or in other words which objectives it is most important to attain, is the subject of never-ending discussion. The most frequent criteria have been summarised in an earlier evaluation study of enrolment to programs in medicine in about twenty European countries. The authors of this study determined eight criteria that are relevant in evaluating testing procedures (Ebach & Trost, 1997),

1) Objectivity
2) Reliability
3) Validity
4) Acceptance
5) Economy
6) Impact on various groups of testees
7) Justice
8) Impact on upper-secondary education

It can be observed in this context that the National Agency for Higher Education's Guidelines for the construction of the national university aptitude test (see Appendix IV) includes most of the above criteria or at least an equivalent wording of them. Where the Swedish national university aptitude test is concerned, this means in practice that each constructor has to consider whether each sub-objective (or criteria) has been attained, which can well be more complex that would appear from a first glance at the guidelines. This complexity is due to the conflict of objectives that may arise if too much weight is given to any one criteria. For instance, to make high prognostic value the first priority in a test like the national university aptitude test would mean forsaking other criteria, especially perhaps criterion 6 in the list above—impact on various groups of testees.

This can be demonstrated by a concrete example. One conflict of objectives arises when vocational experience is added to the score from the national university aptitude test. *In the current enrolment procedure this is done by adding 0.5 to the score for an applicant who can give evidence of 5 years of vocational experience.* It has in fact been demonstrated that this addition does not raise the prognostic value of the test but does, on the other hand, attract new groups of students to higher education. The fact that credit is allowed for vocational experience (with points for vocational experience being added to the national university aptitude test score) should not therefore be seen primarily as a measure intended to increase prognostic value but more of a way of attracting more groups into higher education[1]. In his chapter David Lohman refers to a similar conflict of objectives that

arises in the choice of type of test, as studies have shown that specific subject-related tests normally have higher prognostic value than general ability tests but at the same time they give rise to larger variation between groups with different social backgrounds.

Apart from the existence of conflicts of objectives, which in other words reveal that it is difficult to construct a test that can live up to all the desires expressed by different stakeholders (such as higher education institutions, students, vocational categories) mention should also be made of the contextual factors that have to be taken into account when evaluating a test. The term contextual factors can be used to refer to aspects such as the nature of the student population, the pressure for places in specific programs, the costs incurred in enrolling students who fail to complete their studies, etc. Here again a concrete example will be useful. If the pressure for places is very large, with one in every ten applicants being accepted for instance, and a programme is very expensive, then obviously the prognostic value of the selection instrument plays a major role. Here it is possible to question, as the evaluators do, whether the total score on the national university aptitude test provides enough information about an applicant's capacity to complete the programme. This applies in particular if selection is based on differences of only a few tenths of a norm-related point between applicants. Very high pressure for places in expensive programs is probably a more frequent scenario for universities in the United States than for higher education institutions in Sweden.

Where amendment of the national university aptitude test is concerned, it is of course difficult to get an overall impression of all the consequences of any possible changes in its construction. One method that has proved to be serviceable in this context, or in other words to gauge the impact of possible modifications, is to use simulated enrolments to test what kind of impact a change is likely to have. For example, we can study whether the use of separate norms for subtests in the national university aptitude test has any effect on which students are enrolled. Before separate norms can be used on the current test, however, another subtest will be required as the quantitative elements only amount to 33 per cent of its total content. It can also be added in this context that a simulation recently made by the National Agency for Higher Education reveals that as many as 86 per cent of the students would still be admitted to a degree course in engineering even though only their scores on the quantitative elements of the aptitude test were taken into account.

Even though simulation can never provide a complete answer to the question of what choices students would actually make when new rules come into effect, as they may change the way the choose if there is a change in enrolment conditions, it does give some indication of what the outcome would be. If we want to continue to study separate norms it is also important, as D. Lohman asserts, to go further than merely study performance during the first years of a programme.

[1] A thorough discussion of conflicts of objectives in connection with selection for higher education can be found in Wolming (1999).

This what frequently happens today, mainly because of the desire to follow the same is individuals during the course of their studies, and in the later stages of a programme individuals are more likely to have been replaced by others.

Recently, however, researchers at Gothenburg, using a new statistical method, have shown that it is possible to monitor for a longer period in a meaningful way (Gustafsson, 2001; Svensson, Gustafsson, & Reuterberg, 2001). One interesting finding from following engineering students for five years was that quantitative abilities played a significant role during the first year but that verbal skills became progressively more important later on in the programme. *This finding indicates the importance of selecting relevant criteria for successful study and emphasises the importance of both quantitative and verbal abilities for successful completion of an advanced degree programme in engineering.*

A few reflections have been presented above about the criteria selected for the evaluation, as this subject was raised in the evaluation itself. A number of the other opinions expressed by the assessors could, of course, well be taken up here and commented on, but we are convinced that they will become the subject of discussion when the contents of the report are made public.

Finally, it is worth stating that this evaluation has produced many thought-provoking proposals from three very experienced assessors and we hope that their report will stimulate its readers to continue to discuss the fundamental questions about the most suitable instruments for selecting applicants for higher education.

References

Ebach, J. & Trost, G. (1997). *Admission to medical schools in Europe.* WS Druckerei, Bodenheim.

Gustafsson, J-E. (2001). *Latent variable growth and mixture growth modelling: basic principles and empirical examples.* Paper contributed at the Xth European Conference on Developmental Psychology, 2001.

Svensson, A., Gustafsson, J-E., & Reuterberg, S-E. (2001). *Högskoleprovets prognosvärde. Samband mellan provresultat och framgång första studieåret på civilingenjörs-, jurist- och grundskollärarutbildningarna [The prognostic value of the national university aptitude test. Links between test scores and success during the first year of degree programs in engineering, law and primary school teaching]* . Högskoleverkets rapportserie 2001:19 R. Stockholm: Högskoleverket.

Prop. 2001/02:15: *Den öppna högskolan*. Stockholm: [Government Bill 2001/02:15: Open Higher Education] Riksdagen.

Wolming, S. (1999). Ett rättvist urval? [A fair choice?] *Pedagogisk forskning i Sverige*, *4* (3), p. 245–258.

,

Foreword to the Swedish national aptitude test (SweSAT) evaluation reports

John Fremer, David F. Lohman and Werner W. Wittmann

All three of us were pleased and honored to have this review assignment. The issue of how to select individuals for entry into higher education is a very important one for a society. It is a complex and multi-sided enterprise.

We want to applaud the idea of carrying out an outside "audit" of the SweSAT program. Doing such a careful review and employing seems to us very consistent with modern theories of validation of uses of tests and to be a sound management practice.

Although the three of us did exchange e-mails about the areas that we planned to focus on and John Fremer and David Lohman met briefly at Educational Testing Service as part of a visit by Dr. Lohman to the ETS campus, our reviews were done quite independently.

We conclude that there is a great deal of consistency in our comments. We focused on somewhat different aspects of the SweSAT, but our different perspectives brought us to some common conclusions.

We agree that the SweSAT meets the highest standards for a major testing program. There is an exceptionally fine record of employing research on an ongoing basis to evaluate and enhance the exam and how it is used.

We are all struck by the change in the primary purpose of the exam and of the cohort who sit for it. When there is such a shift in the focus and population for a test, revisiting the foundation of the exam, how it is used, and its impact on students and their education is essential.

Ask a research trained and oriented group to review any substantive academic endeavor and provide them with existing research data and you can be sure that more research will be proposed. Each of us makes suggestions as to areas meriting further study. As has been the case in the evolution of SweSAT to this point, we very much urging trying out materials and approaches systematically before making changes. Sweden has a fine record of collecting data about students and programs. The SweSAT should continue in that tradition.

The folk saying "If it's not broke, don't fix it!" is worth mentioning here. There is a solid foundation of work on the current SweSAT that should be preserved wherever this is consistent with meeting your testing and selection goals. Change whatever will lead to substantive improvement, but keep whatever proves to be working well.

There is a price to be paid for change, beyond monetary costs, and we need to be sure that the price is consistent with the gains achieved.

Keeping the overall goals of SweSAT clearly in mind should help direct research

and operational decisions. What are the qualities that are being sought in students? What is the role of SweSAT in helping to select students with those qualities?

General findings and common recommendations

John Fremer, David F. Lohman and Werner W. Wittmann

1. The SweSAT is a high quality test with a solid research program. We applaud past research and encourage continued research on the test.

2. Uses for which the test was originally designed have changed. The initial purpose of the SweSAT was to provide an estimate of academic potential for young adults who had been in the workforce for several years, and whose academic grades were thought to underestimate the likelihood of their success in university study. Now the test is being used as an alternative selection mechanism for a substantial fraction of the applicant population. This requires a re-thinking of the nature of the test and the ways in which it is used.

3. The decision rule for how SweSAT scores and grades are used should be reconsidered. In the current model, grades and test scores are considered separately. A selection model that simultaneously considers grades and test scores should be investigated.

4. Grade-point-average could be decomposed. The predictive validity of secondary school grades is likely to be higher if universities took into account the subjects in which the grades were earned, possibly using different weightings of grades to predict success in different institutions or programs (e.g., giving more weight to math and science grades for predicting performance in science or engineering programs).

5. Separate scores for components of SweSAT. Predictive validities of SweSAT for particular courses of study are likely to be higher if separate scores were available for at least verbal reasoning and quantitative reasoning.

6. Explore the possibilities of giving diagnostic feedback for counseling and individual score interpretation—at least on the profile of scores that are reported, and possibly on the underlying skill classifications.

7. Consider strategies to establish better proportions of males and females (and minorities) in all areas by using the profile information of grades for different secondary school subjects (or clusters of subjects) and the scores for components of SweSAT. This includes investigating how well the

university curricula in the different areas could be re-designed to accommodate students with different ability profiles.

8. Investigate the possibility of adding more context-based measures of reasoning (i.e., achievement-oriented tests). Both context-reduced and context-rich measures of reasoning abilities can contribute usefully to decisions about the likelihood of future academic success. Assessing both is an especially attractive possibility if the test can be administered adaptively.

9. Explore the use of computer-based test.

10. In conjunction with (9), consider adding a computer-scored measure of writing ability.

11. Anticipate the need to make accommodations for the testing of disabled students.

12. A cost-benefit analysis of the test should be conducted. Information about the contribution of the test to the overall efficiency and success of the educational system need to be part of the discussion about its value.

Three individual evaluations of SweSAT

EVALUATION OF SWESAT BY JOHN FREMER

Introduction

I have spent the bulk of my career on the content/skills side of test development and on test program development and revision. So I will be drawing on my experience as a developer and manager of the revision of the SAT I (Verbal and Math Reasoning) and as a participant in a variety of worker and management roles in a large number of other test and test program development efforts. I have not been involved for many years in the psychometric side of SAT I analyses, for either operational or research work, outside of the area of studies of the validity of the exam for various populations. I will look carefully at test content and think about who takes the test and the role that the results play in decisions. I am not going to focus on analytic methods or any of the part of a review/audit that really needs the training and perspective of someone who is current on statistical and psychometric theoretical and operational work.

Comments

High Quality Test [2]

Reviewing the sample test questions and the background material about the SweSAT, it is clearly a high quality test that meets international standards.

I find the vocabulary items in the WORD portion of the battery a little less "tight" than I am used to seeing with such tests, but the description does say the same or "almost" the same meaning. I like to apply the approach of judging what word or words is likely to be the choice even before I review the options. For a number of the examples we were given, I had to pick the closest choice available, but without much enthusiasm. Some examples:

21 – Is "perishable" that close in meaning to "transitory?"

23 – Does "supply with clothes" mean the same as "equip?"

28 – Are "extreme" and "exaggeration" very close in meaning?

30 – Are "affront" and "disgrace" close in meaning?

I don't see my discomfort here as in any way undermining the function of this test or its overall quality, but there is a slight looseness of the keys that I would be trying to tighten up, if I were on the development team for the exam.

[2] This observation strongly supports David Lohman's Point 1.

Solid Research Program

The test developers and researchers have been carrying out a significant research program over the years. They address questions important to maintaining the vitality of the program and employ the results of their research to refine and enhance the battery and the way it is used. Moreover, the research papers show that the staff involved have been following the work done on other tests and programs and are open to learning from others and to sharing the results of their own studies.

Implications of Dramatic Changes in the Use Made of the SweSAT [3]

The SweSAT began and was used for a number of years primarily as a "second chance" program, affording another avenue to college admissions to older students who had moved on from secondary school without entering college. Now the battery is a basic part of the high school to college transition process and needs to be viewed in that light. We need to ask whether the design, recommendations for use, score reporting, research program, etc. are what we would implement if SweSAT were being initiated now. It is hard to step back and imagine a "fresh start," but this is what I would recommend. We need a plan for implementing a battery of tests to facilitate decisions about students moving from secondary school to college or more generally, decisions about students applying for college, regardless of whether they are moving on directly from secondary school. Who should be involved in developing such a plan? What criteria do we need to meet? These issues are indeed addressed to some extent in materials provided to us as reviewers.

Using SweSAT and Grades in Combination

The basic pattern of using the SAT I in the United States is to employ it either as part of a prediction equation with the student's high school record or as one element in a set of pieces of information that are viewed and weighed as a package. In the actuarial type uses, a college combines SAT I scores with high school grade point average (gpa) or rank-in-class and estimates first year college grades or some similar academic index. Often, in these settings, the only number that is retained for use in subsequent decisions about students is the index that has been computed. The test scores and high school information remains in a computer file not used by those making decisions.

The situation is very different in other US colleges, where a folder is prepared for every applicant into which is placed all the information available, including all test score and grade information, as well as student essays, letters of recommendation, reports of personal interviews, etc. This folder is examined carefully by all participants in decisions about the student.

Without specifying a particular model, using both grades and tests together to make decisions about students, rather than treating the information in two

[3] This observation supports David Lohman's Point 2.

different decision tracks should be investigated for the various groups of students of interest and for the variety of institutions and programs that constitute higher education in Sweden.

Adding Measurement of Writing Ability

The addition of a test of writing ability would broaden the skills coverage of the SweSAT, signal the importance attached to writing, and would also be likely to increase the relationship between test performance and college success. One would not take such a step lightly, as the task of scoring a large number of essays would be a formidable one, but this is the kind of step taken by a number of programs in the US. There are ways, moreover, of applying technology to facilitate the grading enterprise. Choices range from using the Internet to distribute essays to readers in remote sites to actually employing natural language processing systems to grade student papers. The idea of computers grading essays, once viewed as quite futuristic, has turned around dramatically in the US, so that one can now say that the future is now! I will bring some sample materials to our meeting. Other information can be found on a variety of web sites, including www.etstechnologies.com, the web site of a subsidiary of Educational Testing Service.

Adding Other Achievement Measures[4]

When standardized measures first began to be used for college admissions purposes in the US, in 1900, the exams were all subject matter tests, covering major high school subjects of that time. Such subject tests have remained in continuous use for over 100 years. The general ability exam, the SAT I (Verbal and Mathematical Reasoning Test), first introduced in 1926, grew to be the dominant exam in the transition from high school to college, but the use of the subject exams never went away and some of the most selective and prestigious colleges and universities in the US employ these exams. Adding achievement tests to the SweSAT would be another way to broaden what the tests measure. A high-level mathematics content exam would be an obvious addition as would be one or more science or humanities oriented tests. This direction would be one way of dealing with the interest of expanding the value of SweSAT for a variety of different Swedish higher education institutions.

Expanded Score Reporting

There has been a good deal of work in the US on providing answers to the question "what does a score of xxx really mean?" For the SAT I, for the Preliminary SAT / National Merit Scholarship Qualifying Test, for the US National Assessment of Educational Progress and for other exams, the managers of testing programs have striven to define the skills typically associated with performance at particular score levels. It then becomes possible to supplement total scores that depend totally on normative information to acquire meaning with descriptive

[4] This observation supports David Lohman's Point 3b.

score reports related to students' skills. These reports have proven very popular with many users, including students and their families.

Testing of Disabled Students

A significant part of the public debate over standardized testing in the US involves issues related to the testing of students with disabilities. Demonstrating that tests are accessible, that accommodations are available and appropriate, and that assistance with interpretation is provided are among the matters being addressed. It seems logical to review these same issues with respect to the SweSAT. To what extent are these issues being handled well for the program?

One resource that might be employed in investigating these issues is a book just published by the American Psychological Association "The Psychometrics of Testing Individuals with Disabilities." There are also a series of very helpful publications provided by the National Center for Educational Outcomes, http://education.umn.edu/NCEO.

Disentangling Grades[5]

In addition to examining the relationship between SweSAT performance and overall college performance, the link between scores on the battery and individual course grades should be investigated. In the US, it has been found that significantly higher relationships are observed between SAT I scores and individual course grades than is found with grade point averages. This phenomenon is attributable to the tendency of students, aided by guidance from faculty and others, to steer away from courses that they perceive to be "too hard" or otherwise inappropriate for them. The better the job that students and schools do to help students work with their skills and to compensate for their weaknesses, the lower the relationship between test scores and college grades. It is important to recall that this finding is a good one. When a test result suggests that a student might do poorly if thrown into the regular program, without special help, intervening to provide such help is a very constructive step. We want such predictions to turn out wrong and we should celebrate individual and college actions that disconfirm predictions of low student performance.

The study of individual course grades takes out of prediction studies some of the variation in college course attendance that has weaker students gravitating towards easier courses and stronger students taking the relatively harder ones. This helps give a more realistic picture of the relationship between test scores and grades.

Exploring a Computer-Delivered SweSAT

Given that we are carrying out this review in the year 2002, devoting some attention to the possibility of planning for computer-delivery of the SweSAT may be warranted. Advances in the application of technology to testing now open a

[5] This observation supports David Lohman's Point 3c.

variety of ways that SweSAT might be provided in a computer format. Such an action would greatly facilitate changes such as adding a writing test and could also be used to introduce a wide variety of formats that could still be objectively scored. Whether this approach is feasible right now, it seems likely that it will be a real choice soon enough that some time should be spent evaluating what could be gained by such a step. By no means should want move to such a technology "just because it is available," but taking a careful look at what might be the benefits should be the first step. Then the costs and disruption of a transition can be evaluated. If the gain would not be great, there will be no need to explore how to implement such a test. If there are really good outcomes possible, then the task of trying to develop a workable plan is worth the effort.

EVALUATION OF SWESAT BY DAVID F. LOHMAN

The following notes summarize my observations on the SweSAT. Since the goals of the evaluation were somewhat diffuse, I focused on issues that seemed relevant to me. Some of these issues surely are of broad concern; others are more specific. For lack of a better organizational scheme, the points are listed in an order than seems connected. Those that I feel strongly about are preceded by an asterisk (*).

*1. This is an excellent test

It is among the best scholastic aptitude tests in the world. The research studies that have been conducted on it are also first rate. I was particularly impressed with the studies on the effects of incidental selection on group differences (especially those that used longitudinal data sets that reached back to the 1970's), and of the sources of difficulty and individual differences in items on particular tests. All of my remarks must be held against this background.

I am also keenly aware that the construction of every operational test is an exercise in compromise. Even in the best of cases, testing time is constrained, item formats are limited to those that can be scored reliably and efficiently, and methods for accumulating item responses into test scores are limited by the availability of computing resources. In the case of the SweSAT, one can only do so much with five relatively short tests administered in 200 minutes of testing time.

*2. The uses for which the test was originally designed have changed

The initial purpose was to provide an estimate of academic potential for young adults who had been in the workforce for several years, and whose academic grades were thought to underestimate the likelihood of their success in university study. Now the test is being used as an alternative selection mechanism for a substantial fraction of the applicant population. This is troublesome.

*3. Our understanding of the importance of knowledge in adult reasoning has changed since the SweSAT was designed

As students mature, reasoning well in particular domains of study increasingly depends on their accumulated knowledge and skill in those domains. Although learning also depends on students' abilities to reason in more context-reduced ways (indeed, their ability to acquire knowledge in the first place depends on such skills), those who lack fundamental knowledge and skill in a domain find it increasingly difficult to succeed as they move up the educational ladder. The expectation that a measure of more decontextualized reasoning ability (Gf)

should always be an excellent predictor of success in university studies is thus not well supported in modern cognitive psychology. This is not to say that Gf-type abilities are irrelevant, or that they do not contribute significantly to the prediction of success—especially in mathematics, science and engineering and other programs that place heavy demands on students' abilities to solve well-structured but novel problems. This argues for a re-thinking of the role of the SweSAT in admissions decisions, especially a disjunctive admissions policy that considers *only* SweSAT scores for some segment of the population. For example, an alternative rule for this segment of the population would be to consider some combination of SweSAT *and* school grades, or SweSAT *and* performance on one or more domain-specific achievement tests.

There are several corollaries of point 3. I list the three below.

3a. Predictive validities would probably be higher if the SweSAT emphasized Gc and Gq abilities rather than Gf

Gustafsson, Wedman, & Westerlund (1992) claim that SweSAT seems primarily to measure Gf, secondarily a Gc dimension, and a fairly large specific factor. But the predictive validities of Gf tests for college success (defined by GPA or completion of a course of studies) is less than for Gq and Gc tests, which in turn are less than cumulative GPA for prior schooling. Thus, more attention to quantitative reasoning would give slightly better prediction, especially if Gq and Gc scores were differentially weighted for different courses of study. On the other hand, emphasizing Gq would likely exacerbate the problem of sex differences on the test.

Every good ability test measures individual differences that are shared with all other tests in the battery, with some other tests in the battery, and with no other test in the battery. It is the last category that is most troublesome. Some of the specific variance in each test can be attributed to error of measurement., but a larger fraction is systematic variance that is specific to the test. In recognition of this unpleasant fact of life, Thurstone long ago argued that one needed at least three different tests in order to define a factor. In particular, the quantitative reasoning dimension might be better measured if the test had more item formats. This would be especially useful if there is some attempt to develop differential scoring of SweSAT that can be weighted differently for admission to different domains of study.

My own preference is for a profile of reasoning abilities on verbal, quantitative, and figural tasks, each estimated by at least three different item formats. Figural reasoning abilities need not show sex differences. Although they are least predictive of success in school learning, they are helpful in research on adapting instruction to the learning styles of students, and show the smallest social class differences of the three. The profile is also more useful in guidance, although it is important that both students and guidance counselors understand the probabilistic nature of the relationships between test scores and success in learning.

3b. Achievement-oriented tests would likely show even higher predictive validities than SweSAT scores[6]

Domain-specific achievement tests could be developed as supplements to the SweSAT. Test development expenses would be substantial, however. Achievement tests would then also exert influence on the curriculum, which was contrary to the initial design specifications for the SweSAT. However, the SweSAT is now being used in ways that were never initially intended. There are good arguments that a series of achievement tests would better meet these new uses if it sampled objectives of the curriculum in the various domains of secondary education needed as preparation for university study. The mapping need not be so specific that it extends to particular courses of study. For example, a test of the ability to read and comprehend the Swedish language can be linked to many courses of study—as it is on the current SweSAT.

3c. Decomposing grades in a principled way would improve their predictive validity as well

Grades show the same sort of hierarchical structure as abilities. Differentiating grades according to subject matter could improve their usefulness as well. Decision rules, however, need to be sensitive to the possible effects on student study habits. For example, society should not prefer a generation of engineers who are unable to write effectively.

4. High-stakes tests have consequences for the educational system

The consequences of the current SweSAT for the educational goals of educators and study habits of students seem relatively benign. Continued high-stakes use of the test seems likely to change this sate of affairs. It will also elicit even stronger demands that scores not differ between groups of stakeholders. The finding that groups differ in acquired knowledge and skill has different consequences than a finding that they differ on a test that is commonly (albeit erroneously) interpreted as a measure of potential.

---

[6] One interesting suggestion that surfaced in the discussion of the GRE test in Seattle a few weeks ago would be to try to measure 'reasoning in context' by allowing the test taker to opt for parallel versions of some items (or subtests) that differed in the content of the problem. (The formal structure of the problem might be the same). There are some interesting psychometric possibilities (and nightmares) in such a scheme, especially in an IRT framework. But there are good reasons to try to do this. We know that reasoning does not occur in a vacuum. In some sense, the best estimate of a student's ability to reason is given by how well he or she reasons with content that is at least moderately familiar. Perhaps one could estimate a 'general' verbal, a general 'quantitative' and a 'domain-specific' estimate of reasoning ability for each student. Student would pick the domain from among a small set (literature, social studies, physical sciences, etc.). There are lots of interesting issues in how to scale such items (e.g., do you look for items that have poor fit overall but good fit within the population of more knowledgeable test takers?).

5.Nevertheless, a good reasoning test that is somewhat removed (perhaps even more removed than the current SweSAT) from the curriculum still has an important place at the table

Tests of relatively decontextualized reasoning abilities (i.e., Gf) can help identify students who reason well but who, for whatever reason, have not done well in school. This includes individuals who have not had access to quality education, as well as those who have been unable or unwilling to take advantage of educational opportunities offered. This can be important information for decisions about special admissions. It can also usefully supplement other measures of achievement. Admissions committees might make different decisions (in either direction) for two students with similar levels of achievement but different levels of reasoning abilities.

Indeed, data I have gathered for US high school students show that differences between ethnic groups are larger on a good achievement test that is closely tied to the intended curriculum than they are on measures of verbal, quantitative, and figural reasoning that are less directly tied to the curriculum. This does not mean that the achievement test underestimate the academic abilities of minority students. Rather, it means that these students are often more able than they are prepared.

6.Sex and SES differences on existing SweSAT test (especially DTM and DS) could perhaps be reduced by additional study of the sources of item difficulty that moderate such effects

Such effects are often difficult to detect in pools if existing items. Sometimes they can be inferred from patterns of group differences in existing item sets using standard techniques for identifying differential item functioning. In other cases, however, larger item banks must first be developed in which sources of difficulty are varied systematically and orthogonally, if possible. This is because, by design, items on a good ability tests are typically a hodgepodge of sources of difficulty. More complex multivariate procedures such as Tatsuoka's (19) Rule Space method might be helpful, in some cases.

Identification of systematic, but construct-irrelevant individual difference variance is even more difficult. This typically requires the simultaneous administration of large batteries of tests and other scales to large, unselected groups of applicants. For example, it is difficult to estimate the extent to which particular items on the DTM subtest measure spatial ability unless the study includes several good tests of spatial ability (in addition to other abilities). Analyses of within-test covariance matrices for item scores are at best a first approximation. [Parenthetically, a test called "Dial & Table Reading" was a consistent marker for a broad spatial ability factor in the US Aviation Psychology Program in WWII (Guilford & Lacey, 1947).]

Given the difficulty of doing this work, and the number (and quality) of the

studies of this sort already in hand, it might be wiser to consider using additional or alternative item formats rather than hoping to repair existing ones.

*7. The criterion space matters

Credit points and throughput surely matter, but they are not the only criteria of academic success. To paraphrase Berliner, we have measured the predictor with micrometers and the criterion with divining rods. A more differentiated criterion space that included measures of, say, flexibility in problem solving, long-term retention and organization of conceptual knowledge, critical thinking, etc. might show the value of a broader range of predictors. Although there routinely calls for a better understanding of the effects of university education, such research is difficult to conduct. But as Reuterberg, Gustafsson, and others in Sweden have so ably demonstrated, methods of data modeling using missing data techniques can give much better insight into the relationships among variables in incomplete data matrices that result from nonrandom selection and drop out. Of course, multiple predictors and multiple criteria complicate the admissions process. A single rank ordering is easier to use and to defend to the public, even though it can mislead.

*8. Defining the treatment is part of defining the aptitude

If the educational system changes, then the characteristics students need to succeed can also change. Put differently, finding tests that best predict success in a given system may inadvertently help perpetuate a system that is in need of repair. Although I am aware of no evidence that the Swedish higher educational system is ineffective, changes in the student population will most likely make it less effective for a larger segment of the admitted population of students. At the K-12 level, for example, an important use of measures of Gf is to suggest ways in which instruction might be modified better to reduce the correlation between input (Gf) and output (achievement). We know a bit about how to do this. Attempts to make university experience accessible to a broader segment of the population would benefit from a closer look at how it might also be done at the university level as well. The alternatives are increases in the number of remedial/preparatory classes and higher drop-out rates for those who do enroll.

Final remarks

I am particularly concerned about the extent to which the original use (as an aptitude test for those who had long been out of school, and whose school grades might not have been stellar) has now changed (as an aptitude test for most? applicants to higher education). In both cases, one could argue for either a more aptitude-like or for a more achievement-like test. For example, an aptitude-like

test clearly better complements course grades, and thus expands the pool of eligible applicants.. However, much depends on the decision rules for selection, and the selection ratio.

I also think that here, as in every prediction system, the criterion problem looms large. We seem unable to do much about this is the US. Perhaps the Swedes can be encouraged to do more than just think about it. Surely number of credit points accumulated is important (as is our freshman GPA measure) and ultimately, graduation. But the criterion measures used skew the prediction toward similar predictors. I happen to believe that (fluid or critical) reasoning abilities are important outcomes as well, but it is difficult to make this argument when the dependent variable is just "number that survived" or a surrogate.

Put differently, if I had my choice, I'd pick students who were high on both Gf and Gc, not just one or the other. Similarly, I would hope that the education offered would improve both contextualized and decontextualized reasoning abilities. The latter are often touted as the goals of a general liberal arts education. Importantly, fluid reasoning abilities in adulthood may look more like critical thinking than traditional measures of Gf.

References

Gustafsson, J-E., Wedman, I. och Westerlund, A. (1992). The dimensionality of the Swedish scholastic aptitude test. *Scandinavian journal of educational research*, *36* (1).

EVALUATION OF SWESAT BY WERNER W. WITTMANN

Introduction

The SweSAT is under strong scrutiny as regards to goals set by government concerning entrance into higher education in Sweden. There are different objectives against which a selection instrument like the SweSAT can be evaluated. From a historical perspective the SweSAT was conceived as a supplement to the main selection instrument used, namely the school grades.

The goals set by government bill 2001/02:15 (U01.016 November 2001).

Broader recruitment and new paths to Higher Education Objectives for basic higher education:

1) the ability to make independent, critical assessments
2) the ability to identify, formulate and solve problems
3) a preparedness to deal with changes in working life
4) seek and evaluate knowledge at a scientific level
5) follow developments in knowledge, and
6) exchange knowledge with people with no special expertise in the field
7) equal treatment of students
8) greater international mobility.

With these goals in mind I now turn to my review of the SweSAT. I will concentrate mainly on the following issues:

The criterion problem

The criteria of study success in terms of pass/fail and throughput are important ones.

But I'll also point to the problem of the changing demands of the workforce in the sense of Buzz Hunt. Hunt, E.(1995). Will we be smart enough? A cognitive analysis of the coming workforce. New York: Russell Sage Foundation. So dealing with the SweSat one also has to consider these changing demands in terms of the cognitive complexity of the new workplaces and especially spatial/figural reasoning.

The Lubinski and Benbow work with the highly gifted which continues Julian Stanley's projects at Vanderbilt is also relevant here. Dave Lubinski has data that especially the spatial/figural reasoning trait complex (Dick Snow's and Phil Ackerman's terms) are highly relevant but neglected.

The dimensionality of the SweSAT and its consequences for present and future

Given the work of Jan-Eric Gustafsson there is good evidence that the SweSAT maps more than one dimension. Gustafsson labels them as Verbal (WORD, READ, ERC) and Reasoning (DS, DTM). The second one is mainly quantitative numerical reasoning. My understanding of general reasoning extends to numerical, as well as figural and verbal reasoning. Why is that latent not better

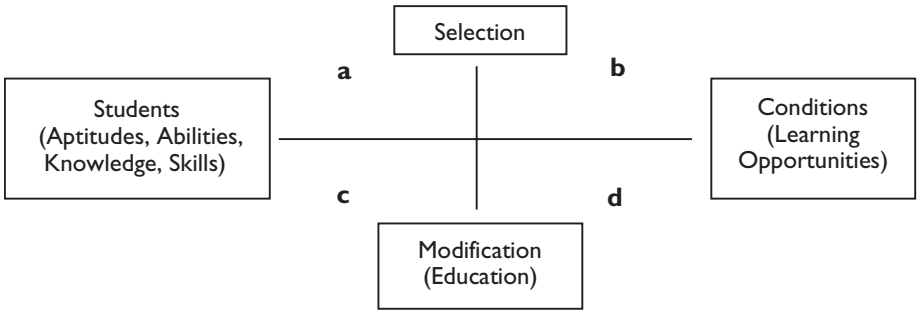
assessed with more and broader tasks (enlarging nomothetic span)? In my own German research I've found that differentiating the reasoning dimension pays off in differential validity for subjects areas related more to natural science contrasted to those more related to humanities, arts and languages (the German distinction between "Natur" and "Geisteswissenschaften". Using two dimensions of the SweSAT would also help not only for selection purposes but also for classification and related counseling and/or recommendations for areas to study.

Gender fairness

Given the overall percentage of 57% of females in the Swedish University system one surely cannot speak about overall unfairness towards females! Yet the gender distribution over different tracks is very uneven and I'll propose strategies to reduce this unevenness. What I found is that the small proportion of females which actually found their way into the technology tracks have in terms of a two-dimensional latent structure of the actual SweSat, the same profile as the males. On the other hand there are females not choosing these tracks which have a profile which would well fit into the demands of these tracks. So better counseling these females on the basis of the SweSat could attract more of them into these programs. I'll use Jan-Eric Gustafssons Streams sample to demonstrate impacts of my proposals. I know that this sample may not be representative despite also having 57% of females. But once my arguments seem to be worthwhile they can be checked at a more representative sample.

A conceptual framework for evaluating the future of the SweSAT

Any reforms in a system like here the higher education can be screened under the perspectives of four different areas of tasks to be done. Crossing two bipolar dimensions 1) selection vs. modification with 2) students vs. conditions leads to these four different areas a–d.



Area a) looks after all the favorable student characteristics which help in predicting success.

The criteria of success are most often tailored to education only in terms of pass/fail rates, or the lowering of dropout rates or accelerating the throughput in

terms of shorter study times. Yet these types of criteria are only intermediate ones, the ultimate criteria are the long term ones, namely the success at the workplace. The SweSat, as given, is mainly tailored to area a, competes with and supplements here school grades from the level of secondary education (Gymnasium).

Area b) asks for selection of the right and appropriate learning conditions and is traditionally the area of counseling and/or placement of students to the right subject area and/or the appropriate institution of higher education. Good and differentiated assessment tools can be used for areas a and b as well. I have found neither evidence nor research whether the SweSat could be used for these tasks. There are only complaints and fears that genders tend to gravitate to some areas but not others, i.e. there are fears about a feminization of the area of education and a lack of females in technology.

Area c) concentrates on what knowledge and skills students should learn to build a workforce for the Swedish society which can compete at an international level and all the demands of the globalization of the economy. Traditionally area c) deals with questions of aptitude x treatment interactions (ATI) in the sense of Cronbach & Snow. The majority of objectives for basic higher education formulated by the government bill above are tailored to that area. The question here is, how do grades and the SweSAT mirror abilities or aptitudes, favorable for achieving the objectives set by the government.

Area d) focuses on the necessities of reforms of the existing system of higher education anticipating the new demands set by changing work places and on the coming workforce for a successful society and economy broadly understood. "The cognitive and social changes that are altering the workplace are driven by a few major technological changes. By far most important, and best publicized, of these is the development of digital computers. But computers are not the only technological innovation involved. Communication technologies have changed as much as computers have. The two technologies are similar in many ways, but they are not identical. For instance, we could have computing without having communication satellites, but we could not have satellites without computers. Both technologies have made impressive changes on their own, and, in addition, they are synergistic."(Hunt, 1995, p. 205).

Another topic to consider here is the distinction of the two cultures in the sense of C.P. Snow. Although Snow complaints about the distinction between arts/language/humanities and science with their different focus on verbal and figural/numerical aptitudes might seem somewhat outdated, they actually are not. Modern societies more and more need decision makers who are educated and speak the languages of both cultures. Any system of higher education must deal with these challenges. The best example for that are the field of biology and the related field of application, namely biotechnology. Hunt's (1995) book title "Will we be smart enough? A cognitive analysis of the coming workforce" nails down the essential points that we have to consider a broader spectrum of abilities and aptitudes of the new incoming workforce than were done earlier.

How does the SweSAT as given, stand up under the challenges and demands of the four fields?

The SweSAT, as given, is mainly tailored to area a). The admission of students to higher education is done by total grades and by total SweSAT scores, 60% and 40% of available places, respectively. This decision strategy does not consider the latest cumulative research on the dimensionality of individual differences laid out in what has now been labeled as the CHC (Cattell, Horn, Carroll) hierarchical organization of dimensions of individual differences in aptitudes and abilities. Extensive research concerning the dimensionality of the SweSAT by the Gothenburg research group of Jan-Eric Gustafsson demonstrates that the SweSat is not one-dimensional and can easily be incorporated into the CHC framework. The school grades also are not one-dimensional but can be organized into a hierarchical framework as well. Developing a profile of aptitudes using grades and SweSAT promises to allow better selection and placement decisions. Thus differentiated profiles to be derived from the assessment devices also will be very helpful for counseling purposes. I have not seen any research concerning the counseling potential of the SweSAT and the grades.

Even research with the very highly gifted done by the Vanderbilt research group founded and established by Julian Stanley and continued by Camilla Benbow and David Lubinski demonstrates the utility of differences in aptitude profiles. In the SMPY (study of mathematically precocious youth) project students had been selected according to their SAT scores (SAT-M and SAT-V). The subjects had taken them several years before the typical age. There are now a couple of papers and reports of a 20-year follow-up of the initial cohort of 1,975 mathematically gifted adolescents who had been assessed at age 12 to 14 and represented the top 1% of their cohort. Even in that highly selected sample it paid off to look at the SAT profiles. Benbow and Lubinski distinguish between three types of profiles, one even level and two tilted ones. In the even profile SAT-M and SAT-V are approximately the same. The tilted ones can either show a higher score on SAT-M or on SAT-V.

In general and irrespective of gender, students with tilted profiles tended to gravitate toward their area of strength. Students with exceptional SAT-M relative to SAT-V gravitated more to mathematics, engineering, and the physical sciences, while those with the inverse profile are more attracted to the humanities, law, and social sciences. Lubinski and Benbow apologize for not having included predominantly spatial/figurative tests, which they now consider very important given the changes in the demands of the workplace.

A SAT with profile information should be very helpful for counseling purposes, especially the students with even profiles are a very interesting group because they should be most flexible and can be recommended to areas where the demands on the workforce are largest.

Respecting and considering the profile information given by tilted ones has some important consequences:

> "Correspondent learning environments foster psychological well-being; discorrespondent learning environments foster psychological pain" (Lubinski and Benbow, 2000,p. 144).

So I recommend considering the SweSat potential for counseling (area b) as well. But its potential could only fully used if its dimensionality is fully exhausted and extended.

Area c) points to the needed change of student's characteristics in terms of knowledge, skills and other competencies. There are many complaints in many countries that schooling and higher education do not consider adequately the most important demands by society and the economy. With respect to higher education many complain about the dominance of rote learning and the aggregation of knowledge pieces of doubtful relevance. University teachers complain about a lack of basics and conceptual thinking. The latest Swedish government bill mirrors these demands in formulating new and better objectives for area c) of our framework.

Yet the results of the recently published PISA study shows that Sweden fares pretty well compared to other countries, e.g. my own, namely Germany! Obviously the Swedish system of secondary education preparing for higher education already has reached many of the objectives set by the Government.

For Sweden it should be interesting to learn of how much the PISA results correlate with the school grades and whether it could still profit to include PISA type tasks into the SweSAT. Here I see an exciting research program for the Umeå group, which is responsible for the SweSAT.

Area d) is the field, which the bill focuses on with its recommendations concerning pedagogical renewal. Much can be learned in this regard from the movement in the United States, which is called physics education research (PER). Here new forms of science education called peer instruction, interactive engagement, Socratic dialogue instruction etc. have been developed demonstrating impressive effect sizes as regards conceptual thinking for physics and non-physics majors as well. Hake (2002) gives an overview about these developments; the leading groups are around Eric Mazur (Harvard), Joe Redish (University of Maryland), David Hestenes (Arizona State) and the Hellers (Univ. of Minnesota) just to name a few. I also noted that Jonte Bernhard a former member of the National Council of Higher Education, now at Linköping University, participates in that movement. From an evolutionary stance one can assume that educational curricula adapted over the long run to the tilted ability and aptitude profiles discussed above. The challenge for pedagogical reform and renewal is whether there are any chances to attract and educate students to a satisfactorily degree in areas where their profiles do not perfectly correspond. Education in natural sciences and the related derived applied technologies gain more and more importance in biology, medicine and even social sciences. These students do not necessarily have the quant-tilted profile, which fits best to the existing curricula in classical natural sciences. The PER movement demonstrates that it is possible to even teach conceptual thinking to a broader cohort of students using interaction, spatial/figural visualization and other activities.

A profiled SweSAT would be of considerable help in better tailoring and evaluating the new pedagogical reforms. Not knowing the ability and aptitude profiles of our students and responding in an aptitude x treatment interaction strategy to them, many reforms are in the severe danger of being doomed to failure.

Biology as a science and the related biotechnology application fields are a prominent example. Biology has made a strong move towards physics, chemistry and math especially in microbiology so the demands have changed considerably. In my own research about intelligence and school grades at the secondary education level (Gymnasium) in Germany it is evident that biology is still taught more like a foreign language and attracts students with tilted verbal profiles. One can easily speculate about difficulties these students have once they choose biology at the university level, which is dominated by the new focus on microbiology. Only when curricula react and adapt to these profiles there is hope that dropout and failure can be minimized.

From the papers I have read and especially from the latest still unpublished research presented by Jan-Eric Gustafsson at the Spearman conference in Sydney Nov. 2001 it seems to me that Sweden again is ahead. The effect sizes especially of the science and technology education at the university level are impressive. I recalculated more than half a standard deviation on fluid intelligence (gf) and visualization (gv), which are very basic ingredients of conceptual thinking. Science education in Sweden as it has been done in the last decade cannot be bad and I would be happy if my own country could compete with it. So I recommend having a closer look at the causes of these success stories in order not to misdirect the new pedagogical reforms to be done in the future.

There are also new developments in the United States which must be considered in discussing and solving the challenges related to areas c) and d). The old occupational titles have been removed, reformulated and improved by the O*Net taskforces. Phil Ackerman (1996) has proposed a theoretical framework called PPIK and concentrates on the relationship between abilitities, aptitudes, personality, interests and knowledge in a research program about trait complexes. He distinguishes between intelligence as process (P),intelligence as knowledge (K) and demonstrates how personality (P) and interests (I) are important mediators in the very process of the investment of fluid intelligence and the crystallizing results of knowledge. Ackerman's(in press) trait complexes mirror the different areas of higher education and the related area of the workplaces. Lubinski and Benbow(2000) give that theory which capitalizes on the Cattell/Horn investment theory a prominent place and hint to its pay offs which the theory of work adjustment by Dawis and Lofquist (1984). Holland's (1985,1996) RIASEC model of vocational preferences is incorporated in all these theories and approaches. Any initiatives of pedagogical reforms are well advised to consider these impressive developments in individual differences research and the related field of industrial/organizational psychology.

Fairness to gender and minorities

Fairness to gender and minorities is a very important topic in every democratic society. There is considerable and excellent research on that topic in Sweden. The SweSAT shows substantial differences in the nonverbal parts favoring males. It is suspected that these differences are a possible bias of test construction. Thus depending too much on the SweSAT as given for selection purposes might discriminate against women.

To my astonishment little research is to be found concerning a similar possible bias of school grades. In many countries, not only in Sweden, girls have the better grades for whatever reasons I will not discuss at the moment.

Looking at the proportion of females entering and accepted to higher education in Sweden from the reports of the National Agency of Higher Education, I learned that this proportion is a 57% favoring females. Given that number one cannot conclude that the SweSAT introduces bias towards females. Quite the opposite seems to be true, the SweSat at least somewhat ameliorates bias against males! A different story is the gender imbalance in the different subject areas, the doctoral and the staff level. But these differences cannot be attributed to the SweSAT!

A new look at the gender fairness problem

What the SweSAT assesses with its nonverbal parts, i.e. interpreting diagrams, figure and tables is very important for the changing new workplaces. Looking only at mean differences and introducing a kind of affirmative action to secure a proportional balance is in my eyes the wrong strategy and misleading, because it neglects individual differences. Despite these mean differences there is substantial overlap in the distribution of both sexes. There are many females which have higher scores on these nonverbal parts than the mean male. So assuming that there are at the moment maybe only 30% of females which have the corresponding quant-tilted profile one has to ask how many of that proportion actually find their way into science and technology. I fear that this potential is not really exhausted. Therefore strategies of encouragement based on SweSAT and other information have to be developed to attract women with the corresponding profile to these disciplines. They same can be done with males who have the corresponding profiles for the nowadays female dominated areas like education and psychology! Once more females and males are counseled to chose their appropriate field they will function as role models and attract in the long run more of their same sexed peers to the field. I am confident that such a policy will finally better solve the fairness problem than any policies of affirmative action.

These recommendations are focused on the areas a) and b) only. Additionally there is hope that new types of curricula as discussed above as regards areas c) and d), like the *per* movement, can further ameliorate gender (and minority) differences, once they are implemented and better consider the abilities and aptitudes of these students.

Summary and recommendations

1) The SweSAT should be expanded to consider the cumulative evidence from international research on individual differences like the CHC-theory.
2) The SweSAT should be developed to give profile information helpful for selection and counseling.
3) The test parts of the SweSAT which show gender, minority and other mean group differences should not be removed or abolished. They should be analyzed according to the demands on the new workforce and the overlap information should be used to better attract those who really fit to these areas. In the longer run this will better contribute to fairness.
4) The school grades should also be dimensionalized and their profile potential be combined with the SweSAT profiles.
5) The school grades should undergo the same amount of research as regards bias and unfairness as was directed toward the SweSAT.
6) The SweSat and school grade profile should be known and used by all educators and researchers at institutions of Higher Education to better tailor the curricula to abilities and aptitudes of the new cohorts of students.
7) All pedagogical reforms should be systematically evaluated. A profiled SweSAT should play a prominent role in these evaluations to learn about the baselines and benchmark any improvements.
8) Sweden is obviously very successful in secondary education demonstrated by the PISA studies and probably as well in its three year track science and technology programs. Research disentangling the causes of these successes should help and guide any new pedagogical reforms.
9) It should be checked whether PISA type tasks can be adapted for the SweSAT.
10) The cost-benefit of the SweSAT should be researched. I assume that such a type of analysis will unravel that the economic return of such a selection instrument is heavily underestimated.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interest, and knowledge. *Intelligence*, *22*, 227–257.

Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment*. Minneapolis: University of Minnesota Press.

Holland, J. L. (1985). *Making vocational choices* (2nd ed., pp. 419–451). New York: Wiley.

Hunt, E. (1995). *Will we be smart enough? A cognitive analysis of the coming workforce*. New York: Russell Sage Foundation.

Lubinski, D., & Benbow, C. P. (2000). States of excellence. *American Psychologist*, *55*, 137–150.

Appendix 1

by Werner W. Wittmann

Gender and different profiles

The following calculations are for demonstration purposes only, due to not having a representative data set available. What is used here is the demonstration data set delivered by Jan-Eric Gustafsson's Streams software package, at least its gender proportion is similar to what is reported by the annual reports of the National Agency of Higher Education.

A categorical variable labelled DIFF3KAT was computed from a two factor varimax rotated solution of the SweSAT subtests based on the total samples of 579 students. The first factor represented the verbal SweSAT and the second one the quant/spatial one in z-scores. The difference scores from these two factors have been used to categorize three groups. The even group had scores between –0.5 and +0.5 standard deviation (sd) on it. The verbal-tilted group had scores half a sd higher on the verbal factor than on the quant/spatial one. The quant/spat-tilted group had scores half a sd higher on the quant/spat factor than on the verbal one.

Here are the results of that categorization for the total sample:

```
Frequencies
 Values for DIFF3KAT
```

```
   verbal even quant/spat  Total
   tilt     tilt
   +-------------------+
   | 184  213  182 |    579
   +-------------------+
```

```
Percents of total count
 Values for DIFF3KAT
```

```
    verbal  even  quant/spat  Total   N
    tilt       tilt
    +-------------------------+
    | 31.779 36.788 31.434 | 100.000   579
    +-------------------------+
```

Here are the results cross-tabulated by gender:

Frequencies
 GENDER (rows) by DIFF3KAT (columns)

```
    verbal even quant/spat  Total
    tilt     tilt
    +-------------------+
male  |  45  101  100 |   246
female | 139  112  82 |   333
    +-------------------+
Total   184  213  182     579
```

Percents of total count
 GENDER (rows) by DIFF3KAT (columns)

```
      verbal  even  quant/spat Total   N
      tilt      tilt
     +-------------------------+
male   | 7.772 17.444 17.271 | 42.487   246
female  | 24.007 19.344 14.162 | 57.513   333
     +-------------------------+
Total   31.779 36.788 31.434  100.000
 N      184   213   182         579
```

In the following we cross-tabulate the three profiles by area chosen (or admitted)

broken by type of gender:

```
>TABULATE GENDER * PROG * DIFF3KAT
```

Percents of total count
PROG (rows) by DIFF3KAT (columns)

GENDER = 1 (male)

	verbal tilt	even	quant/spat tilt	Total	N
Business	5.691	12.602	12.195	30.488	75
Humanities	0.000	0.000	0.000	0.000	0
Social	4.065	8.943	8.537	21.545	53
Science	3.659	5.285	2.846	11.789	29
Technology	4.878	14.228	17.073	36.179	89
Total	18.293	41.057	40.650	100.000	
N	45	101	100	246	

GENDER = 2 (female)

	verbal tilt	even	quant/spat tilt	Total	N
Business	7.207	12.312	6.907	26.426	88
Humanities	5.405	0.901	0.601	6.907	23
Social	7.808	6.306	6.907	21.021	70
Science	19.520	12.012	6.006	37.538	125
Technology	1.802	2.102	4.204	8.108	27
Total	41.742	33.634	24.625	100.000	
N	139	112	82	333	

Gender proportions selected to higher education:

Females : 333/579 = .575
Males : 246/579 = .425

This is not an indication of unfairness towards females!

Looking at the proportion of females having a quant/spat-tilted profile, we find that only 82/333 = .246, i.e. one fourth of the females selected have the quant/spat-tilted profiles which may be favorable for technology education with the curricula as given.

Looking at the proportion of males having a quant/spat-tilted profile we find that a much larger proportion 100/246= .4065, i.e. forty percent of males have that quant/spat-tilted profile.

Most of the discussion of possible unfairness of the SweSAT is related to the fact that using it, one finds only 24.6% of females against 40.65% males with the quant/spat-tilted profile.

The proponents of unfairness of the SweSAT towards males could easily counterattack such a claim by pointing to the fact that verbal abilities are very important and that the existing selection instruments are biased towards males!

Looking at the proportion of females having a verbal-tilted profile we find that this is 139/333= .417 out of the selected females; and looking for the proportion of males we find that this is only 45/246 = .183 out of the selected males.

One could also seriously doubt that out of the cohort of males only 18% have a verbal-tilted profile!

The new look at gender fairness should concentrate for the moment at the fact, why out of the 82 females (24.6% of all females) with a favorable profile, only 14 (17.07% of the 82 females) are in the technology programs, whereas out of the 100 males (40.60% of all males) with a favorable profile there are at least 42 (42% of the 100 males) in these programs.

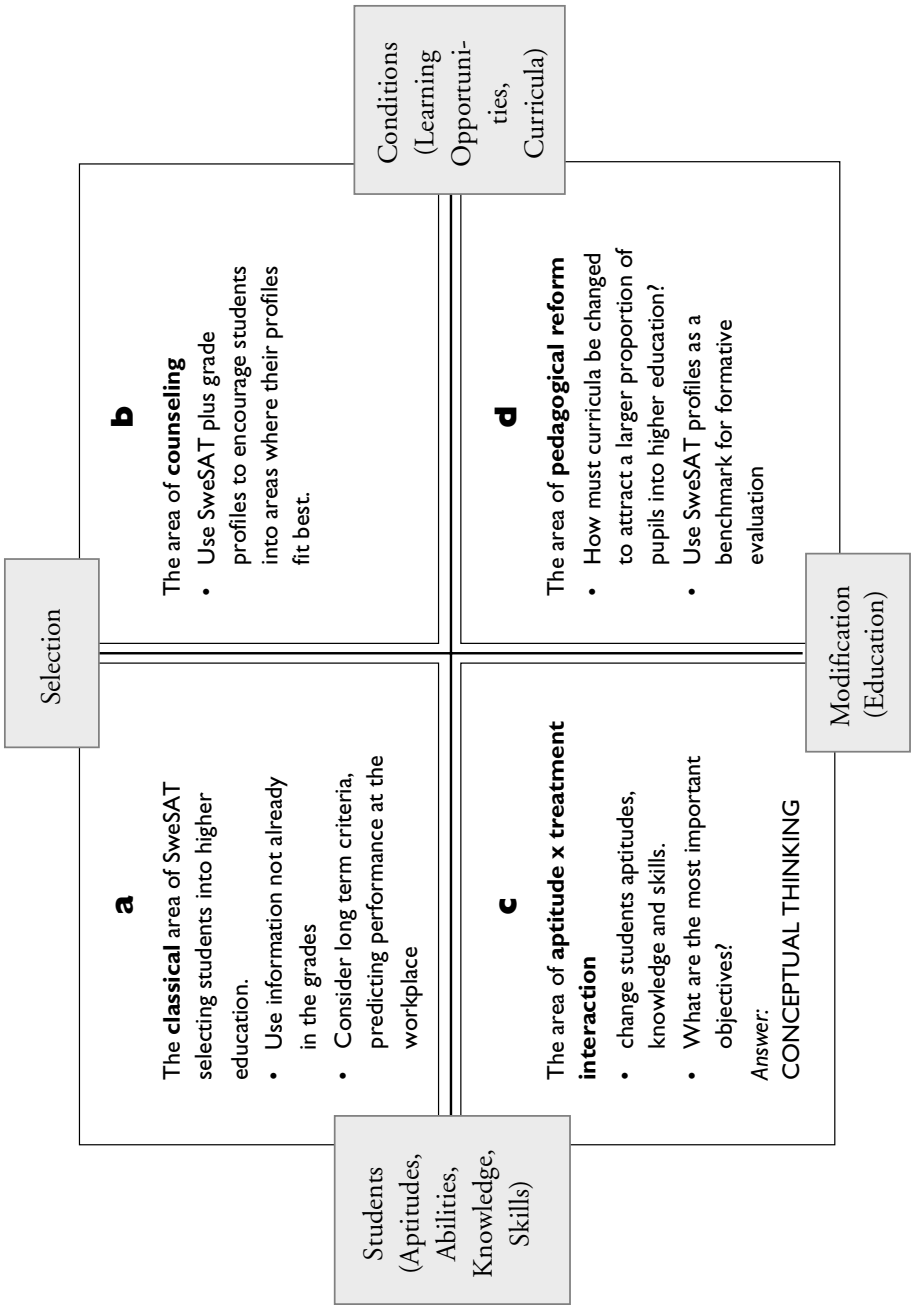
On the other hand it is amazing that only 27 out of 246 males selected for higher education are in the science programs (27/246 = 10.98%), whereas 125 out of 333 females are in these programs (37.54%). I assume that education with its dominant female population is included here under the label of science. Again here it is very unplausible that only females are qualified for the science areas and/or education within it.

Surely these differences towards males or females cannot be attributed to the SweSAT but maybe more a function of stereotypes, interests, learning opportunities, the dominant use of total grades with their heavy weighting of verbal aptitudes etc. and especially due to a lack of counseling and marketing of the different programs!

Appendix II

by Werner W. Wittmann

Four different areas to be considered for the future of the SweSAT



Appendix III

by Werner W. Wittmann

The effect of adding more ability-type tests, especially figural/spatial ones, and the consequences for gender differences

Here some analysis is presented concerning the dimensionality of the SwesSAT, the effect of adding more ability-type tests, especially figural/spatial ones, and the consequences for gender differences.
>PEARSON WORD READ GI ERC DS DTM SYNONY VIZUAL FIGRES NUMSER*SWESAT_V , >SWESAT_QF SWESAT2_V SWESAT2_QF

Pearson correlation matrix (these are actually the correlations of the varimax rotated factor scores with the original tests.)
SweSAT_V and SweSAT_QF were derived from using the SweSAT tests (Word-DTM) only.
The SweSAT2_V and SweSAT2_QF were derived from all tests (WORD-Numser).

This last solution checked what would happen when more ability like measures like synony, visual, figres and numser would be added to the battery.

	SWESAT_V	SWESAT_QF	SWESAT2_V	SWESAT2_QF
WORD	0.863	0.174	0.864	0.149
READ	0.810	0.244	0.779	0.178
GI	0.571	0.227	0.709	0.198
ERC	0.831	0.256	0.802	0.206
DS	0.238	0.861	0.338	0.728
DTM	0.221	0.869	0.325	0.703
SYNONY	0.687	0.206	0.812	0.144
VIZUAL	0.161	0.480	0.129	0.722
FIGRES	0.149	0.524	0.114	0.792
NUMSER	0.086	0.378	0.052	0.623

Number of observations: 579

In the following these four factors are tested for gender differences:

```
>TEST SWESAT_V SWESAT_QF SWESAT2_V SWESAT2_QF * GENDER

Two-sample t test on SWESAT_V grouped by GENDER
 Group        N     Mean       SD
     1      246    0.014    1.004
     2      333   -0.010    0.998

  Separate Variance t=0.287 df=526.4  Prob=0.774
  Difference in Means=0.024  95.00% CI=-0.141 to 0.190

  Pooled Variance t=   0.287 df=577   Prob=0.774
  Difference in Means= 0.024 95.00% CI=-0.141 to 0.189

Two-sample t test on SWESAT_QF grouped by GENDER
 Group        N     Mean       SD
     1      246    0.393    0.882
     2      333   -0.291    0.983

  Separate Variance t=8.778 df=555.7  Prob=0.000
  Difference in Means=0.684  95.00% CI=0.531 to 0.837

  Pooled Variance t=  8.636 df=577  Prob=0.000
  Difference in Means=0.684  95.00% CI=0.528 to 0.839

Two-sample t test on SWESAT2_V grouped by GENDER
 Group        N     Mean       SD
     1      246    0.071    0.979
     2      333   -0.052    1.013

  Separate Variance t=1.473 df=537.5  Prob=0.141
  Difference in Means=0.123  95.00% CI=-0.041 to 0.287

  Pooled Variance t=  1.466 df=577  Prob=0.143
  Difference in Means=0.123  95.00% CI=-0.042 to 0.288

Two-sample t test on SWESAT2_QF grouped by GENDER
 Group        N      Mean       SD
     1      246     0.282    0.971
     2      333    -0.208    0.971

  Separate Variance t=6.004 df=527.8  Prob=0.000
  Difference in Means=0.490  95.00% CI=0.330 to 0.651

  Pooled Variance t=  6.004 df=577  Prob=0.000
  Difference in Means=0.490  95.00% CI=0.330 to 0.650
```

Now we calculate the gender differences in terms of Cohen's d for the old SweSAT_QF factor mainly loaded by the DS and the DTM subtests:

>calc (.392+.291)/sqr((246*.882*.882+333*.983*.983)/579)

Cohen's <u>d = 0.726</u> using SweSAT_QF

and then Cohen's d based on the SweSAT2_QF factor on which additionally to DS and DTM the ability (intelligence) tests VIZUAL,FIGRES and NUMSER have substantial loadings.
>calc (.282+.208)/sqr((246*.971*.971+333*.971*.971)/579)

Cohen's <u>d = 0.505</u> using SweSAT2_QF

This demonstrates what Dave Lohman claimed in his remarks, that including more ability-like measures for spatial/figural reasoning does not necessarily widen the gender gap. In this example quite the opposite seems to be true.

Appendix IV

Guidelines for the construction of the Swedish national university aptitude test (SweSAT)

Construction of the national university aptitude test is to follow the methods previously used in which the fundamental requirements are as follows:

- The national university aptitude test is a selection test that must be able to rank students in accordance with the success they are likely to have in their studies.
- The test must be constructed so that it is considered suitable by higher education institutions for the selection of applicants to their programs.
- From the spring of 1996 and onwards, the test is to consist of five subtests that measure the comprehension of words and concepts (WORD), mathematical reasoning ability (DS) reading comprehension in the broad meaning of the term (READ), the ability to interpret information in diagrams, tables and maps (DTK) and English reading comprehension (ELF) as laid down in a decision issued by the National Agency for Higher Education on 17 November 1995 (reg. no. 81-1221-95).
- The test is to agree with the objectives and content of higher education and have relevance for the higher education sector.
- It must be impossible for individuals to improve their test scores by prior mechanical practice or by learning any special principles for working out solutions.
- The tests must be constructed so that the testees feel that they are a suitable instrument for selection to higher education programs.
- Every effort should be made in constructing the test to ensure that there is no discrimination because of social origin or gender.
- The test must be capable of being graded rapidly using objective and cost-effective methods.

Högskoleverkets rapportserie från år 2000 och framåt

Rätt juristutbildning? Utvärdering av juristutbildningar
Högskoleverkets rapportserie 2000:1 R

Forskarskolor – ett regeringsuppdrag
Högskoleverkets rapportserie 2000:2 R

Journalistutbildningarna i högskolan
Högskoleverkets rapportserie 2000:3 R

Högskolestudier och funktionshinder
Högskoleverkets rapportserie 2000:4 R

Utbildningar inom vård och omsorg
– en uppföljande utvärdering
Högskoleverkets rapportserie 2000:5 R

Utvärdering av Socionomutbildningar
Högskoleverkets rapportserie 2000:6 R

Förnyad granskning och bedömning av kvalitetsarbetet vid Högskolan i Jönköping
Högskoleverkets rapportserie 2000:7 R

Lärosätenas arbete med jämställdhet, studentinflytande samt social och etnisk mångfald
Högskoleverkets rapportserie 2000:8 R

Goda exempel
Hur universitet och högskolor kan arbeta med jämställdhet, studentinflytande och social och etnisk mångfald
Högskoleverkets rapportserie 2000:9 R

Tentamen: "Plussning" och begränsning av antalet tillfällen
Högskoleverkets rapportserie 2000:10 R

Designutbildningar i Sverige. En utredning och utvärdering.
Högskoleverkets rapportserie 2000:11 R

Högskoleprovet – Gårdagens mål och framtida inriktning
Högskoleverkets rapportserie 2000:12 R

Eldsjälar och institutionell utveckling
Högskoleverkets rapportserie 2000:13 R

Antagning till högskolan – erfarenheter och visioner
Högskoleverkets rapportserie 2000:14 R

Att leda universitet och högskolor. En uppföljning och analys av styrelsereformen 1998
Högskoleverkets rapportserie 2000:15 R

Högskolornas tillämpning av EG-direktiv i sjuksköterskeutbildningen och barnmorskeutbildningen
Högskoleverkets rapportserie 2000:16 R

Sexuella trakaserier mot studenter – högskolornas åtgärder
Högskoleverkets rapportserie 2000:17 R

Livlångt lärande som i dé och praktik i högskolan
Högskoleverkets rapportserie 2001:1 R

Nationella ämnes- och programutvärderingar
Högskoleverkets rapportserie 2001:2 R

Vilken betydelse har utländsk bakgrund för resultatet på högskoleprovet?
Högskoleverkets rapportserie 2001:3 R

Examensrättsprövning
– utgångspunkter och tillvägagångssätt för Högskoleverkets examensrättsprövning
Högskoleverkets rapportserie 2001:4 R

Förnyad granskning och bedömning av kvalitetesarbetet vid Lunds universitet
Högskoleverkets rapportserie 2001:5 R

Förnyad granskning och bedömning av kvalitetesarbetet vid Uppsala universitet
Högskoleverkets rapportserie 2001:6 R

Karriär genom befordran och rekrytering
Högskoleverkets rapportserie 2001:7 R

Högskoleverkets utvärderingar – från bedömning av kvalitetsarbete till bedömning av kvalitet
Högskoleverkets rapportserie 2001:8 R

From quality audit to quality assessment
The New Evaluation Approach for Swedish Higher Education
Högskoleverkets rapportserie 2001:9 R

Internationell jämförbarhet & nationell styrning
– aktuella perspektiv på högskolans examensordning
Högskoleverkets rapportserie 2001:10 R

National Rewiew of Subjects and Programmes
Högskoleverkets rapportserie 2001:11 R

Forskarskolor i Sverige – en sammanställning
Högskoleverkets rapportserie 2001:12 R

Utvärdering av datavetenskapliga/datalogiska utbildningar i Sverige
Högskoleverkets rapportserie 2001:13 R

Förnyad granskning och bedömning av kvalitetsarbetet vid Kungl. Musikhögskolan, Mälardalens högskola, Karlstads universitet samt Örebro universitet
Högskoleverkets rapportserie 2001:14 R

Tid för studier – en jämförelse mellan fyra yrkesutbildningar
Högskoleverkets rapportserie 2001:15 R

Förnyad granskning och bedömning av kvalitetsarbetet vid högskolan i Skövde
Högskoleverkets rapportserie 2001:16 R

Granskning och bedömning av kvalitetsarbetet vid Malmö högskola
Högskoleverkets rapportserie 2001:17 R

Pedagogisk skicklighet och pedagogiska meriter – historik och praktik
Högskoleverkets rapportserie 2001:18 R

Högskoleprovets prognosvärde
Högskoleverkets rapportserie 2001:19 R

Core curriculum – en bildningsresa
Högskoleverkets rapportserie 2001:20 R

Akademisk frihet – en rent akademisk fråga?
Högskoleverkets rapportserie 2001:21 R

Förnyad granskning och bedömning av kvalitetsarbetet vid Högskolan i Kalmar, Högskolan i Trollhättan/Uddevalla, Karolinska Institutet samt Stockholms universitet
Högskoleverkets rapportserie 2001:22 R

Förnyad granskning och bedömning av kvalitetsarbetet vid högskolan Kristianstad
Högskoleverkets rapportserie 2001:23 R

Utvecklingen av högskolans samverkansuppdrag
Högskoleverkets rapportserie 2001:24 R

Utvärdering av medie- och kommunikationsvetenskapliga utbildningar vid svenska universitet och högskolor
Högskoleverkets rapportserie 2001:25 R

Studenterna i Sverige. Om livet och tillvaron som student vid sekelskiftet 2000
Högskoleverkets rapportserie 2001:26 R

20 åtgärder för att stärka studenternas rättssäkerhet – Högskoleverkets redovisning av ett regeringsuppdrag
Högskoleverkets rapportserie 2001:27 R

Studenternas resultat
Högskoleverkets rapportserie 2001:28 R

Högskolelärares bisysslor
– hur fungerar högskolans information och kontroll?
Högskoleverkets rapportserie 2001:29 R

Kollegialitet eller rättssäkerhet
Högskoleverkets rapportserie 2001:30 R

Uppföljning av särskilt urval vid antagning
Högskoleverkets rapportserie 2002:1 R

Befordringsreformen 1999: Hur har det gått?
Högskoleverkets rapportserie 2002:2 R

Utvärdering av ämnena latin, grekiska och nygrekiska vid svenska universitet
Högskoleverkets rapportserie 2002:3 R

Utvärdering av teologiska och religionsvetenskapliga utbildningar vid svenska universitet och högskolor
Högskoleverkets rapportserie 2002:4 R

Utvärdering av matematikutbildningar vid svenska universitet och högskolor
Högskoleverkets rapportserie 2002:5 R

Kvalitetsutveckling och kvalitetsmodeller för högskolans bibliotek – en förstudie
Högskoleverkets rapportserie 2002:6 R

Utbildning anpassad efter arbetsmarkandens behov – så arbetar högskolan
Högskoleverkets rapportserie 2002:7 R

Utvärdering av utbildningar i svenska/nordiska språk i Sverige
Högskoleverkets rapportserie 2002:8 R

Utvärdering av utbildning i nationalekonomi vid svenska universitet och högskolor
Högskoleverkets rapportserie 2002:9 R

Utvärdering av utbildning i företagsekonomi vid svenska universitet och högskolor
Högskoleverkets rapportserie 2002:10 R

Utvärdering av grund- och forskarutbildning i lingvistik och teckenspråk i Sverige
Högskoleverkets rapportserie 2002:11 R

Förnyad granskning och bedömning av kvalitetsarbetet vid Danshögskolan
Högskoleverkets rapportserie 2002:12 R

Förnyad granskning och bedömning av kvalitetsarbetet vid Dramatiska institutet, Högskolan Dalarna, Kungliga Konsthögskolan och Sveriges lantbruksuniversitet
Högskoleverkets rapportserie 2002:13 R

Förnyad granskning och bedömning av kvalitetsarbetet vid Högskolan i Halmstad
Högskoleverkets rapportserie 2002:14 R

Högskoleverkets tillsyn
Viktigare beslut 2000–2001
Högskoleverkets rapportserie 2002:15 R

Förnyad granskning och bedömning av kvalitetsarbetet vid Chalmers tekniska högskola, Göteborgs universitet och Kungl Tekniska högskolan
Högskoleverkets rapportserie 2002:16 R

Granskning och bedömning av kvalitetsarbetet vid högskolan på Gotland
Högskoleverkets rapportserie 2002:17 R

Studiemedelssystemets tidsgräns
– en sammanställning av lärosätenas synpunkter
Högskoleverkets rapportserie 2002:18 R

Ledarutveckling – ett styrinstrument?
Högskoleverkets rapportserie 2002:19 R

Metautvärdering av Högskoleverkets modell för kvalitetsbedömning av högre utbildning. Hur har lärosäten och bedömare uppfattat modellen?
Högskoleverkets rapportserie 2002:20 R

Studentspegeln 2002
Högskoleverkets rapportserie 2002:21 R

Högskoleverkets rapportserie 2002:22 R
ISSN: 1400-948X

Högskoleverket är en central myndighet för frågor som rör universitet och högskolor. Verket arbetar med kvalitetsbedömningar, tillsyn, uppföljningar, utveckling av högre utbildning, utredningar och analyser, bedömning av utländsk utbildning och studieinformation.

HÖGSKOLEVERKET
National Agency for Higher Education