Paper 12000-2016

Analysis of Grades for University Students Using Administrative Data and the IRT Procedure

Sara Armandi, University of Copenhagen

ABSTRACT

This study is the first to implement item response theory on an entire educational program rather than on a single test. I use the new IRT procedure, introduced in SAS/STAT® 13.1, to compare the difficulty and use of grading scale across compulsory courses at the bachelor's program in economics at the University of Copenhagen. Further, the latent ability traits estimated for all students in the sample by PROC IRT are used as predictors in a logistic regression model. The hypothesis is that students who have a lower ability trait will have a greater probability of dropping out than students with a higher ability.

In the paper, three unidimensional item response models, two dichotomous and one polytomous, are applied on data from one cohort of students (n=236). The results suggest that in order to receive the highest possible grade, the highest level of student ability is needed for the course exam in the first-year course *Descriptive Economics A*. In contrast, the third-year course *Econometrics C* is the easiest course in which to receive a top grade. Additionally, I find that as student ability decreases, the probability of a student dropping out of the bachelor's program increases drastically. However, contrary to expectations, some students with high ability levels also end up dropping out.

INTRODUCTION

Universities have always had great impact on a nation's innovation and development. In a desire to increase the contribution from universities to the economy, the university system in Denmark has initiated a new reform, the "Study Progress Reform". The Danish Parliament passed this reform in 2013. The purpose is to get students through the educational system and into the job market as fast as possible in order to increase the number of taxpayers and decrease the expenses on education. The universities are facing a dilemma; on one hand they receive grants for each student completing a degree, on the other hand they are penalized if they do not succeed in decreasing the total study duration of its students. Even though it may be appealing for the universities to expel low ability and lazy students who might increase the total study duration, it could turn out to be unprofitable as some students might eventually graduate even though they do not have the best prerequisites.

By realizing which courses are the most difficult ones, and by examining whether the ability of students affects the study progress, it is possible to provide explicit information to the universities on how they can improve the performances and the study progress of their students and, eventually, increase earnings. This is where the large amounts of data, collected for purely administrative purposes in the field of education, can be utilized. The potential insights stored in the data could help the universities in understanding the educational process of the students.

Classical test theory (CTT) is one possible way of measuring the difficulty of a course and the abilities of students. However, one of the main drawbacks of CTT is that student characteristics and characteristics of courses cannot be separated. This is not very convenient, as it is troublesome to compare students who partake in different courses and difficult to compare courses using different groups of students. In order to overcome these weaknesses *item response theory* (IRT) is used to quantify the properties.

ITEM RESPONSE THEORY

In the United States, the first steps towards IRT were made by Allan Birnbaum who contributed with some chapters about IRT in Lord and Novick's "Statistical Theories of Mental Test Scores" in 1968. A separate but highly related line of development is traced to the Danish mathematician, George Rasch. Independently of the work by Birnbaum, he developed IRT models in Denmark to measure reading ability to devise tests for the military. With these new models some of the shortcomings of classical measurement models were solved. Since the 1960s IRT models have been further developed and have rapidly become mainstream as the theoretical basis for measurement. All models both evaluate student ability and properties of the test questions, such as the difficulty and discrimination capability (Embretson and Reise 2000).

The usual practice when applying IRT models to data is to consider a number of questions which together constitutes a test. An example of such a test could be a mathematics test with different questions within subtraction and multiplication. From this test it would be possible to estimate properties of the questions such as difficulty. Further, all students participating in the test are evaluated and assigned ability measures based on their answers to the questions.

The procedure used here is similar; however, this paper is the first to implement IRT on an entire educational program instead of merely using a single test. Hence, the test considered is more general and not as specific as in the mathematics test. The questions - usually denoted *items* in the literature - used in this paper are 17 end-of-course exams at the bachelor's program in economics at the University of Copenhagen. Together these 17 items form the test, measuring an overall ability, which reflects how good the students are in completing the bachelor's program in economics.

The focus of this paper is to investigate and identify the most difficult courses for the students. First, I investigate whether different courses are equally easy to pass and whether the grading scale is used consistently across courses. Second, I examine the study progress of the students by looking at the dropout rates as well as the completion time. Finally, I briefly describe student profiles by examining which characteristics have a significant influence on the ability estimates.

BACKGROUND

Denmark is a welfare state, which means that all citizens have equal access to various public services. The welfare state aims to provide security and equality of opportunity for all citizens. The available services are beneficial, and, among other things, include free healthcare and care for the elderly. Further, the majority of educational institutions in Denmark are provided to citizens at no cost. In Denmark all universities are funded based on the “taximeter system”. This system implies that finances of the education are based on the academic activity of students, and hence, directly linked to the number of students who pass their exams.

The government is interested in increasing the labor force in order to maintain balanced budgets, and is increasingly interested in saving on the expenditures in the public budgets. This has led to large changes within the university system. One example is study completion, which in 2004 was reflected in the taximeter system. The study completion entails that a completion bonus is established and only triggered when students complete a bachelor’s or master’s program. Additionally, a new completion bonus was introduced in 2009, which was conditional upon the duration of the study. Hence, the universities only receives the completion bonus if the student completes his study within a specified period. If students fail their exams or do not take exams, the universities do not receive any kind of compensation for that student (The Danish Ministry of Education, 2015).

In 2013 the government further imposed the *study progress reform*, a reform to get students through the educational system as fast as possible and into the labor market. This reform pushes the universities further, by making the completion bonuses even stricter. A very important feature of the new reform is that universities are additionally fined, if they do not manage to decrease the total completion time. As the most important source of funding for the universities is the taximeter subsidies provided by the state, the universities are very eager to comply with the demands of the government. This puts the universities in a difficult situation. On one hand they are granted for each student completing a degree, while on the other hand, not succeeding in decreasing the total study duration of its students, would result in loosing large amounts of money as they will be fined. Hence, the universities are attaching remarkable importance to the educational process of its students. They try to figure out what can be done to continue receiving enough money, but still be able to maintain a good quality of the education. As students are the building blocks of the community of tomorrow, the quality of the education is essential. After graduation, students should namely be able to contribute to the economic growth of Denmark.

THEORY

The following sections present the fundamental features of IRT models. Further, the assumptions of the models are discussed.

ABILITY AND ITEM CHARACTERISTIC CURVES

An important goal of educational measurement is the determination of how much of a latent trait a student possesses. A latent trait, or a latent variable, is an unobservable entity that influences observable variables such as test scores or item responses (Embretson and Reise 2000). In IRT the term "ability" is used to refer to the latent variable (Baker, 2001).

To measure the ability of a student, an arbitrary underlying ability scale is defined. This ability scale has a midpoint of zero and a range from negative infinity to positive infinity. Using this scale the amount of ability a student possesses can be determined, and abilities of different students can be compared. The main purpose of IRT is to place items and students on the same measurement scale.

Figure 1 shows three students (*A*, *B*, and *C*) who are all located along the same measurement scale according to their ability. Students with a low ability are located on the left of the scale while students with a higher ability are located on the right. Items are located on the same scale according to their difficulty. These are indicated by the thick lines numbered from *1* to *14* in Figure 1. Less difficult items are located on the left of the scale while the most difficult items are located on the right. A feature of the scale is that a one unit change in the scale means the same at different parts of the scale. In general, students who have an ability, which is higher than the difficulty of the item, can solve the item. Looking at Figure 1, student *A* represents a student with a low amount of ability since he is only expected to succeed in the three easiest items. Student *B* has an intermediate ability which means he is most likely to succeed in item *1* to *8*. However, it is not very likely that student *B* succeeds in item *9* to *14*. Finally, student *C* has the highest amount of ability of the three and is expected to succeed on all items except item *14*, which is the most difficult one.

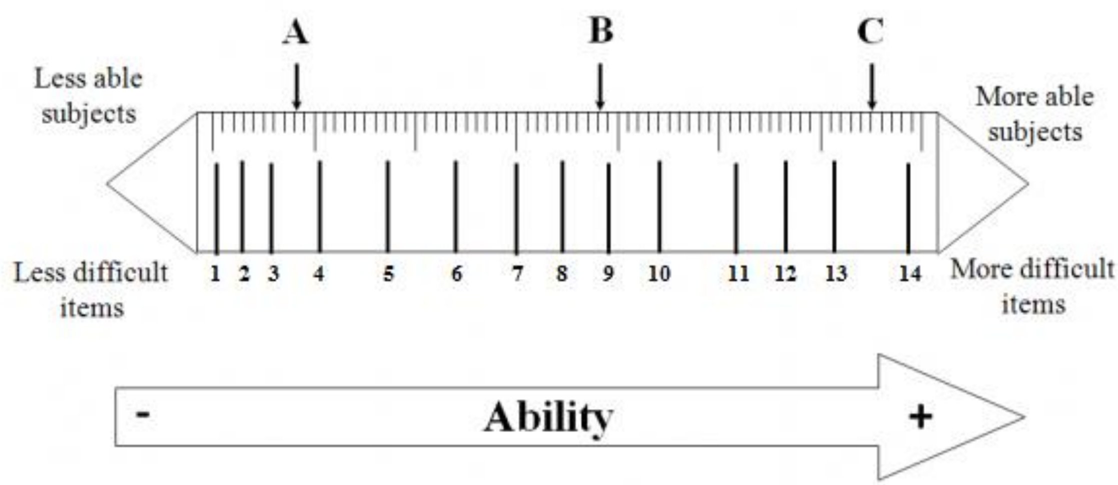

Figure 1. Subject Ability and Item Difficulty Scale
Source: Rehab-Scales.org.

To put the students on the scale, the latent ability traits are measured, usually by a number of test items, each measuring some facet of the particular ability of interest. Students responding to an item will possess some amount of the underlying ability, which can be considered as a numerical score value. This score, denoted by θ, places the student somewhere on the ability scale (Baker, 2001).

At each ability level, θ, there will be a certain probability, $P(\theta)$, that a student with that ability will give a correct answer to an item. Generally, students with low ability will have a small probability of correctly answering the item whereas high ability students will have a large probability. Plotting $P(\theta)$ as a function of ability will result in a smooth S-shaped curve, as the one depicted in Figure 2. For the lowest levels of ability the probability of succeeding in an item is close to zero. The probability increases until it reaches the highest levels of ability where the probability of succeeding approaches one. The curve illustrates the relationship between the probability of succeeding in an item and the ability measured by the test. In IRT it is known as the *item characteristic curve* (ICC) which is a unique curve, as each item has its own.

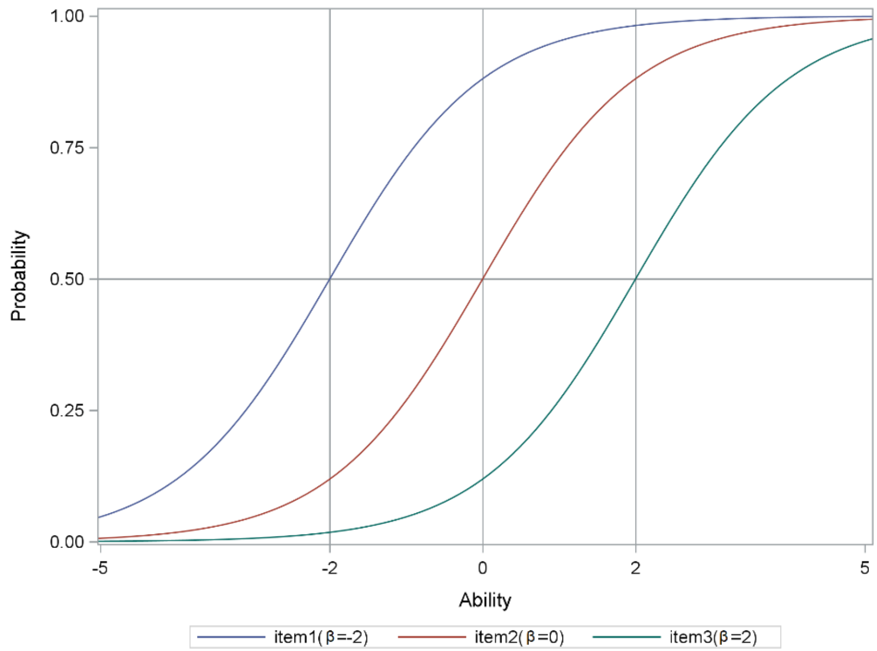

Figure 2. Item Characteristic Curves
Source: An et al., 2014

The simplest ICC can be formulated as in Equation (1), which is additionally the formulation of the simplest IRT model. This model is called the *Rasch* model or the *one-parameter logistic* (1PL) model. The simplicity occurs as the model assumes that test outcomes are binary and further, that only one item characteristic influences student performance. The influencing characteristic is *item difficulty* and is denoted by the parameter β in Equation (1). The Rasch model calculates the probability of a response and is given by:

$$\mathrm{P}\left(x_{ij} = 1 \middle| \theta_i, \beta_j\right) = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}} \tag{1}$$

where x_{ij} refers to a response made by student i to item j. When $x_{ij} = 1$ a "correct" response is observed whereas $x_{ij} = 0$ corresponds to an "incorrect" response. The trait or ability level for student i is given by θ_i, and β_j refers to the item parameter describing the difficulty of item j. This means that $\mathrm{P}\left(x_{ij} = 1 \middle| \theta_i, \beta_j\right)$ is the probability that a randomly chosen student i with ability θ_i answers item j with difficulty β_j correctly. Difficulty, β_j, is one of the properties of the ICC indicating the position of the ICC in relation to the ability scale. Three different β_j parameters are shown in Figure 2. The greater the value of the β_j parameter, the harder the item. Hence, greater ability is required for a student to have a 50% chance of getting the item right for items located at the right end of the ability scale than for items located at the left end of the ability scale (Hambleton et al. 1991). These ICCs are the basic building blocks of IRT as they connect a student's probability of success on an item to the trait measured by the test items (Baker 2001).

By extending the Rasch model with another item characteristic, the *two-parameter logistic* (2PL) model is obtained. A further extension of the 2PL model to encompass ordinal data yields the *graded response* (GR) model. These are described in more detail later in this paper.

ASSUMPTIONS

IRT is based on a set of assumptions that, if they are met, ensure the validity of the estimates. The two basic assumptions in the IRT models are unidimensional and locally independent item responses. *Unidimensionality* implies that the number of latent factors equals one. Thus, the set of items only assesses one single underlying trait dimension in accordance to explaining student performance. However, this assumption can never be strictly met because several cognitive, personality and test-taking factors (e.g. motivation, test anxiety, ability to work quickly, tendency to guess when in doubt and so on) always affect test performance. For the unidimensionality assumption to be met adequately the presence of a "dominant" component or factor that influences test performance is required.

The second assumption is local independence. This impose that, conditional on latent factors, items are independent. In other words, if only one ability is assumed to determine success on each item, then the ability level for a given student is the only thing that systematically affect item performance. This implies that when considering all students' abilities, there must be no relationship between students' responses to items.

METHOD

This section describes the data used for the analyses. Subsequently, the statistical procedures, including the SAS® IRT procedure, is reviewed.

DATA

The analyses are based on an extraction from the Danish Student Administrative System (STADS). Developed by the Danish Ministry of Education, the Danish Ministry of Research, and a number of companies, STADS is a cross-university system used by all universities in Denmark as well as a range of other educational institutions. STADS is designed to handle anything from admission, enrollment, group formation, registration for courses and examinations to registration of grades. The database contains detailed information about the study history on a very large number of students and new information is constantly being added. The amount of data are excessive, and there are numerous possibilities of using data.

In this work, administrative data from one cohort of first-year students in the economics program at the University of Copenhagen are used. A total of 262 students were admitted to the program in the spring of 2010. Of these students, 236 participated in at least one exam at the end of the first semester and are therefore included in the analyses.

Courses

The economics program at the University of Copenhagen consists of a bachelor's program and a master's program. The prescribed study duration at the bachelor's program is three years. This part of the education mainly introduces basic and compulsory courses, which are meant to provide students with strong analytical skills within economic theory.

The items considered in this paper are examination grades from the compulsory courses at the bachelor's degree in economics. Each of these end-of-course exams corresponds to an item, all measuring some facet of the particular ability of interest. The overall ability measured in this paper, is a general intelligence within the fields of economics. In total, 17 compulsory courses, which together constitutes the majority of the bachelor's program in economics at the University of Copenhagen, are examined.

The Grading Scale

The Danish marking scale is a 7-point grading scale. An explanation of the different marks as well as their appearance in the data are shown in Table 3.

Mark	ECTS Mark	Explanation of the mark	Ordinal data	Binary data
12	A	For an excellent performance	5	1
10	B	For a very good performance	4	1
7	C	For a good performance	3	1
4	D	For a fair performance	2	0
02	E	For an adequate performance	1	0
00	Fx	For an inadequate performance	0	0
-3	F	For an unacceptable performance	.	.

Table 1. The Danish Marking Scale
Notes: The grades 02 and 00 both begins with a "0" as grades sometimes are handwritten, and it should not be possible to change the grades easily to 10 and 12 respectively.
Source: The Danish Ministry of Education

Two different types of data are considered in this paper. In the ordinal data six different marks are used, ranging on a scale from 0 to 5, where 0 represent the lowest possible value and 5 the highest. Hence, less ability is required of a student to receive the lower marks while more ability is needed to receive the highest. The grades are ordered categorical implying that the polytomous model, which uses ordinal data, contains six different response categories. In the binary data the six grade categories from the ordinal data are reduced to two, 0 and 1, where 0 represent the category in which students have the lowest amount of ability. The line dividing the two categories is between the marks 4 and 7 as these are often considered the grades distinguishing the good students from the bad. The reason why the dividing line between the two categories is not chosen to distinguish the failing grades (-3 and 00) from the non-failing grades (02, 4, 7, 10 and 12), is because the number of students who are actually failing within one course is very limited.

The grade −3 is considered a missing observation as there are some problems associated with this lowest possible grade. It is worth emphasizing that −3 is very rarely used, if a student has actually tried to provide an answer to an exam question. Often, the grade -3 is given if a student choses to fail an exam on purpose.

STATISTICAL ANALYSIS

All statistical analyses in this paper have been performed using SAS® version 9.4. Two different statistical approaches are used. The first estimates the IRT models and the second approach is a logistic regression, which utilizes the ability estimates obtained from the IRT models.

The IRT Procedure

The IRT models specify that the probability of a student answering a given item correctly depend on the student's abilities and the characteristics of the item. The problem of estimation is hence to determine the value of the latent ability trait, θ, for each student and the item parameters, α and β, from the item responses.

The variables used in this analysis are item1 to item17, respectively, indicating the grades for each of the 17 compulsory courses of the bachelor's degree. The 17 variables are compiled differently depending on whether binary or ordinal data are considered. In the case of the binary, the variables only take values 0 and 1, whereas the variables take six different values for the ordinal data.

As all items are designed to measure student ability at the economics program at the University of Copenhagen, it is reasonable to start with applying unidimensional IRT models, and examine whether these models fit the data. The IRT procedure, which was new in SAS/STAT® 13.1, is used to fit the models and determine the parameters, by using the code sample below:

```
proc irt data=irt out=irt_out itemstat itemfit plots=all;
var item1-item17;
model item1-item17/resfunc=rasch;
run;
```

Values for the item parameters are shown in output tables due to the ITEMSTAT option, while estimates of the θ parameter for each student are exported to an output dataset, IRT_OUT. The ITEMFIT option ensures that the item fit statistics are computed and displays in a table. The display option, PLOTS=ALL, provides graphical output, for instance all figures displayed in the forthcoming results section.

PROC IRT yields marginal maximum likelihood estimates of item parameters and maximum likelihood estimates (conditional on the item parameter estimates) of students' ability scores (SAS® Institute Inc.). The value that maximizes the maximum likelihood function is found by using the fact that at the point where the function reaches a maximum, the slope of the function (the first derivative) is zero. This equation can however not be solved directly, and hence, the Newton-Raphson approximation method is used.

The IRT procedure supports several response models for binary and ordinal responses, which can be specified by the RESFUNC option. In the code above RESFUNC=RASCH specifies that the data should be fitted to the Rasch model. The two other models considered in this paper, the 2PL model and the GR model, are specified by modifying the RESFUNC option to =TWOP and =GRADED, respectively. To change the dimensionality of the estimated model the NFACTOR option should be used. This will result in an exploratory item response model estimating the number of factors specified in the NFACTOR option.

The Logistic Regression Model

Two logistic regressions are performed to test to what extent there is a correlation between the value of the estimated latent ability trait, θ, and, respectively, whether the students drop out of the bachelor's program and the master's program in economics at the University of Copenhagen. To find the degree of correlation a new variable is defined. The variable, DropOut, is a binary variable taking the value 1 if a student drops out of the program. If a student has completed, or is still enrolled in the program, the variable takes the value 0. The LOGISTIC procedure is used to test the interaction. This procedure models the probability of some event occurring as a linear function of a set of predictor variables. Using the θ estimates (calculated by the IRT procedure) as the independent variable, the following code estimates the probability of the dependent variable, DropOut, taking the value 1:[1]

```
proc logistic data=irt_out desc DESCENDING plots=all;
model DropOut = Factor;
run;
```

The code sample above shows how the logistic regression model is estimated using SAS®. By default the probability of the lowest value of the dependent variable is modeled. As the interest lies in finding the probability of a student dropping out (DropOut=1), the DESC option is added, which provides the desired outcome as it reverses the sorting order of the response variable.

RESULTS

This section presents the results of the model estimations. As data are strongly suggestive of a single factor only, purely unidimensional models are considered in this paper. Initially, the results from the dichotomous 2PL model, which uses binary data, are displayed. Following, the results from the polytomous GR model, utilizing ordinal data, are shown.

[1] The θ estimates from PROC IRT are assigned the variable name *Factor*.

THE DICHOTOMOUS MODELS

The Rasch model and the 2PL model are both dichotomous models, as they use binary data. The difference lies in weather one or two item characteristics are included in the model. Since these models are quite similar, the results of the Rasch model are not displayed. In Equation (2) the formulation of the 2PL model is given. It predicts the probability of success ($x = 1$) for a student as follows:

$$P(x = 1|\theta, \alpha, \beta) = \frac{e^{\alpha(\theta-\beta)}}{1 + e^{\alpha(\theta-\beta)}} \quad (2)$$

where θ_i refers to ability of the student i, β_j is the difficulty of item j and α_j is the discrimination of item j. When setting the discrimination parameter α_j equal to one, the 2PL model reduces to the Rasch model.

In Figure 3, ICCs for six of the 17 end-of-course exams are displayed. The β_j value from Equation (2) corresponds to the point on the x-axis (the ability scale) where the ICC is steepest. Additionally, it represent the difficulty of the item. The ICCs for the two most difficult end-of-course exams, *Macroeconomics C* and *Macroeconomics A*, are given in the frames in the left side of Figure 3. These courses have the highest estimated β values. In contrast, the two ICCs in the middle frames of Figure 3 have the lowest β values. Hence, the end-of-course exams in *Business Economics* and *Economic History* are respectively estimated as the two easiest courses.

The α_j parameter is an estimate of the slope of the ICC. It indicates how good the item is at discriminating between different students. In the right hand side of Figure 3 the ICCs of the two courses with the highest discrimination parameter, *Macroeconomics B* and *Microeconomics A*, are given. In addition to being the easiest course, *Business Economics* also has the lowest discrimination indicated by the relatively flat slope of the ICC. This implies that it is easier to discriminate between student abilities in *Macroeconomics B* and *Microeconomics A* than *Business Economics*, as the probability of a correct response at the low ability levels are more identical to the high ability levels in *Business Economics*.

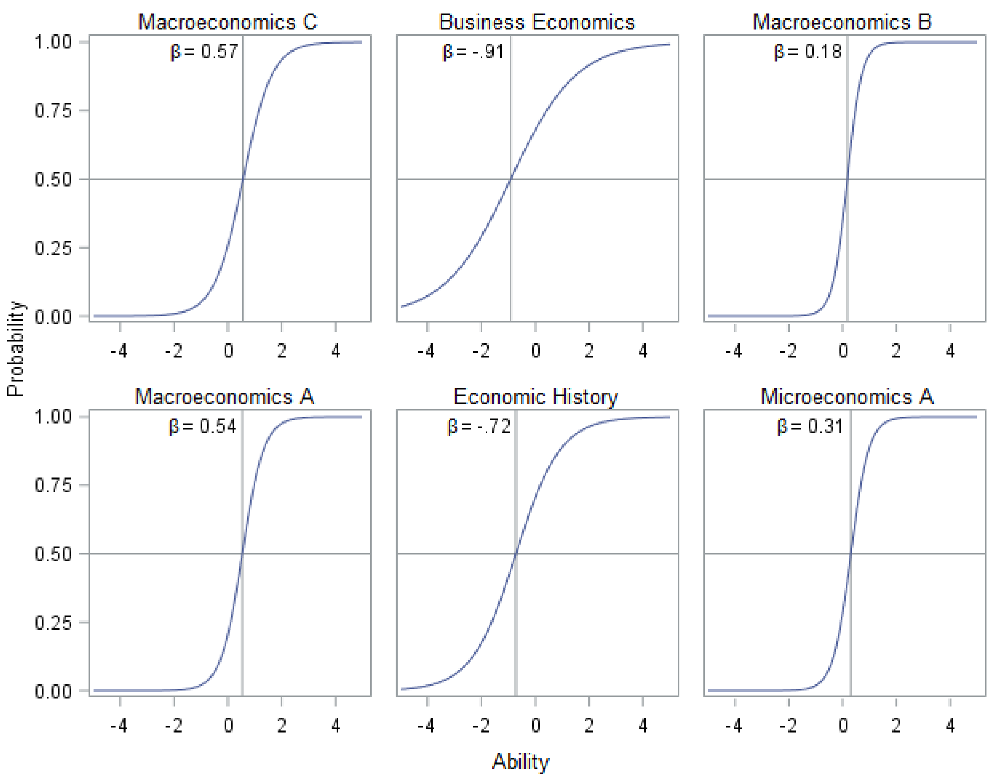

Figure 3. Item Characteristic Curves for the 2PL Model
Source: Output from PROC IRT, SAS ® 9.4.

In the dichotomous models the number of categories is limited. However, when imposing a polytomous model, the number of grade categories increases from two to six, leading to an increased amount of information. This might result in a model, which is even more precise in estimating ability and has an even better fit of the ability distribution.

THE POLYTOMOUS MODEL

The GR model is an extension of the 2PL model. It belongs to the category of polytomous models, as ordinal data are used. Essentially, what occurs in the GR model is that the item is treated as a series of dichotomies and 2PL models are estimated for each dichotomy (Embretson and Reise 2000, p.99).

The probability that a student with a given ability will score in a particular grade category is illustrated by the *category response curves* (CRC) displayed in Figure 4.

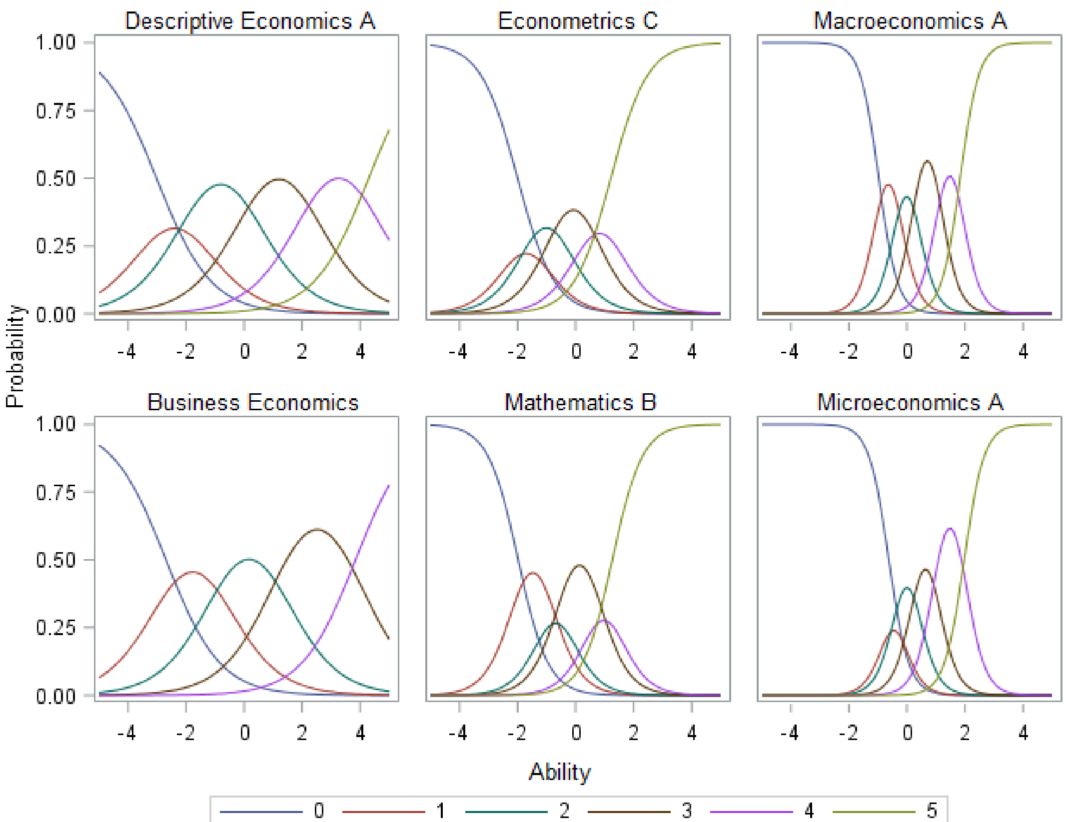

Figure 4. Category Response Curves for the GR Model
Source: Output from PROC IRT, SAS ® 9.4.

The two frames at the left side of Figure 4 display the courses where it is most difficult to obtain a high grade, whereas the two frames in the middle are examples of courses where it is relative easy to obtain the highest grade. A difficult course is in this case referring to the ability levels required to obtain a good grade. The two frames at the right side in Figure 4 illustrate courses with high discrimination, which can be seen because the curves are peaked, making it easier to determine exactly which abilities are associated with each category. These can be compared to the two courses at the left side, which, in addition to being the most difficult courses, also are worst at discriminating between ability estimates, indicated by the flat and wide curves.

Figure 4 shows that in *Descriptive Economics A* students need an ability higher than 4 to have equal probability of being assigned the top grade 12 as opposed to any other possible grade. In the opposite end of the grading scale for this course, a very low ability (below -3) is needed to have a 50% chance of passing the exam in *Descriptive Economics A*. Hence, in this course almost all students pass, but only a few are assigned the top grade.

A surprising result appears when considering which end-of-course exams requires the lowest ability to have a 50% chance of receiving the highest grade. When looking at the two middle frames of Figure 4 these courses turn out to be *Econometrics C* followed by *Mathematics B*, where the abilities required are only slightly above 1.

As Figure 4 illustrates, the grade distributions varies quite a lot across. This indicate, that the grading scales are not used consistently across courses and that the courses varies quite a lot in difficulty. These results provide specific information about which courses are hard for the students, and for which courses the grade distributions are different from the rest. Hence, the university obtain information about which specific courses that might need to be examined in more detail.

EDUCATION STATUS

Not all students who undertake an education manage to graduate. In this section the education status of the students will be examined and measured against the estimated ability parameters from the GR model. This will be done by using a logistic regression model. Initially, the education status for the bachelor's degree is examined followed by a short presentation of the education status at the master's degree.

Bachelor's Degree Status

To find the amount of student ability required for a student to have a high probability of completing his education, a logistic regression model is estimated. The relationship between the latent ability trait and the bachelor's degree status is illustrated in Figure 5. This figure predicts the probability of a student dropping out of the degree. A clear tendency is seen in Figure 5, showing that students with a lower level of ability have a much larger probability of dropping out of the economics program. Students with an ability level below -1 only have a 50% chance of completing their education. An ability level below -2 results in close to 100% certainty that a student will not manage to graduate. The interesting thing is, as the ability level increases to a value above 1, there is almost no chance that a student interrupts the program. However, there are still a few students with an ability above 0 who drop out. These students might have experienced a change of interests, and have dropped out of the economics program to attend medical school, for instance.

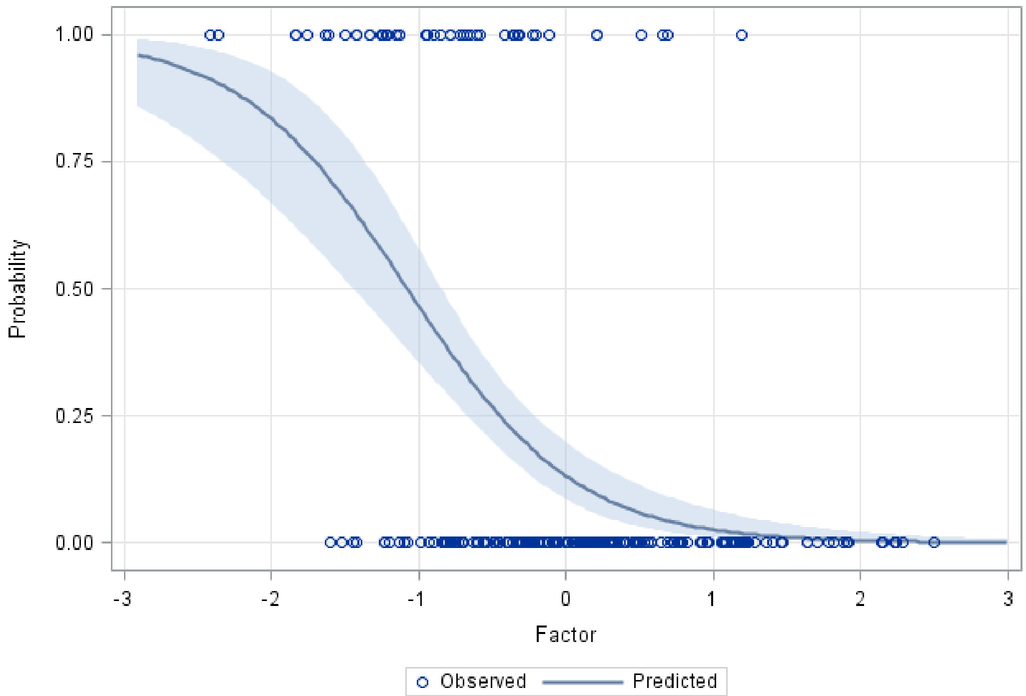

Figure 5. Predicted Probabilities for Drop Outs at the Bachelor's Degree
Notes: The circles represent the observed students. The light gray area is the 95% confidence limits.
Source: Output from PROC IRT, SAS ® 9.4.

As can be seen from the confidence limits in Figure 5, the deviations between the predicted model and the observed observation are largest for the low ability levels.

Master's Degree Status

As for the bachelor's degree, a logistic regression is conducted in order to find the relationship between student ability and the probability of dropping out of the master's program. The ability estimates used are the ones obtained from the GR model. Figure 6 is a graphical illustration of the relationship. As indicated by the confidence limits, the deviations between the predicted model and the observed observations are considerably larger for high ability levels. This is mainly caused by the limited number of students, choosing to interrupt their education. The reason why the high ability students might choose to drop out could be because of a greater number of opportunities. It should be noted that the model has a hard time in predicting the education status for students with a high level of ability.

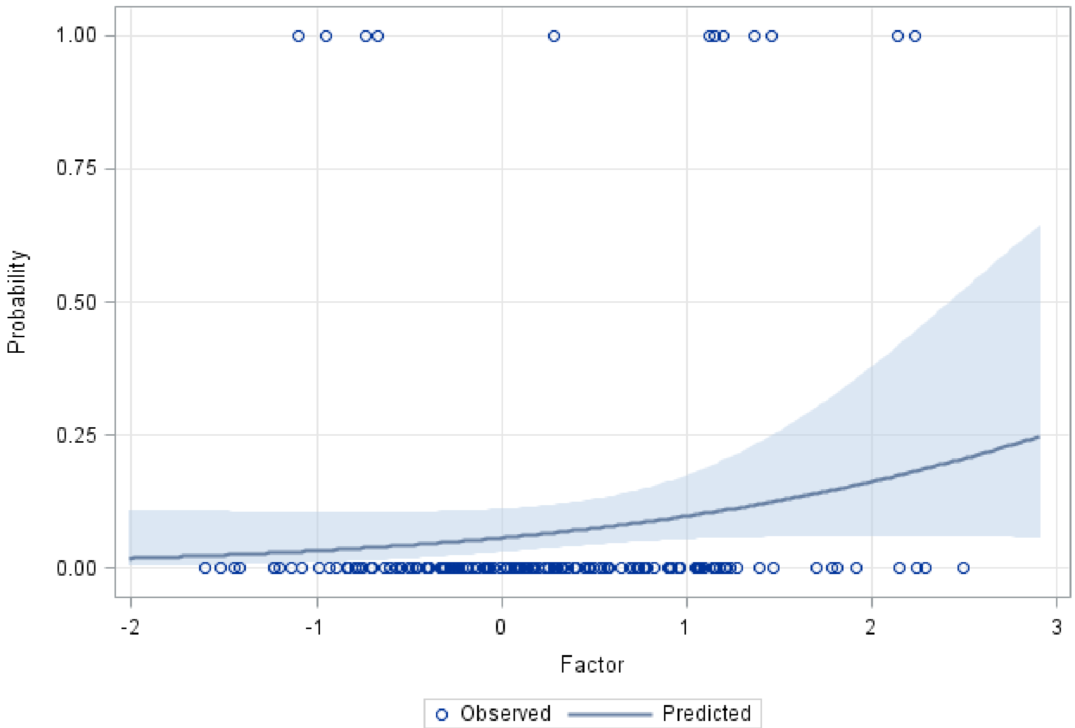

Figure 6. Predicted Probabilities for Drop Outs at the Master's Degree
Notes: The circles represent the observed students. The light gray area is the 95% confidence limits.
Source: Output from PROC IRT, SAS ® 9.4.

Length of Bachelor's Degree

Applying the ability estimates from the GR model, a linear regression is performed to see how the level of ability relates to the completion time of the bachelor's degree. The calculations are only based on those students who have completed their degree. Table 2 shows the linear regression parameter estimates for the length of the bachelor's degree. When the factor level increases by one, the time a student uses to complete the education decreases by a bit more than a month.

Parameter	DF	Estimate	Standard Error	t Value	Pr > ChiSq
Intercept	1	3.305	0.029	115.91	<.0001
Factor 1	1	-0.097	0.020	-4.93	<.0001

Table 2. Parameter Estimates for Length of Bachelor's Degree
Source: Output from PROC IRT, SAS® 9.4.

STUDENT PROFILING

To examine which characteristics have a significant influence on the ability levels, a linear regression model is fitted to the data for all students. The result of the regression is shown in Table 3. The grade point average (*GPA*) from the students' secondary education is found to be an important characteristic. When the GPA increases by one unit the ability estimate increases significantly by 0.3. The gender of the student (*Gender*) also turns out to have a significant effect. Given a student is female, the ability level decreases with -0.2. Finally, the age at which the student matriculates (*Start Age*) is borderline significant. When a student's age at matriculation increases by one year, the ability level also slightly increases.

Parameter	DF	Estimate	Standard Error	t Value	Pr > ChiSq
Intercept	1	-3.419	0.639	-5.35	<.0001
GPA	1	0.307	0.032	9.67	<.0001
Gender	1	-0.205	0.116	-1.77	0.0773
Start Age	1	0.045	0.027	1.67	0.0956

Table 3. Parameter Estimates for Ability Influence
Source: Output from PROC IRT, SAS® 9.4.

To sum up, a profile of a student, which the university should keep an extra eye on when matriculating, is a young female. Of course, the most important and significant background characteristics is the GPA from the secondary education. This variable should definitely be considered, if the university wants to know where to implement new initiatives and to whom more guidance should be provided. When helping students with the lowest ability, the dropout rate as well as the completion time (both important for the economy of the university) can, and most certainly will, be influenced.

CONCLUSION

This paper analyzes student administrative data on 236 economics students who enrolled in the bachelor's degree at the University of Copenhagen in 2010. The aim is to figure out where new initiatives can be implemented in an attempt to improve the educational process for the students.

The approach used in this paper is of course not flawless. Generally, it is very difficult to construct a statistical model capable of predicting human behavior. However, there are advantages of the approach, as it provides easily interpretable and understandable results, which are simple to communicate to interested parties. All results indicate that the ability of students play an important role in the study process. When investigating which student characteristics influences the ability level, the GPA from the student's secondary education turns out to play an important role.

This paper recommends that the university take a closer look at the courses that are troublesome for the students. Further, it is suggested that the university, as soon as possible after student enrollment, take care of the weakest students in order to improve the study progress, and eventually, to strengthen the economy of the university.

REFERENCES

An, X. & Yung, Y. (2014). Item Response Theory: What It Is and How You Can Use the IRT Procedure to Apply It, Cary, NC: SAS® Institute Inc.

Baker, F. B. (2001). The Basics of Item Response Theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.

The Danish Ministry of Education. (2015, October 7) Funding for education. From http://ufm.dk/en/education-and-institutions/higher-education/danish-universities/the-universities-in-denmark/economics-of-university-sector/funding-for-education (accessed 18.03.2016)

Embretson, S. E. & Reise, S. P. (2000). Item Response Theory for Psychologists. Mahwah, NJ: Lawrence Erlbaum Associates.

Hambleton, R.K., Swaminathan, H.H. & Rogers, H.J. (1991). Fundamentals of item response theory (Measurement methods in the social sciences ; 2). Newbury Park, Calif: Sage Publications.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests (Studies in mathematical psychology 1). København: Danmarks Pædagogiske Institut.

Templin, J. (2016). Item Response Theory - Methods for the Analysis of Discrete Survey Response Data. ICPSR Summer Workshop at the University of Michigan, from http://jonathantemplin.com (accessed 18.03.2016)

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor, Anders Milhøj, whose expertise, knowledge, and assistance added considerably to my graduate experience. Anders Milhøj is the one teacher who truly made a difference in my educational life. It was because of him I developed a focus and became interested in applied statistics and analytical economics. He provided me with direction, technical support and became more of a mentor and friend, than a lecturer and employee at the university.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Sara Armandi
University of Copenhagen
+45 22272271
sara@armandi.dk