Prediction of study success

- should selection instruments measure cognitive or non-cognitive factors?

Christina Wikström, Per-Erik Lyrén & Magnus Wikström

Umeå University, Sweden

Abstract

This study compares the subtests of a traditional admissions test (the SweSAT) with high school grades in verbal and quantitative subjects in terms of construct and predictive strength as selection instruments to higher education. The purpose is to find out if and how the difference in predictive validity can be related to the properties of these instruments. The data is analysed by factor analysis, structural equation modelling (SEM) and regression analysis. The findings show that there is a common verbal factor in the verbal subtests and the English grade, and a quantitative factor shared by the quantitative subtests and also the maths grade. A third and separate factor is also detected in the grades in Swedish. This is assumed to be a grade related factor incorporating non-cognitive skills. When comparing the predictive strength of these instruments, with credits from the freshman year in a Swedish economics programme, the pattern is not entirely clear, but tests and grades loading on the quantitative factor as well as the x-factor can explain some of the variation on the criterion for academic performance (first year credits). None of the verbal sub-scores in the SweSAT correlates positively with academic achievement.

Introduction

In Sweden, there are two instruments with the purpose to select applicants to higher education; upper secondary school grades (GPA) and an admissions test called the SweSAT. The usefulness, i.e predictive strength, of the admissions test and the upper secondary school grades have been debated for a number of years, in Sweden but also internationally when it comes to similar tests. Previous research generally shows that the upper secondary school grades are better at predicting academic success. The hypothesis is generally that the GPA represents cognitive as well as non-cognitive aspects, both important in education, while the test does not capture motivation, industriousness, social skills etc (Gustafsson, 2003). The SweSAT is now about to be revised. The direction is clear, however. It has been suggested that the test should be more directed towards school curricula, like similar tests internationally, and hence capture similar aspects as the grades. It has also been suggested that the test should be changed the other way, by being more clearly focused on verbal and quantitative factors like traditional aptitude tests. The decision to this date is to make a first revision where more emphasis is given to the quantitative part of the test, and this way increase the predictive validity of the test and the relevance for different types of university programmes by making it possible to select students on the basis of verbal or quantitative sub-scores (Wikström, 2008; Lyrén, 2009). However, most of the arguments regarding how the test should be modified are mainly based on assumptions. It is not very clear what the school grades from upper secondary school and the subtests of the SweSAT really measure. There has been limited research on the construct validity of these instruments and how elements within these instruments contribute to the prediction of study success in higher education. There are however several studies comparing the predictive validity of the SweSAT and the GPA, generally showing that neither instrument can explain much of the variance, but that the GPA from upper secondary school correlates higher with study success in higher education than the test does, hence being the basis for the criticism against the SweSAT (see, for instance, Svensson, Gustafsson & Reuterberg, 2001; Gustafsson, 2003; Svensson & Nielsen, 2004). Among the few studies focusing on the construct validity of the subtests, there is a study by Lexelius (2005) that analyses the relationship between the DS-test and math-grades. He finds that these instruments are basically measuring similar constructs. Åberg-Bengtsson (1999) has studied items in the DTM test and found that it partly measures a quantitative factor shared by the DS test. Studies of the verbal tests and grades are scarcer. Gustafsson (2003) has identified a factor that he calls ”X”, which *is a property of the school grades. According to Gustafsson, this factor is a complex of personality traits, interests and social background (e. g., motivation, effort, industriousness, interests, social and communication skills), accounts for much of the general grades factor. (…) A grade average is a better predictor of subsequent achievement than measures of Gf and Gc, which probably is because grades represent factor X, along with Gf and Gc.*

This study is partly built on this claim. The hypothesis is if this x-factor has been found in the GPA, it may also be found in the separate grades, and consequently that instruments with this factor, either it is a grade or a test score, would be a good predictor of academic achievement. This would also be important to know in the discussion about how to revise the admissions instruments and especially the test/s. Hence, the purpose of this study is to find out to what degree grades and sub-tests are measuring common factors, and if there are grades that are more sensitive to non-cognitive aspects than other grades. Furthermore, the study will investigate the relationship between isolated grades, sub-scores and performance in higher education. The aim is to find out if there are differences in predictive strength between grades and subtests that can be explained by the construct of the instrument, i.e if it can be linked to cognitive or non-cognitive factors.

Data and research strategy

When a student applies for a university course or programme, grades from upper secondary school (or equivalent) are used for eligibility. If there are more applicants than study places, he or she will be selected on the basis of his or her GPA or SweSAT score. The instruments are never combined as in many other countries, but applicants are ranked in different quota groups depending on which instrument he or she is applying with. An applicant with both a GPA and a test score from the SweSAT will be admitted on the basis of the instrument that is most favourable. The GPA is the average grade from upper secondary school. When calculating the GPA, all grades included have, until now, been of equal weight. It is basically a criterion-referenced system, but when the GPA is calculated, the letter grades are quantified and calculated, making the GPA a relative measure, with the sole purpose of ranking students (Wikström, 2005). The SweSAT is also a norm-referenced instrument, used for ranking the students. At present, the SweSAT includes five subtests: Swedish vocabulary (WORD); Swedish reading comprehension (READ); English reading comprehension (ERC); Data sufficiency (DS) and Diagrams, tables and maps (DTM). The test is optional but open for all, and administered twice a year. There is no limit to the number of times the test can be taken. A test score is valid for five years and it is the best score that counts. This means that the student has little to lose by taking and also by re-taking the test (Henriksson & Bränberg, 1994; Törnkvist & Henriksson, 2002). The results are public, which means that the items and the test forms are changed for each administration. To ensure comparability, the test is standardised and equated over time (Stage, 2003).

This study is based on empirical data. The focus will be on grades from courses that can be linked to the subtests of the SweSAT, such as grades in Swedish, English and Mathematics. It should be noted that not all students have taken the SweSAT, but the proportion is high among students on academic track. The analysis is based on descriptive statistics, regression analysis, factor analysis

and structural equation modelling. The analyses are carried out with the software SPSS and AMOS. The purpose of the factor analysis is to identify common factors in grades and test scores. When the main factors are identified, analyses with structural equation modelling is carried out, in order to control for explanation variables and find grades and/or test scores that include other factors than those that can be linked to verbal or quantitative factors. Finally, the predictive validity of grades with-, and grades without non-cognitive factors will be compared. The predictive validity analysis will be carried out through linear regression analysis. First year credits at university level will serve as criterion for academic success.

The data used in this study has been collected by Statistics Sweden (SCB) and includes information about all students who graduated from upper secondary school in 1997-2002 (N=209 758). The data comprises information about the students' background, such as birth year, sex, socioeconomic background, immigrant status, previous study results, SweSAT scores, their grades and GPA from elementary and upper secondary school. Grades take the values 0, 10, 15 or 20, and the GPA from upper secondary school 0-20.00, following an approximate normal distribution. The SweSAT scores on a normed scale from about 0.0 to 2.0. In the first part of the study, focusing on the construct validity of the grades and the subtests, the data includes a random selection of all the students how graduated from upper secondary school 1997-2002 (n= 20 874). 13 416 of these have taken the SweSAT at least once. In the second part, where the predictive validity of the grades and the subtests is investigated, only the students who entered a university programme in Economics or business administration in the autumn of 1999 are included (N=3159), and when missing data is removed, 1 653 remains. The purpose with this selection is to have an as homogeneous group as possible, who also have been studying at university level within the same time frame, not too long after they graduate from upper secondary school. The economics programme is regarded appropriate since there is a selection to the programmes, there is usually a fairly even composition in terms of gender and socio-economic background among the admitted, and there is a variation in terms of how many credits the students achieve, which here serves as criterion for academic performance. For a more thorough description of the data, see the appendix.

Analysis

Construct validity

In order to find out the degree of common factors between the SweSAT subtests and the grades, an explorative factor analysis is carried out. The variables included in the analysis are (1) scores from each subtest in the SweSAT (DS, DTK, READ, WORD and ERC), where the scores are equated in order to be comparable over time, and (2) grades in the compulsory subjects Swedish, English and Mathematics (from the courses Swedish A, Swedish B, English A and Mathematics A). Based on how the factors are loading on the different variables, it will be possible to make

inferences about the underlying ”abilities” affecting test scores and school grades. Two basic models are tested. In the first model the components to be analysed are assumed to have either a test factor or a grade factor. In the other model the components are assumed to have either a verbal factor or a numerical factor. The results from the analysis show that there are three factors with an eigenvalue lager than 1, explaining almost 73 percent of the total variance. It is also obvious that these three factors that are of most importance, since there is considerable drop in eigenvalue between the third and forth component. The rotated factors explain the variance to almost the same degree (28,0, 23,6 and 21,4 percent). The factor loadings for the rotated solution (Table 1 below) show that the first factor are loading highly on the SweSAT subtests WORD, READ, ERC and the school grade English A. It is also loading a little on the subtest DTM (correlation over 0,30). Based on these loadings the factor is interpreted as a verbal factor. The other factor is loading on the subtests DS and DTM, and the grade in Mathematics A, and is therefore labelled a quantitative factor. It is also loading a little on the subtest READ, which may be interpreted as the test is requiring analytical and logical thinking. The last factor is mostly loading on the two grades in Swedish and to some extent on the English grade. It also loads on the Mathematics grade. It is assumed that this factor is grade related, but not strictly verbal or quantitative.

Table 1. Factor loadings for the rotated solution.

	Component		
	1	2	3
WORD	,828	,147	,166
READ	,739	,307	,174
ERC	,835	,227	,126
EngA	,584	,135	,516
DS	,261	,833	,057
DTM	,338	,785	,031
MathA	,059	,770	,392
SweA	,189	,157	,835
SweB	,159	,099	,854

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Following the factor analysis, SEM (Structural equations modelling) is then used for comparing and extending the models above. The model contains the factors that are given by the factor analysis above and explanatory variables. The first model builds on the same assumption as

above, i.e. that there is a test-factor and a grade-factor, which can explain the difference between grades and SweSAT. The other model is based on the assumption that there is a verbal factor and a quantitative factor that can explain the variation in test scores and grade outcome. However, from the basis of the RMSEA-values and AIC-values of the two models it can be concluded that the Verbal–Quantitative model is better adapted to data than the Test–Grade model, but also that none of the models fits the data as well as would be desired (see the appendix for details). After modifying the model several times from the perspective that three factors (Verbal-Quantitative, and the third factor "X") can explain the main part of the variation, and noticing that the error terms correlates with each other, an additional factor is included in the model. This factor is common for all three factors (see figure 1). Three background variables are also included, that are expected to contribute to explaining and controlling for the variation in the factors. One variable is socioeconomic background, the other is gender and the third is the graduation year from upper secondary school. A forth variable was also tested –which type of school the student graduated from - but this did not improve the model fit and is therefore not included in the final model.

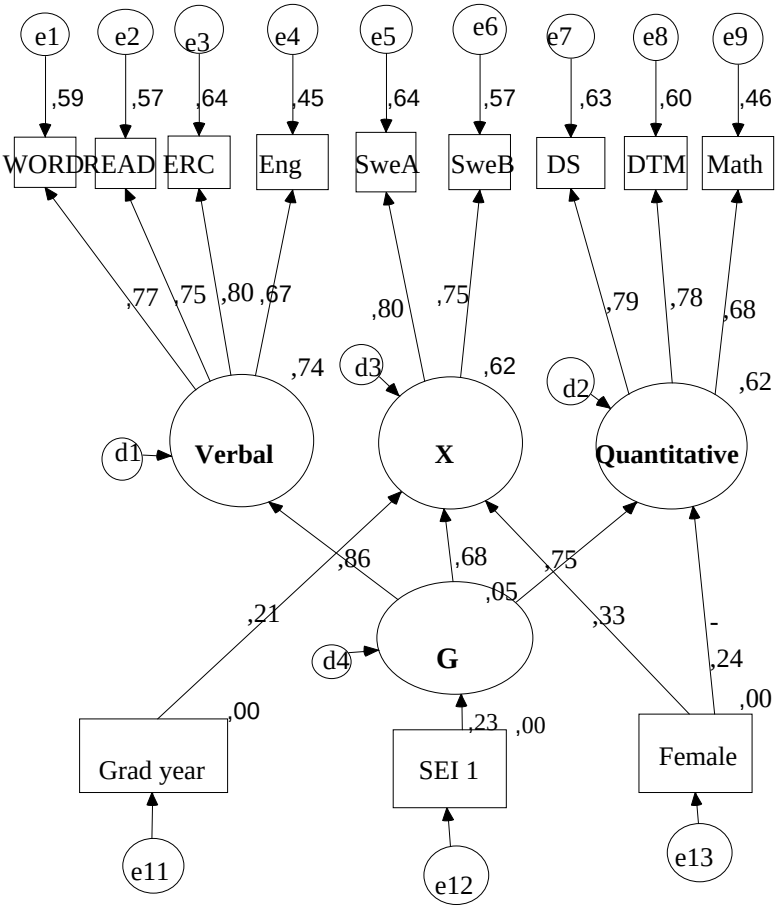

Figure 1. Model based on the factor analysis; version with best model-data-fit with standardised parameter estimates.

The model with the best fit is illustrated in figure 1 above. The forth factor is here labelled “G” as it is assumed to be a general ability factor. The explanatory variable Graduation year is linked to the X-factor, the gender variable (Female) is linked to the numerical factor and the x-factor, and shows that females are advantaged in the x-factor, but not when it comes to the quantitative factor. Socioeconomic background is connected to the G-factor, which, in turn, is explaining the variation in the other three factors, and especially the verbal factor, thereafter the numerical factor, and to the least extent to the x-factor. The x-factor is mostly affected by graduation year and gender. Since it is well known that female students receive higher grades than male students (everything else equal), and that graduation year is affected by grade inflation, it strengthens the assumption that the x-factor is grade related, and could be similar to the factor discovered in earlier studies of the grades (see previous research).

Predictive validity

As a second step in this analysis, the grades and test scores analyses above will be analysed in terms of their predictive strength as selection instruments to higher education. The method used for this analysis is linear regression. As dependent variable is a measure of academic performance, in this case the number of credits taken at the first year in an economics or business administration programme at a Swedish university. It should be noted that, as previously described, the population for this analysis is smaller and somewhat different than the population behind the factor analysis and SEM above. The reason is practical: it is necessary to obtain a criterion that is common for all included in the analysis, and combining credits taken at various programmes of very different character would be problematic. Furthermore, in order to make as adequate comparisons as possible among the students, only one cohort (the earliest) is selected.

There are a number of explanatory variables included in the regression, all known to be important for explaining academic success. These variables are; gender, ethnic background (immigrant), socioeconomic background, previous type of education (school type and programme type), here coded as dummy variables taking the value 0 or 1, as well as graduation year from upper secondary school and number of times SweSAT has been taken. The other variables are the grades and subtests that previously have been analysed in this study.

Table 2. Regression estimation results: dependent variable First year credits

Dependent variable: First year credits (n=1653)		
	Coefficient	Std.Error
(Constant)	3260,85**	813,32
Female	*0,49*	*0,64*
Immigrant	-2,52*	1,29
SEI1	-0,43	0,60
Soc.sc progr	3,15**	0,67
Private school	-3,46*	1,69
Grad year	-1,62**	0,41
#SweSAT	-0,92**	0,23
SweA	*0,09*	*0,11*
SweB	0,43**	0,10
WORD	-0,12*	0,07
READ	*-0,08*	*0,12*
ERC	*0,13*	*0,11*
Eng A	-0,33**	0,11
MathsA	0,43**	0,10
DTM	0,27**	0,11
DS	0,18	0,10
R^2 0,09		

Note: Values in table are unstandardized coefficients. ** denotes that coefficients are significantly determined on the 99% level of significance, * denotes that the coefficient is significant on the 95% level.

The analysis shows that the variables entered in the regression only partly contribute to explaining the variation in academic achievement. However, most of the coefficients are significant, and shows that there are differences in performance on the criterion that can be explained by student characteristics as well as performance on the grades and subtests of interest in this study. Surprisingly, there is no significant difference between students of different gender and social background, which is highly unusual in educational contexts. However, the students in this programme are likely to be fairly homogeneous and the group self-selected, which means that they may not be representative for a larger population of university students. Still, there is a difference between students of Swedish-non-Swedish origin, showing that students who are first or second generation immigrants perform somewhat lower than other students. The students' previous education matters and those who have graduated from a social science oriented programme perform higher than other students. The earlier graduation year from upper secondary

school, the higher achievement in terms of credits, which is a little surprising, given the fact that all students included here entered the particular programme at the same time. Number of times the test is taken is negative, which is expected since it is expected to be related to a gambling-behaviour that also to a higher degree makes these student change track or something similar.

The instruments (grades and subtests) are sorted according to their common factors in the table. The pattern is not entirely clear, but the analysis shows that it is the grade(s) loading on factor x, and the grades and subtests loading on the quantitative factor that contribute to the prediction of academic achievement. The grades and test loading on the verbal factor are negative, or non-significant. This is rather surprising, since it could have been expected that instruments measuring verbal abilities would be important for a university programme in the social sciences area, and especially the freshman year.

Conclusion

This study has investigated the characteristics of the upper secondary school grades in Swedish, English and Mathematics (basic level courses) together with the SweSAT subtests. The purpose has been to find information about what these instruments measure and if a non-cognitive factor, or "factor x" could be identified. The purpose was also to see if there are differences in the instruments predictive strength that could be related to common factors.

The analysis resulted in a model where three factors were identified, plus a forth factor that had importance for the other three, and was identified by SEM. Since factor one and two seem to be characterised in terms of content rather than format, the first factor is labelled verbal and the other quantitative, in line with the content of the instruments. The third factor is labelled x, and assumed to be at least partly a non-cognitive factor, in line with similar previous findings. The forth factor is assumed to be a general factor G since it is of importance for all other factors.

The conclusion is that the grades and subtests are partly measuring the same constructs. There is an element of non-cognitive aspects, here illuminated through the x-factor, loading on grades in Swedish. This may of course be a factor that is specific for the Swedish subject, but the fact that it does not relate to the verbal factor, like the English grade, or tests measuring vocabulary, reading comprehension etc suggests that it is something very different. Since there is evidence that the factor is grade-related, it supports that this factor is similar to the so called x-factor found in other studies, and has to do with students social skills and industriousness (Gustafsson, 2003). Instruments containing this factor, as well as instruments containing the quantitative factor, seem to contribute to the prediction of academic achievement, as opposed to the other verbal tests and grades. These findings suggest that especially the verbal instruments may need to be revised to better serve their prediction purposes. It also raises the question whether it is appropriate to select students on the basis of this factor x in a meritocratic admissions system, and, if so, how such

abilities could be measured by means of other instruments than the grades. The results from this study should however be interpreted with caution. It should be noted that it is limited when it comes to population and criterion. Future research should investigate the constructs of the verbal instruments more thoroughly, in order to find out why the verbal instruments are unsuccessful in the prediction. It would also be useful to learn more about other grades than those included here. Further prediction studies should also focus on other populations or criteria, since the predictive strength of selection instruments are likely to vary depending on university programme and how academic achievement is measured.

References

Gustafsson, J-E., (2003). *The predictive validity of grades and aptitude tests in higher education*. Presentation at the AEA-Europe conference in Lyon, France, November 2003. Presentation retrieved from www.aea-europe.net

Henriksson, W. & Bränberg, K. (1994). The effects of practice on the Swedish Scholastic Aptitude Test (SweSAT). *Scandinavian Journal of Educational Research, 28*(2), 129-148.

Lexelius, A. (2005). *Högskoleprovet och de målrelaterade betygen* (BVM-report). Umeå: Umeå University, Department of educational measurement.

Lyrén, P-E. (2009). *A perfect score. Validity arguments for College Admission Tests.* (doctoral thesis). Umeå: Umeå University, Department of educational measurement.

Stage, C. (2003). *Normering, ekvivalering eller kalibrering av delar av högskoleprovet* (Pm nr 185). Umeå: Umeå University, Department of educational measurement.

Svensson, A., Gustafsson, J-G., & Reuterberg, S-E. (2001). *Högskoleprovets prognosvärde.* [the prognostic value of the SweSAT]. 2001:19R. Stockholm: Agency for higher education.

Svensson, A., & Nielsen, B. (2004). *Urvalet till juristutbildningen. Skillnader mellan dem som antas i betygs- respektive provurvalet.* [the selection to the law program. Differences between student admitted by tests and grades] (IPD-report nr 2004:11). Göteborg: Göteborgs University, Department of education.

Tabachnich, B.G, & Fidell, L.S. (2001). *Using multivariate statistics* (4th edition), Boston: Allyn & Bacon.

Törnkvist, B. &. Henriksson, W. (2002). *The Effects of Repeated Test Taking in Relation to the Test Taker and the Rules for Selection to Higher Education in Sweden.* (EM Nr 41). Umeå: Umeå University, Department of educational measurement.

Wikström, C. (2005). *Criterion-referenced measurement for educational evaluation and selection.* (doctoral thesis) Umeå: Umeå University, Department of educational measurement.

Wikström, C. (2008). *Urvalsinstrument ur ett svenskt och internationellt perspektiv:* (BVM 2008:35) Umeå, Sweden: Department of Educational Measurement, Umeå University.

Åberg-Bengtsson, L. (1999). Dimensions of performance in the interpretation of diagram, tables, and maps: Some gender differences in the Swedish Scholastic Aptitude Test. *Journal of Research in Science Teaching, 36*, 565–582.