

## Measuring academic learning and exam self-efficacy at admission to university and its relation to first-year attrition: an IRT-based multi-program validity study

Tine Nielsen<sup>a</sup>, Ida Sophie Friderichsen<sup>a</sup>, Bjarke Tarpgaard Hartkopf<sup>b</sup>

<sup>a</sup>University of Copenhagen, Denmark

<sup>b</sup>The Danish Evaluation Institute, Denmark

Article received 4 June 2019 / Article revised 13 August / Accepted 30 August / Available online 02 October

### Abstract

*Self-efficacy is associated with both academic performance and attrition in higher education. Whether it is possible to measure students' academic self-efficacy after admission and prior to commencing higher education (i.e. pre-academic self-efficacy) in a valid and reliable way has hardly been studied. Aims: 1) to evaluate the construct validity and psychometric properties of two short scales to measure Pre-Academic Learning Self-Efficacy (PAL-SE) and Pre-Academic Exam Self-Efficacy (PAE-SE) using Rasch measurement models, 2) to investigate whether pre-academic self-efficacy was associated with half-year attrition across degree programs and institutions. Data consisted of 2686 Danish students admitted to nine different university degree programs across two institutions. Item analyses showed both scales to be essentially objective and construct valid, however, all items from the PAE-SE and two from the PAL-SE were locally dependent. Differential item functioning was found for the PAL-SE relative to degree programs. Reliability of the PAE-SE was .77, and varied for the PAL-SE from .79 to .86 across degree programs. Targeting was good only for the PAL-SE, thus we proceeded with the PAL-SE. PAL-SE was found to be associated with half-year attrition: A difference in PAL-SE from minimum to maximum was associated with a difference in half-year attrition of approximately 7%. This association was found both in the bivariate model and in the multivariate models with control of degree program, and with control of degree program and individual covariates such as earlier educational achievement and social background variables. Results thus also indicate that PAL-SE has a causal effect on half-year attrition.*

**Keywords:** Pre-academic self-efficacy; construct validity; differential item functioning; graphical loglinear Rasch model; attrition



## 1. Introduction

Self-efficacy, at the general level, refers to the belief in one's own capability to plan and perform necessary actions to attain a certain outcome (Bandura, 1997). Students who feel efficacious when learning or performing a task, participate more readily, work harder, persist longer when they encounter difficulties, and achieve at a higher level of academic performance (Schunk & Pajares, 2002).

Self-efficacy is a central and much used construct in many fields, not least in education and mental health research and the combination of the two: The positive correlational relationship between self-efficacy and academic performance is well-established (Ferla et al., 2009; Luszczynska et al., 2005; Zimmerman et al., 1992). Low self-efficacy has been found to be associated with increased risk of depression and anxiety (Luszczynska, et al., 2005; Muris, 2002; Tahmassian & Jalali Moghadam, 2011), and self-efficacy has been found to mitigate the negative effect of stress on life satisfaction (Burger & Samuel, 2017). Self-efficacy, self-esteem and psychological distress were found to be the most important predictors of stress in university students (Saleh et al, 2017). Recent meta-analyses concluded that self-efficacy is the strongest (non-ability) predictor of grade point average (GPA) in higher education (HE) above personality traits, motivation and various learning strategies (Bartimote-Aufflick et al., 2015; Richardson et al., 2012). An ongoing and unresolved discussion in self-efficacy research is whether self-efficacy is a general or a specific construct, or something in between (Bandura, 1997; Pajares, 1997; Scherbaum, Cohen-Charash, & Kern, 2006). Thus, there is a plethora of self-efficacy instruments measuring from the altogether general self-efficacy, over the domain-specific, the context-specific, to course- or even task-specific self-efficacy. While the General Self-Efficacy Scale (GSES; Schwartzer & Jerusalem, 1995), or adaptations of it, is probably the most commonly used in health related research, in educational research it is more common to construe self-efficacy as at least domain-specific; e.g. the General Academic Self-efficacy Scale (GASE; Nielsen et al., 2018) and the Physics Self-Efficacy Questionnaire (PSEQ; Lindstrøm & Sharma, 2011), both for the academic domain and both based on the GSE. In between the general and very specific instruments, are a range of sub-domain-, context- and course-specific self-efficacy instruments, e.g. the Computer Self-Efficacy Scale (Scherer & Siddiq, 2015), the Psychologist and Counsellor Self-Efficacy Scale (Watt et al., 2019), the Diabetes Management Self-Efficacy Scale (Williams et al., 2014), and the commonly used Specific Academic Self-Efficacy Scale from the Motivated Strategies of Learning Inventory (MSLQ; Pintrich et al, 1991).

Given the importance of self-efficacy for educational outcomes, a large body of research has focused on investigating differences in the levels of self-reported self-efficacy for student subgroups, e.g. gender, age, academic discipline, institution and so on. However, as pointed out by Nielsen and colleagues (2018), hardly any studies comparing gender groups of students have determined whether the self-efficacy scales they use in fact are measurement invariant and thus whether self-efficacy scores for these groups can be compared without bias. The formal definition of measurement invariance (MI), according to Mellenbergh (1989) and Meredith (1993), states that MI holds when the probability of observed scale scores depends only on the estimated latent variable for the measure and not any other characteristic of the persons being measured (i.e. relevant demographic and background variables such as gender, age, etc.). In terms of self-efficacy this would mean that the observed raw scores for any students should only be determined by their latent level of self-efficacy and not also on whether they are, for example, male or female, or whether they are entering one higher educational degree program or another. Within Item Response Theory, including the subgroup of Rasch models, MI is more often formulated as a requirement of absence of Differential Item Functioning (DIF), which is defined in exactly the same way as at the item level. Thus, no DIF is formally defined as that the probability of an item responses should depend only on the estimated latent variable, and not additionally on any background variables (or exogenous variables) (Kreiner, 2013). We can then say that items are conditionally independent of the background variable given the estimated latent variable (or person parameter), and thus items function in the same way for all respondents belonging to any of the subgroups defined by the exogenous variable, and the scores will not be biased for some groups of students. Avoiding bias caused by



DIF (i.e. having MI) is important, as some students would otherwise obtain unfair scores compared to other students (i.e. artificially low or high scores obtained due to subgroup membership and not because they have higher or lower self-efficacy). For research purposes this is equally important, because DIF, depending on its magnitude, can cause statistical comparisons of student self-efficacy to be confounded by their subgroup status (Kreiner, 2007). In the literature search for the current study, we found that none of these looked into the issue of invariant measurement across the subgroups that were compared. A recent meta-analysis suggested student self-efficacy to be heterogeneously associated with gender as the relationship is moderated by academic discipline (Huang, 2013). Young, Wendel, Esson and Plank (2018) found that self-efficacy levels were higher for men than for women, but that levels of student self-efficacy did not differ according to the included academic disciplines (science, technology, engineering, and mathematics). Whether levels of self-efficacy vary across students from different HE institutions has been left largely un-investigated. To the authors' knowledge, only one study has looked into this: Ginsborg, Kreutz, Thomas and Williamon (2009) found that the level of general self-efficacy in music performance students from HE music conservatories was significantly lower than in university health students from nursing and biomedical science.

The predictive power of pre-higher education (pre-HE) self-efficacy for academic outcome has hardly been studied internationally: Yong (2010) found no discipline-specific differences in general self-efficacy level in a study of 105 pre-university engineering and business students admitted to a private Malaysian university. Van Herpen et al. (2017) found that pre-university academic self-efficacy did not predict academic success in the first year in general (sample sizes were too small to look into single disciplines), when assessed at the time of application to university. Aryee (2017) found that high levels of pre-HE mathematics self-efficacy were associated with student persistence and the completion of college, while low levels of pre-HE mathematics self-efficacy were associated with a diminished likelihood of attaining any degree or credential.

The relationship of self-efficacy (measured after the commencement to HE) and student persistence and attrition is well studied within the educational research field. A meta-analysis concluded that student self-efficacy beliefs were associated with a variety of academic persistence outcomes ("number of academic terms completed", "number of items or tasks completed", and "amount of time spent on the performance") (Multon, Brown & Lent, 1991). Furthermore, academic self-efficacy in HE students has been found to be associated with student attrition: De Clercq, Galand and Frenay (2017) found that low levels of academic self-efficacy in students during their first year at university were associated with different combinations of achievement predictors including an increased risk of attrition. Devonport and Lane (2006) also found an association between low levels of self-efficacy (on a self-developed academically related instrument) in first year university students, and an increased risk of withdrawal. The relationship between academic self-efficacy and student attrition has also been studied through students' self-perceived intention to persist in HE and several studies have found a positive association between high levels of academic self-efficacy and an increased completion intention (Elliot, 2016; Lent et al., 2016; Thomas, 2014; Vuong, Brown-Welty & Tracz, 2010; You, 2018). However, none of the studies investigating self-efficacy either pre-HE or after commencement to HE looked into the issue of measurement invariance across subgroups such as academic discipline, degree program or institution. In addition, it remains unanswered whether pre-academic self-efficacy can indeed predict outcomes, and specifically attrition.

## 1.1 The present study

The present study seeks to fill the aforementioned gaps in the existing research on academic self-efficacy: 1) to investigate the relationship between pre-academic self-efficacy and first semester attrition in university. In order to do this, we thus needed 2) to obtain a construct valid and psychometrically sound instrument for the measurement of pre-academic self-efficacy, suitable for use before students commence the university degree programs to which they have been admitted; and within the psychometric analysis 3) to investigate whether the instrument is measurement invariant across institutions, degree programs and other relevant subgroup divisions of students. None of these elements has been researched earlier.



As we could not identify an appropriate instrument, we, as an initial step, adapted an existing self-efficacy instrument to be an instrument of Pre-Academic Learning and Exam Self-Efficacy. The aims of the present study were then two-fold: Firstly, to conduct a thorough investigation of the construct validity and psychometric properties of the instrument, using cutting edge item response theory models; i.e. graphical loglinear Rasch models. These are models where both differential items functioning and a lack of local independence between items can be adjusted for, as long as these are uniform, in order to obtain unbiased scores for subgroups, while retaining sufficiency of the summed score. In the psychometric analyses, we included investigations of measurement invariance across both institutions and degree programs and other subgroupings of students, in order to prevent subsequent confounding of the relationship between pre-academic self-efficacy and attrition by institution or degree program. Secondly, to employ the instrument in an analysis of the association between the two measures of pre-academic self-efficacy and attrition within the first half-year. More specifically, we investigated the following research questions:

RQ1: Is it possible to obtain valid, reliable and well-targeted measures of pre-academic learning self-efficacy and/or pre-academic exam self-efficacy of undergraduate higher education students after admission and prior to commencing education, across multiple degree programs and institutions?

In these analyses, a major focus was measurement invariance in relation to degree programs and institutions, as well as across subgroups of students defined by gender, age, and priority of program at the time of application.

For student groups where psychometrically sound and invariant measures of self-efficacy could be established, we also investigated the research question:

RQ2: How is pre-academic self-efficacy at admission associated with half-year attrition? How is this association between pre-academic self-efficacy and attrition changed when variation between educational programs and individual covariates are taken into account? And how can these results be interpreted in terms of causality?

## 2. Methods

### 2.1 Data

The data for the current study was collected by the National Evaluation Institute (EVA) within the framework of the ongoing national longitudinal panel on student attrition encompassing several year cohorts of undergraduate students admitted to higher education (HE). This attrition panel is conducted in the population of students admitted to HE degree programs through the Coordinated Admission (KOT) in Denmark. Among students who were admitted to a higher educational program in 2017 a cohort of students participate in a total of four waves of data collection time-wise spanning from admission notice to 13 months after commencing the degree program. Student participation was solicited through the national digital mail system [e-boks] in the first wave, where consent was also given. Data was collected during two weeks, starting 10 days after students had received their admittance letter and ending before commencement of the HE programs in Denmark. The first wave for the 2017-cohort included the two scales intended to measure academic learning and exam self-efficacy in relation to the students' future first semester courses, which are utilized in the current study.

Sample Selection Criteria:

1. For validity studies of new measurement scales or adapted versions of existing scales, it is desirable to employ a controlled and maximally comparable data sample across subgroups, in order to be able to identify properly the source of any measurement issues arising. The current sample was thus selected



from the first wave of data from the 2017 cohort (c.f. the above) as student cases with complete self-efficacy data which fulfilled three a-priori criteria:

2. students who were admitted to the two largest universities in Denmark; the University of Copenhagen (UCPH) and the University of Aarhus (UAA) for the summer start of programs,
3. academic programs existing at both universities,
4. academic programs who admitted at least 200 students in one of the included universities. However, as the initial selection of programs did not include any from the faculty of arts, we decided to include also the largest language and largest “other humanities” programs as well.

### 2.1.2 The current data sample

The resulting data sample consisted of complete self-efficacy and background information from, on average, 54% of the students admitted to nine large academic programs (Table 1).

Table 1.

#### *Distribution of the study sample on institution and degree programs*

University and academic Discipline	Admitted N	Complete study data <sup>a</sup>	
		N	% of admitted
University of Copenhagen			
Medicine	559	297	53%
Law	764	362	47%
Psychology	230	129	56%
Economy	330	169	51%
Computer science	242	114	47%
Biology	219	115	53%
Political science	264	156	59%
Danish language	194	80	41%
History	159	74	47%
sub-total	2961	1496	51%
University of Aarhus			
Medicine	481	289	60%
Law	420	244	58%
Psychology	215	126	59%
Economy	180	111	62%
Computer science	156	89	57%
Biology	118	59	50%
Political science	275	160	58%
Danish language	102	54	53%
History	109	58	53%
sub-total	2056	1190	58%
grand total	5017	2686	54%

*Notes.* <sup>a</sup> Of the 2777 students from the included programs who responded to the survey, 91 had completely missing self-efficacy data. Of these, 56 were from the University of Copenhagen and 35 from the University of Aarhus, which was not a significant difference across universities (Fischer’s exact  $\chi^2$  1,219, df 1,  $p=0.285$ ). Looking further into the pattern of missing data in each of the 18 programs, we found that between 0 and 11 students had missing self-efficacy data, with the highest numbers being in the largest programs. Thus, the missing pattern appears to be random. The 91 cases were excluded from the analysis, which was thus done on 2686 students.

The University of Copenhagen (UCPH) sample consisted of 1496 students. Their mean age was 21.7 years (SD 4.8), 61.3 % (n = 917) were women, and 92.5 % (n = 1384) were admitted to the academic program that was their first priority when they applied. The University of Aarhus (UAA) sample consisted of 1190 students. Their mean age was 21.2 years (SD 3.2), 59.5 % (n = 708) were women, and 92.1 % (n = 1096) were admitted to the academic program that was their first priority.





## 2.2 Instrument & administration

The self-efficacy scales employed in the current study was adapted from the Specific Academic Learning Self-Efficacy Scale (SAL-SE) and Specific Academic Exam Self-Efficacy Scale (SAE-SE) (Nielsen et al., 2017). The SAL-SE and the SAE-SE scales are subscales of the self-efficacy scale in the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1991), translated into Danish and with an adapted response scale. The MSLQ is intended to measure students' motivational orientation and learning strategies in high school and HE (Pintrich et al., 1991), and has been widely used and translated into several other languages (Credé & Philips, 2011; Duncan & McKeachie, 2005). Nielsen and colleagues (2017), with samples of psychology students, found that the SAL-SE and the SAE-SE scales each fit a Rasch model and as such the sum scores of each scale are sufficient statistics for the latent Rasch score. It was clearly rejected that the two subscales formed a single unidimensional scale. Both the SAL-SE and the SAE-SE subscales had sufficient reliability for the investigated samples to be used both in surveys at the group level and to assess the SE levels of individual students as intended with the MSLQ (SAL-SE  $r = .87$ , SAE-SE  $r = .89$ ). Targeting of the SAL-SE and the SAE-SE subscales to the sample of psychology students were not optimal, in the sense that students were located more towards the higher end of the scales while most information could be obtained more towards the lower end of the scales. The targeting of the SAL-SE subscale was the better of the two.

The SAL-SE and the SAE-SE scales (Nielsen et al., 2017) are course specific measures of self-efficacy. However, it was not possible to refer to specific courses in the current student population, as the students had not yet commenced the degree program that they had been admitted to. Accordingly, all items had been adapted so that instead of referring to "this class/course", they referred to "the first courses in my program", while the response categories were not changed (Table A in the appendix shows the Danish as well as an English version of the items). The adaptation made the subscales somewhat more general measures of learning and exam self-efficacy than the SAL-SE and SAE-SE scales, as well as measures of what might be termed pre-academic self-efficacy as the students had not yet entered academia. As a consequence, we have renamed the two scales to the more appropriate Pre-Academic Learning Self-Efficacy Scale (PAL-SE) and Pre-Academic Exam Self-Efficacy Scale (PAE-SE). The item texts of the PAL-SE and the PAE-SE are provided in both Danish and English in the Appendix (Table A1).

All PAL-SE and PAE-SE items were administered in the same relative order as in the original MSLQ (Pintrich et al., 1991), and they were mixed in with items from other scales.

## 2.3 Rasch analysis and graphical Loglinear Rasch models

The family of Rasch models, which is a subgroup of item response theory models (van der Linden & Hambleton, 1997) will be used for the item analysis. Fit to a Rasch model (Rasch, 1960) provides optimal measurement properties of the scale in question (Kreiner, 2007, 2013; Mesbah & Kreiner, 2013) including: local independence of items (i.e. responses to a self-efficacy item depends only on the level of self-efficacy and not on responses to any of the other self-efficacy items). Absence of differential item functioning (no DIF) (i.e. responses to a self-efficacy item depends only on the level of self-efficacy and not on persons' membership of subgroups such as gender, age, degree program, etc.). Optimal reliability, as items are conditionally independent. Score sufficiency (i.e. the sum score is a sufficient statistic for the latent self-efficacy score), which is a property only provided by the Rasch model. The property of sufficiency is desirable when using the summed raw score of a scale, as it is the usual case with the present self-efficacy scales. Thus, fit to the Rasch model allows the use of the summed raw score or the estimated person parameters (sometimes called Rasch-scores) subsequently. Thus, the choice of using the summed raw score or the estimated person parameters has to be made in each case, as it depends on both the purpose for using the score (i.e. for statistical analysis or individual assessment), the length of the scale, the targeting and the reliability of the scale, and so on. One additional factor to take into account in relation to the choice of using sum scores or person parameters is the



interpretability of these; the first is easily interpreted in relation to item scores, while the latter is not as it is a logit scale. These considerations extend to the graphical loglinear Rasch models described below.

Specifically, we used the partial credit model (PCM; Masters, 1982), which is a generalization of the Rasch model for ordinal items – thus this is denoted Rasch model (RM) throughout the article. DIF will to some degree cause bias in the scores (sum or person parameters) of some groups relative to other groups, thus affecting the validity of the scales and limiting their use in future studies where comparisons are made, as such comparisons would be confounded (Kreiner, 2007; Nielsen & Kreiner, 2013). Therefore, adjustment for DIF is particularly important in relation to subsequent comparisons of the self-efficacy of students across degree programs and institutions. Local dependence between items will not confound subsequent comparisons of subgroups, but will instead inflate reliability, if not taken into account. In order to accommodate these issues, we used graphical loglinear Rasch models (GLLRMs; Kreiner & Christensen, 2002, 2004, 2007), when fit to the Rasch model was rejected, as GLLRMs allow adjustments for local dependence between items and of differential item functioning (DIF or item bias), when these departures are uniform (i.e. are of the same strength at all levels of the latent trait), while retaining close to optimal measurement (Kreiner & Christensen, 2007).

### 2.3.1 Item analysis

Overall tests of fit (i.e. tests of global homogeneity by comparison of item parameters in low and high scoring groups) and of no global DIF were conducted using Andersen's (1973) conditional likelihood ratio test (CLR). Effectively, this tests the hypothesis that the item parameters were the same for those students with a low level of self-efficacy and those with a high level of self-efficacy (i.e. overall fit), and the hypothesis that the item parameters were the same for subpopulations of students defined by university, program gender and so on (i.e. no global DIF). The fit of individual items was tested by comparing the observed item-rest-score correlations with the expected item-rest-score correlations under the model (i.e. the RM or the specified GLLRM) (Kreiner & Christensen, 2004). The presence of LD and DIF in GLLRMs was tested using two tests; Kelderman's (1984) conditional likelihood ratio test of local independence (i.e. no DIF, no LD), and conditional independence tests using partial Goodman-Kruskal gamma coefficients for the conditional association between item pairs (presence of LD) or between items and exogenous variables (presence of DIF) given scores (Kreiner & Christensen, 2004). Specifically, we tested for DIF in relation to institution (University of Copenhagen, University of Aarhus), degree program (medicine, psychology, political science, economy, law, biology, computer science, Danish language, history), program priority (1<sup>st</sup>, 2<sup>nd</sup> or less), gender (male, female), age groups (20 years or younger, 21 years or older). Evidence of overall fit and no global DIF found in the overall tests (CLR) was rejected if this was not supported by lack of evidence of both LD and DIF and evidence of individual item fit.

Reliability was estimated using Hamon and Mesbah's (2002) Monte Carlo method for reliability in Rasch scales, as this method can overcome violations of the assumption of locally independent items, in contrast to Cronbach's Alpha which tends to overestimate reliability in such cases. Targeting was assessed numerically by two indices as well as graphically, to allow for both graphical and numerical evaluation of targeting (Kreiner & Christensen, 2013). Numerically, we calculated the test information target index as the mean test information divided by the maximum test information for theta, and the root mean squared error (RMSE) target index as the minimum standard error of measurement divided by the mean standard error of measurement for theta. Both indices should preferably have a value close to one. In addition, we estimated the target of the observed score and the standard error of measurement (SEM) of the observed score. Graphically, we plotted so-called items maps with the distribution of the item threshold locations against weighted maximum likelihood estimations of the person parameter locations as well as the person parameters for the population (assuming a normal distribution) and the information function.

The Benjamini-Hochberg (1995) procedure was used to adjust for false discovery rate (FDR) due to multiple testing, whenever appropriate. As recommended by Cox et al. (1977), we did not apply a critical limit



of 5% for p-values as a deterministic decision criterion. Instead, we distinguished between weak ( $p < .05$ ), moderate ( $p < .01$ ) and strong ( $p < .001$ ) evidence against the model.

## 2.4 Pre-Academic Learning self-efficacy and half-year attrition

In order to answer the second research question, the association between the measures of academic self-efficacy and attrition was investigated using the adjusted sum score scales resulting from the item analysis described above.

The linear probability model (LPM) was applied using a dichotomous outcome variable indicating dropout within the first half year (1) or not (0) (i.e. the dependent variable). LPM models the probability of dropout as a linear function of covariates, and present the conditional expectation of the outcome (Greene, 2011). When we use LPM,  $\beta$ -coefficients can be interpreted as differences in the probability of attrition in percentage points corresponding to a +1 difference in the independent variables (e.g. PAL-SE). This straightforward interpretation is one of the reasons why the linear probability model has been an increasingly popular choice for modelling binary outcomes (Breen et al, 2018: 4.12). The PAL-SE variable has been rescaled to 0-1 for the LPM analysis, in order to make findings more intuitive.

Three models were applied to answer the question about how Pre-Academic Self-efficacy relates to attrition:

- 1) A bivariate model including academic self-efficacy and half-year attrition. This is in order to investigate whether PAL-SE can be used to predict differences in half-year attrition without taking any other covariates into account.
- 2) A model including academic self-efficacy, half-year attrition and a nominal variable indicating educational program as a covariate. This was in order to investigate whether a correlation between PAL-SE and attrition could simply be due to differences in the type of students between programs that also relates to both PAL-SE scores and dropout, regardless of student's educational experience. This is to answer the question whether PAL-SE can be used to predict differences in half-year attrition if differences between programs are taken into account. From an institutional point of view, this can be valuable information although it does not imply causality.
- 3) A model including academic self-efficacy, half-year attrition, a nominal variable indicating educational program and a range of individual covariates as for example earlier educational attainment (GPA) and a range of social background variables (parents education, income etc.). We thus investigated the correlation between PAL-SE and attrition in a stratified comparison comparing students with different values of PAL-SE who are otherwise as similar as possible. This was to get an indication as to whether a correlation between PAL-SE and attrition could be explained by differences in background variables or whether it alternatively could be interpreted as a causal effect of PAL-SE on attrition.

## 2.5 Software

All item analyses by RM and GLLRMs were conducted in DIGRAM (Kreiner, 2003; Kreiner & Nielsen, 2013), while R was used for some of the graphical illustrations. The analysis of the effect of Pre-Academic Learning self-efficacy on half-year attrition was conducted in STATA.

## 3. Results

The Pre-Academic Learning self-efficacy subscale (PAL-SE) did not fit the Rasch model, nor did the Pre-Academic Exam self-efficacy subscale (PAL-SE). However, for both subscales it was possible to establish





fit to a GLLRM, though of different complexity (Figure 1). Global tests-of-fit and DIF for the two subscales are presented in Table 2, while item fit statistics are provided in Table 3.

The PAL-SE subscale fitted a GLLRM with strong local dependence between items 1 (I'm certain I can understand the most difficult material presented in the readings for the first courses in my program) and 3 (I'm confident I can understand the most complex material presented in the first courses in my program), as well as very weak local dependence between items 2 (I'm confident I can understand the basic concepts taught in the first courses in my program) and 4 (I'm certain I can master the skills being taught in the first courses in my program). In addition, item 2 (see above) was found to function differentially in relation to the degree programs. This meant that the student responses to item 2 differed systematically in how likely they were to feel described as future students by the statement dependent on the degree program they were admitted to, no matter their level of pre-academic learning self-efficacy.

The PAE-SE subscale also fitted a GLLRM, but in this case all items were locally dependent. The strongest local dependence was found between items 2 (*I'm confident I can do an excellent job on the assignments and tests in the first courses in my program*) and 3 (*I expect to do well in the first courses in my program*), while the remaining instances of local dependence were both weak. No evidence of DIF was found in relation to the PAE-SE scale.

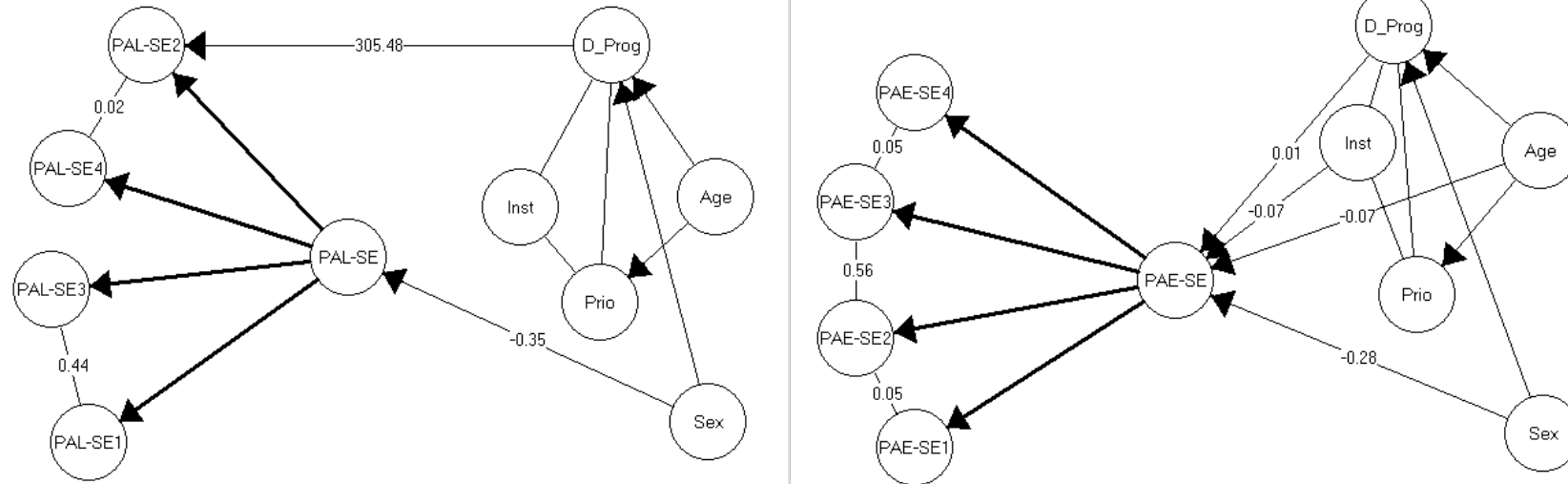


Figure 1. The final graphical loglinear Rasch models for the Pre-Academic Learning self-efficacy subscale (left side) and the Pre-Academic Exam self-efficacy subscale (right side).

Notes. Correlations are partial Goodman & Kruskal's gamma coefficients. In the GLLRM for the PAL-SE scale no gamma coefficient is available to describe the DIF of item two relative to degree program, as this is a nominal variable the chi-square value is shown instead. Similarly, in the GLLRM for the PAE-SE scale the correlation between degree program and the PAE-SE score is so low (.01) because the association is not an ordered one across the nine nominal degree programs.



Table 2.

*Global Tests-of-fit and differential item function for the Pre-Academic Exam self-efficacy and the Pre-Academic Learning self-efficacy subscales for the total sample*

Tests	PAE-SE (RM)			PAE-SE (GLLRM) <sup>a</sup>			PAL-SE (RM)			PAL-SE (GLLRM) <sup>b</sup>		
	CLR	df	<i>p</i>	CLR	df	<i>p</i>	CLR	df	<i>p</i>	CLR	df	<i>p</i>
Global homogeneity	88.1	15	<.0005 <sup>+</sup>	39.6	52	.90	106.2	15	<.0005 <sup>+</sup>	91.0	77	.13
DIF relative to:												
University	17.3	15	.30	49.9	52	.56	23.5	15	.08	88.6	77	.17
Degree program	119.3	120	.50	347.1	416	.99	177.9	120	<.0005 <sup>+</sup>	357.6	336	.20
Priority	20.1	15	.17	31.6	52	.99	18.7	15	.23	75.4	77	.53
Gender	14.2	15	.51	56.6	52	.31	26.6	15	.03 <sup>+</sup>	89.6	77	.16
Age groups	13.8	15	.54	48.5	52	.61	18.7	15	.23	100.3	77	.04 <sup>++</sup>

*Notes.* PAL-SE: Pre-Academic Learning Self-Efficacy; PAE-SE: Pre-Academic Exam Self-Efficacy; RM: Rasch model; GLLRM: Graphical loglinear Rasch model; CLR: Conditional Likelihood Ratio; df: degrees of freedom; *p*: *p*-value; DIF: differential item function. Global homogeneity test compares items parameters in approximately equal-sized groups of high and low scoring parents. The critical limits for the *p*-values after adjusting for false discovery rate were: <sup>+</sup> 5% and 1% limit unaltered. <sup>++</sup>5% limit *p* = .0083, 1% limit *p* = .0017.

<sup>a</sup> The GLLRM for the PAE-SE subscale assumed that some items pairs are locally dependent (items 3 and 4, and items 3 and 2, and items 2 and 1).

<sup>b</sup> The GLLRM for the PAL-SE subscale assumed that items 1 and 3 are locally dependent and that item 2 functions differentially relative to Degree program.



Table 3.

*Item fit statistics for the Pre-Academic Exam self-efficacy and the Pre-Academic Learning self-efficacy subscales under the respective RM and the GLLRMs*

Subscale & items	Observed $\gamma$	Item-restscore correlations				
		Fit to RM		$p$	Fit to GLLRM <sup>a</sup>	
		Expected $\gamma$			Observed $\gamma$	Expected $\gamma$
<b>SAE-SE</b>						
PAE-SE1	0.718	0.749	.017 <sup>+</sup>	0.718	0.702	0.271
PAE-SE2	0.826	0.770	<.0005 <sup>+</sup>	0.826	0.839	0.204
PAE-SE3	0.833	0.759	<.0005 <sup>+</sup>	0.833	0.811	0.062
PAE-SE4	0.690	0.756	<.0005 <sup>+</sup>	0.690	0.703	0.391
<b>SAL-SE</b>						
PAL-SE1	0.741	0.714	.025 <sup>++</sup>	0.741	0.764	0.050
PAL-SE2	0.657	0.700	<.0005 <sup>++</sup>	0.657	0.659	0.868
PAL-SE3	0.795	0.713	<.0005 <sup>++</sup>	0.795	0.762	0.004 <sup>+++</sup>
PAL-SE4	0.685	0.707	0.101	0.685	0.669	0.276

*Notes.* PAL-SE: Pre-Academic Learning Self-Efficacy; PAE-SE: Pre-Academic Exam Self-Efficacy;  $\gamma$  = Goodman & Kruskal’s gamma coefficients; RM: Rasch model; GLLRM: Graphical loglinear Rasch model. The critical limits for the p-values after adjusting for false discovery rate were: <sup>+</sup> 5% limit unaltered, 1% limit  $p = .0092$ . <sup>++</sup> 5% limit  $p = .0375$ , 1% limit  $p = .0058$ . <sup>+++</sup> 5% limit  $p = .0197$ , 1% limit  $p = .0023$ .

<sup>a</sup> The GLLRM for the PAE-SE subscale assumes that some items pairs are locally dependent (items 3 and 4, and items 3 and 2, and items 2 and 1), while the GLLRM for the PAL-SE subscale assumes that items 1 and 3 are locally dependent and that item 2 functions differentially relative to degree program.

### 3.1 Effect of degree program DIF

A single item in the PAL-SE scale suffered from DIF relative to degree program. Thus, to be able to use either the summed scale scores or the estimated person parameters in subsequent statistical analysis, the DIF first has to be taken into account by adjusting scores accordingly. In the Appendix, we provide conversion tables providing both the necessary information for converting the summed scale scores to estimated person parameters, the estimated person parameters for all the different subgroups affected by the DIF, and DIF-adjusted scale scores for these subgroups where appropriate (Table A3 parts 1 to 3 for the PAL-SE scale, and Table A4 for the PAE-SE scale). Using the summed and the DIF-equated scores for the PAL-SE scale it is possible to investigate to which extent the degree program DIF would confound subsequent statistical comparison, if not adjusted for. The results of such an overall comparison is shown in Table 4.



Table 4.

*Comparison of observed and DIF-adjusted mean Pre-Academic Learning self-efficacy scores in the Degree program subgroups affected by differential item functioning*

Degree program (n)	Observed scores		Adjusted scores		Bias
	Mean	SE	Mean	SE	
Medicine (586)	12.65	.12	12.65	.12	.00
Psychology (255)	12.33	.17	12.47	.17	-.13
Political science (316)	12.69	.15	12.63	.16	.06
Economy (280)	13.14	.16	13.19	.16	-.04
Law (606)	12.35	.11	12.45	.11	-.10
Biology (174)	12.74	.20	12.60	.20	.14
Computer science (203)	13.70	.22	13.68	.22	.02
Danish language (134)	12.28	.27	12.30	.27	-.02
History (132)	13.59	.24	13.61	.24	-.02

Notes. SE: Standard error. Overall differences in observed mean scores ( $\chi^2(8) = 59.1, p < .001$ ). Overall differences in adjusted mean scores ( $\chi^2(8) = 52.1, p = .001$ ).

In addition to the overall tests of difference in the observed mean PAL-SE scores and the DIF-adjusted mean PAL-SE scores of students admitted to the nine different degree programs, we conducted post-hoc step-wise analyses of pairwise collapsibility of the nine nominal degree program categories (see supplement file for details). In the case of the observed scores, the analysis of pairwise collapsibility resulted in four groups of degree programs with significantly different mean PAL-SE scores: medicine, political science & biology (mean 12.68, se .087); psychology, law & Danish language (mean 12.34, se .089); economy (mean 13.14, se .156); and finally computer science & history (13.65, se .162). The pairwise collapsibility analysis of the DIF-adjusted scores resulted in only two groups of degree programs with significantly different mean PAL-SE scores: medicine, political science, biology, psychology, law & Danish language (mean 12.54, se .063); economy, computer science & history (13.41, se .114).

### 3.2 Targeting and reliability

Targeting of the PAL-SE and PAE-SE scales was very different (Table 5). Targeting of the PAL-SE scale was considered good for all students, though it varied slightly across subgroups of students admitted to the nine different degree programs (ranging from 73% to 81% of the maximum information obtained for students admitted to biology and economy, respectively). Targeting of the PAE-SE scale was the same for all students and not good (52% of the maximum information obtained). The differences in targeting are also illustrated by the item maps showing the relative locations of person parameters and item thresholds along the two theta scales as well as the test information curves (Figure A1 in the appendix file). The item map for the PAE-SE scale shows that even though person parameters and item thresholds are both spread out along the entire scale, maximum information is located at the lower end of the scale where hardly any students are located, and there is much less information where most of the students are concentrated. For the PAL-SE scale, the separate item maps for students admitted to the nine different degree programs also show that person parameters and item thresholds are both spread out along, more or less, the entire scale. However, for the PAL-SE scale, the maximum information is found closer to more of the student locations, and there is much more information where most of the students are located than was the case with the PAE-SE scale.

The reliability of the PAE-SE scale (all students) was .77, while the reliability of the PAL-SE scale varied slightly across subgroups of students admitted to the nine different degree programs; from .79 for students admitted to biology and economy, and .86 for students admitted to computer science (weighted average for all students .81) (Table 5). Thus, by conventional standards, the reliability of both scales was at acceptable levels.





Table 5.

*Targeting and reliability of the Pre-Academic Exam self-efficacy and the Pre-Academic Learning self-efficacy subscales*

Subscales and Degree program subgroups <sup>a</sup> (n)	Target	Mean	TI mean	TI max	Theta			RMSE mean	RMSE min	RMSE target index	Target	Score		r
					TI Target index	Mean	SEM							
PAL-SE subscale														
Medicine (586)	-1.85	.29	1.519	1.988	.764	.818	.709	.867	9.10	12.65	1.23	.83		
Psychology (255)	-1.91	.19	1.519	1.987	.765	.819	.709	.866	8.89	12.33	1.23	.80		
Political science (316)	-1.88	.26	1.491	1.894	.787	.827	.727	.879	9.21	12.69	1.22	.80		
Economy (280)	-2.10	.66	1.439	1.770	.813	.833	.752	.902	8.88	13.14	1.20	.79		
Law (606)	-1.95	.16	1.549	2.027	.764	.811	.702	.866	8.80	12.35	1.24	.80		
Biology (174)	-2.41	.23	1.527	2.106	.725	.816	.689	.845	8.21	12.74	1.23	.79		
Computer science (203)	-2.00	1.05	1.458	2.001	.728	.825	.707	.857	8.81	13.70	1.20	.86		
Danish language (134)	-2.27	.10	1.489	1.965	.758	.825	.713	.864	8.44	12.28	1.22	.84		
History (132)	-2.43	.93	1.420	1.855	.765	.837	.734	.878	8.33	13.59	1.19	.81		
PAE-SE subscale														
All programs ( <i>no DIF</i> ) (2686)	-3.16	.46	1.312	2.531	.518	.891	.629	.705	2.91	8.93	1.14	.77		

Notes. PAL-SE: Pre-Academic Learning Self-Efficacy; PAE-SE: Pre-Academic Exam Self-Efficacy; TI: Test information; RMSE: The root mean squared error of the estimated theta score; SEM: The standard error of measurement of the observed score; r: reliability.

<sup>a</sup>. Targeting and reliability is provided for groups defined by DIF variables.



### 3.3 Pre-Academic Learning Self-Efficacy and half-year attrition

Based on the results of the item analyses of the PAL-SE and the PAE-SE scales (RQ1), particularly the marked difference in the targeting of the two (Table 5 and Figure A1 in the appendix), we decided to proceed to RQ2 (How is academic self-efficacy at admission associated with half-year attrition) using only the PAL-SE scale. We report the result of the regression analysis using the DIF-adjusted sum score of the PAL-SE, as there were no noteworthy differences in the results using the estimated person parameters, and the results using the adjusted sum score are more readily interpreted as they relate directly to the instrument.

The analysis showed a significant association between PAL-SE and half-year attrition; the higher students scored on PAL-SE the lower was the probability for them to drop out within the first half year in the degree program (Figure 2 and Table A5 in the appendix). The association was even stronger when program was included in the model as a covariate, and could thus not be explained by differences in the type of students who were admitted to the different programs. This leads to the conclusion that PAL-SE can predict significant differences in students' probability to drop out within the first half year. More precisely the probability of dropping out within the first half year was 7.2% lower for students with the highest possible level of PAL-SE compared to students with the lowest possible level of PAL-SE, when degree program was included as a covariate. This is considered a substantial difference, considering the overall dropout percentage in the sample was 3.9 % during the first half year.

Adding individual controls did not change the correlation substantially. Hence, the correlation between PAL-SE and attrition also could not be explained by differences in the included individual covariates indicating that PAL-SE has a causal effect on attrition. The fact that the correlation did not change substantially when we added covariates further supports a causal interpretation (Heckman, Humphries et al. 2016). The design is however not a strong design in terms of potential selection bias stemming from unobservable heterogeneity. Overall, in terms of causality, the results thus indicate that PAL-SE does have an effect on student attrition, although especially the size of this effect is associated with uncertainty.

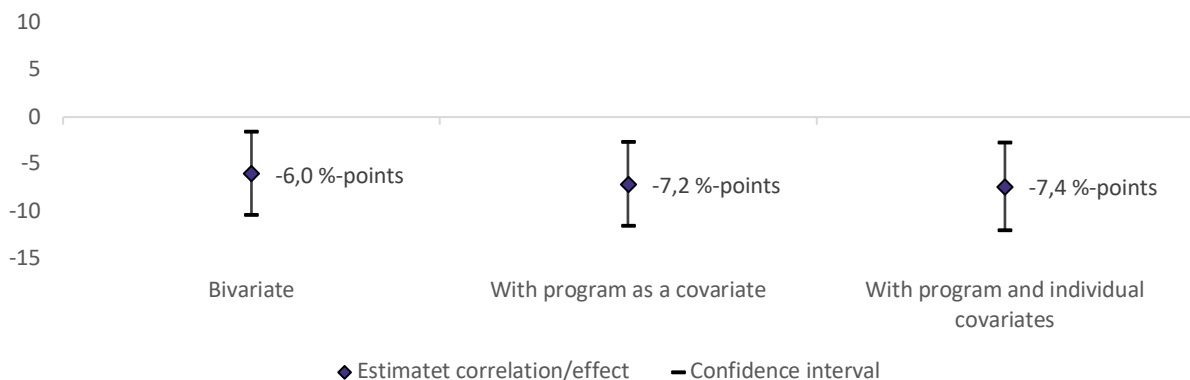


Figure 2. Results from regression analysis by Linear Probability Model: Association between Pre-Academic Learning Self-Efficacy on half-year attrition in models with different covariates included.

Notes. The estimated effect corresponds to a difference in PAL-SE from minimum to maximum, since PAL-SE is measured on a scale ranging from 0-1. We use the adjusted raw score from the item analysis to measure PAL-SE. Individual controls included as covariates: 1) Priority of program in application 2) Sex 3) Age 4) Grade point average from high school [Danish gymnasium] 5) Parents highest completed educational level 6) Parents income 7) Ethnicity 8). N: bivariate; 2.340, N: with program as covariate; 2.340, N: with program and individual covariates: 2.195.



#### 4. Discussion and future directions

This aim of the study was firstly to investigate the construct validity and psychometric properties of the proposed PAL-SE and PAE-SE scales in a large sample of university students just admitted to the same nine degree programs in two large universities, using Rasch models. Secondly, and depending on the results of the psychometric part of the study, the aim was to investigate how self-efficacy before commencement of degree program was associated with half-year attrition.

With regard to the first aim, the construct validity of the PAL-SE and PAE-SE scale was to a large extent confirmed, though with different psychometric properties. Thus, neither scale fit a Rasch model. The PAL-SE scale was found to be, psychometrically speaking, the better of the two scales. We found evidence of differential item functioning relative to degree program, but it was possible to address this by adjusting the score. We did not find any evidence for differential items functioning for the PAE-SE scale, only evidence of local dependence among items, this we also found to some degree in the PAL-SE scale. The main reason that we find the PAL-SE scale to be the better of the two scales are the differences in targeting; the targeting of the PAL-SE scale was good, while the targeting of the PAE-SE scale was poor. The items in both scales are spread out across the latent scale, as are the person estimates, however the level of information in the PAE-SE scale is substantially lower than in the PAL-SE scale. Thus, the items in the PAL-SE scale provide more information on the students in the study sample than it is the case with the PAE-SE items. Furthermore, the variations in the obtained information with the PAL-SE scale for students from the different degree programs vary very little, even though the admission GPA across degree programs range from “all admitted” (i.e. no admission GPA) to 11.5 (the highest is 12). This finding adds substantially to the field, as no previous attempts have been made to investigate the construct validity and psychometric properties of an instrument for pre-academic self-efficacy using the framework of graphical loglinear Rasch models.

When considering that the scales are intended to measure learning and exam self-efficacy, respectively, after admission but before commencement of university studies, we find the difference in targeting to be further evidence of construct validity. All things being equal, it is reasonable to assume that the PAL-SE items should be better suited to the study population, as the students probably have a more well-developed idea about their future learning capacity, as they have applied to a program most likely of their interest and have obtained the admission GPA necessary to be admitted, while they do not have as well-developed a sense of the exams in the university system, and thus the items carry less information for this student population.

Another entirely new finding is that we have established that it is indeed possible to obtain invariant measures of both Pre-Academic Learning Self-efficacy and Pre-Academic Exam Self-efficacy for students admitted to a wide range of university degree programs and at two different universities, when making the proper adjustments. This new finding will prove valuable in future studies on self-efficacy, whether pre-academic self-efficacy or academic self-efficacy measured during studies, as it paves the way for more studies assuring the application of unbiased self-efficacy scores.

The LPM regression analysis adds to the current evidence that self-efficacy is strongly correlated to educational outcomes, such as attrition. It more specifically points out that this is also the case, even though we measured the level of self-efficacy before the students started on the university program and as a measure that did not relate to a specific course but to the upcoming courses in the first semester as a whole. Van Herpen and colleagues (2017), using an adaptation of the self-efficacy scale of the MSLQ (Pintrich et al., 1991) where they changed items from being course-specific to referring in general to “the first year of study”, found no association between pre-university self-efficacy and academic success. The instrument they used was the most similar to ours, but they treated it as a single scale, did not investigate its measurement properties, and their sample size was too small to take academic discipline into account in the analysis. In a study on discipline-specific differences in general self-efficacy prior to commencing university, Yong (2010) found no such differences with small samples of future engineering and business students. The only study we located, which investigates pre-higher education self-efficacy and attrition was that of Aryee’s (2017), which found that high levels of pre-higher education *mathematics* self-efficacy were associated with completion of college, while low levels were associated with a diminished likelihood of attaining a degree. Our findings are thus an



expansion of the very scarce research on pre-higher education self-efficacy, and the even scarcer research on its association with attrition, as it represents the first multi degree program study on this relationship. Furthermore, our findings are in line with the research where self-efficacy is measured during degree programs, typically during the first semester, as they find that low self-efficacy is associated with increased attrition, increased risk of withdrawal, decreased intention to persist, and decreased intention to complete (De Clercq, Galand & Frenay, 2017; Elliot, 2016; Lent et al., 2016; Thomas, 2014; Vuong, Brown-Welty & Tracz, 2010; You, 2018).

Even though we still need further evidence relating to the causal link between Pre-Academic Learning Self-Efficacy and attrition, as we do not have a very strong design in terms of causality, the results do indicate that PAL-SE has a causal effect on attrition. The results are promising, since self-efficacy to a large extent is regarded as context specific and malleable, thus providing an opportunity for teachers and other educational professionals to promote students' self-efficacy in order for them to achieve higher learning outcomes and reduce attrition. The current study is the first to show that even prior to commencing the degree program students admitted to university are learning-wise self-efficacious with regard to the upcoming courses and that their level of Pre-Academic Learning Self-Efficacy is associated with attrition. We therefore call for three lines of further studies. Firstly, studies to replicate the findings with a stronger causal design. Secondly, applied studies to show how end how much teachers and other educational professionals are able to affect student's self-efficacy positively. Thirdly, controlled intervention studies showing the effect of pedagogical strategies to enhance self-efficacy on attrition, intention to persist and complete degree programs.

## Keypoints

- Short scales measuring Pre-Academic Learning Self-efficacy (PAL-SE) and Pre-Academic exam Self-efficacy (PAE-SE) were developed from existing scales.
- Validation data was drawn from a national sample of students admitted to, but prior to starting at, university, using a-priory defined criteria.
- Using graphical loglinear Rasch models the PAL-SE and PAE-SE were found to be construct valid.
- The PAL-SE was used for investigating the relationship between pre-academic self-efficacy and half-year attrition.
- Results from restricted models indicate that Pre-Academic Learning Self-efficacy has a substantial effect on half-year attrition.



## Acknowledgments

The authors would like to thank Pedro Henrique Ribeiro Santiago for providing us with the R-code for the items maps preliminary to its release.

## References

- Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38(1), 123–140. DOI: 10.1007/BF02291180
- Aryee, M. (2017). *College students' persistence and degree completion in science, technology, engineering, and mathematics (STEM): The role of non-cognitive attributes of self-efficacy, outcome expectations, and interest*. (Doctoral dissertation). Available from: ProQuest Dissertations and Theses databases. (UMI No. 10264673)
- Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York, NY: Freeman. DOI: 10.5860/CHOICE.35-1826
- Bartimote-Aufflick, K., Bridgeman, A., Walker, R., Sharma, M., & Smith, L. (2015). The study, evaluation, and improvement of university student self-efficacy. *Studies in Higher Education*, (ahead-of-print), 1-25. DOI: 10.1080/03075079.2014.999319
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x
- Breen, R.; Karlson, K. B. & Holm, A. (2018). Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models. *Annual Review of Sociology*, 44, 4.1–4.16. DOI: 10.1146/annurev-soc-073117-041429
- Burger, K. & Samuel, R. (2017). The Role of Perceived Stress and Self-Efficacy in Young People's Life Satisfaction: A Longitudinal Study. *Journal of Youth and Adolescence*, 46, 78–90. DOI: 10.1007/s10964-016-0608-x
- Cox, D. R.; Spjøtvoll, E.; Johansen, S.; van Zwet, W. R.; Bithell, J. F. & Barndorff-Nielsen, O. (1977). The Role of Significance Tests [with Discussion and Reply]. *Scandinavian Journal of Statistics*, 4, 49–70.
- Credé, M., & Phillips, L. A. (2011). A meta-analytic review of the motivated strategies for learning questionnaire. *Learning and Individual Differences*, 21(4), 337–346. DOI: 10.1016/j.lindif.2011.03.002
- De Clercq, M.; Galand, B. & Frenay, M. (2017). Transition from high school to university: a person-centered approach to academic achievement. *European journal of psychology of education*, 32(1), 39-59. DOI: 10.1007/s10212-016-0298-5
- Devonport, T. J. & Lane, A. M. (2006). Relationships between self-efficacy, coping and student retention. *Social Behavior and Personality: an international journal*, 34(2), 127-138. DOI: 10.2224/sbp.2006.34.2.127
- Duncan, T. G. & McKeachie, W. J. (2005). The making of the motivated strategies for learning questionnaire. *Educational Psychologist*, 40(2), 117–128. DOI: 10.1207/s15326985ep4002\_6.
- Elliott, D. C. (2016). The impact of self beliefs on post-secondary transitions: The moderating effects of institutional selectivity. *Higher Education*, 71(3), 415-431. DOI: 10.1007/s10734-015-9913-7
- Ferla, J.; Valcke, M. & Cai, Y. (2009). Academic self-efficacy and academic self-concept: Reconsidering structural relationships. *Learning and Individual Differences*, 19(4), 499-505. DOI: 10.1016/j.lindif.2009.05.004
- Ginsborg, J.; Kreutz, G.; Thomas, M. & Williamon, A. (2009). Healthy behaviours in music and non-music performance students. *Health Education*, 109(3), 242-258. DOI: 10.1108/09654280910955575
- Greene, W. H. (2011). *Econometric Analysis. Seventh Edition*. Upper Saddle River, Prentice Hall.
- Hamon, A. & Mesbah, M. (2002). Questionnaire reliability under the Rasch model. In: Mesbah, M.; Cole, B. F. & Lee, M. T. (Eds.). *Statistical Methods for Quality of Life Studies*. Dordrecht: Kluwer Academic Publishers, pp. 155-68. DOI: 10.1007/978-1-4757-3625-0\_13
- Heckman, J.J.; Humphries, J.E. and Veramendi, G. (2016). Returns to Education: The Causal Effects of Education on Earnings, Health and Smoking. *NBER Working Paper, May 2016 (22291)*.





- Huang, C. (2013). Gender differences in academic self-efficacy: a meta-analysis. *European Journal of Psychology of Education, 28*(1), 1–35. DOI: 10.1007/s10212-011-0097-y
- Kelderman, H. (1984). Loglinear Rasch model tests. *Psychometrika, 49*, 223-245. DOI: 10.1007/bf02294174
- Kreiner, S. (2003). *Introduction to DIGRAM*. Copenhagen: Department of Biostatistics, University of Copenhagen.
- Kreiner, S. (2007). Validity and objectivity. Reflections on the role and nature of Rasch Models. *Nordic Psychology, 59*, 268-298. DOI: 10.1027/1901-2276.59.3.268
- Kreiner, S. (2013). The Rasch model for dichotomous items. In Christensen, K. B.; Kreiner, S. & Mesbah, M. (Eds.) *Rasch models in health*. London: ISTE Ltd, Wiley, pp. 5–26. DOI: 10.1002/9781118574454.ch1
- Kreiner, S. & Christensen, K.B. (2002). Graphical Rasch models. In: Mesbah, M.; Cole, B. F. & Lee, M. T. (Eds.) *Statistical methods for quality of life studies*. Dordrecht: Kluwer Academic Publishers, pp. 187–203. DOI: 10.1007/978-1-4757-3625-0\_15
- Kreiner, S. & Christensen, K. B. (2004). Analysis of local dependence and multidimensionality in graphical loglinear Rasch models. *Communication in Statistics – Theory and Methods, 33*(6), 1239–1276, DOI: 10.1081/STA-120030148
- Kreiner, S. & Christensen, K. B. (2007). Validity and Objectivity in health-related Scales: Analysis by Graphical Loglinear Rasch models. In von Davier, M. & Carstensen, C. H. (Eds.) *Multivariate and Mixture Distribution Rasch Models*, New York, Springer, pp. 329-346. DOI: 10.1007/978-0-387-49839-3\_21
- Kreiner, S. & Christensen, K. B. (2013). Person Parameter Estimation and Measurement in Rasch Models. In Christensen, K. B.; Kreiner, S. & Mesbah, M. (Eds.) *Rasch models in health*. London: ISTE Ltd, Wiley, pp. 63–78. DOI: 10.1002/9781118574454.ch4
- Kreiner, S. & Nielsen, T. (2013). *Item analysis in DIGRAM 3.04. Part I: Guided tours. Research report 2013/06*. University of Copenhagen, Department of Public Health.
- Lent, R. W.; Miller, M. J.; Smith, P. E.; Watford, B. A.; Lim, R. H. & Hui, K. (2016). Social cognitive predictors of academic persistence and performance in engineering: Applicability across gender and race/ethnicity. *Journal of Vocational Behavior, 94*, 79-88. DOI: 10.1016/j.jvb.2016.02.012
- Lindstrøm, C. & Sharma, M. D. (2011). Self-efficacy of first year university physics students: Do gender and prior formal instruction in physics matter? *International Journal of Innovation in Science and Mathematics Education (formerly CAL-laborate International), 19*(2), 1–19.
- Luszczynska, A.; Gutiérrez-Doña, B. & Schwarzer, R. (2005). General self-efficacy in various domains of human functioning: Evidence from five countries. *International Journal of Psychology, 40*(2), 80-89. DOI: 10.1080/00207590444000041
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174, DOI: 10.1007/BF02296272
- Mellenbergh, G. J. (1989). Item Bias and Item Response Theory. *International Journal of Educational Research, 13*, 127-143. DOI: 10.1016/0883-0355(89)90002-5
- Meredith, W. (1993). *Psychometrika, 58*(4), 525-543. DOI:10.1007/BF02294825
- Mesbah, M. & Kreiner, S. (2013). The Rasch model for ordered polytomous items. In Christensen, K. B.; Kreiner, S. & Mesbah, M. (Eds.) *Rasch models in health*. London: ISTE Ltd, Wiley, pp. 27-42. DOI: 10.1002/9781118574454.ch2
- Multon, K. D.; Brown, S. D. & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of counseling psychology, 38*(1), 30-38. DOI: 10.1037/0022-0167.38.1.30
- Muris, P. (2002). Relationships between self-efficacy and symptoms of anxiety disorders and depression in a normal adolescent sample. *Personality and Individual Differences, 32*(2), 337–348. DOI: 10.1016/s01918869(01)00027-7
- Nielsen, T.; Dammeyer, J.; Vang, M. L. & Makransky, G. (2018). Gender fairness in self-efficacy? A Rasch-based validity study of the General Academic Self-efficacy scale (GASE). *Scandinavian Journal of Educational Research, 62*(5), 664-681. DOI: 10.1080/00313831.2017.1306796
- Nielsen, T. & Kreiner, S. (2013). Improving items that do not fit the Rasch model: exemplified with the physical functioning scale of the SF-36. *Annales de L’I.S.U.P. Publications de L’Institut de Statistique de L’Université de Paris, Numero Special, 57*(1-2), 91-108.



- Nielsen, T.; Makransky, G.; Vang, M. L. & Dammeyer, J. (2017). How specific is specific self-efficacy? A construct validity study using Rasch measurement models. *Studies in Educational Evaluation*, 57, 87-97. DOI: 10.1016/j.stueduc.2017.04.003
- Pajares, F. (1997). Current directions in self-efficacy research. In M. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement (Vol. 10)*. Greenwich, CT: JAI Press, pp. 1–49.
- Pintrich, P.R.; Smith, D.A.F.; Garcia, T. & McKeachie, W. J. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Technical Report No. 91-8-004. The Regents of The University of Michigan.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Danish Institute for Educational Research.
- Richardson, M.; Abraham, C. & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353-387. DOI: 10.1037/a0026838.
- Saleh D.; Camart N. & Romo L. (2017). Predictors of Stress in College Students. *Frontiers in Psychology*, 8(19). DOI: 10.3389/fpsyg.2017.00019
- Scherbaum, C. A.; Cohen-Charash, Y. & Kern, M. J. (2006). Measuring general self-efficacy: A comparison of three measures using item response theory. *Educational and Psychological Measurement*, 66(6), 1047–1063. DOI:10.1177/0013164406288171
- Scherer, R. & Siddiq, F. (2015). Revisiting teachers' computer self-efficacy: A differentiated view on gender differences. *Computers in Human Behavior*, 53, 48–57. DOI:10.1016/j.chb.2015.06.038
- Schunk, D. & Pajares, F. (2002). The development of academic self-efficacy. In A. Wigfield, & J. Eccles (Eds.), *Development of achievement motivation*. San Diego: Academic Press, pp. 15-31. DOI: 10.1016/b978-012750053-9/50003-6
- Schwarzer, R. & Jerusalem, M. (1995). Generalized self-efficacy scale. Measures in health psychology: A user's portfolio. *Causal and Control Beliefs*, 1, 35–37.
- Tahmassian, K. & Jalali Moghadam, N. (2011). Relationship between self-efficacy and symptoms of anxiety, Depression, worry and social avoidance in a normal sample of students. *Iranian Journal of Psychiatry and Behavioral Sciences*, 5(2), 91–98.
- Thomas, D. (2014). Factors that influence college completion intention of undergraduate students. *The Asia-Pacific Education Researcher*, 23(2), 225-235. DOI: 10.1007/s40299-013-0099-4
- van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of modern item response theory*. Springer-Verlag, New York. DOI: 10.1007/978-1-4757-2691-6
- van Herpen, S. G.; Meeuwisse, M.; Hofman, W. A.; Severiens, S. E. & Arends, L. R. (2017). Early predictors of first-year academic success at university: pre-university effort, pre-university self-efficacy, and pre-university reasons for attending university. *Educational Research and Evaluation*, 23(1-2), 52-72. DOI: 10.1080/13803611.2017.1301261
- Vuong, M.; Brown-Welty, S. & Tracz, S. (2010). The effects of self-efficacy on academic success of first-generation college sophomore students. *Journal of college student development*, 51(1), 50-64. DOI: 10.1353/csd.0.0109
- Watt, H. M. G.; Ehrich, J.; Stewart, S.E.; Snell, T.; Bucich, M.; Jacobs, N.; Furlonger, B. & English, D. (2019). Development of the Psychologist and Counsellor Self-Efficacy Scale. *Higher Education, Skills and Work-Based Learning*, (Early cite). DOI: 10.1108/HESWBL-07-2018-0069
- Williams, B. W.; Kessler, H. A. & Williams, M. V. (2014). Relationship among practice change, motivation, and self-efficacy. *Journal of continuing education in the health professions*, 34(S1), 5-10. DOI: 10.1002/chp.21235.
- Yong, F. L. (2010). A study on the self-efficacy and expectancy for success of pre-university students. *European Journal of Social Sciences*, 13(4), 514-524.
- Young, A. M., Wendel, P. J., Esson, J. M., & Plank, K. M. (2018). Motivational decline and recovery in higher education STEM courses. *International Journal of Science Education*, 40(9), 1016-1033. DOI: 10.1080/09500693.2018.1460773



Zimmerman, B. J.; Bandura, A. & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American educational research journal*, 29(3), 663-676. DOI: 10.2307/1163261

**Appendix**  
**The self-efficacy scales**

Table A1.

*Items and response scales of the Danish and English versions of the Pre-Academic Learning Self-efficacy (PAL-SE) and Pre-Academic Exam Self-efficacy (PAE-SE) scales adapted from Nielsen et al. (2018)*

Items	Danish version	English version
PAL-SE1	Jeg er sikker på, at jeg kan forstå det sværeste litteratur i de første fag på min uddannelse	I'm certain I can understand the most difficult material presented in the readings for the first courses in my program.
PAL-SE2	Jeg er sikker på, at jeg kan forstå de grundlæggende begreber, der undervises i i de første fag på min uddannelse	I'm confident I can understand the basic concepts taught in the first courses in my program.
PAL-SE3	Jeg er sikker på, at jeg kan forstå de mest komplekse emner, jeg bliver præsenteret for i de første fag på min uddannelse	I'm confident I can understand the most complex material presented in the first courses in my program.
PAL-SE4	Jeg er sikker på, at jeg kan mestre de færdigheder, der undervises i i de første fag på min uddannelse	I'm certain I can master the skills being taught in the first courses in my program.
PAE-SE1	Jeg tror, jeg vil få rigtig gode karakterer i mine første fag	I believe I will receive an excellent grade in the first courses in my program.
PAE-SE2	Jeg er sikker på, at jeg vil klare mig godt i opgaver og til eksamen i de første fag på min uddannelse	I'm confident I can do an excellent job on the assignments and tests in the first courses in my program.
PAE-SE3	Jeg forventer, at jeg klarer mig godt i de første fag på min uddannelse	I expect to do well in the first courses in my program.
PAE-SE4	Hvis man tager sværhedsgraden af de første fag, underviserne og mine egne evner i betragtning, tror jeg, at jeg vil klare mig godt i mine første fag	Considering the difficulty of this course, the teacher, and my skills, I think I will do well in the first courses in my program.
Response scale	1 = slet ikke, 2 = I ringe grad, 3 = I nogen grad, 4 = I høj grad, 5 = helt perfekt	1 = not at all, 2 = to a small degree, 3 = to some degree, 4 = to a large degree, 5 = perfectly

*Notes.* Question prompt for the items was: How well do the following statements describe you as a future student?

Items were administered mixed with items from other scales. Here they are shown by subscale, not the order they were administered in.



**Post-hoc step-wise analyses of pairwise collapsibility of the nine nominal degree program categories**

In the stepwise analyses, the two categories with the largest categories were collapsed if the p-value was larger than the critical p-value. Critical p-values were determined while controlling the false discovery rate at 0.05 relative to the total number of tests performed at the end of each step, using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995; Kreiner & Nielsen, 2013). The steps and decisions in the analysis of the observed mean PAL-SE scores and the DIF-adjusted mean PAL-SE scores are shown in Table S-Æ, while the collapsed groups of non-different degree programs are provided in the results section.

Table A2.

*Steps and decisions in the analyses of pairwise collapsibility*

Degree program categories compared	Critical $p$	Largest $p$	Decision
<b>Observed scores</b>			
2 & 5	0.0222	0.9358	Collapsed
1 & 3	0.0214	0.8667	Collapsed
2+5 & 8	0.0225	0.8297	Collapsed
1+3 & 6	0.0239	0.7398	Collapsed
7 & 9	0.0250	0.7377	Collapsed
4 & 7+9	0.0259	0.0247	Collapse not possible
<b>DIF-adjusted scores</b>			
2 & 5	0.0250	0.9215	Collapsed
3 & 6	0.0245	0.9106	Collapsed
7 & 9	0.0250	0.8432	Collapsed
1 & 3+6	0.0262	0.8375	Collapsed
2+5 & 8	0.0264	0.6074	Collapsed
1+3+6 & 2+5+8	0.0267	0.1125	Collapsed
4 & 7+9	0.0269	0.0445	Collapsed
1+3+6+2+5+8 & 4+7+9	0.0272	0.0000	Collapse not possible

Notes: 1 = medicine, 2 = psychology, 3 = political science, 4 = economy, 5 = law, 6 = biology, 7 = computer science, 8 = Danish language, 9 = history.



**Conversion tables**

Table A3 – part 1.

*Weighted maximum likelihood estimates of person parameters and DIF-equating of the sum score for the Pre-Academic Learning Self-Efficacy scale*

sum score	p.param	<u>Medicine</u>			DIF-adjusted score	<u>Psychology</u>				DIF-adjusted score	<u>Political science</u>			
		p.param SEM	bias	score SEM		p.param	p.param SEM	bias	score SEM		p.param	p.param SEM	bias	score SEM
4.00	-6.226	1.027	0.573	0.55	4.00	-6.052	1.015	0.570	0.56	4.00	-7.389	1.204	0.641	0.49
5.00	-4.560	1.040	0.072	0.88	5.09	-4.411	1.014	0.066	0.90	4.62	-5.257	1.230	0.094	0.74
6.00	-3.519	0.902	-0.003	1.12	6.10	-3.436	0.886	-0.004	1.14	5.68	-3.745	1.010	-0.018	1.02
7.00	-2.870	0.803	-0.010	1.29	7.10	-2.816	0.793	-0.009	1.31	6.75	-3.002	0.865	-0.019	1.23
8.00	-2.371	0.744	-0.002	1.38	8.10	-2.327	0.739	-0.001	1.39	7.78	-2.475	0.780	-0.004	1.34
9.00	-1.902	0.714	0.003	1.41	9.09	-1.860	0.714	0.004	1.41	8.83	-1.985	0.734	0.004	1.38
10.00	-1.415	0.718	0.003	1.40	10.10	-1.371	0.724	0.005	1.39	9.88	-1.473	0.730	0.003	1.37
11.00	-0.918	0.751	0.007	1.36	11.11	-0.863	0.761	0.009	1.34	10.92	-0.954	0.762	0.006	1.34
12.00	-0.378	0.796	0.015	1.28	12.13	-0.297	0.806	0.014	1.25	11.95	-0.408	0.804	0.015	1.27
13.00	0.334	0.838	0.007	1.17	13.14	0.451	0.850	0.002	1.15	12.97	0.299	0.845	0.009	1.16
14.00	1.195	0.864	-0.017	1.15	14.15	1.306	0.875	-0.014	1.14	13.99	1.189	0.869	-0.017	1.14
15.00	1.909	0.869	-0.008	1.18	15.18	2.019	0.877	-0.002	1.17	15.00	1.909	0.872	-0.008	1.18
16.00	2.578	0.855	0.004	1.17	16.21	2.749	0.853	0.000	1.15	16.00	2.582	0.856	0.004	1.17
17.00	3.369	0.847	-0.016	1.14	17.20	3.554	0.844	-0.025	1.16	17.00	3.373	0.847	-0.016	1.14
18.00	4.101	0.864	-0.035	1.13	18.15	4.210	0.851	-0.033	1.16	18.00	4.104	0.864	-0.036	1.13
19.00	4.770	0.831	-0.083	1.03	19.08	4.846	0.814	-0.078	1.05	19.00	4.770	0.831	-0.083	1.03
20.00	5.781	0.766	-0.449	0.70	20.00	5.857	0.761	-0.450	0.70	20.00	5.783	0.766	-0.449	0.70

Continues....





Table A3 – part 2.

*Weighted maximum likelihood estimates of person parameters and DIF-equating of the sum score for the Pre-Academic Learning Self-Efficacy scale*

<u>Economy</u>					<u>Law</u>					<u>Biology</u>				
equated score	p.param	p.param SEM	bias	score SEM	equated score	p.param	p.param SEM	Bias	score SEM	equated score	p.param	p.param SEM	bias	score SEM
4.00	-6.908	1.649	1.571	0.31	4.00	-5.957	1.013	0.570	0.56	4.00	-5.813	0.989	0.564	0.57
4.20	-5.504	1.005	0.572	0.56	5.15	-4.320	1.000	0.063	0.91	5.19	-4.225	0.944	0.054	0.97
5.49	-3.860	0.953	0.041	0.97	6.17	-3.382	0.874	-0.005	1.16	6.12	-3.420	0.829	-0.008	1.24
6.60	-3.093	0.862	-0.005	1.20	7.15	-2.788	0.783	-0.009	1.33	6.97	-2.915	0.755	-0.007	1.39
7.65	-2.550	0.793	0.003	1.31	8.13	-2.314	0.730	-0.000	1.41	7.83	-2.493	0.715	0.005	1.45
8.75	-2.033	0.755	0.009	1.33	9.10	-1.859	0.708	0.005	1.42	8.76	-2.060	0.700	0.011	1.43
9.88	-1.472	0.756	0.004	1.32	10.09	-1.378	0.718	0.005	1.40	9.77	-1.554	0.716	0.004	1.38
11.01	-0.905	0.789	0.006	1.29	11.10	-0.873	0.756	0.009	1.35	10.82	-1.007	0.761	0.002	1.33
12.09	-0.303	0.823	0.013	1.22	12.11	-0.308	0.801	0.014	1.26	11.86	-0.448	0.805	0.010	1.27
13.13	0.465	0.849	0.001	1.15	13.11	0.427	0.841	0.002	1.16	12.88	0.245	0.838	0.007	1.17
14.12	1.286	0.860	-0.014	1.16	14.09	1.262	0.860	-0.015	1.16	13.87	1.098	0.856	-0.015	1.15
15.10	1.968	0.858	-0.003	1.19	15.08	1.956	0.862	-0.005	1.19	14.85	1.807	0.864	-0.008	1.19
16.09	2.644	0.845	0.004	1.17	16.07	2.634	0.848	0.003	1.17	15.84	2.450	0.859	0.008	1.17
17.07	3.430	0.842	-0.019	1.15	17.06	3.424	0.843	-0.018	1.15	16.86	3.241	0.855	-0.012	1.12
18.05	4.133	0.857	-0.034	1.15	18.04	4.129	0.858	-0.034	1.14	17.91	4.044	0.880	-0.041	1.11
19.02	4.789	0.825	-0.082	1.04	19.02	4.787	0.825	-0.082	1.03	18.96	4.736	0.848	-0.089	1.02
20.00	5.800	0.764	-0.449	0.70	20.00	5.797	0.764	-0.449	0.70	20.00	5.750	0.771	-0.449	0.69

Continues....



Table A3-part 3.

*Weighted maximum likelihood estimates of person parameters and DIF-equating of the sum score for the Pre-Academic Learning Self-Efficacy scale*

<u>Computer science</u>					<u>Danish language</u>					<u>History</u>				
equated score	p.param	p.param SEM	bias	score SEM	equated score	p.param	p.param SEM	Bias	score SEM	equated score	p.param	p.param SEM	bias	score SEM
4.00	-6.061	1.013	0.568	0.56	4.00	-6.036	0.992	0.561	0.57	4.00	-6.162	0.972	0.550	0.58
5.08	-4.424	1.010	0.066	0.90	5.06	-4.452	0.986	0.064	0.92	4.95	-4.636	0.984	0.069	0.92
6.07	-3.460	0.881	-0.004	1.15	6.00	-3.530	0.867	-0.005	1.17	5.83	-3.696	0.876	-0.001	1.14
7.05	-2.850	0.789	-0.009	1.31	6.91	-2.945	0.783	-0.008	1.33	6.71	-3.081	0.797	-0.006	1.30
8.02	-2.369	0.736	0.000	1.40	7.84	-2.476	0.738	0.003	1.40	7.65	-2.587	0.757	0.004	1.36
9.00	-1.908	0.711	0.005	1.41	8.83	-2.010	0.719	0.008	1.39	8.68	-2.089	0.740	0.010	1.35
10.01	-1.418	0.718	0.004	1.39	9.88	-1.487	0.733	0.004	1.36	9.80	-1.520	0.753	0.002	1.31
11.02	-0.908	0.753	0.007	1.35	10.95	-0.940	0.775	0.005	1.32	10.94	-0.939	0.792	0.003	1.29
12.02	-0.355	0.794	0.013	1.27	12.02	-0.361	0.819	0.013	1.24	12.04	-0.338	0.827	0.012	1.22
13.01	0.348	0.830	0.004	1.18	13.06	0.388	0.856	0.005	1.14	13.08	0.424	0.852	0.003	1.15
13.97	1.169	0.848	-0.015	1.17	14.09	1.267	0.878	-0.015	1.14	14.08	1.260	0.862	-0.014	1.16
14.93	1.856	0.856	-0.007	1.20	15.12	1.987	0.877	-0.005	1.17	15.06	1.949	0.861	-0.004	1.19
15.91	2.500	0.851	0.007	1.18	16.14	2.697	0.854	0.001	1.15	16.06	2.621	0.847	0.004	1.17
16.92	3.286	0.849	-0.013	1.13	17.14	3.498	0.845	-0.022	1.16	17.04	3.407	0.843	-0.018	1.15
17.94	4.062	0.873	-0.039	1.12	18.10	4.174	0.854	-0.033	1.16	18.03	4.121	0.859	-0.035	1.14
18.97	4.747	0.841	-0.088	1.02	19.05	4.818	0.819	-0.079	1.04	19.01	4.781	0.826	-0.083	1.03
20.00	5.757	0.769	-0.449	0.70	20.00	5.829	0.762	-0.449	0.70	20.00	5.790	0.764	-0.449	0.70

Notes. p.param = Person parameter. SEM = standard error of measurement



Table A4.

*Weighted maximum likelihood estimates of person parameters for the Pre-Academic Exam Self-Efficacy scale*

Sum score	p.param	p.param SEM	bias	score SEM
4.00	-4.380	0.553	0.350	0.94
5.00	-3.781	0.563	0.088	1.36
6.00	-3.440	0.607	0.036	1.54
7.00	-3.136	0.625	0.026	1.59
8.00	-2.780	0.638	0.019	1.54
9.00	-2.304	0.690	0.005	1.45
10.00	-1.828	0.772	0.015	1.39
11.00	-1.335	0.830	0.039	1.26
12.00	-0.430	0.869	-0.013	1.04
13.00	0.637	0.942	-0.061	1.14
14.00	1.253	0.947	-0.004	1.17
15.00	1.942	0.944	0.040	1.09
16.00	3.096	0.921	-0.014	1.02
17.00	3.979	0.910	-0.039	1.14
18.00	4.579	0.911	-0.016	1.15
19.00	5.263	0.901	-0.042	1.00
20.00	6.627	0.890	-0.519	0.61

Notes. p.param = Person parameter. SEM = standard error of measurement



### Item maps

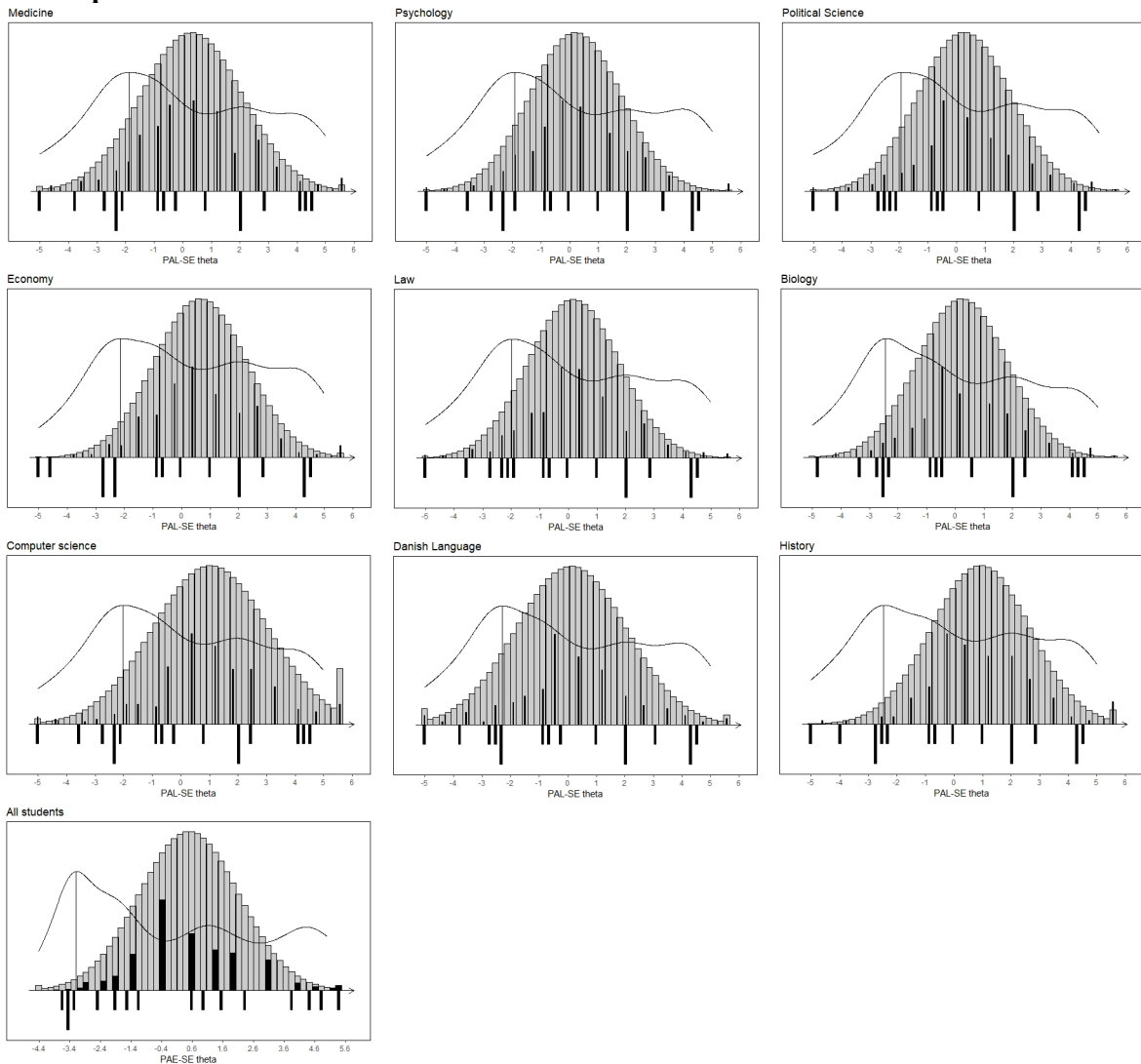


Figure A1. Item maps with distributions of person parameter locations and information curve above item threshold locations.

Notes. Person parameters are weighted maximum likelihood estimates and illustrate the distribution of these for the study sample (black bars above the line) and for the population under the assumption of normality (grey bars above the line), as well as the information curve, relative to the distribution of the item thresholds (black bars below the line). For the PAL-SE scale item maps are shown for students admitted to each of the nine degree programs, as the scale functioned differentially relative to degree program (top nine graphs). For the PAE-SE scale a single item map is shown for all students (bottom graph).



Table A5.

*Association between Pre-Academic Learning Self-Efficacy (independent variable) and probability to drop out within the first half year (dependent variable), using linear probability models with different covariates*

	Model 1: Bivariate LPM (n= 2.340)	Model 2: LPM regression including program as covariate (n= 2.340)	Model 3: LPM regression including program and individual covariates (n= 2.195)
PAL-SE	-0.06 ** (-2.67)	-0.07 ** (-3.15)	-.07 ** (-3.13)
Program indicator:			
Medicine, KU		Ref.	Ref.
Computer science, KU		0.07** (2.78)	0.06* (2.18)
Biology, KU		0.08*** (3.27)	0.06* (2.47)
Danish Language, KU		0.01 (0.48)	-0.00 (-0.12)
History, KU		0.03 (1.24)	0.02 (0.87)
Psychology, KU		0.01 (0.41)	0.00 (0.42)
Law, KU		0.04 (1.88)	0.03 (1.43)
Political Science, KU		0.05* (2.07)	0.05* (2.15)
Economy, KU		0.06** (2.90)	0.05* (2.23)
Medicine, AU		0.00 (0.02)	-0.00 (-0,17)
Biology, AU		0.04 (1.20)	0.02 (0,74)
Computer Science, AU		0.08** (2.90)	0.05 (1.63)
History, AU		0.13*** (4.24)	0.10*** (3,22)
Danish Language, AU		0.05 (1.72)	0.04 (1,40)
Law, AU		0.03 (1.66)	0.02 (0,89)
Psychology, AU		-0.00 (-0.13)	-0,00 (-0,06)
Political Science, AU		0.03 (1.42)	0.02 (0,90)
Economy, AU		0.10*** (3.96)	0.08** (2.97)
Priority <sup>a</sup>			0.02 (1.58)
Sex			-0.00 (-0.36)
GPA			0.00 (-1.49)
Age (group indicator)			-0.00 (-0.61)
Parents income in DKK			0.00 (0.98)
Parents highest educational level:			
No registered education			Ref.
Primary school or high school			-0.04 (-0.67)
Vocational education			-0.05 (-0.95)
Higher education (short)			-0.06 (-1.03)
Higher education (medium)			-0.05 (-0.85)
Higher education (long or ph.d.)			-0.06 (-1.12)
Immigrant			0.02 (0.47)
Descendant			-0.01 (-0.36)
Cons	.07 *** (5.55)	.04 (1.92)	0.05 (0.57)

Notes: values are  $\beta$ -values. t-values are reported in parenthesis. \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$ . <sup>a</sup> in application for program. GPA = Grade point average (gymnasium/high school). All results were confirmed in models using robust standard errors. This made p-values a bit larger, but results remained significant. In the bivariate model the self-efficacy p-value changed from 0.008 to 0.016. In the model with program as covariate the p-value changed from 0.002 to 0.005, and in the model with individual covariates the p-value changed from 0.002 to 0.006.