

Predicting study performance for one academic year at university

A comparison of the predictive validity of the Finnish matriculation examination,
entrance exam and fluid intelligence

Jonathan Rasmus

Master's Thesis in Psychology

Supervisor: Matti Laine

Faculty of Arts, Psychology and Theology

Åbo Akademi University

Turku 2016

**ÅBO AKADEMI UNIVERSITY – FACULTY OF ARTS, PSYCHOLOGY AND
THEOLOGY**

Subject: Psychology	
Author: Jonathan Rasmus	
Title: Predicting study performance for one academic year at university: A comparison of the predictive validity of the Finnish matriculation examination, entrance exam and fluid intelligence	
Supervisor: Matti Laine	
Abstract: The present study addressed the predictive validity of three common types of selection methods used in student selection in Finland by examining their ability to predict one-year study performance at university. The three predictors were the Finnish matriculation examination, entrance exam, and a fluid intelligence task. In order to tap various aspects of study performance, three outcomes were measured; grade point average, amount of credit, and a composite of the two. The sample consisted of 100 students from Åbo Akademi University. Based on hierarchical multiple regression analyses, the matriculation examination scores significantly predicted grade point average, fluid intelligence showed only a weak positive relationship to grade point average, and the entrance exam failed to predict any of the outcome measures. The implications of the results for student selection and follow-up at university level are discussed.	
Keywords: Finnish matriculation examination, entrance exam, fluid intelligence, academic performance	
Date: 17.5.2016	Page count: 30

ÅBO AKADEMI – FAKULTETEN FÖR HUMANIORA, PSYKOLOGI OCH TEOLOGI

Ämne: Psykologi	
Författare: Jonathan Rasmus	
Titel: Predicting study performance for one academic year at university: A comparison of the predictive validity of the Finnish matriculation examination, entrance exam and fluid intelligence	
Handledare: Matti Laine	
<p>Abstract: Syftet med denna studie var att undersöka studentexamensbetygets och inträdesprovets prediktiva validitet på ett års studieprestation vid universitet. Därutöver undersöktes ifall ett mått på flytande intelligens kunde erbjuda inkrementell validitet över de traditionella antagningskriterierna. Tre beroende variabler användes för att undersöka olika aspekter av studieprestation: medelvitsord, antalet studiepoäng och en sammansättning av dessa två variabler. Samplet bestod av 100 studerande från Åbo Akademi. Den prediktiva validiteten av variablerna analyserades med hierarkisk multipel regression.</p> <p>Studentexamensbetyget visade en signifikant prediktion av medelvitsordet, flytande intelligens en svag positiv relation med medelvitsord, och inträdesproven visade ingen relation till de mätta beroende variablerna. Resultatets innebörd för antagningsprocessen till universitet diskuteras.</p>	
Nyckelord: Studentexamen, inträdesprov, flytande intelligens, akademisk prestation	
Datum: 17.5.2016	Sidoantal: 30

Acknowledgments

I would like to express my gratitude to my advisor Matti Laine for the interest, knowledge and guidance he offered. I am also grateful for the help Mira Karrasch and Pekka Santtila offered at different stages of this work.

My friends and co-workers, Eric Nyfors, Christopher Harju and Oscar Damén all need a special mention because of their unique and intelligent contribution to this project.

I am also very grateful for the positive attitude and help that many persons in the administrative staff at Åbo Akademi University provided during this project.

Last but not least, I thank Anne, my family and friends who offer the much needed balance between work and life.

Table of contents

Abstract

Acknowledgments

1 Introduction	1
1.1 Predictive validity of common international admission measures and theoretical background	2
1.2 Student selection and predictive validity of admission measures in Finland	4
1.3 Aims of the study	6
2 Method	6
2.1 Predictive measures	6
2.2 Criterion measures	8
2.3 Participants	9
2.4 Procedure	10
2.5 Statistical analyses	11
3 Results	11
3.1 The fluid intelligence measure	12
3.2 Predicting study performance at university	12
4 Discussion	14
4.1 Predictive validity of the three predictors	14
4.2 Limitations of the study	15
4.3 Implications and future studies	17
4.4 Conclusion	19
5 Swedish Summary – Svensk sammanfattning	19
6 References	26
Appendix	
A Examples of similar items to the ones used in the actual measure	
B Distribution of scores on entrance exam and the intelligence measure	
C Scoring of the matriculation examination results	
Pressmeddelande	

1 Introduction

Universities around the world face the challenge of having far more applicants than can be admitted. Therefore, universities have to create their admission requirements upon which the selection of students is based. Specific requirements differ between countries, universities, faculties and programs. The creation of these admission requirements, their implementation and the selection of students tend to require considerable resources.

Some tests and measurements are widely used as admission requirements in certain countries and can thus be seen as a golden standard within the country. These include, for example, the SAT in the USA, the SweSAT in Sweden, and the Finnish matriculation examination in Finland. In addition to these national golden standards, a large variety of admission tests and requirements is used in less standardized forms, such as program-specific entrance exams.

A basic issue in admission measurements concerns aptitude versus achievement (Atkinson, 2004). Aptitude is a person's potential for learning (Kaplan & Saccuzzo, 2009), whereas achievement refers to a person's accomplishment in a specific area (Stemler, 2012). Concerning admission procedures, at issue is whether aptitude tests, measuring e.g. general or fluid intelligence, predict academic performance better than achievement tests that are designed to tap subject-specific achievement, say for example in economics or psychology.

To examine the predictive validity of various tests and measurements in the context of academic performance, one needs to conduct follow-up studies where one attempts to predict future academic performance on the basis of prior tests and measurements. The most commonly used statistical methods in predictive validity studies are correlation and regression analyses (see e.g. Burton & Ramist, 2001). Some measures, such as an individual's admission test score are natural candidates for predictive variables, but there could be other measures or variables that could give a more accurate estimate of future performance.

Kuncel, Hezlett and Ones (2001) suggest that in order to examine the validity of a selection measure, three aspects of validation should be addressed: methodological, statistical and theoretical. Theoretical aspects concern whether the measurements used capture all relevant abilities and help explain why a certain measure should predict performance. Statistical aspects concern features of the predictive and criterion measures used, such as the reliability of the predictive measures and restriction of range. The latter issue is relevant when participants have been selected according to certain criteria prior to the study. For example, using an admission test for selection limits the variance of test scores in the study, because the ones with the lowest admission test scores have not obtained a study place at the university and cannot therefore participate in the predictive validity study. Another statistical issue concerns the insufficiency of using a single criterion measure. Grade point average (GPA) is perhaps the most commonly used criterion variable of academic performance (Stemler, 2012), but does not alone capture all relevant aspects of performance.

1.1 Predictive validity of common international admission measures and their theoretical background

In the USA, several admission tests are widely used and many predictive validity studies have been conducted with these tests. Common admission tests in the USA are the SAT I, the SAT II and the American College Test (ACT). The SAT I consists of a verbal part and a mathematical part, and it is intended to be an aptitude test which gives information on students' capacity to learn (Geiser & Studley 2002). The SAT II, on the other hand, can be seen as an achievement test that assesses a person's current mastery of a subject. The SAT I has been shown to have high predictive power for academic performance as measured by GPA (DeBerad, Julka & Speilmans, 2004). Burton and Ramist (2001) reviewed studies on the SAT I and found a medium size weighted average correlation between the SAT verbal test and GPA, as well as between the SAT Math and GPA. However, Geiser and Studley (2001) reported higher predictive validity for the SAT II than for the SAT I. According to these studies, both the aptitude test version and the achievement test version of the SAT predict academic performance.

In Sweden, the two most widely used selection instruments for higher education are secondary school GPA and the Swedish Scholastic Assessment Test (SweSAT). The SweSAT is an optional test, since the admission to university is either based on the SweSAT or previous GPA. The SweSAT consists of five subtests: Swedish reading comprehension, Vocabulary, English reading comprehension, Data Sufficiency, and Diagrams, tables and maps. In a review, Lyrén (2008) found that the SweSAT scores predict academic performance to some extent, but he concluded that its predictive validity differs depending on study program. In three of the reviewed correlational studies, the predictive validity was higher in study programs such as law, engineering physics and medicine, but lower in programs such as social work, teacher education and business administration (Cliffordson, 2004; Svensson, Gustafsson & Reuterberg, 2001; Wolming, 1999). The size of the correlations varied mostly between .10 and .30. Burton and Ramist (2001) reported somewhat higher correlation coefficients for the SAT I in the USA, namely between .20 and .60. However, the studies included in these reviews used different criterion variables: the Swedish studies employed mostly the amount of credit while the US studies used mostly GPA.

As noted earlier, an important aspect of predictive validity is its theoretical basis, in other words, why is there a correlation between an admission test score and performance level at university. Spearman (1904) gave an early theoretical framework for intelligence by describing a general intelligence factor “g” consisting of more specific intelligence factors that correlate with the “g” factor. Why general intelligence is an important concept in this context is that it predicts performance on cognitively demanding tasks (Hunter & Hunter, 1984) and relates to acquisition of skills and knowledge (Lohman, 1999; Schmidt, 2002). Naturally enough, academic studies offer cognitively demanding tasks in various amounts, but general intelligence has been shown to predict performance in tasks of low, medium and high cognitive complexity (Hunter, 1980). Nevertheless, the importance of general intelligence in predicting academic performance is well established (Deary, Strand, Smith & Fernandes, 2007; Kuncel, Hezlett & Ones, 2001). Taking a longer perspective, general intelligence predicts not only academic performance, but future job performance as well (Schmidt, 2002), which can be seen as another argument for using measures of intelligence in the selection process.

With regard to more specific factors of intelligence, Horn and Cattell (1966) introduced the concepts crystallized intelligence and fluid intelligence. Crystallized intelligence can briefly be defined as acquired knowledge, whereas fluid intelligence refers to novel problem solving which is not culturally bound. Fluid intelligence correlates highly with general intelligence and according to Furnham (2012), measures of fluid intelligence are better predictors of academic performance than measures of crystallized intelligence. Also, measures of general intelligence tend to correlate with scores on the SAT I (Frey & Detterman, 2004)

1.2 Student selection and predictive validity of admission measures in Finland

In Finland, each university is allowed to decide upon its specific selection criteria. For the academic year 2013-2014, approximately 90 000 persons applied for a study place at a university and 26 091 (29 %) were selected. During the past four years, the number of applicants has risen by about 10 000 while the number of selected students rose only by around 500. Thus, an increasing number of applicants must be excluded in the selection procedure.

The golden standard for admission requirement is the matriculation examination (ME), which is one of the main criteria used by Finnish universities to select students. The majority of university students, but not all, have completed the ME, which usually takes place at the end of high school (Finnish lukio). The examination consists of subject-specific exams, where one has to pass at least four exams. One can complete an exam in nearly all subjects taught at high school. Languages and mathematics are offered as either a short or a long version. The exams are scored and standardized at a national level by the Matriculation Examination Board. The highest grade is *laudatur* and the lowest is *approbatur* on a six level grading scale. The grading is normally distributed.

The rationale for the extensive usage of ME might be that past educational performance predicts future educational performance. To take an example from a Finnish study, Vaahtera (2007) found that grades from Finnish primary school predicted performance in the matriculation examination. The conclusion that past educational performance predicts future educational performance is supported by many studies conducted in different countries (Larson & Scontrino, 1976; DeBerad,

Julka & Spielmans, 2004; Olani 2008; Robbins, et al., 2004; Utriainen, 2011). However, only a few studies have addressed the predictive validity of the Finnish matriculation examination concerning study performance at university level. Utriainen (2011) reported a statistically significant medium sized ($r = .45$) correlation between ME scores and grades at university, but the association between ME scores and amount of credit failed to reach significance.

Entrance exams (EE) are another common selection method. However, these are often specific to a university and a study program, although some programs have a nationwide entrance exam. At the Åbo Akademi University, where the present study was conducted, most entrance exams are created and organized by the programs and differ both in content and level of difficulty. It has been argued that entrance exam scores are not valid predictors of academic performance (Olani, 2008), but it is difficult to make generalizations regarding the predictive value of subject-specific entrance exam results because of their differences.

With regard to the contents of the ME and entrance exams, it can be assumed that the language exams of the ME tap, at least partly verbal intelligence, having some similarities to the SAT verbal test. Entrance exams in subjects such as history and geography can be argued to measure mainly crystallized intelligence since subject specific knowledge is required. All in all, the ME and entrance exams are achievement measures, perhaps more comparable with the SAT II than the SAT I. As mentioned earlier, both achievement and aptitude measures have shown to correlate with academic performance and, therefore, the ME and entrance exams were expected to do so also in the present study. Since measures of achievement are influenced by motivation, an aptitude test measuring fluid intelligence was added to the present study in order to examine whether it would provide any incremental predictive value.

The university culture in Finland differs from many other countries, which makes it difficult to generalize international research results regarding the predictive validity of achievement and aptitude measures to university studies. Firstly, there is no tuition fee for studying at a Finnish university and students receive financial support from the state. This might lower the pressure for completing a degree within the targeted time, since additional years spent at a university do not

bring an extra annual fee. Secondly, the university studies might be less grade-focused compared with the USA.

1.3 Aims of the study

The aim of this study was to evaluate the predictive validity of common selection criteria employed by Finnish universities by using Åbo Akademi University as the study object. Åbo Akademi University is the only completely Swedish-speaking university in Finland with campuses in Turku, Vaasa and Pietarsaari. In 2014, there were altogether 6 139 students at the university, around 70 % of them located in Turku and 20 % in Vaasa, 58 % female and 42 % male.

The predictive variables included both achievement and aptitude scores to see whether an intelligence test as an aptitude task could add incremental predictive value over the traditional criterions. Furthermore, the aim was to study which of the three predictive variables, the matriculation examination, entrance exam, or fluid intelligence, offers the best predictive value. The outcome variables for this one-year follow-up were grade point average, amount of credit, and a composite of these two variables. According to prior research and theoretical considerations, it was hypothesized that all three predictive variables are significantly associated with the outcome measures.

2 Method

2.1 Predictive measures

Matriculation examination. The participant's ME grades (as well as the entrance exam scores and grades from university courses) were gathered from the digital student registry at the Åbo Akademi University Computing Center in September 2015. ME exams are evaluated on a six level scale, from highest to lowest: *laudatur*, *eximia*, *magna*, *cum laude*, *lubenter* and

approbatur. The scoring system was the same as used by the Åbo Akademi University (Appendix C), and the sum of all exam scores was counted. ME scores were identified for 78 participants.

Entrance exam. All applicants are not required to participate in an entrance exam when applying to the university. Some applicants may acquire maximum or close to maximum admission points just from their ME scores and would thus not need to participate in the entrance exam. Most of the entrance exams were essays or/and multiple choice questions based on selected literature that the applicants are to study beforehand. Since the program-specific entrance exam scores had different amounts of maximum points, scores were computed as the percentage of correct answers of the maximum score possible. Entrance exam scores were identified for 59 out of 100 participants.

Fluid intelligence measure. The test used for measuring fluid intelligence was constructed by four master's students in psychology, including the present author. It was designed to tap fluid intelligence through its resemblance to other measures of the same construct, especially Raven's progressive matrices (RPM; Raven, Raven & Court 2000). RPM was used as a model since it has shown predictive validity on academic performance, as well as concurrent validity (Rohde & Thompson, 2007). The present measure consisted of 60 matrices with a time limit on 30 minutes. Participants had to identify a rule by which eight figures were arranged in a box and deduce which of the 6 alternatives fit in the empty box following the same logical rule (Appendix A). The 60 matrices were categorized in five different series, each series containing different type of logical rules. Each series began with easier matrices and thereafter advancing to more difficult ones. Each series was on average somewhat more difficult than the previous series. Items were made progressively more difficult by adding more changing components and logical rules. Pilot studies were conducted in order to arrange the matrices according to level of difficulty to see whether alternative solutions were found, and to decide an appropriate time limit.

Each correct response earned one point and an incorrect one gave zero points. For the items left unanswered when the time limit was reached, an estimate was made of how likely a person would be to answer it correctly, depending on the difficulty of the item and the proficiency of the participant. The probability was calculated by applying Item Response Theory and the guidelines

of Yu (2011). This was done in order to get a more accurate estimate of a participant's true ability, that is the score which a participant could have achieved with unlimited time. Proficiency stands for the average amount of correct answers by a participant. If a participant answered correctly on 40 out of 60 items, the proficiency level would equal 0.66. Difficulty of an item stands for the average amount of incorrect answers given on an individual item. If nine out of ten participants answered incorrectly on an item, the difficulty level would equal 0.9. The probability of an individual to answer an unanswered item correctly was then counted with the following equation (Yu, 2011):

$$P = 1/(1+\exp(-(\text{proficiency} - \text{difficulty})))$$

There were several reasons for designing a new intelligence measure for the present study. First, some established intelligence tests, such as the Raven task have been leaked to the internet where correct answers can be found, and some participants may be familiar with the tests. Second, no appropriate intelligence test was available for mass administration at the university. Third, there was interest to develop a measure that could be used in future selection processes for predicting academic performance.

2.2 Criterion measures

Grade point average and amount of credit are common measures of academic performance, and they were also used in the present study as criterion measures. Grades from university studies were collected in autumn 2015 for the previous academic year (1.9.2014 – 31.7.2015). Three different criterion variables were used: *grade point average (GPA)*, *amount of credit (AC)*, and a composite of both GPA and AC that is coined here as *academic performance (AP)*. The rationale behind this was that these three variables might behave differently, as different factors may underlie AC and GPA. Also, as the use of individual criterion variables for academic performance has received some criticism (Kuncel, Hezlett & Ones, 2001), using a composite of two criteria may give a more valid estimate of true performance.

Performance on a course is evaluated on a scale from 1 to 5, with a higher score meaning better performance. Some courses are simply scored as pass or fail, and these were not included in the GPA, only in the AC. If less than two courses taken during the year were evaluated on the scale 1 to 5, the participant was excluded from the analysis in order to avoid unreliable GPA. This led to an exclusion of five participants. Credits from a course are given according to the European Credit Transfer and Accumulation System (ETCS). Both AC and GPA were transformed to z-scores and then combined to yield the dependent variable AP. GPA was transformed to z-score individually for each faculty since there were faculty-wise differences in GPA, probably due to different grading criteria rather than actual differences in performance of students (table 2).

2.3 Participants

For the aptitude test with the new fluid intelligence measure, one hundred Åbo Akademi University students were recruited in the autumn of 2014. The recruitment was conducted by email, posters, and by informing about the study at several introductory seminars during the first week of the new academic year. The participants had a chance of winning a movie ticket and all of them could receive feedback on their test results. Students from both Turku and Vaasa campuses were recruited. The mean age of the participants was 22.2 years ($SD = 4.8$) and the majority (73.7 %) were first year students while the rest were students at different stages in their studies. In terms of geographic location and gender, the sample is representative for the university as a whole. The sample consists of 79 % from Turku (70 % of all students located in Turku), 21 % from Vaasa (20 %), 29.5 % males (42 %) and 70.5 % female (58 %). More descriptive information is provided in Table 1. The participants gave consent for their present and future study performance to be tracked.

As noted above, five of the 100 students were excluded due to unreliable GPA. Out of the remaining 95 participants, intelligence scores, matriculation examination results, entrance exam scores, grades and course credits could be identified for 50 students. These formed the final sample that was entered into multiple regression analyses.

The present study was approved by the ethics committee of the Department of Psychology and Logopedics. The Student Administration of the Åbo Akademi University provided the necessary data on the participants.

Table 1

Descriptive information of the whole sample and those who were included in the multiple regression analyses

	Whole sample (<i>n</i> =95*)	Final sample (<i>n</i> = 50)
Turku	79.0 %	68.0 %
Vaasa	21.0 %	32.0 %
Year of study		
First year student	73.7 %	65.3 %
Other	26.3 %	34.7 %
Gender		
Male	29.5 %	12.0 %
Female	70.5 %	88.0 %

Note. Five participants with unreliable GPA were excluded

2.4 Procedure

All in all, twelve group test sessions were held between August 2014 and November 2014 at different faculties, during different times and days. This way the students could participate in the session that best fitted their own schedule. Three sessions were held in Vaasa and nine in Turku, and each session lasted approximately two hours. At first, the participants filled in a questionnaire regarding demographic factors, well-being and lifestyle variables like exercise, alcohol use and motivation. The second part consisted of the new intelligence measure with 60 matrices. The participants were instructed to select one alternative of the six possibilities that would logically fit in the sequence. They were asked to complete as many matrices as possible within 30 minutes. The third part consisted of a personality questionnaire. Everyone started with each part at the same time. A standardized instruction manual was used at every session. The present study will focus on the cognitive predictors and therefore only use data from the intelligence measure. The survey and the personality questionnaire will be analyzed in another master's thesis.

2.5 Statistical analyses

Three separate hierarchical multiple regression analyses were conducted to assess the predictive validity of the matriculation examination, entrance exam scores, and the fluid intelligence measure. In each model the same three predictors were used, but the criterion variable was changed. The matriculation examination was the first predictor to be entered in the regression model, at the second stage fluid intelligence was entered and entrance exam was entered at stage three. If a model was found significant, the individual contribution of each predictor was reported in more detail.

3 Results

Results are presented in three parts. First, descriptive statistics on the predictors and the criterion variables are provided for the whole sample and divided by the four current faculties in Table 2. Second, the fluid intelligence measure designed for this study is briefly analyzed. Third, multiple regression analyses are conducted to examine the predictive validity of the independent variables.

Table 2

The total and faculty-wise means (standard deviations) on the independent and dependent variables

Measure	Included in regression (<i>n</i> = 50)	All (<i>n</i> = 95)	Arts, Psychology and Theology (<i>n</i> = 36)	Science and Engineering (<i>n</i> = 23)	Education and Welfare (<i>n</i> = 19)	Social Sciences and Business (<i>n</i> = 17)
GPA	3.6 (0.7)	3.5 (0.8)	3.7 (0.7)	3.2 (0.8)	3.8 (0.6)	3.4 (0.9)
AC	61.1 (15.7)	58.8 (15.6)	56.8 (15.0)	54.8 (12.4)	65.4 (17.7)	60.7 (16.7)
AP	.11 (1.3)	.00 (1.2)	-.06 (1.2)	-.13 (1.2)	.21 (1.4)	.06 (1.4)
ME	30.5 (9.2)	31.3 (9.0)	33.4 (9.4)	31.9 (8.4)	25.7 (6.6)	32.4 (9.2)
EE	.69 (0.2)	.69 (0.2)	.67 (0.1)	.60 (0.2)	.84 (0.2)	.59 (0.2)
FI	45.2 (5.4)	45.7 (5.8)	44.7 (5.3)	47.6 (7.6)	44.6 (5.0)	46.8 (4.8)

Note. Due to differences in GPA they were transformed to z-scores individually for each faculty. Academic performance reflects z-scores from AC and faculty-specific z-scores from GPA. Entrance exam scores represent percentage of correct answers. Intelligence measure show average points of the maximum 60. FI stands for Fluid Intelligence

3.1 The fluid intelligence measure

Internal consistency of the intelligence test was assessed using Cronbach's alpha (α). The reliability coefficient was $\alpha = .86$ which can be considered as good. No single item in the measurement would have considerably lowered or raised the reliability coefficient (see figure 1). The intelligence measure showed a significant correlation ($r=.34$, $p<0.01$) with science subjects scores (mathematics, chemistry and physics) from the ME, but not with other ME scores.

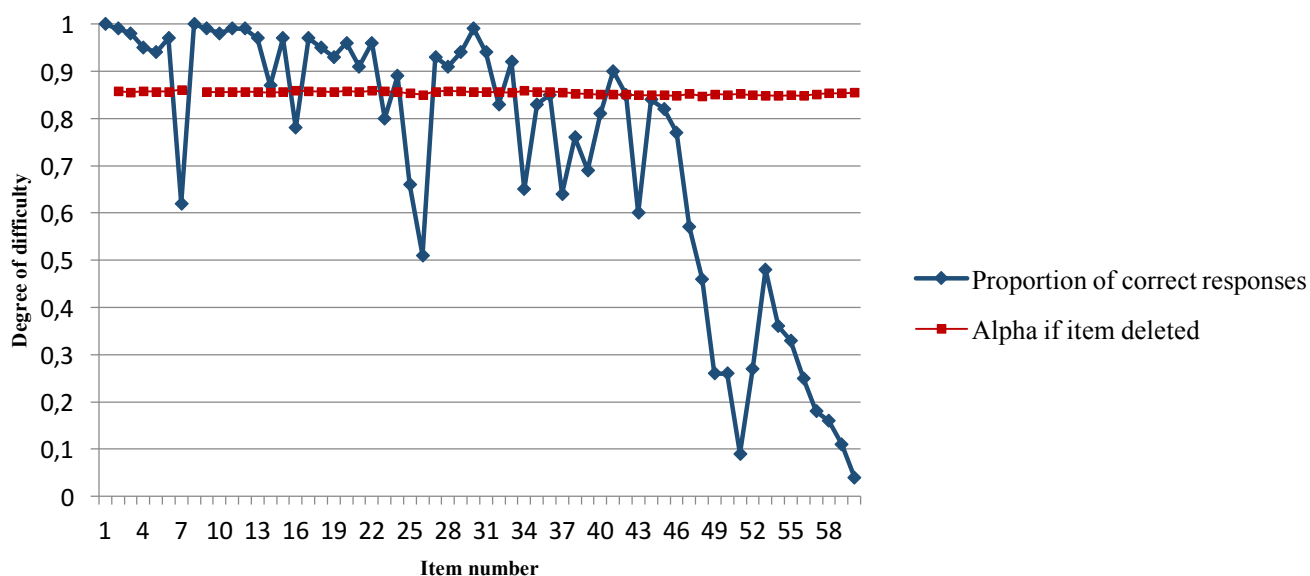


Figure 1. Proportion of correct responses on each item, where 1 = 100 %. Alpha if item deleted illustrates the overall reliability if that particular item would be deleted. All participants answered items 1 and 8 correct, therefore the alpha if item deleted is missing for these two items.

3.2 Predicting study performance at university

The independent variables were tested for multicollinearity. To provide incremental value, an independent variable should preferably correlate with the dependent variable and only weakly, or not at all, with the other independent variables used in the model (Kuncel, Hezlett & Ones, 2001). The model showed no risk for multicollinearity since the highest correlation between independent variables was .11, found between ME and the intelligence measure. It is suggested that a predictor should be removed if it correlates with another predictor at a level greater than .70 (Zizzi, 2005). The Variance inflation factor (VIF) was between 1.00 and 1.01 for all models,

which also suggests a lack of multicollinearity. However, it should be noted that entrance exam score did not correlate positively with the dependent variables and despite this, it was included in the regression models. Also, the distribution of entrance exam scores was negatively skewed (Appendix C). Correlations between all variables are presented in Table 3.

Table 3

Correlation matrix of the studied variables

	1.	2.	3.	4.	5.	6.
1. Matriculation Examination	-					
2. Entrance Exam	-.07	-				
3. Fluid Intelligence Measure	.07	.05	-			
4. Grade Point Average	.46**	-.09	.17	-		
5. Amount of Credit	.04	-.09	.09	.45**	-	
6. Academic Performance	.37**	-.11	.16	.94**	.74**	-

Note. * = $p < 0.05$, ** = $p < 0.01$.

Multiple regression analyses were conducted in three separate models with different dependent variables, namely GPA, AC and AP, while keeping the predictors constant. The multiple regression model with *AP as the criterion variable* was significant, $F(3, 46) = 2.97$, $p = 0.04$, with an overall R^2 of .16. The matriculation examination alone provided a R^2 of .13 and intelligence added .02 as R^2 change, while entrance exam did not add any value ($R^2 = .00$).

The second model, with same predictors but *GPA as the criterion variable* reached statistical significance $F(3, 46) = 4.762$, $p = 0.01$, and explained 24 % of the variance ($R^2 = .24$). The matriculation examination alone provided a R^2 of .21 while intelligence added .02 as R^2 change and entrance exam .01.

The third model using *AC as the criterion variable* did not reach significance $F(3, 46) = 0.293$, $p = 0.83$. The model explained only 2 % of the variance ($R^2 = .02$) in amount of credit.

Standardized beta values for each model are provided in Table 4.

Table 4

Standardized beta values for each predictor in three multiple regression models

Criterion variable	Academic Performance	Grade Point Average	Amount of Credit
	<i>B</i>	<i>B</i>	<i>B</i>
Constant	-3.06	-2.71	47.93
Matriculation Examination	0.35*	0.45*	0.03
Fluid Intelligence Measure	0.15	0.14	0.10
Entrance Exam	-0.09	-0.07	-0.10

Note. * = $p < 0.05$

4 Discussion

The aim of this study was to examine whether common methods for selecting university students in Finland predict one-year study performance at university, and to clarify whether a newly designed fluid intelligence measure would provide additional predictive value in the present context. To provide a better view on study performance, three different criterion variables were used, namely grade point average, amount of credit and a composite of these two. The study was conducted at one university, the Åbo Akademi University.

4.1 Predictive validity of the three predictors

In the multiple regression models, the matriculation examination emerged as the sole important predictor. The ME scores significantly predicted GPA and explained 21 % of the variance in the model. The fluid intelligence scores added 2 %, and the entrance exam 1 % to the variance explained, hence the latter two did not provide any significant incremental value. The significant positive correlation ($r = .51$) between the matriculation examination and GPA was similar to what has been found between the SAT I and GPA (Bridgeman, McCamley-Jenkins & Erwin, 2000). The ME scores were also the sole important predictor of AP. In the regression model with AP as the criterion variable, ME accounted for 13% of the variance, with a further non-significant two percent increase stemming from the fluid intelligence test. The significant relationship between AP and the predictors is almost solely due to the relationship between ME and GPA.

Entrance exam, even though widely used, did not predict AP, GPA or AC. It did not provide any incremental value in the regression models and the correlational coefficients to predictors were

close to zero on the negative side. There are some problems with the present entrance exams variable. The exams differed in both content and difficulty between programs. Entrance exam scores were not normally distributed, but instead negatively skewed (Appendix B). Nevertheless, the extensive use of entrance exams as a method for selecting students should be given a second thought, since there seemed to be no relationship between the exams and one-year study performance at university. The entrance exams did not provide any incremental value over the ME, suggesting that the ME alone would be sufficient as a selection criterion.

The fluid intelligence measure showed good reliability and a medium size significant correlation with scores from science subjects in the matriculation examination. Similar correlations between intelligence measures and performance in science subjects have been found in other studies as well (Furnham & Chamorro-Premuzic, 2004). On the other hand, the correlation between the fluid intelligence measure and GPA was weak ($r = .15$), and did not quite reach significance. Fluid intelligence did not significantly add explained variance to the prediction of AP, GPA or AC.

None of the predictors, either together or alone, correlated with the amount of credit gathered during a year. The correlation coefficients were close to zero on the negative side.

AC did nevertheless correlate significantly ($r = .45$) with GPA, indicating a relationship between grades and credits.

4.2 Limitations of the study

One limitation of the current study is the sample size. The final sample that was included in the multiple regression analysis was quite small. Green (1991) recommends a minimum sample size of 50 plus 8 for each predictor, meaning that 74 would be the minimum for regression models with three predictors, while the present sample was 50. Furthermore, all participants were volunteers, which increases the risk of not obtaining a fully representative sample amongst university students. The less active students, with perhaps somewhat lower grades and amount of credit, might not have participated in the study.

The study performance was followed up only for one academic year. This is just a part of the whole study period, and it may be a less representative one especially for university freshmen. The majority of our sample consisted of first year students who completed on average somewhat lower amount of credit (58 ECTS) than other students did (62 ECTS) during the academic year. Also, some programs offer a number of courses that allow their students to exceed the goal of 60 credit points per year if they wish so, while programs with fewer students may only offer relevant courses worth 60 credit within a year.

Restriction of range is one of the main statistical concerns in predictive studies and might contribute to a lack of significant relationship between predictors and outcome. Restriction of range occurs when the sample has previously been selected based on some criteria. In the present case, persons with the lowest entrance exam scores, ME scores, and perhaps lowest fluid intelligence scores have all been excluded in the selection process to the university. Also, the ones with the highest scores on these measures may apply to other universities, e.g. to a medical school. Therefore, the present study is based on a sample with limited variance in the measures used, and it may not give a full picture of the overall relationship between the variables of interest. For example, there could be a perfect linear relationship between matriculation examination scores and study performance up to a certain point, after which the relationship evens out. In many studies where the researchers have been able to adjust for range restriction, the association between predictors and performance has been notably strengthened (Burton & Ramist, 2001; Ramist, Lewis, and McCamley-Jenkins, 1994). Thus the predictive validity results in the present study might be underestimations. Moreover, restriction of range not only happens when a group has already been excluded but also by the way in which students select their programs. Willingham (1985) argues that the best prepared students usually choose programs in disciplines that are graded stringently whereas the least prepared students select more leniently graded programs.

Another limitation in the present study is the usage of a fluid intelligence measure with no data about its convergent validity with other intelligence measures. However, the test was designed to resemble other measures of fluid intelligence, such as the Raven's Progressive Matrices, and thus has face validity. It is unclear whether the results indicate a lack of a relationship between "fluid

intelligence” and AP, or just a lack of a relationship between the measure used and AP. Due to the range restriction issue, it is unclear how the present fluid intelligence measure would fit as selection instrument. Despite relatively low predictive validity, it nevertheless showed more predictive value than entrance exam scores.

Some considerations regarding the criterion variables should be mentioned. Despite the standardized ECTS evaluation system, there might be different amount of work in courses at different faculties, thus creating a bias for the AC variable. Grades are neither an objective or direct measure of performance (Stelmer, 2012). Ramist, Lewis, and McCamley (1990, p .261) found that on a 4-point grading scale, there was more than one grade point difference between the most stringently and leniently graded courses. Grades may also be influenced by the teachers’ perception of a student’s temperament (Mullola et.al., 2010). The criteria for grades may differ somewhat between courses, since different abilities and talent may contribute to grades in various degrees depending on the discipline.

4.3 Implications and future studies

Predictors tapping achievement and aptitude do not function independently and lead directly to academic performance. Rather, they function in interaction with a number of variables including psychological variables (personality, motivation) and university environment. Therefore, a measure can be a good predictor of performance in one setting but worse in another. Researchers have noted that when standardized tests and earlier academic performance are used alone as predictors, they show relatively low predictive validity (Ting & Sedlacek, 1998). This has led to suggestions of using psychological variables together with traditional cognitive measures for better predictions of academic performance (Le et al. 2005). Many researchers have noted the incremental value of personality over cognitive measures when predicting academic performance (Leeson, Ciarrochi & Heaven, 2008; Nofle & Robbins, 2007). Nevertheless, the extent to which the student selection measure itself predicts future academic performance remains a central issue.

One important factor to keep in mind regarding generalizations to other Finnish universities is that the Åbo Akademi University gives scores based on every written exam in the ME, and that is

how the ME was scored in the present study as well. However, many other Finnish universities give scores for only four subjects of the exam.

The criteria used for selecting students to the Åbo Akademi University not only differ between faculties but also between the programs of the same faculty. It could be more cost-effective and reliable to use a more standardized way of selecting students. Many students change their main subject and might have to participate in an entrance exam many times, even though they have previously proved their ability to acquire knowledge and apply it. Given the current tight economies at the Finnish universities, renewal of student selection systems could save resources by moving from university- and subject-specific entrance exams, which seem to lack empirical evidence in terms of their predictive value, to standardized nationwide tests. Focusing the selection on existing scores, such as the matriculation examination, would save time and money both for the university and the applicant.

Burton and Ramist (2001) recognize some changes in the demographics of university students, including more ethnical diversity and increasing amounts of international and older students. Changing applicant characteristics are important to keep in mind in order to offer an equal possibility to all applicants. Most participants in the present study were students who completed the matriculation examination less than 5 years before applying to their current studies at the university. However, it is unclear if matriculation scores are accurate estimates of applicants' current ability if the examination was completed 10 or 20 years ago, and whether the measure has predictive validity over a longer time-span.

Despite the strengths and limitations of the selection criteria mentioned before, there has been less discussion about admission tests in a broader perspective. Universities serve multiple functions in society (Stemler, 2012). Schmitt (2012) presents some mission statements endorsed by many universities; knowledge and mastery of general principles, continuous learning and intellectual interest and curiosity, artistic and cultural appreciation, multicultural appreciation, leadership, interpersonal skills, social responsibility, physical and psychological health, and ethics. With these aspects in mind, GPA and AC seem like narrow definitions of success at university. However, the selection instruments can be either good or bad depending on the

desired outcome. Also, when taking a student's perspective on academic performance, it might mean something else than GPA or AC. For a student, good performance at university could for example mean personal development, networking, or acquiring relevant knowledge for future or current work.

4.4 Conclusion

The matriculation examination offered the best predictive value for one-year academic performance and grade point average. The explained variance is similar to many other selection instruments used internationally. The fluid intelligence measure seemed to be weakly correlated with GPA. Entrance exam scores were unrelated to the measured outcomes. The results are broadly in line with those of Utriainen (2011) who examined the predictive validity of the matriculation examination and entrance exams in Finland.

5 Swedish Summary – Svensk sammanfattning

Prediktion av ett års studiest prestation vid universitet:

En jämförelse av studentexamensvitsord, inträdesprov och flytande intelligens

Introduktion

Universitet runtom i världen möter liknande utmaningar när det finns fler sökanden till universitet än tillgängliga studieplatser. Därmed bör universitet skapa kriterier för att välja vilka sökande som får en studieplats. Det finns standardiserade test som används i stor utsträckning inom enskilda länder, t.ex. SAT i USA, SweSAT i Sverige och studentexamensbetyget i Finland. Utöver dessa standardiserade test används även mindre standardiserade test; till denna kategori hör flera programspecifika inträdesprov.

För att undersöka testens prediktiva validitet utförs uppföljningsstudier där man utifrån tidigare poäng i ett test försöker förutsäga prestation i akademiska studier. Vanliga statistiska analyser för

att undersöka den prediktiva validiteten är korrelation och regressionsanalys (Burton & Ramist 2001).

Vanliga test i USA är bland annat SAT I, som mäter talang eller begåvning, och SAT II, som mäter prestation i ett specifikt ämne. Både SAT I och SAT II har visats ha prediktiv validitet för prestation i akademiska studier (Geiser & Studley 2002). I Sverige är antagningen till universitet baserad endera på tidigare medelvitsord eller på testpoäng från SweSAT. Enligt en kunskapsöversikt av Lyrén (2008) kan man med poäng från SweSAT förutsäga framtida akademiska prestationer, men validitetens styrka beror på vilket ämne som studerades. I inriktningar såsom juridik-, ingenjör- och läkarutbildningen hittades en starkare samband än i inriktningar så som lärare, socialarbete och handelsrelaterade ämnen. I en del av de svenska studierna användes medelvitsord som mått på akademisk prestation medan det i andra användes antal studiepoäng. Inom forskningen är medelvitsord det vanligaste måttet på akademisk prestation (Stemler, 2012).

En viktig aspekt när det gäller användningen av antagningstest och deras prediktiva validitet är den teoretiska basen. Spearman (1904) beskrev intelligens som bestående av en generell intelligens faktor ”g”, som består av mer specifika intelligensfaktorer. Generell intelligens förutsäger prestation i kognitivt krävande uppgifter (Hunter & Hunter, 1984) och är relaterad till inläring av kunskap och förmågor (Lohman, 1999; Schmidt, 2002). Flera studier bekräftar den generella intelligensens prediktiva validitet för akademiska prestationer (e.g. Kuncel, Hezlet, & Ones 2001). Horn och Cattell (1966) introducerade begreppen kristalliserad intelligens och flytande intelligens. Kristalliserad intelligens definieras som inlärd kunskap, medan flytande intelligens syftar på problemlösningsförmåga i nya situationer; en förmåga som är kulturellt obunden.

Inför det akademiska året 2013–2014 gjordes omkring 90 000 studieplatsansökningar till universitet, varav 29 % blev accepterade. De vanligaste antagningskriterierna i Finland är studentexamensbetyget och inträdesprov. En av orsakerna till användningen av studentexamensbetyget som antagningskriterium är att tidigare studieprestationer förutsäger framtida studieprestationer, vilket har bekräftats i flera studier i flera länder (Larson & Scontrino,

1976; DeBerad, Julka, & Spielmans, 2004; Olani 2008; Robbins, et al., 2004). Endast ett fåtal studier har direkt undersökt den prediktiva validiteten av studentexamensbetyg på framtida akademiska studieprestationer. Utriainen (2011) rapporterade en statistiskt signifikant korrelation ($r = 0,45$) mellan studentexamensbetyg och vitsord vid universitet, men relationen mellan studentexamensbetyg och antal studiepoäng nådde inte signifikans.

Inträdesproven vid Åbo Akademi består till största delen av programspecifika prov som skiljer sig åt både i innehåll och svårighetsgrad mellan olika ämnen. Dessa prov har fått mindre uppmärksamhet i forskningen. Olani (2008) fann ingen prediktiv validitet av inträdesprov, men det är svårt att göra generaliseringar utifrån resultat av internationella studier där de undersökta inträdesproven varit annorlunda.

Eftersom både inträdesprov och studentexamensbetyg kan ses som mått på prestation i specifika ämnen, och resultatet därmed är påverkat av motivation, inkluderades i denna studie ett mått på flytande intelligens, för att undersöka om det medförde inkrementell validitet till prediktionen. Målet med denna studie är att undersöka den prediktiva validiteten av antagningskriterier som används vid Åbo Akademi, och huruvida ett mått på flytande intelligens medför inkrementell validitet utöver de traditionella måtten.

Metod

År 2014 fanns det 6 139 studerande vid Åbo Akademi, 70 % i Åbo, 20 % i Vasa och resten i Jakobstad, varav 58 % var kvinnor och 42 % män. Under hösten 2014 deltog 100 studerande i denna undersökning, 79 % från Åbo och 21 % från Vasa, varav 30 % var män och 70 % kvinnor. Studerande vid universitetet informerades om studien via e-post och vid flera introduktionsseminarier. Sammanlagt hölls tolv testsessioner, nio i Åbo och tre i Vasa. Deltagaren skulle utföra ett intelligensstest samt i fylla en bakgrundsblankett och ett personlighetsformulär.

Studentexamensbetygs-poäng, inträdesprovspoäng och medelvitsord samlades från ICT-service vid Åbo Akademi i september 2015. Samma poängsättning av studentexamensbetyget användes

som Åbo Akademi använder vid antagningsprocessen. Studentexamensbetyget hade registrerats för 78 deltagare, och dessa inkluderades i studien.

Ämnena vid Åbo Akademi gav olika mängd maximala poäng i inträdesproven. Därmed transformerades inträdesprovspoängen till procent rätt svar av maximala antalet rätt. Inträdesprovspoäng hade registrerats för 59 deltagare. Information om alla variabler hittades för 50 personer vilka utgör det slutliga samplet i studien.

Måttet på flytande intelligens som användes i studien konstruerades av fyra magisterstuderande i psykologi. Testet var ämnat att efterlikna Ravens progressiva matriser, eftersom det är ett mått på flytande intelligens som visats ha prediktiv validitet för studieprestation vid universitet (Rhode & Thompson, 2007). Det konstruerade måttet bestod av 60 matriser. Uppgiften var att identifiera logiska regler om hur åtta figurer var arrangerade och bland sex alternativ välja vilken den rätta nionde figuren var. Deltagarna hade 30 minuter på sig att utföra testet. För rätt svar gavs ett poäng, medan fel svar inte gav några poäng. Tre orsaker påverkade beslutet att konstruera ett nytt mått på flytande intelligens. För det första kan rätta svar på flera intelligenstest hittas på internet. För det andra fanns det inget lämpligt mått på flytande intelligens för massadministration vid universitetet och för det tredje fanns det intresse att utveckla ett mått som kan användas i antagningsprocessen till universitet.

Tre beroende variabler användes: *medelvitsord*, *antalet studiepoäng* och en kombination av dessa som kallas *studieprestation*. Detta gjordes eftersom olika faktorer kan påverka vitsord och antalet studiepoäng. Användningen av endast ett mått på studieprestation har även fått kritik av forskare (Kuncel, Hezlett & Ones, 2001).

Resultat

Den prediktiva validiteten av studentexamensbetyg, inträdesprov och flytande intelligens undersöktes med tre separata hierarkiska multipla regressioner. I varje analys användes samma prediktorer, men den beroende variabeln ändrades.

Måttet på flytande intelligens visade god reliabilitet $\alpha = 0,86$. Inget individuellt item sänkte betydligt den interna konsistensen av måttet. Flytande intelligens visade en signifikant korrelation ($r = 0,34$, $p < 0,01$) med poäng på matematik-, kemi- och fysikprovet i studentexamen, men inte med prov i andra ämnen.

De prediktiva måtten testades för multikollinearitet och ingen risk för detta hittades, eftersom den högsta korrelationen fanns mellan studentexamensbetyget och flytande intelligens ($r = 0,11$). Multipla regressionen med *studieprestation* som beroende variabel var signifikant, $F(3, 46) = 2,97$, $p = 0,04$ med en R^2 på 0,16. Studentexamensbetyget bestod av en R^2 på 0,13 och flytande intelligens tillade 0,02 som R^2 medan inträdesproven inte erbjöd något inkrementellt värde ($R^2 = 0,00$). Den andra modellen med samma prediktorer men med *medelvitsord* som beroende variabel nådde signifikans $F(3, 46) = 4,762$, $p = 0,01$ och förklarade 24 % av variansen i vitsorden ($R^2 = 0,24$). Studentexamensbetyget förklarade största delen av variansen ($R^2 = 0,21$), medan flytande intelligens medförde R^2 på 0,02 och inträdesprovet ett R^2 på 0,01.

Diskussion

Målet med studien var att undersöka den prediktiva validiteten av de antagningskriterier som används vid Åbo Akademi och undersöka ifall ett mått på flytande intelligens medför inkrementellt värde till prediktionen.

I regressionsanalysen fanns studentexamensbetyget vara den mest valida prediktorn. Studentexamensbetyget förklarade 21 % av variansen i medelvitsord, och dessa två korrelerade signifikant med varandra ($r = 0,51$). Korrelationen ligger på samma nivå som korrelationen mellan SAT I och medelvitsord (Bridgeman, McCamley-Jenkins, & Erwin, 2000). Studentexamensbetyget predicerade även studieframgång genom att förklara 13 % av variansen, med en medelstor signifikant positiv korrelation. Denna relation förklaras enbart av relationen mellan studentexamensbetyget och medelvitsord, eftersom ingen signifikant relation hittades mellan studentexamensbetyget och antalet studiepoäng. Inga av de kognitiva prediktoren i denna

studie förklarade antalet studiepoäng under ett år. Korrelationerna var nära noll på den negativa sidan.

Inträdesproven, trots deras omfattande användning, predicerade varken studieprestation, medelvitsord eller antalet studiepoäng. Inträdesproven medförde ingen inkrementell validitet i regressionsmodellerna, och korrelationen med de beroende variablerna var nära noll på den negativa sidan. Resultatet tyder på att studentexamensbetyget skulle vara ett tillräckligt antagningskriterium, ifall man vill predicera studieprestation.

Måttet på flytande intelligens visade god reliabilitet och en medelstor positiv korrelation med poäng i matematik-, fysik- och kemiproven i studentexamen. Måttet visade en svag korrelation med medelvitsord ($r = 0,15$), men nådde inte signifikans. Flytande intelligens medförde ingen signifikant inkrementell validitet gällande prediktionen av studieprestation, medelvitsord eller antalet studiepoäng.

Begränsningar och implikationer

Samplet som inkluderades i regressionsanalyserna var litet. Green (1991) rekommenderar ett minimum på 74 deltagare i regressionsanalyser med tre prediktorer. Alla deltagare var frivilliga, vilket kan innebära att de som är minst aktiva vid universitetet och kanske även har lägre medelvitsord och färre studiepoäng inte deltog i studien i lika hög grad som mer aktiva studerande.

Restriction of range är ett statistiskt problem som ofta påverkar studier om prediktiv validitet. *Restriction of range* inträffar när samplet i studien har valts utifrån något kriterium före deltagandet i studien. I denna studie handlar det om att personerna med lägst inträdesprovspoäng och poäng från studentexamensbetyget inte fått en studieplats och därmed även exkluderats från denna studie. Därmed är denna studie baserad på ett sampel med begränsad varians i poäng på de prediktiva måtten. Den prediktiva validiteten av måtten i denna studie kan vara underskattningar, eftersom den prediktiva validiteten har ökat betydligt i flera studier där det varit möjligt att

kontrollera för *restriction of range* (Burton & Ramist, 2001; Ramist, Lewis, & McCamley-Jenkins, 1994).

Andra begränsningar i studien är att det inte finns data om konvergent validitet av måttet på flytande intelligens som användes i studien. Därmed är det oklart om resultatet tyder på en begränsad relation mellan flytande intelligens och studieprestation eller mellan det mått som användes och studieprestation. Det fanns även problem med de beroende variablerna.

Inträdesprovspoängen var inte normalfördelade och det kan finnas bias i måtten på vitsord och studiepoäng.

Studentexamensbetyget förklarade 21 % av variansen i medelvitsorden. Forskare har föreslagit användning av psykologiska variabler i kombination med kognitiva variabler för att förbättra prediktionen av studieprestation (Le et al. 2005).

Antagningskriterierna och inträdesproven som används vid Åbo Akademi skiljer sig åt både mellan fakulteterna och inom fakulteterna. Det kunde vara mer kostnadseffektivt att använda ett mer standardiserat urvalssystem. Flera studerande byter huvudämne under studietiden och kan då bli tvungna att delta i flera inträdesprov, även om de redan tidigare visat sin förmåga att tillägna sig kunskap och använda sig av den. Förnyandet av antagningskriterierna genom att införa standardiserade test eller betona existerande prestationer, såsom studentexamensbetyget, skulle spara tid och pengar både för ansökanden och för universiteten. Användningen av resurskrävande programspecifika inträdesprov får inte stöd i denna avhandling.

Slutsats

Av de undersökta variablerna predicerade studentexamensbetyget bäst ett års studieprestation och medelvitsord. Förklaringsgraden är på samma nivå som det av flera andra internationella antagningsinstrument. Flytande intelligens visade en svag korrelation med medelvitsord. Inträdesproven var orelaterade till de undersökta utfallen.

6 References

- Atkinson, R. C. (2004). Achievement versus aptitude in college admissions. *Rethinking the SAT: The Future of Standardized Testing in University Admission*, New York, London: Routledge-Falmer, 15-23.
- Baddeley, A. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science*, 10, 341-342.
- Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning test* (College Board Report No. 2000-1). New York, NY: College Entrance Examination Board.
- Burton, N. W., & Ramist, L. (2001). *Predicting success in college: SAT studies of classes graduating since 1980* (College Board Research Report No. 2001-2). New York: College Board.
- Cliffordson, C. (2004). De målrelaterade gymnasiebetygens prognosförmåga [The predictive validity of goal-related grades from upper secondary school; in Swedish]. *Pedagogisk Forskning i Sverige*, 9(2), 129-140.
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13-21.
- DeBerard, S. M., Julka, D.L., & Spielmans, G. I. (2004). Predictors of academic achievement and retention among college freshmen: A longitudinal study. *College Student Journal*, 38, 66-85.
- Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment of g? *Psychological Science*, 15, 373-377.
- Furnham, A., & Chamorro-Premuzic, T. (2004). Personality and intelligence as predictors of statistical examination grades. *Personality and Individual Differences*,

37, 943 – 955.

Geiser, S., & Studley, W. R. (2002). UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. *Educational Assessment*, 8(1), 1-26.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, 26, 499–510.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of educational psychology*, 57(5), 253.

Hunter, J. E. (1980). *Validity Generalization for 12,000 Jobs: An Application of Synthetic Validity and Validity Generalization to the General Aptitude Test Battery (GATB)*. US Department of Labor, Employment Service.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72-98.

Kaplan, R. M., & Saccuzzo, D. P. (2009). Standardized tests in education, civil service, and the military. *Psychological testing: Principles, applications, and*, (7), 325-327.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychological bulletin*, 127(1), 162

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of personality and social psychology*, 86(1), 148.

Larson, J. R., & Scontrino, P. M. (1976). The consistency of high school grade point average and of the verbal and mathematical portions of the SAT of the CEE Board, as predictors of college performance: An eight-year study. *Educational and Psychological Measurement*, 36(2), 439-443.

Leeson, P., Ciarrochi, J., & Heaven, P. C. (2008). Cognitive ability, personality, and academic performance in adolescence. *Personality and Individual Differences, 45*(7), 630-635.

Lohman, D. F. (1999). Minding our p's and q's: On finding relationships between learning and intelligence. *Learning and individual differences: Process, trait, and content determinants, 55-76*.

Mullola, S., Ravaja, N., Lipsanen, J., Hirnistö-Snellman, P., Alatupa, S., & Keltikangas-Järvinen, L. (2010). Teacher-perceived temperament and educational competence as predictors of school grades. *Learning and Individual Differences, 20*(3), 209-214.

McGrath, M., & Braunstein, A. (1997). The prediction of freshmen attrition: An examination of the importance of certain demographic, academic, financial, and social factors. *College Student Journal, 31*, 396-408.

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big Five correlates of GPA and SAT scores. *Journal of Personality & Social Psychology, 93*, 116–130.

Olani, A. (2009). Predicting first year university students' academic success. *Electronic Journal of Research in Educational Psychology, 7*(3), 1053-1072.

Pekkarinen, T. (2014). Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations. *Journal of Economic Behavior & Organization*.

Ramist, L., Lewis, C., & McCamley, L. (1990). Implications of using freshman GPA as the criterion for the predictive validity of the SAT.

Ramist, L., Lewis, C., & McCamley-Jenkins, L. (1994). Student group differences in predicting college grades: Sex, language, and ethnic groups. *ETS Research Report Series, 1994*(1), i-41.

Raven, J., & Raven, J. C. (2000). Court, JH (2000). *Manual for the Raven's Progressive Matrices and Vocabulary Scales. Section 3: The Standard Progressive Matrices.*

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*(1), 83-92.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human performance, 15*(1-2), 187-210.

Schmitt, N. (2012). Development of rationale and measures of noncognitive college student potential. *Educational Psychologist, 47*(1), 18-29.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology, 15*(2), 201-292.

Stemler, S. E. (2012). What should university admissions tests predict? *Educational Psychologist, 47*(1), 5-17.

Svensson, A., Gustafsson, J.-E., & Reuterberg, S.-E. (2001). *Högskoleprovets prognosvärde. Samband mellan provresultat och framgång första året vid civilingenjör-, jurist- och grundskollärautbildningarna* [The prognostic value of the SweSAT. Relation between test result and firstyear academic performance at the civil engineering, law and elementary school teacher programmes; in Swedish] (Högskoleverkets rapportserie No. 2001:19 R). Stockholm: Högskoleverket.

Utriainen, J. (2011). Valintakokeen ja lukion opintomenestyksen merkitys yliopisto-opintojen menestyksen ennustajina.

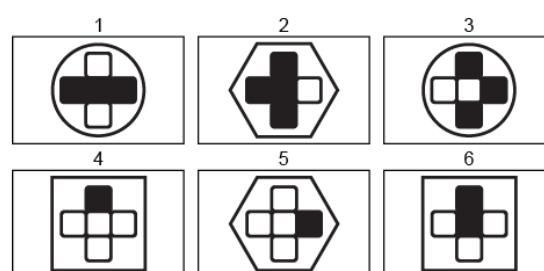
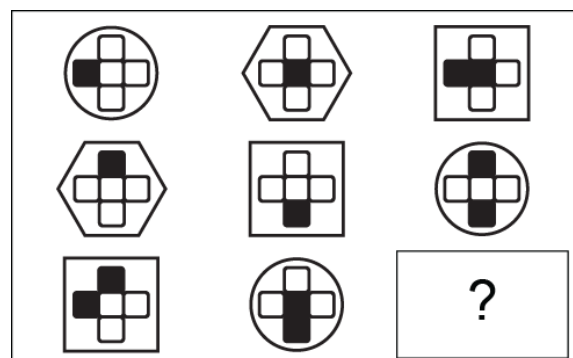
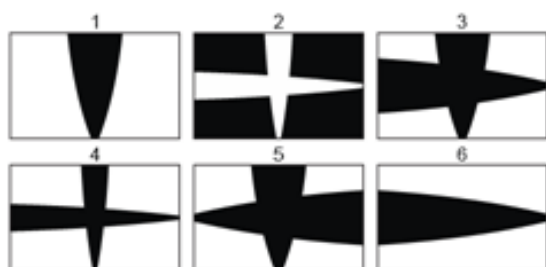
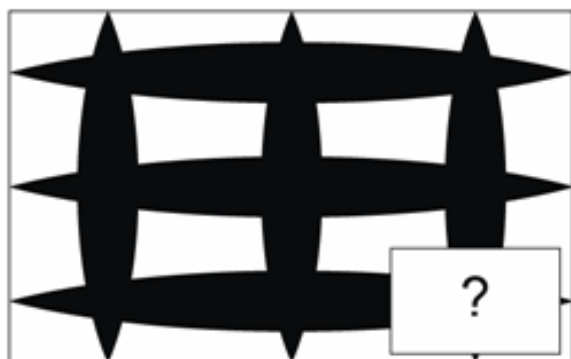
Willingham, W. W. (1985). *Success in College: The Role of Personal Qualities and Academic Ability*. College Board Publications, Box 886, New York, NY 10101.

Wolming, S. (1999). Validity issues in higher education selection: A Swedish example. *Studies in Educational Evaluation*, 25(4), 335–351.

Yu, C. H. (2011). A Simple Guide to the Item Response Theory (IRT) and Rasch Modeling. Retrieved 31 January 2016, from www.creative-wisdom.com/computer/sas/IRT.pdf

Zizzi, S. (2005). *SS 726 – multiple regressions*. Virginia University.

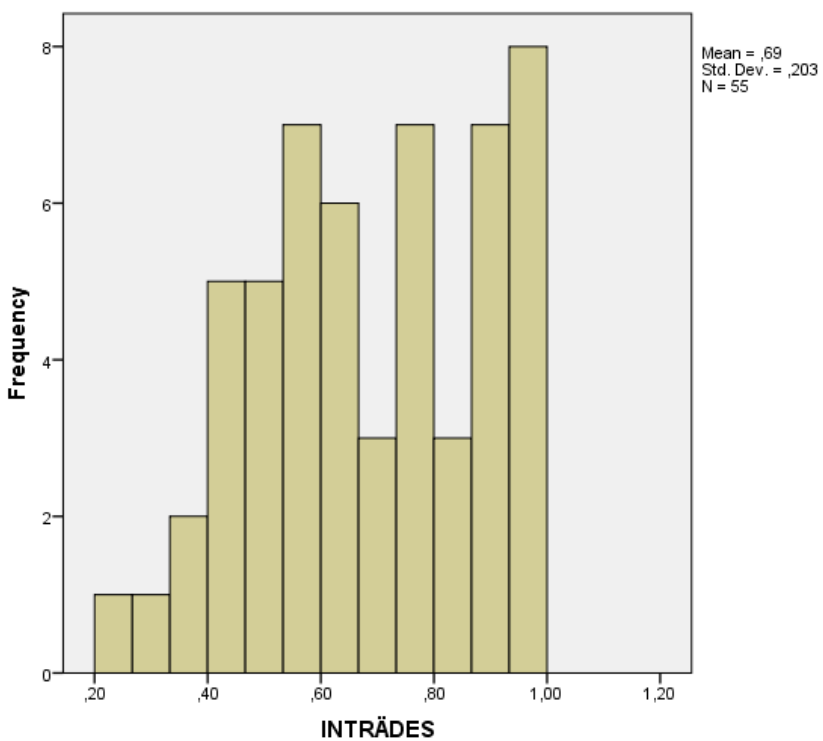
Appendix A
Examples of similar items to the ones used in the actual measure



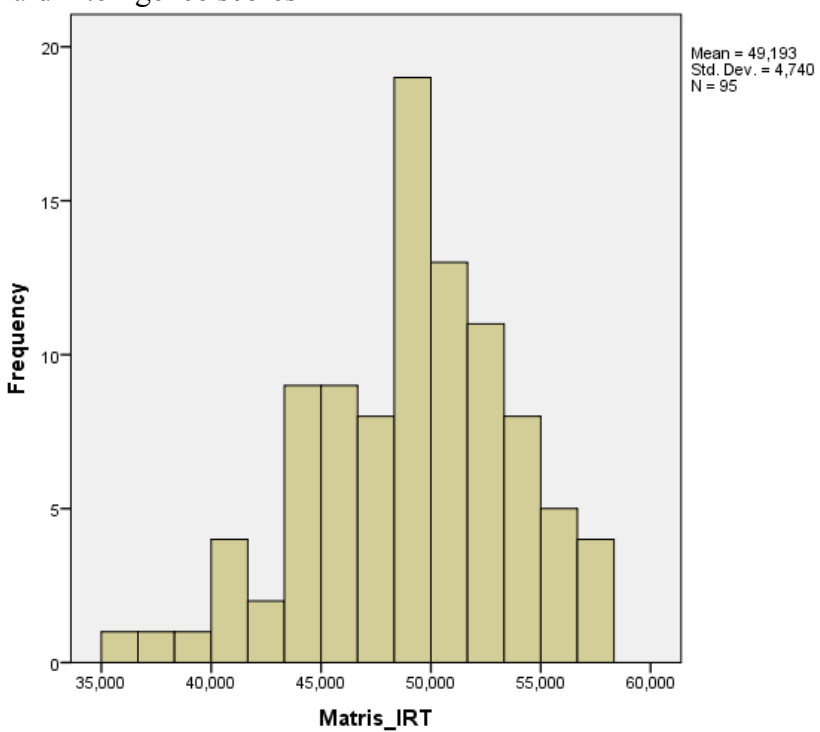
Appendix B

Distribution of scores on entrance exam and the intelligence measure

Entrance exam scores



Fluid intelligence scores



Appendix C
Scoring of the matriculation examination results

Record	Long	Short
Laudatur	7	5
Eximia	6	4
Magna	5	3
Cum Laude	4	2
Lubenter	3	1
Approbatur	2	0

PRESSMEDDELANDE

Vitsord i studentexamen förutsäger prestation vid universitet

Pro gradu-avhandling i psykologi

Fakulteten för humaniora, psykologi och teologi

Resultatet i studien visar att det kan finnas skäl att överväga universitetens antagningskriterier. De deltagare som presterade högt i studentexamen hade i medeltal även högre vitsord i sina universitetsstudier. Däremot hittades inga skillnader i universitetsvitsord mellan de som fick lägre och högre poäng i inträdesprov. De som presterade högt i intelligensmättet hade i genomsnitt lite högre vitsord i universitetsstudierna, men relationen var svag. Inget av de ovannämnda måtten förutsåg antalet studiepoäng som studerande tog under ett år. Resultatet innebär att ifall man med antagningskriterierna vill förutsäga vilka personer kommer att klara av universitetsstudierna väl, kunde man betona studentexamensbetyget framom inträdesproven. Detta skulle fungera som en utmärkt sparåtgärd för både universitet och ansökande.

Runt 90 000 ansökan om studieplats vid finska universitet görs årligen, men endast kring 29 % blir accepterade. Under de senaste åren har mängden ansökan stigit mera än antalet tillgängliga platser. Vanliga antagningskriterier som används i finska universiteten är vitsord från studentexamensbetyg och inträdesprov. Betydelsefulla val görs på basen av resultaten från dessa prov, men finns det vetenskapligt stöd för användningen av både studentexamensbetyget och inträdesprovet som antagningskriterier?

Jonathan Rasmus undersökte i sin avhandling ifall studentexamensbetyget och inträdesprovet förutsäger hur en studerande presterar i sina studier vid universitet. Därutöver undersöktes ifall ett mått på intelligens förutsäger prestationen lika väl som de traditionella antagningskriterierna. Ett hundra studerande från Åbo Akademi deltog i undersökningen. Deras vitsord, samt antalet studiepoäng under ett år insamlades.

Ytteligare information fås av:

Jonathan Rasmus
Psykologi/Åbo Akademi
jrasmus@abo.fi

Matti Laine
Psykologi/Åbo Akademi
matlaine@abo.fi