Predicting University Dropouts: Evidence on the Value of Student Expectations and Motivation*

Thomas F. Epper[a], Kristoffer Ibsen[b,c], Alexander Koch[b,c,d], and Julia Nafziger[b,e]

[a]CNRS, IESEG School of Management, Univ. Lille, UMR 9221 – LEM – Lille Economie Management, F-59000 Lille, France

[b]Aarhus University, Department of Economics and Business Economics, Universitetsbyen 51, , 8000 Aarhus, Denmark

[c]TrygFonden's Centre for Child Research, Aarhus University

[d]IZA

[e]CEPR

Abstract

University dropout is costly, making it a policy priority to identify factors that predict dropout. Using a survey experiment with incoming first-year students linked to long-run administrative outcomes, we assess which information improves dropout prediction beyond standard university records. A small number of targeted, study-specific survey items—especially motivation and expectations about degree completion—substantially improve predictive performance. By contrast, widely used measures of general preferences and traits (such as grit and self-control) add little incremental value—a result that we qualitatively replicate in a large population. Our findings suggest inexpensive, scalable ways to improve dropout predictions.

JEL codes: I23, D91

Keywords: dropout, non-cognitive skills, motivation, economic preferences, beliefs, education, machine learning.

*We thank Eric Bettinger, Felix Chopra, Leonie Gerhards, and Erik Sørensen, as well as numerous conference and seminar audiences for helpful comments. Financial support from the Aarhus University Research Foundation (AU IDEAS 2011 grant AUFF-E-201-FLS-1-17) and the Danish Independent Research Fund (grants 12-124835 and 8091-00069B) is gratefully acknowledged. Thomas Epper gratefully acknowledges the financial support received from the Métropole Européenne de Lille (MEL).

1 Introduction

Dropout from higher education remains a pervasive and costly issue. In the United States, roughly 40 percent of students fail to complete a degree within six years of enrollment (de Brey et al., 2021), and dropout rates across OECD countries average around 32 percent (OECD, 2022). While some students may benefit from dropping out—particularly when it reflects a well-informed match correction—policymakers and educational institutions are primarily concerned with the broader costs of dropout for individuals,[1] higher education institutions,[2] and society.[3] Identifying factors that predict dropout is therefore important for informing policies aimed at reducing dropout.

Previous research has singled out past academic performance as the most important predictor of educational achievement and dropout (e.g., Rothstein, 2004; Cyrenne and Chan, 2012; Bettinger et al., 2013). Over the last two decades, non-cognitive skills have also been recognized as being as predictive of lifetime outcomes as IQ and various achievement tests (Heckman et al., 2006; Borghans et al., 2008; Almlund et al., 2011). This evidence has motivated an increasing interest in collecting and using measures of non-cognitive skills at entry to identify students at risk.

However, it remains unclear which types of measures – among non-cognitive skills, socioeconomic background, and indicators of students' motivation and expectations

[1]Dropout without any formal university degree is associated with worse economic outcomes, including lower lifetime earnings (Zimmerman, 2014; Ost et al., 2018); and any kind of dropout is associated with higher debt burden (e.g. Chatterjee and Ionescu, 2012).

[2]Universities are vulnerable to students dropping out of a program or leaving the institution without a degree because, next to fees paid by students, public funding depends on the number of enrolled students and often also graduation rates (for an overview Kalamova et al., 2020). Dropout also affects the reputation of study programs and institutions, and causes mismatches between student body and capacity.

[3]Obtaining a higher education degree is associated with larger tax payments, lower crime, better health, and lower mortality (e.g., Lange and Topel, 2006; Oreopoulos and Petronijevic, 2013; Heckman et al., 2018) as well as other non-pecuniary benefits (e.g. Oreopoulos and Salvanes, 2011) and may lower income inequality (Hershbein et al., 2020). While on average returns to a university degree are high, this is not necessarily the case for all (e.g., Heckman et al., 2018; Webber, 2016) and students might only learn about mismatches after experiencing whether a university program aligns with their expectations (e.g., Stinebrickner and Stinebrickner, 2014a). While allowing for some degree of experimentation might be optimal (e.g., Stinebrickner and Stinebrickner, 2014b), information frictions account for a non-trivial share of dropout (Arcidiacono et al., 2025).

– are actually useful for predicting dropout once the information already available in administrative records, especially prior grades, is taken into account. This question matters for two reasons relevant for policy and practice. First, university administrations already observe some baseline information for incoming students; additional data collection is only justified if it adds meaningful predictive content beyond these readily available inputs. Second, many non-cognitive batteries are lengthy and can raise privacy concerns or suffer from low response rates, which limits their value.

We address the question of which types of measures help to predict dropout by combining an entry survey experiment for 616 first-year students enrolled in business and social science degrees at Aarhus University[4] with Danish national administrative records that allow us to follow educational outcomes for eight years after enrollment. The dataset includes questionnaire measures and incentivized tasks capturing economic preferences and behavioral traits, as well as study-related motivation, expectations, and socioeconomic characteristics. We assess which of these measures improve the prediction of dropout beyond the information typically available to a university administration. Throughout, we use the term *non-cognitive skills* broadly to encompass preferences, behavioral biases, and other characteristics, to distinguish them from cognitive skills.

Our analysis builds on data typically available to university administrations, including gender, age, high-school grade point average (GPA), priority ranking of programs, and, of course, the program enrolled in. We refer to this as the Basic information set and extend it in three different ways: First, the information set Noncog augments Basic with measures of non-cognitive skills, encompassing (i) a broad set of preferences and traits (such as self-control, grit, overconfidence, loss aversion, narrow bracketing, competitiveness, strategic thinking, and cognitive reflection), and (ii) self-regulation strategies related to studying. Second, the information set Expect augments Basic with a range of survey questions on students'

[4]Aarhus University is Denmark's second-largest university and consistently ranks among the top 150 universities worldwide (Rector's Office, 2024).

expectations about their future earnings and study completion, reasons for their major choice, and their motivation for the chosen studies. Third, the information set BACKGR augments BASIC with survey data on students' socioeconomic background and their financial situation. Finally, ALL combines all information sets.

For each of these information sets, we build predictive models and compare their performance in predicting dropout. In addition, we identify the importance of individual variables that are included in the broader sets. We rely on machine learning techniques because they capture non-linear relationships and interactions among variables and provide robust predictions in settings like ours, where the ratio of predictors to observations is high (e.g., Mullainathan and Spiess, 2017).

Our main analysis focuses on dropout from the bachelor program. Yet, unlike most of the previous literature, the link between university administrative records and Danish national administrative data allows us also to disentangle three major categories of dropout and test robustness of our results to different dropout definitions (see, e.g., Tinto, 2012): *program dropout* (dropping the degree program in which the student originally enrolled), *institution dropout* (leaving the university originally enrolled in without a degree), and *system dropout* (leaving the higher education system without obtaining any bachelor's degree).

Across these alternative dropout definitions, the overall findings are that (i) previous academic performance (high school GPA) is by far the most important predictor of dropout, (ii) simple, self-reported measures of study expectations and motivation add substantial predictive value—most clearly for program dropout and, for institution dropout, mainly relative to models that use GPA alone, (iii) study-specific self-regulation strategies provide complementary information for distinguishing students who are least likely to persist in their studies, and (iv) behavioral traits and preferences (such as, for example, self-control, grit, competitiveness or loss aversion) add little incremental value for predicting dropout once basic administrative information is included.

This last finding is surprising for two reasons. First, one might expect general measures of preferences and traits to subsume context-specific questions about motivation, expectations and study habits. Second, the literature documents robust correlations between traits such as grit or self-control and academic performance, especially grades. Our data confirm these correlations for both prior grades and university grades. Yet, these measures show little correlation with dropout itself. This contrast highlights an important distinction: predicting grades and predicting dropout are different challenges, likely because dropout reflects additional margins—fit, constraints, and persistence—that are not well summarized by broad trait batteries.

We also address two potential concerns about interpretation. One concern is that administrative information may already absorb much of the predictive content of preferences and trait measures, for instance if traits are largely reflected in prior grades or other readily observed inputs. Another is selection: survey participation and university enrollment may generate a sample in which broad non-cognitive batteries have limited variation or limited relevance. We find little evidence that any of these mechanisms account for the main patterns in our results. In particular, we replicate the qualitative finding that broad preference and trait measures, such as grit and self-control, add little predictive value in an additional, much larger dataset from a broader population, using a related educational completion outcome.

Taken together, the practical implication is that predicting dropout may be easier, and less intrusive, than is often assumed. Institutions can largely rely on existing administrative information and gain additional accuracy by adding a small number of short, targeted questions on motivation and expectations about completion, and possibly few study-specific self-regulation questions. Because such questions can be collected during the routine communication that institutions already have with incoming students, this approach is scalable at low cost and avoids privacy concerns associated with lengthy batteries that collect more sensitive information.

Finally, to shed light on why a simple motivation measure is predictive of dropout, we draw on an additional dataset linking motivation to study time. Many students invest fewer hours than programs typically require, and study effort is correlated with self-reported motivation. A plausible mechanism consistent with our findings is that students who report low initial motivation invest fewer study hours, which can generate early knowledge gaps and ultimately raise dropout risk. This perspective aligns with the view that reducing dropout may require interventions that effectively increase student effort (Oreopoulos and Petronijevic, 2019).

The remainder of the paper is organized as follows. Next we discuss related literature. Section 2 describes the institutional setting. Section 3 presents the data and experimental design. Section 4 outlines the empirical strategy, Section 5 presents the results, Section 6 discusses mechanisms and robustness, and Section 7 concludes.

Related literature We contribute to the burgeoning literature investigating the role of non-cognitive skills in predicting educational outcomes, and in particular dropout. Our machine learning approach allows us to systematically study the predictive power of a broad set of non-cognitive skills, motivation and expectations that may relate to university dropout.[5] Previous studies either did not take past academic performance into account (Burks et al., 2015), focused on different educational settings (Backes-Gellner et al., 2021; Almås et al., 2016a) or on a narrow set of non-cognitive skills, namely measures of grit, conscientiousness, and locus of control (Caviglia-Harris and Maier, 2020; Saltiel, 2020; Duckworth et al., 2019). By contrast, we jointly consider a broad battery of preference and trait measures together with study-specific motivation and expectation measures, while drawing on detailed administrative information.

Our results are in line with Caviglia-Harris and Maier (2020), who find that non-cognitive skills do not relate to university dropout, but to GPA. In contrast, Saltiel (2020) finds a relation between dropout and non-cognitive skills; and Duckworth

[5]Machine learning has also been used to predict educational outcomes, see Peña-Ayala (2014) for an overview of this literature and Berens et al. (2018) for a recent example. However, non-cognitive skills are typically not incorporated in this literature.

et al. (2019) find a relation between grit and timely graduation for students in a highly selective elite military academy.

Our research is also related to the literature that examines the impact of past education performance and non-cognitive skills on academic outcomes other than dropout (e.g., Nichols et al., 2018; Beattie et al., 2019; Yu et al., 2020; Bjerre-Nielsen et al., 2021). Bjerre-Nielsen et al. (2021) rely on several data sources to examine which data best predict grades at a Danish university. Next to administrative data, they rely on a range of survey measures, such as the Big Five, locus of control, self-esteem, and self-efficacy, as well as cell phone network data. They find that adding the behavioral survey measures and network data does not improve predictability relative to using only the administrative data. Yu et al. (2020) follow a similar approach, but do not rely on validated psychological scales or elicited preferences for the non-cognitive skills. They also find little support for the predictive power of the survey measures (or network data). Beattie et al. (2019) focus only on the best and worst performing students. For these selected student samples they find that low effort provision and procrastination are good predictors of low academic performance. Nichols et al. (2018) follow a similar strategy as Beattie et al. (2019), but consider a more restrictive set of non-cognitive skills (Big Five, grit, and locus of control).

Taken together, this literature suggests that while non-cognitive skills are often correlated with academic performance, their incremental value for predicting dropout––especially beyond prior grades and administrative information––remains unclear. Our paper directly addresses this gap in a unified predictive framework

2 Institutional background

Three key aspects of the Danish higher education system warrant attention. First, all Danish universities are public and offer tuition-free education to students from European Union (EU) member states. Danish students additionally receive around $1,000 USD in monthly financial support through the State Educational Grant (SEG).[6]

Second, admission capacities in Danish universities are constrained. There are two admission systems, 'quota 1' and 'quota 2', with the former relying solely on high school GPA, while the latter also considers additional criteria. Universities admit around 10 percent of students via the 'quota 2' system. The application process is centralized with students ranking their top eight institution-degree pairs (such as "Economics at Aarhus University"). After the application deadline, candidates are sorted according to their GPA and admitted to the highest-ranked program on their list that is still available, if any. This is the sole offer that applicants receive, and to enroll within the current academic year, they must accept this offer.

Third, most bachelor programs in Danish universities last three years and consist of two semesters each year. Students enroll full-time and typically follow a fixed course structure for the first two years. Transferring to another study program requires going through the centralized annual application procedure, which limits mid-year switching due to the structured curriculum.

3 Data and experimental design

3.1 Data

3.1.1 Sample

All 2,747 first-year students at the Faculty of Business and Social Sciences at Aarhus University in Fall 2013 were invited to participate in our survey experiment.[7] Stu-

[6]The SEG is limited to 70 months of support. Students must not fall behind in their studies by more than 12 months and face a cap on how much they can earn besides the SEG. EU students are eligible for the SEG if they work more than 40 hours per month.

[7]Our target sample matches the actual cohort with the exception of a few students who were enrolled after we were given the list of incoming students by the university. Our list included a few

dents were enrolled in Economics, a variety of Business programs, Law, Political Science, or Psychology. A total of 616 students completed the survey (a 22.42 percent response rate). In Appendix A.6, we present data describing selection into the survey. While we observe certain differences, such as participants being somewhat less likely to drop out and more likely to complete a master's degree, these patterns align with previous studies (e.g., Epper et al., 2022; Hvidberg et al., 2023), which find that individuals who choose to participate can differ from those who do not. Accordingly, we employ several approaches to test the robustness of our results to selection, as described in Section 6.

3.1.2 Student records and national administrative data

We link the data from the survey experiment to two other administrative data sources. First, student records from Aarhus University provide us with students' semester-by-semester enrollment status and the set of background variables that a university administration would typically have access to for incoming students, including gender, age, high school grade point average (GPA), high school type (or foreign degree), high school graduation date, and whether the study program was ranked highest by the student in the university application form or was assigned a lower rank. We also construct program fixed effects to account for unobserved differences across study programs (for example, due to different GPA admission cutoffs). Second, national administrative data from Statistics Denmark allow us to track students' educational trajectories for more than eight years after enrollment, until the end of 2021, even if they drop out of their initial study program.

3.1.3 Outcomes

We consider the three major categories of dropout as our outcome variables (see, e.g., Tinto, 2012). First, *program dropout* occurs if a student leaves the study pro-

students who then did not end up starting their studies. These are excluded.

gram originally enrolled in without obtaining a bachelor's degree.[8] This is our main outcome variable. Second, *institution dropout* occurs if a student leaves Aarhus University without obtaining a bachelor's degree. Finally, *system dropout* occurs if a student does not complete any bachelor's program of at least three years' duration at a Danish university or a university college. Because institution and system dropout are less frequent, they are secondary outcome variables.

3.2 Experimental design

The survey experiment consisted of several incentivized tasks and a questionnaire. The full instructions of the survey are reproduced in Appendix A.9.[9] We first motivate the different measures collected and then describe the procedure used to elicit them.

3.2.1 Motivation for the different measures

The set Noncog consists of several measures of non-cognitive skills. The first subset of non-cognitive skills broadly relate to perseverance, including psychological scales on grit (Duckworth et al., 2007) and self-control (Tangney et al., 2004) that are important correlates of academic outcomes in different educational settings (Shoda et al., 1990; Duckworth and Seligman, 2005; Duckworth et al., 2007, 2012; Eskreis-Winkler et al., 2014; Golsteyn et al., 2014; Figlio et al., 2019). Because studying often requires costly effort to secure gains that lie in the future, we measure willingness to exert effort in a tedious real-effort task that generates payment several weeks later. To manage self-control problems in studying, students may employ self-regulation strategies, such as goal setting (see, e.g., Locke and Latham, 1990; Clark et al., 2020), motivational bracketing (Koch and Nafziger, 2016, 2019), or hard commitment de-

[8] To transfer to another study program (even within the same university) the student must again go through the annual centralized application process described in Section 2.

[9] The study was notified to the Central Denmark Region Committees on Health Research Ethics (the IRB in charge), who did not require it to undergo ethical review (Reference number 1-10-72-20-13 85/2013).

vices (Bryan et al., 2010).

Our second subset of non-cognitive skills encompasses several preference measures. First, we elicit participants' risk preferences, and, more specifically, their loss aversion. Karle et al. (2022) observe that loss aversion relates to worse exam strategies and outcomes. Since repeated low exam performance may lead to dropout, more loss-averse students may be more likely to withdraw. Second, we elicit the willingness of students to select into competition (Niederle and Vesterlund, 2007). Participation in competition not only reflects risk preferences and confidence (Gillen et al., 2019; Van Veldhuizen, 2022), but is also predictive of exam performance (Ors et al., 2013; Pekkarinen, 2014; Jurajda and Münich, 2011), education and occupation choices (Buser et al., 2014, 2017, 2022, 2024; Almås et al., 2016b), as well as income or income expectations (Reuben et al., 2017; Buser et al., 2024).

Our third subset of non-cognitive skills includes behavioral biases that induce suboptimal choices. First, we elicit a measure of overconfidence. Overconfidence may lead students to exert suboptimal study effort or to choose a poorly matched major. Second, we include measures of narrow bracketing due to cognitive limitations, which may lead to suboptimal decisions (e.g., Rabin and Weizsäcker, 2009). For example, individuals who see a bad grade in isolation from the rest of their study performance, might be more demotivated and hence more likely to drop out.

Our fourth subset of non-cognitive skills includes measures of strategic thinking and cognitive reflection. Fe et al. (2022) observe a relation between strategic sophistication and educational outcomes. Cognitive reflection (Frederick, 2005) is related to a variety of outcomes and preferences (e.g., Taylor, 2020). Students who score higher in these domains may better anticipate the consequences of their choices and thus be less likely to drop out.

The information set Expect augments Basic with a range of survey questions on students' expectations about their future earnings and study completion, reasons for their major choice, and their motivation for the chosen studies. For example, Kunz and Staub (2020) report a relationship between beliefs about completing an education and actual completion. Several studies document associations between university enrollment or major choice and beliefs about pecuniary and non-pecuniary benefits of education (Wiswall and Zafar, 2015; Patnaik et al., 2021; Boneva and Rauh, 2017; Boneva et al., 2022). However, according to a study by Arcidiacono (2004) sorting appears mainly driven by preferences—differences in monetary returns matter very little.

The information set Backgr augments Basic with survey data on students' socioeconomic background and their financial situation, incorporating elements identified as influential for educational success, such as parental college education (e.g., Almås et al., 2016a), access to Danish student grants and loans, other means of financing the studies, and liquidity constraints (e.g., Joensen and Mattana, 2021; Dynarski et al., 2023).

3.2.2 Non-cognitive skills

Competitiveness. Following the literature based on Niederle and Vesterlund (2007), we elicited this measure using a real effort task. Specifically, students were required to count the number of zeros in tables with zeros and ones (Abeler et al., 2011). In the first stage, participants had three minutes to count as many tables as possible and earned DKK 0.5 (about USD 0.09) per table. In the second stage, students were informed that they would again have three minutes to count the number of zeros in up to 40 tables. Before counting, they could choose whether to be paid based on their own performance, or to compete with others' performance. If they chose the first option (No Competition), they received DKK 0.5 per correctly counted table. If they chose the second option (Competition), they received DKK 1 per correctly counted table if they correctly counted more tables than one randomly selected student in round 1. Ties were randomly resolved.

Overconfidence. After the choice between Competition and No Competition, and before doing the actual counting task, we also asked students to guess the percentile rank of their performance in round 1 relative to the performance of the other students in round 1. They received DKK 5 if they guessed their rank correctly. Comparing their believed rank to their actual rank provides us with a measure of overconfidence.

Effort measure. The real effort task also gives us a measure of how willing a student is to invest immediate effort in a boring task in exchange for a payment a couple of weeks later. We include in our analysis the first round of counting, which is uncontaminated by the competition choice.

Self-control and grit. Several questions measured how self-controlled and gritty students are and which self-regulation strategies they employ. First, we included two psychological scales: the 13-item brief self-control scale (Tangney et al., 2004) and the 8-item grit scale (Duckworth et al., 2007). Second, we included 21 questions on general and study-specific self-regulation strategies that students may use or endorse (such as personal goals and deadlines, or hard commitment devices like mandatory hand-in requirements). Based on a principal component analysis of these 21 questions, parallel analysis (Horn, 1965; Patil et al., 2008) suggests five components, which we extract and include in our analysis (see Table A.7).

Risk preferences. We elicited risk preferences using incentivized choice lists to recover certainty equivalents (Farquhar, 1984; Bruhin et al., 2010). In our analysis, we use a non-parametric index of loss aversion. Full details on the elicitation procedure and the construction of the index are provided in Appendix A.1.

Narrow bracketing. We included a set of questions on narrow bracketing.[10] The first question is based on Tversky and Kahneman (1981), where decision makers face two decisions that jointly determine their payoff and need to choose between two lotteries for each decision. The decision was incentivized.[11] Typically, a large fraction of decision makers choose a lottery combination that is first-order stochastically dominated by another possible combination (Tversky and Kahneman, 1981; Rabin and Weizsäcker, 2009). Such choice errors can be explained by narrow bracketing together with loss aversion.

Further, students were asked which kind of small scale insurance (cycle, phone, baggage, travel, computer/laptop) they ever bought. One explanation for buying small scale insurance is narrow bracketing in conjunction with expectation based reference dependence preferences (Sydnor, 2010). Similarly, whether students divide their monthly budget into several separate budgets or not is indicative of narrow bracketing in terms of narrowly defined mental accounts (Heath and Soll, 1996).

The final question in this block builds on Kahneman and Tversky (1984)'s "lost ticket versus lost money" scenario. They observed that many participants were willing to repurchase a concert ticket after losing an equivalent amount of cash (here DKK 100), but far fewer were willing to do so after losing the ticket itself. This suggests that buying a ticket creates a narrow, topical mental account for the concert, where the cost of the lost ticket is charged—but not the lost cash.

Cognitive reflection and strategic sophistication. Students completed the three-item cognitive reflection test by Frederick (2005), consisting of simple math exercises where the impulsive answer is wrong. Students received DKK 2 for each correct answer.

Further, we elicited a measure of strategic sophistication using a beauty contest

[10]These data were also used in Koch and Nafziger (2019) to examine theories that classify narrow bracketing phenomena either as errors due to cognitive limitations or strategies to achieve self-control.

[11]One randomly chosen student was paid for both decisions. To cover any possible losses, this participant received an extra 100 kr. endowment.

game (e.g., Nagel, 1995). Students chose a number between 0 and 100, knowing that the other participants did the same. The average of all entered numbers was then computed and multiplied by two thirds. The student whose entry was closest to this number won DKK 200. Applying iterated elimination of strictly dominated strategies leads to the unique Nash equilibrium, where all players choose 0. A lower chosen number thus is indicative of deeper strategic reasoning.

3.2.3 Expectations and motivation related to the study program

The survey included a range of questions about how students selected their study subject (interest, job opportunities, recommendations, fits talents, did not know), whether it was their most desired subject, whether students were certain about their choice, as well as their motivation and satisfaction with the studies. In addition, we asked about expectations and beliefs associated with future outcomes, such as future earnings after graduating, expected study length for the respective degree, the relationship between obtained grades and income, and the likelihood that a student obtains a bachelor's or master's degree in the respective subject.

3.2.4 Socioeconomic background and appearance

This part of the survey collected information on students' socioeconomic background that may help predict dropout, such as whether parents have attended college, access to Danish student grants and loans, other means of financing the studies, and the extent to which a student is liquidity constrained. In addition, we included questions related to self-perception of attractiveness and physical strength, height and weight (which we aggregate as body mass index) because prior research links these characteristics to labor market outcomes (e.g., Rooth, 2011; Scholz and Sicinski, 2015).

3.3 Experimental procedures

Students were invited by email and informed that they could earn money by completing an online survey on their own computer in either Danish or English and that it would take approximately 60 minutes to complete the study.[12] Invitations were sent in five consecutive waves within the first weeks of the semester. Students who did not start or did not complete the survey received reminders.

Average earnings were DKK 148 (USD 27). Upon accessing the study, students were presented with an overview of the study and its procedures. Additionally, the students were requested to grant their consent to allow us to link their data with student records and national administrative data. In order to receive payment, students had to complete the entire survey, which consisted of several incentivized tasks in addition to a fixed payment of DKK 50.[13] Students were informed that payments would be made two to six weeks after completion, via a standard system through which public institutions can transfer money using the recipients' social security number.

4 Empirical strategy

We frame student dropout prediction as a binary classification task and use machine learning algorithms to estimate models that predict dropout risk. We carry out a comprehensive analysis by comparing the performance of five distinct machine learning algorithms. Our objective is predictive rather than causal. Accordingly, we focus on out-of-sample performance (using the nested cross-validation procedure described below) rather than structural interpretation of individual coefficients.

[12] A software filter restricted smartphone access.

[13] Students who completed the survey could earn an additional DKK 200 by participating in a follow-up study eliciting time preferences with the procedure outlined in Augenblick and Rabin (2019) and described in Epper et al. (2018). These measures are not used in this study because only 314 students completed them.

4.1 Predictive models

For conciseness, we focus in the main text on results from *XGBoost* (Extreme Gradient Boosting; Chen and Guestrin, 2016). The algorithm successively builds models to minimize prediction errors from previous iterations. XGBoost is widely used in research and industry because it balances accuracy with computational efficiency. Appendix A.2 provides further details on XGBoost and the four alternative models—a standard logistic regression model, LASSO, neural network, and random forest. These models differ in their flexibility to capture linear and non-linear relationships, allowing us to assess whether more complex algorithms outperform simpler models in predictive accuracy.

4.2 Nested cross-validation

We apply nested cross-validation (CV)—also known as double cross-validation (Stone, 1974)–when estimating all models. Nested CV mitigates overfitting by separating model selection, hyperparameter tuning, and performance evaluation, thereby providing a more reliable estimate of the model's generalizability (Varma and Simon, 2006). Details of the procedure and hyperparameter tuning are provided in Appendices A.3 and A.4.

4.3 Evaluation of the model performance

Evaluating the performance of predictive models involves comparing predicted outcomes to realized outcomes while balancing Type I (false positive) and Type II (false negative) errors. We rely on the *Area Under the Curve* (AUC), a widely adopted metric for binary classification tasks. The AUC equals the probability that the model assigns a higher predicted dropout risk to a randomly selected student who drops out than to a randomly selected student who does not.[14] Because all models are

[14]Specifically, the AUC measures the area under the *Receiver Operating Characteristic* (ROC) curve. A point on the ROC curve represents the true positive rate for a given false positive rate. AUC

evaluated on the same individuals, their ROC curves (and hence their AUC estimates) are statistically dependent. We therefore use DeLong's test for two correlated ROC curves, which compares paired AUCs using a nonparametric estimate of the covariance structure of the underlying U-statistics (DeLong et al., 1988; Rainio et al., 2024).

In practice, each individual contributes a single out-of-sample predicted dropout risk per model, obtained from the outer fold in which the individual serves as part of the validation set. We pool these out-of-fold predictions across individuals to compute AUCs and apply DeLong's test to compare paired predictions across models.

We further corroborate the conclusions drawn from the AUC analysis in Appendix A.5.2.1 using precision-recall curves and the corresponding *Precision-Recall Area Under the Curve* (PRAUC) metric, which is robust to imbalanced outcomes.

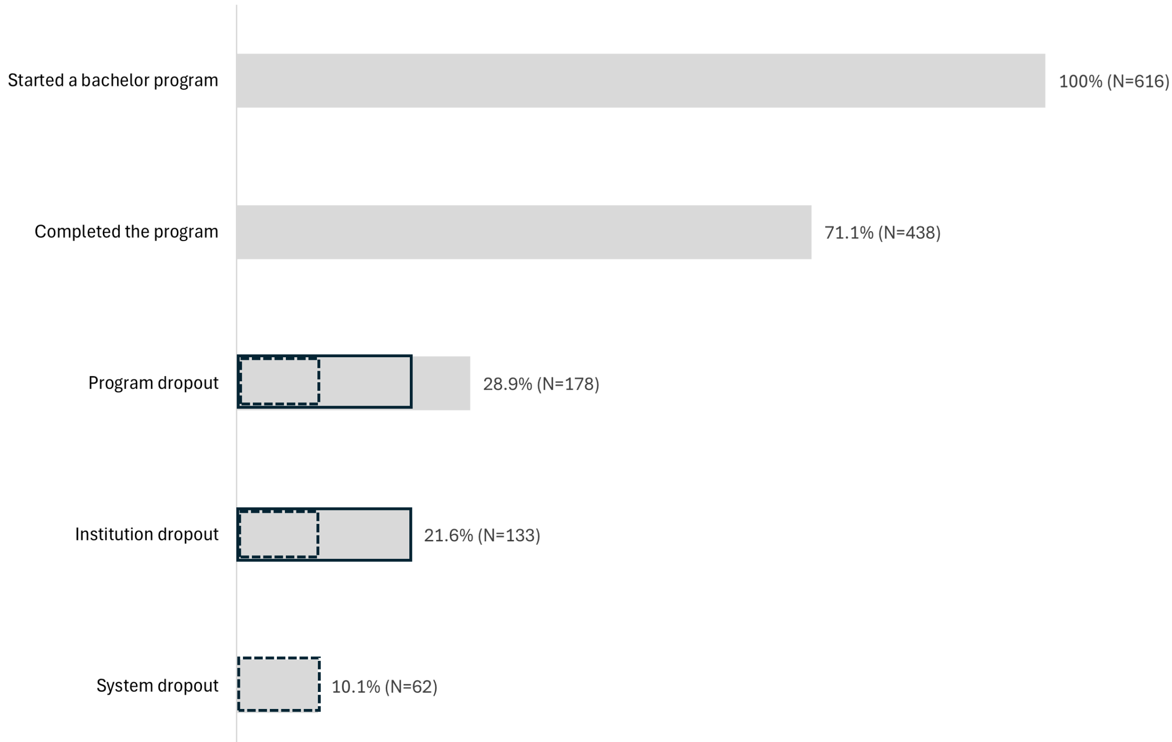
5 Results

5.1 Descriptive statistics on dropout

Figure 1 provides descriptive statistics on the different categories of dropout for the 616 students who took the survey. Of these, 438 students (71.1 percent) obtained a bachelor's degree in the program for which they were enrolled, whereas 178 students (28.9 percent) dropped out of their program. Of these, 45 students completed another bachelor's degree at Aarhus University. Hence, considering dropout at the institution level, 133 students (21.6 percent of the 616 students) left Aarhus University before obtaining a degree. Of these, 70 students completed a bachelor's program of at least three years' duration at another Danish university or university college. This leaves 62 students (10.1 percent of the 616 students) who dropped out without completing any university degree. Given the sample size, our main focus is

aggregates over all such possible classification thresholds. See Appendix A.5.1 for further details.

Figure 1: Categories of dropout



Notes: Categories of dropout: *Program dropout* (a student drops the study program originally enrolled in), *institution dropout* (a student leaves the university without a degree), and *system dropout* (a student leaves the Danish higher education system without any bachelor's degree). 616 students completed the survey—among 2,747 students starting a bachelor program at the Faculty of Business and Social Sciences at Aarhus University in the fall of 2013 (a 22.42 percent response rate). Study areas include several business and economics programs, law, political science, and psychology.

on program dropout, but we also report results for institution and system dropout to illustrate robustness and provide a more complete picture.

5.2 Predictive performance

This section compares the gains in predictive performance when additional information sets (i.e., groups of variables or "features") are included in the models beyond high school GPA and indicators for the enrolled bachelor program. We start with the Basic information set that includes variables typically available to university administrations. Next, we separately add the information sets listed in Table 1. We focus on the results for XGBoost. The other machine learning algorithms perform similarly to XGBoost (see Appendix A.5.2).

Table 1: Information sets

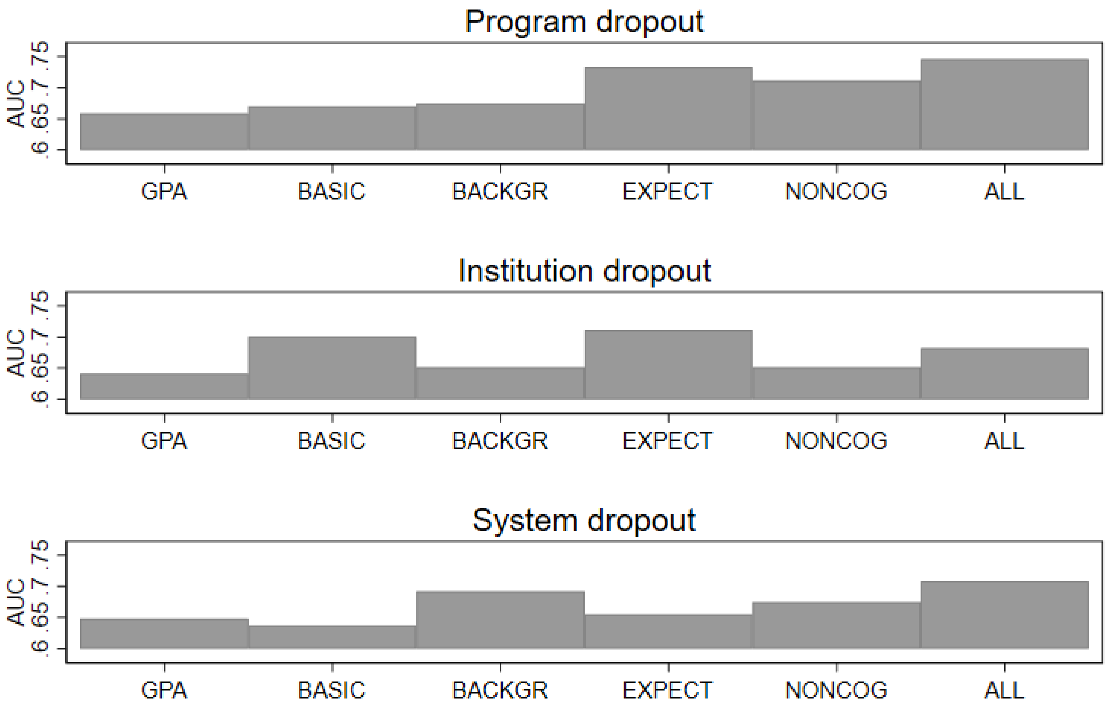
Set	Explanation
GPA	High school grade point average and dummies for the bachelor program entered
Basic	GPA (including program dummies) + additional features typically available to universities for incoming students (e.g., gender, age, type of high school, priority ranking of the program)
Backgr	Basic + additional information from survey questions on students' background (e.g., education of mother and father, appearance)
Expect	Basic + additional information from survey questions on students' expectations and study motivation (e.g. see footnote 17)
Noncog	Basic + additional measures related to students' cognitive and non-cognitive skills (e.g., Brief Self-Control Scale, cognitive reflection, grit, loss aversion)
All	All features combined in one set

Predicting dropout

Figure 2 compares the AUC across the different information sets, separately for each of the three dropout categories. Table 2 reports pairwise AUC comparisons based on the DeLong test (DeLong et al., 1988; Rainio et al., 2024). The top panel of Figure 2 shows results for *program dropout*. Adding the Basic information set to the baseline model with only high school GPA and program dummies increases the AUC from 0.66 to 0.67, yielding no statistically significant improvement. Adding the student background characteristics in Backgr yields a similar AUC (0.67). In contrast, Expect increases the AUC to 0.73; similarly, Noncog increases the AUC to 0.71—both statistically significant improvements. The highest AUC of 0.75 is obtained with All.

The middle panel of Figure 2 examines the predictive performance for *institution dropout*. Here, the Basic set significantly improves prediction over the GPA baseline, increasing the AUC from 0.64 to 0.70. None of the other information sets offer statistically significant gains beyond Basic. Backgr and Noncog even slightly reduce predictive performance relative to Basic, suggesting that these additional variables

Figure 2: Model performance with different information sets (XGBoost)



Notes: Model: XGBoost. All information sets to the right of BASIC include the BASIC features (see Table 1). AUC captures the probability that the model ranks a randomly selected student who drops out more highly in terms of dropout risk than a randomly selected student who does not drop out. It takes values between 0 (all predictions are wrong) and 1 (all predictions are correct).

add limited signal in this setting and may introduce noise or overfitting.

The bottom panel displays results for *system dropout*. Here, only ALL yields a (borderline) statistically significant gain relative to the GPA baseline, reaching an AUC of 0.71. Relative to BASIC, both BACKGR (with an AUC of 0.69) and ALL provide a (borderline) significant improvement.

Overall, for institution and system dropout, adding the NONCOG information set does not improve predictive performance relative to BASIC. Because institution and system dropout are rarer outcomes, these null effects should be interpreted with caution. For program dropout, NONCOG *does* add predictive power, but the analysis in Section 5.4 shows that the gain is driven primarily by study-related self-regulation components and overconfidence rather than by the broad preference and trait measures (self-control, grit, competitiveness, risk preferences, cognitive reflection, strategic sophistication, or narrow bracketing). This pattern, especially the

Table 2: Test of pairwise AUC differences (XGBoost)

	AUC	GPA	Basic	Backgr	Expect	Noncog	All
				Program dropout			
GPA	0.6583	–	0.011	0.016	**0.074*****	**0.053****	**0.088*****
Basic	0.6698		–	0.005	**0.063****	**0.041***	**0.076*****
Backgr	0.6745			–	**0.058***	**0.037***	**0.071****
Expect	0.7327				–	-0.022	0.013
Noncog	0.7112					–	0.035
All	0.7458						–
				Institution dropout			
	AUC	GPA	Basic	Backgr	Expect	Noncog	All
GPA	0.6415	–	**0.059****	0.009	**0.069****	0.010	0.041
Basic	0.7008		–	**-0.050****	0.010	**-0.050***	-0.019
Backgr	0.6510			–	**0.060****	0.000	0.031
Expect	0.7107				–	**-0.060***	-0.028
Noncog	0.6510					–	0.031
All	0.6822						–
				System dropout			
	AUC	GPA	Basic	Backgr	Expect	Noncog	All
GPA	0.6479	–	-0.010	0.044	0.007	0.027	**0.060**$^{+}$
Basic	0.6377		–	**0.054**$^{+}$	0.017	0.037	**0.070**$^{+}$
Backgr	0.6921			–	-0.037	-0.017	0.016
Expect	0.6548				–	0.020	0.053
Noncog	0.6747					–	0.033
All	0.7078						–

Notes: Each panel reports the Area Under the ROC Curve (AUC) in column 2 and the pairwise differences in the AUC between XGBoost models, based on DeLong's test for two correlated ROC curves, in columns $3-8$. Cells above the diagonal show AUC(column) $-$ AUC(row); positive values therefore indicate that the model in the column achieves higher predictive performance than the model in the row. GPA, Basic, Backgr, Expect, and Noncog refer to the information sets defined in Table 1. All includes all features. Significance levels: $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

limited contribution of the preference/trait measures, contrasts with prior evidence linking these measures to other academic outcomes, which we discuss in Section 6.

In contrast, the results in Figure 2 and Table 2 show that survey-based measures of students' expectations and motivations for their specific study program provide important additional information for predicting program dropout beyond the information already captured in the BASIC information set. For institution dropout, these measures improve predictive performance relative to a model using only high school GPA and program dummies, but they do not yield statistically significant gains beyond BASIC.

Since the expectation and motivation measures are tailored to a particular institution, it is perhaps unsurprising that they offer little predictive value for system dropout, where students leave higher education entirely.

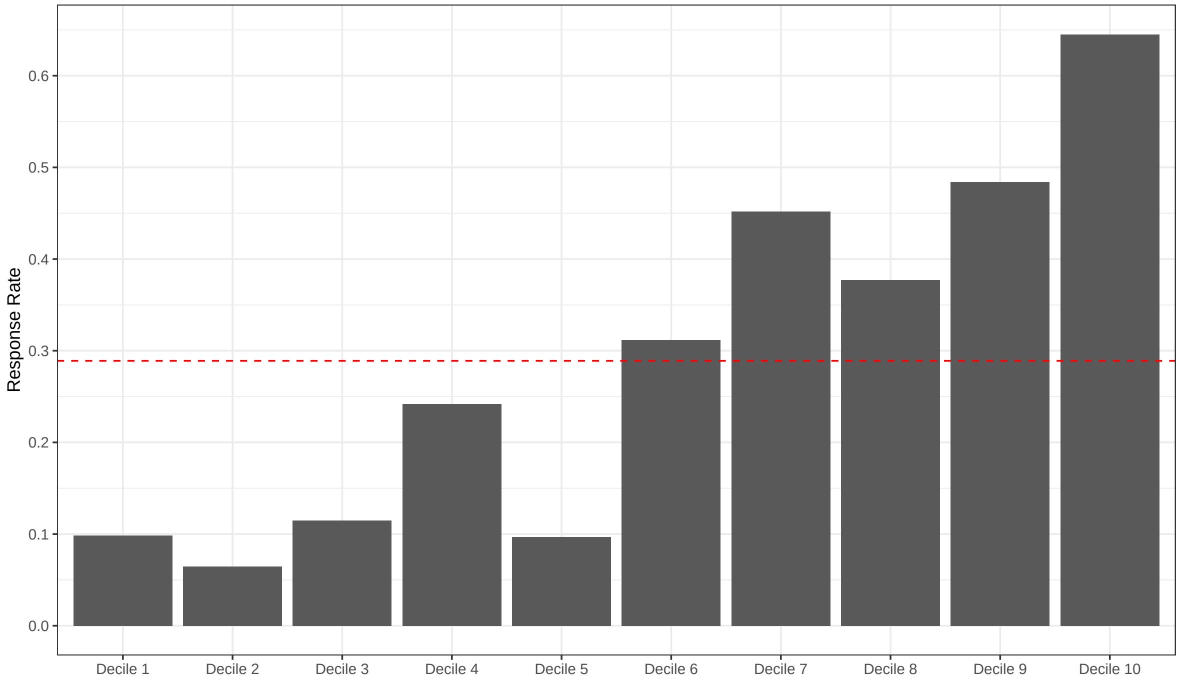
Another notable result is that the BACKGR information set does not improve predictive performance over the BASIC set. The only exception is for system dropout, where it yields a marginally significant gain. The latter suggests that socioeconomic background may play a more important role in explaining complete withdrawal from higher education than in predicting within-institution outcomes.

5.3 Identification of at-risk students

To provide a sense of how well the machine learning models perform when ranking students in terms of dropout risk, Figure 3 shows the extent to which the risk classification by the predictive model correlates with actual dropout from the program. The x-axis captures the risk ranking assigned by XGBoost. Decile 1 contains those students predicted to be among the 10 percent with the lowest dropout risk and decile 10 contains those students predicted to be among the 10 percent with the highest dropout risk. The y-axis measures the proportion of actual cases of dropout (true positives) in a decile (the response rate).

The dotted line at 28.9 percent marks the overall average program dropout rate. That is, a model that perfectly identified dropout would assign all dropout cases to the three highest risk deciles. In this case, the response rate would be 1 for the deciles 10 and 9 and close to 1 for decile 8. What we actually see is that among the students assigned to the highest risk decile 10, around 65 percent do drop out of the program.

Figure 3: Response rates for program dropout (XGBoost)



Notes: XGBoost model using the full feature set All (see Table 1). Deciles capture the risk ranking assigned by XGBoost: decile 1 contains the 10 percent of students with the lowest predicted dropout risk and decile 10 contains the 10 percent of students with the highest predicted dropout risk. The graph shows the response rate, which measures the proportion of actual cases of dropout (true positives) in a decile. The dotted line shows the 28.9 percent dropout rate at the program level for the entire sample.

Further, the response rate generally declines from the highest-risk decile 10 to the lowest-risk decile 1, although not strictly monotonic. Overall, students assigned a higher predicted risk rank exhibit substantially higher realized dropout rates.

Appendix A.5.3 complements this analysis with lift curves—another way of visualizing the ability of a predictive model to identify at risk students. Overall, these analyses suggest that machine learning models applied to our data achieve economically meaningful out-of-sample predictive performance. In Appendix A.5.3, we also report corresponding analyses for institution and system dropout and compare the different machine learning algorithms. Due to the smaller sample sizes, the predictive models are less reliable in predicting dropout in these settings, in particular, for system dropout.

5.4 Feature importance

Machine learning models are often criticized as opaque 'black boxes' (e.g. Burrell, 2016). To unveil the complex layers of machine learning models, the concept of feature importance has emerged (e.g., Doshi-Velez and Kim, 2017; Lundberg et al., 2020). It assesses and ranks the variables, or 'features', in a dataset based on their impact on the predictions of the model. *Permutation Feature Importance* (PFI) and *SHapley Additive exPlanations* (SHAP) are two widely applied measures that have the advantage of being model independent. That is, they can be applied across diverse machine learning models, providing a consistent way to interpret and compare models. For conciseness, we focus here on the *SHAP* measure introduced by Lundberg and Lee (2017), as it has the advantage over *PFI* that it takes into account the interaction effects between features.[15]

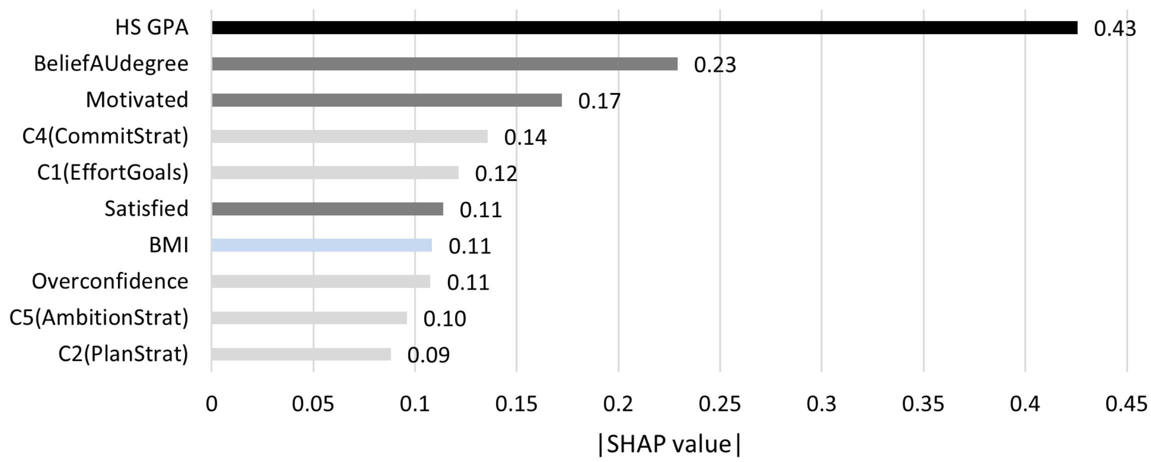
A variable with a higher *SHAP* feature importance has a stronger impact on the model's prediction.[16] Figure 4 shows the ranking of *SHAP* feature importance for the XGBoost model predicting *program dropout* using the full feature set ALL, measured by the absolute *SHAP* value. It represents the magnitude of a feature's contribution to the model's prediction, irrespective of the direction of the influence (positive or negative). The figure displays the ten features that contribute most to the prediction of *program dropout*.

High school GPA (*HS GPA*) is by far the most important feature—confirming the general insight from the previous literature (see the introduction). Two survey measures from the EXPECT information set are highly influential ($|SHAP|$ = 0.24 and 0.19,

[15]The general patterns are robust to using the PFI measure as an alternative (see Appendix A.5.5).

[16]*SHAP* builds on a concept from cooperative game theory that Shapley (1953) developed for assigning to each player a value based on the contribution that the player has to the total payoff when taking into account all possible coalitions that players may form. In the case of *SHAP*, a 'player' is a feature, the 'total payoff' is the prediction for a particular coalition, and a 'coalition' is a subset of features. The Shapley value of a feature is computed as the average of its marginal contributions across all possible coalitions of all features. The marginal contribution is the difference in the prediction when we include the feature and when we do not. This means that *SHAP* not only considers the main effect of a feature (i.e., its contribution to the prediction in isolation) but also interaction effects with other features. In this way, *SHAP* allocates the value for the total prediction among the input features.

Figure 4: Feature importance for predicting program dropout (XGBoost)



Notes: Model: XGBoost using the full feature set All (see Table 1). Top 10 feature importance in terms of absolute *SHAP* value for program dropout. Components of the full feature set All: Black is included in the GPA / Basic set, light blue is included in Backgr, dark gray is included in Expect, and light gray is included in Noncog.

respectively): students' certainty about completing a degree at Aarhus University (*BeliefAUdegree*) and their motivation for the enrolled study program (*Motivated*).[17] These two simple, self-reported measures of study motivation and degree completion expectations add substantially more to the predictive power of the model than any of the variables capturing students' background or non-cognitive skills.

Questions related to specific, study-related self-regulation strategies (included in Noncog) also contribute to the prediction of *program dropout*. These are captured by some of the five principal components ($C1, C2, \ldots, C5$) derived from responses to 21 questions, relating, for example, to the use of, or preference for, external commitment devices (such as mandatory hand-in requirements and project deadlines) and to setting specific goals for study behavior (for example, attendance at lectures, participation in study groups, and completion of homework). More broadly these measures capture good study habits and an understanding of what studying at university entails. Among the other non-cognitive skills, only overconfidence is

[17]BeliefAUdegree: *"I am very certain that I will finish my studies at Aarhus University with a bachelor's or master's degree"*. Motivated: *"I am very motivated for my studies"*. Both are assessed on a 5-point Likert scale: Not like me at all, Not much like me, Somewhat like me, Mostly like me, Very much like me.

important for predicting dropout.

How satisfied a student is with the study program in which the student is enrolled (*Satisfied*) and body mass index (*BMI*) also contribute to dropout prediction. This latter finding aligns with empirical studies linking physical attractiveness to labor market success (e.g., Rooth, 2011; Scholz and Sicinski, 2015).

In Appendix A.5.4, we report corresponding analyses for *institution* and *system* dropout. Overall, these results show that across all three dropout definitions (i) academic performance (GPA) dominates dropout prediction, (ii) the simple self-reported expectation measures from footnote 17 add substantial predictive value, especially for program dropout and institution dropout, (iii) self-regulation strategies related to studying provide additional, complementary information that can help distinguish students who are least likely to persist in their studies, and (iv) measures of preferences and traits (such as grit and self-control) add little value in predicting dropout. The next section examines potential mechanisms and presents further robustness analyses.

6 Discussion and mechanisms

6.1 The role of motivation

A central pattern in our results is that measures of expectations and study-related motivation are among the most informative predictors of program dropout. This section discusses an interpretation that is consistent with that pattern and presents auxiliary evidence on one plausible channel, effort investment, using a separate cohort for which study hours are observed.

One interpretation is that study motivation predicts dropout because it is related to subsequent study effort. In our main survey experiment, we do not observe study hours, so we cannot directly test this channel within the original cohort. To provide suggestive evidence, we administered the same motivation measure to a different cohort of students at the same faculty on the first day of their studies. In the final weeks of the first semester, we asked these students how many hours they studied per week on average.[18] Motivation at study start is positively correlated with self-reported weekly study hours several months later in a bivariate regression ($\beta = 2.18^{***}$, SE=0.61, N=442). This estimate implies roughly 9 additional study hours per week when moving from the lowest scale point (1) to the highest scale point (5) on the motivation measure. Because this evidence is correlational and drawn from a different cohort, we interpret it as corroboration of a plausible channel rather than as a causal mechanism in the main sample.

We also observe that average study effort is low. Students report spending around 32 hours per week on lectures and self-study combined, whereas full-time students should spend around 37-40 hours per week according to the European ECTS system. Thus, the average student studies almost one full working day less per week than program requirements imply. Moreover, around 44 percent of students study 30 hours or less per week. Such low effort plausibly contributes to early knowledge gaps and, in turn, to higher dropout risk.

The broader idea that the effort margin is important aligns with Oreopoulos and Petronijevic (2019). They document that students study on average five to eight hours less than what they think is optimal and discuss that the gap relative to program requirements can be even larger.[19] They evaluate a set of soft interventions aimed at increasing study effort and find some positive effects on students' well-being and ambitions. However, interventions increase average study time by only about two hours per week, an increase on the order of 10 hours per week would likely be needed to generate economically meaningful improvements in grades according to Oreopoulos and Petronijevic (2019). Together, these findings suggest that

[18]As we do not have a measure for number of work hours per week in the dataset for our survey experiment, we rely on survey data from the 2022 student cohort in the Business program at Aarhus University. These data were collected as part of a different study.

[19]Oreopoulos and Petronijevic (2019) cite evidence that students spend approximately 15 hours per week on self-study outside of lectures, whereas study programs typically stipulate at least 25 hours.

reducing dropout may require interventions that materially increase study effort rather than marginal nudges.

Finally, the effort-investment interpretation is not the only explanation consistent with our findings. Motivation, and related measures of study self-regulation, may also proxy for unobserved constraints (e.g., time, financial stress, competing work demands) or for match quality between student and program. Under this alternative view, these measures predict dropout not because they causally raise effort, but because they capture early signals of mismatch or constraints that also reduce effort. Distinguishing between these interpretations would require either richer time-varying data on effort and constraints in the main cohort, or exogenous variation in inputs that plausibly shifts effort.

6.2 The role of expectations

Students' entry expectations about completing their degree (see footnote 17) are among the most informative predictors of subsequent dropout. While it may seem intuitive that stated completion beliefs predict realized completion at the same institution, the result is not tautological: expectations are often biased, and individuals frequently mispredict their own future behavior and outcomes (e.g., DellaVigna and Malmendier, 2006; Williams, 2008; Stinebrickner and Stinebrickner, 2014a). Consistent with Wiswall and Zafar (2021), we nonetheless find that elicited expectations contain substantial predictive content. The predictive power of entry expectations in our setting suggests that the expectation measure captures information about perceived fit, anticipated constraints, or other determinants of persistence that are not fully reflected in administrative variables.

These findings raise the question of why students enroll in a program for which they report relatively low completion expectations or low motivation. One candidate explanation is limited access to preferred programs.[20] However, the BASIC

[20]In our data, 86.7 percent of the students got admitted to their first priority, 8.6 percent of students

information set includes the priority ranking of the enrolled program, and priority is not among the most important predictors once expectations and motivation are included (see Section 5.4). Moreover, priority is only weakly correlated with motivation ($\beta = -0.08^{**}$, SE=0.03) and not significantly correlated with expectations ($\beta = -0.06$, SE=0.04) and the $R^2 < 0.01$ in both cases. This pattern suggests that expectations and motivation are not merely proxies for not being admitted to a first-choice program; instead, they likely reflect perceived match and/or anticipated constraints that are orthogonal to the priority measure.

Taken together, the evidence indicates that elicited expectations at entry summarize a distinct component of dropout risk, one that is informative even when expectations are imperfect and even when administrative measures and simple program-choice proxies are observed.

6.3 The role of preferences and traits

Study effort and habits plausibly matter for dropout. Yet, it is surprising that the preference and trait measures in the non-cognitive skills set NONCOG contribute little to predicting dropout in our setting. This pattern appears both in the SHAP values in Section 5 and in the bivariate correlations between the preference/trait measures and dropout (Table A.8). One might have expected these measures to be more helpful for predicting dropout than the survey questions on study motivation and expectations, because preferences and traits are general constructs that should be able to explain, for example, motivation.[21] Our results, however, suggest that for dropout the specificity of the motivation and expectations questions allows them to capture nuances that the more general constructs do not capture.

Our finding that traits like grit and self-control do not improve the prediction of

to their second priority, 2.8 percent of students to their third priority, and 2.0 percent of students to their fourth or lower priority.

[21]Indeed, we observe that motivation is correlated with grit ($\beta = 0.57^{***}$, SE=0.06) and self-control ($\beta = 0.47^{***}$, SE=0.06) with an R^2 of 0.14 and 0.11, respectively. That is, at least to some extent, motivation measures a similar construct as these general behavioral traits.

dropout (and do not correlate with it) also appears to contrast with evidence linking these traits to other academic outcomes (see the introduction). This relation has mainly been established for grades among adolescents. In line with the literature, we also find in our sample that grit and self-control correlate with high school grades (Table A.8). This suggests that the measures behave as expected in our sample and that it is unlikely that our dropout results are driven by, for example, grit being perceived differently in Denmark than in the US or by our sample being special (e.g., Danish university students being selected in a specific way).

It should be noted that the relation between grit and outcomes is established to a lesser extent for academic outcomes of university students than for adolescents. In particular, there has been little research on the association between university dropout and, for example, grit. Consistent with our findings, Caviglia-Harris and Maier (2020) also find that grit and conscientiousness do not relate to dropout. Duckworth et al. (2019) do find a correlation between grit and timely completion, but timely completion is not the same as dropout, and their sample–students of the elite West Point military academy–may differ substantially from ordinary university students.

Overall, our measures of grit and self-control behave as expected with respect to grades, suggesting validity rather than noise. A potential reason why they are not predictive for dropout from the university could be that those students who drop out come mostly from the lower end of the high school GPA distribution, and that for these students ability is the key driver and non-cognitive skills matter less for the decision to drop out. This explanation is consistent with high school GPA being the most important variable for predicting dropout. Another, complementary explanation could be that high school GPA (or, more broadly, the variables in the Basic information set) is a sufficient statistic for the preference measures and traits, and that the latter therefore provide no new information to the prediction model. It could also be that our sample simply is specific, or that university students in general are special because they are a selected group.[22] We investigate this further in the next sections and find little evidence for either of these potential explanations.

6.3.1 Is high school GPA a sufficient statistic?

A potential explanation of our result that measures of preferences and traits add little value to predicting dropout is that high school GPA (or certain variables of the Basic information set) are sufficient statistics for these measures. For example, prior research pointed out that math grades correlate with IQ (e.g., Deary et al., 2007). Yet math (and other) grades also correlate with traits, such as self-control (e.g., Duckworth and Seligman, 2005). Table A.8 shows that several of the measures of preferences and traits are correlated with high school GPA in our study. However, we find that the R^2 in regressions of high school GPA on individual non-cognitive skills is generally low (all below 0.03).

To assess whether variables from the Basic information set (such as high school GPA or whether the study program was the first priority of the student) absorb the information in the non-cognitive skills variables, we also test how predictive performance changes when we include only age, gender and program dummies in addition to the non-cognitive skills (Noncog w/o Basic information set) and compare its performance to the model using the Basic information set.

Table A.9 shows that, compared to a model including just age, gender, and program dummies, predictive performance does not significantly improve when using the Noncog w/o Basic information set, which adds the non-cognitive skills variables but not high school GPA and other variables from Basic such as priority. For *program* and *institution* dropout, the Noncog w/o Basic model performs significantly worse than the Basic model, whereas for *system* dropout predictive accuracy is on par with

[22] Lira et al. (2022) observe that how participants answer the grit scale also depends on their reference group. Such a bias might also explain our findings. Yet, they observe that the reference group bias does not arise for incentivized measures of perseverance. When considering the feature importance we have seen that in our context neither the incentivized, nor the unincentivized measures have much predictive power—suggesting against reference group bias as an explanation behind our findings.

the Basic model.

Hence, rather than the Basic features being a sufficient statistic for the preference measures and traits, the reason for the limited predictive power of preferences and traits appears to be simpler: these measures are largely unrelated to dropout in our sample (see Table A.8)

6.3.2 Is selection into the survey driving the results?

Another possible reason why measures of preference and traits do not add value in predicting dropout could be selection into the survey, as discussed in Section 3.1.1. Students who select into the survey are less likely to drop out (we can observe education trajectories from administrative data even for non-participants). It is also possible that students who select into the survey have, for example, higher self-control and grit than those who do not respond. If the relationship between these skills and dropout differs between survey participants and the full population, selection could attenuate the predictive value of traits in our estimation sample. Note, however, that survey participants also have higher GPA and might be more motivated for their studies, and these variables still matter for predicting dropout.

To directly address selection, we run logit regressions where we reweigh each observation based on propensity scores estimated on observed covariates (see also Epper et al. (2020)). Doing so creates a weighted sample that matches the covariate distribution of non-participants on observed characteristics. Figures A.21 - A.32 compare the predicted dropout probabilities for the original sample with the weighted sample and show that both closely correspond to each other. Hence, we conclude that it is unlikely that selection into the survey is driving our findings.

6.3.3 Is selection into university driving the result?

Beyond selection into the survey, selection into university could in principle explain why preference measures and traits add little predictive value: university students may be more homogeneous in these traits than the general population, so there is less usable variation left to predict dropout. At the same time, our earlier results show that several traits, such as self-control and grit, still correlate with high school GPA for our university sample. This suggests that the measures capture meaningful variation and are not trivially "washed out" by selection.

To assess directly whether the distribution of these measures looks unusually compressed, we compare the distributions of grit and self-control in our sample to (i) a general population sample from the US and (ii) a large Danish sample of 9th graders.[23] Figure A.34 shows that the distribution of grit among university students is shifted to the right relative to the broader populations–which is expected given selection into university. Yet, the overall range and shape are similar. The distribution of self-control in our sample is also comparable to the US general population sample (see Figure A.33). Both distributions are lower than for the 9th graders, which likely reflects age differences rather than extreme selection. Overall, these comparisons speak against the idea that selection into university mechanically eliminates variation in grit and self-control to an extent that would explain their weak predictive content for dropout.

A second, complementary concern is that our main sample size might not be large enough to detect the incremental predictive value of traits and preferences if these effects are small. To address both concerns–selection into university and statistical power–we replicate a closely related exercise in a much larger and less selected population sample. Specifically, we use data from Hvidman et al. (2024), who conducted and evaluated an intervention targeting pupils in 9th grade of Danish lower secondary schools. They administered a survey among all 9th graders[24] at the participating schools for three cohorts in Spring 2017, 2018, and 2019 (N=16,233), of

[23]US sample: 299 participants, representative of the US population in terms of sex and age, collected on Prolific Academic in June 2024. 9th graders: 13,698 Danish 9th grade pupils from Hvidman et al. (2024) (see below for further details on this sample).

[24]9th grade is the penultimate year of the 10-year compulsory schooling. In Denmark, it would be referred to as 8th grade, because counting starts from grade zero instead of 1st grade.

whom 13,698 completed the survey.[25] The survey is linked to Danish national administrative data containing subsequent educational outcomes through the end of 2024.

The survey included a broad set of non-cognitive skills measures that partly overlaps with our NONCOG information set (including grit, self-control, and survey measures of risk preferences).[26] The measures are based on questionnaires (not incentivized tasks). While this sample is too young to evaluate university dropout, it contains a closely related educational completion outcome: whether a pupil completed academic upper secondary school within four years of graduating from lower secondary school. This is the case for 64.13 percent of the sample; the main alternative outcome within four years is vocational education (26.63 percent).[27]

We replicate our main approach by constructing three information sets, BASIC, NONCOG, and BACKGR, and testing whether NONCOG or BACKGR separately increase predictive power beyond BASIC.[28] Table A.10 reports the results. Given the large sample size, even tiny differences are statistically significant. However, the substantive pattern mirrors our main analysis: the non-cognitive skills variables in NONCOG add very little predictive value for completion of academic upper secondary school within four years, and the same is true for adding background variables from BACKGR. In contrast, adding BASIC information on academic performance to age and gender increases predictive accuracy more than thirty times as much as addi-

[25]The study involved a pre- and post-survey. We use data from the post-survey and impute post-survey values from the pre-survey in case of missing observations. Without imputation, we have 12,305 observations.

[26]Specifically, the study of Hvidman et al. (2024) includes: grit, the self-control scale for children Tsukayama et al. (2013), mindsets Dweck (2006), self-concept Judge et al. (2003), the strengths and difficulties questionnaire for adolescents Goodman et al. (1998), as well as a survey-based measure of risk preferences Dohmen et al. (2011), and questions on time preferences from Vischer et al. (2013) and the GSOEP.

[27]Upper secondary school lasts for most programs 3 years in Denmark. Yet, a large fraction of pupils (48% in our sample) take a voluntary additional year of schooling after lower secondary school—either at a school or at boarding schools ("efterskole") that in addition to a basic academic curriculum often focus on specific activities like sports or music.

[28]Similar to the classification in Table 1, BASIC contains information on age, gender, school starting age, and academic performance (the grade point average in the final year of lower secondary school and national standardized tests from 6th grade in math and Danish); NONCOG contains the non-cognitive skills measures described in footnote 26; and BACKGR contains information on the educational attainment of the mother and father.

tionally including the preference/trait measures in Noncog, and around fifty times as much as additionally including the background variables in Backgr.

Taken together, these results make it unlikely that our main finding is driven primarily by selection into university. Instead, the evidence is consistent with a broader conclusion: for educational completion outcomes, measures of prior academic performance contain substantially more predictive content than additional batteries of general preferences and traits, even in much larger and less selected populations. This reinforces the paper's central implication for measurement design: if the goal is early risk stratification, the highest returns come from prior performance measures and targeted, domain-specific elicitation (e.g., expectations and motivation), rather than from expanding broad trait and preference batteries.

7 Conclusion

This paper asks a practical and policy-relevant question: what information improves the early identification of students at risk of dropping out of university, beyond what administrations already observe at entry? Using an entry survey experiment for 616 first-year students at Aarhus University linked to Danish administrative records that track educational outcomes for eight years, we compare the predictive content of a set of realistic "information sets". Starting from GPA (high-school GPA and program dummies) and Basic (standard administrative information), we evaluate what is gained by additionally collecting Expect (expectations and study motivation), Noncog (a broad set of skills, preferences, traits, and study self-regulation measures), Backgr (socioeconomic and financial background), and All (the union of all measures).

Three main conclusions emerge. First, prior academic performance is the single most important predictor of dropout, consistent with a large literature. Second, a small number of targeted, study-specific questions, especially motivation and expectations about completing the degree, adds substantial predictive information beyond administrative records, particularly for dropout from the initially enrolled program. Third, and more unexpectedly, widely used measures of general preferences and traits (such as grit, self-control, and risk preferences) add little incremental value for predicting dropout once basic administrative information is included. The predictive gains from the broader NONCOG information set are driven primarily by measures closer to the study process, self-regulation strategies and overconfidence, rather than by the trait and preference measures. These patterns also appear in our replication exercise using a larger, independent dataset and related educational completion outcome, suggesting that the findings are not an artifact of the specific sample or measurement context.

Universities can improve dropout predictions at low cost by augmenting existing administrative data with a handful of short questions about students' expectations and motivation (and, potentially, a small number of items capturing study self-regulation). This approach is scalable and avoids the logistic and privacy challenges associated with lengthy batteries that collect sensitive information. At the same time, the results caution against interpreting broad non-cognitive batteries as universally informative for all academic outcomes: measures that predict grades need not predict dropout, which plausibly reflects additional margins such as perceived fit, constraints, and persistence.

Several limitations point to directions for future work. First, our evidence speaks to prediction, not the causal effects of interventions triggered by predictions; the next step is to test whether support targeted using the most informative low-cost measures can reduce dropout in a cost-effective way. Second, while the replication supports the qualitative patterns, additional work across institutions and cohorts would help characterize external validity. Third, because prediction-based targeting can generate distributional and ethical concerns, an important agenda is to evaluate how early-warning systems perform across student subgroups and how targeting rules should be designed to balance efficiency, equity, and transparency.

Overall, the paper's contribution is to clarify which inputs matter for forecasting dropout in a realistic institutional setting: existing administrative data go a long way, but inexpensive, study-specific measures of expectations and motivation contain decisive additional information, whereas standard preference and trait batteries contribute little incremental value.

References

Abeler, J., A. Falk, L. Götte, and D. Huffman (2011): "Reference Points and Effort Provision," *American Economic Review*, 101, 470–492.

Almås, I., A. W. Cappelen, K. G. Salvanes, E. Ø. Sørensen, and B. Tungodden (2016a): "What explains the gender gap in college track dropout? Experimental and administrative evidence," *American Economic Review*, 106, 296–302.

——— (2016b): "Willingness to compete: Family matters," *Management Science*, 62, 2149–2162.

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011): "Personality psychology and economics," in *Handbook of the Economics of Education*, Amsterdam: Elsevier, vol. 4, 1–181.

Arcidiacono, P. (2004): "Ability sorting and the returns to college major," *Journal of econometrics*, 121, 343–375.

Arcidiacono, P., E. Aucejo, A. Maurel, and T. Ransom (2025): "College attrition and the dynamics of information revelation," *Journal of Political Economy*, 133, 53–110.

Augenblick, N. and M. Rabin (2019): "An experiment on time preference and misprediction in unpleasant tasks," *Review of Economic Studies*, 86, 941–975.

Backes-Gellner, U., H. Herz, M. Kosfeld, and Y. Oswald (2021): "Do preferences and biases predict life outcomes? Evidence from education and labor market entry decisions," *European Economic Review*, 134, 103709.

Beattie, G., J.-W. P. Laliberté, C. Michaud-Leclerc, and P. Oreopoulos (2019): "What sets college thrivers and divers apart? A contrast in study habits, attitudes, and mental health," *Economics Letters*, 178, 50–53.

Berens, J., K. Schneider, S. Görtz, S. Oster, and J. Burghoff (2018): "Early detection of students at risk–predicting student dropouts using administrative student data and machine learning methods," CESifo Working Paper.

Bergstra, J. and Y. Bengio (2012): "Random search for hyper-parameter optimization." *Journal of Machine Learning Research*, 13.

Bettinger, E. P., B. J. Evans, and D. G. Pope (2013): "Improving college performance and retention the easy way: Unpacking the ACT exam," *American Economic Journal: Economic Policy*, 5, 26–52.

Bjerre-Nielsen, A., V. Kassarnig, D. D. Lassen, and S. Lehmann (2021): "Task-specific information outperforms surveillance-style big data in predictive analytics," *Proceedings of the National Academy of Sciences*, 118.

Boneva, T., M. Golin, and C. Rauh (2022): "Can perceived returns explain enrollment gaps in postgraduate education?" *Labour Economics*, 77, 101998.

Boneva, T. and C. Rauh (2017): "Socio-economic gaps in university enrollment: The role of perceived pecuniary and non-pecuniary returns," *CESifo Working Paper Series*.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008): "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43, 972–1059.

Breiman, L. (2001): "Random Forests," *Machine Learning*, 45, 5–32.

Bruhin, A., H. Fehr-Duda, and T. Epper (2010): "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion," *Econometrica*, 78, 1375–1412.

Bryan, G., D. S. Karlan, and S. Nelson (2010): "Commitment Devices," *Annual Review of Economics*, 2, 671–698.

Burks, S. V., C. Lewis, P. A. Kivi, A. Wiener, J. E. Anderson, L. Götte, C. G. DeYoung, and A. Rustichini (2015): "Cognitive skills, personality, and economic preferences in collegiate success," *Journal of Economic Behavior & Organization*.

Burrell, J. (2016): "How the machine 'thinks': Understanding opacity in machine learning algorithms," *Big Data & Society*, 3, 2053951715622512.

Buser, T., M. Niederle, and H. Oosterbeek (2014): "Gender, Competitiveness, and Career Choices*," *The Quarterly Journal of Economics*.

——— (2024): "Can competitiveness predict education and labor market outcomes? Evidence from incentivized choice and survey measures," *Review of Economics and Statistics*, 1–45.

Buser, T., N. Peter, and S. C. Wolter (2017): "Gender, competitiveness, and study choices in high school: Evidence from Switzerland," *American economic review*, 107, 125–130.

——— (2022): "Willingness to compete, gender and career choices along the whole ability distribution," *Experimental economics*, 25, 1299–1326.

Caviglia-Harris, J. and K. Maier (2020): "It's not all in their heads: the differing role of cognitive factors and non-cognitive traits in undergraduate success," *Education Economics*, 28, 245–262.

Chatterjee, S. and F. Ionescu (2012): "Insuring student loans against the financial risk of failing to complete college," *Quantitative Economics*, 3, 393–420.

Chen, T. and C. Guestrin (2016): "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Clark, D., D. Gill, V. Prowse, and M. Rush (2020): "Using goals to motivate college students: Theory and evidence from field experiments," *Review of Economics and Statistics*, 102, 648–663.

Cyrenne, P. and A. Chan (2012): "High school grades and university performance: A case study," *Economics of Education Review*, 31, 524–542.

Davis, J. and M. Goadrich (2006): "The relationship between Precision-Recall and ROC curves," in *Proceedings of the 23rd international conference on Machine learning*, 233–240.

de Brey, C., A. Zhang, and S. Duffy (2021): "Digest of Education Statistics, 2021," .

Deary, I. J., S. Strand, P. Smith, and C. Fernandes (2007): "Intelligence and educational achievement," *Intelligence*, 35, 13–21.

DellaVigna, S. and U. Malmendier (2006): "Paying not to go to the gym," *american economic Review*, 96, 694–719.

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson (1988): "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, 837–845.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011): "Individual Risk Attitudes: Measurement, Determinants, And Behavioral Consequences," *Journal of the European Economic Association*, 9, 522–550.

Doshi-Velez, F. and B. Kim (2017): "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly (2007): "Grit: perseverance and passion for long-term goals." *Journal of Personality and Social Psychology*, 92, 1087–1101.

Duckworth, A. L., P. D. Quinn, and E. Tsukayama (2012): "What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades." *Journal of Educational Psychology*, 104, 439–451.

Duckworth, A. L., A. Quirk, R. Gallop, R. H. Hoyle, D. R. Kelly, and M. D. Matthews (2019): "Cognitive and noncognitive predictors of success," *Proceedings of the National Academy of Sciences*, 116, 23499–23504.

Duckworth, A. L. and M. E. Seligman (2005): "Self-discipline outdoes IQ in predicting academic performance of adolescents," *Psychological Science*, 16, 939–944.

Dweck, C. S. (2006): *Mindset: The new psychology of success*, Random house.

Dynarski, S., L. Page, and J. Scott-Clayton (2023): "College costs, financial aid, and student decisions," in *Handbook of the Economics of Education*, Elsevier, vol. 7, 227–285.

Epper, T., E. Fehr, H. Fehr-Duda, C. T. Kreiner, D. D. Lassen, S. Leth-Petersen, and G. N. Rasmussen (2020): "Time Discounting and Wealth Inequality," *American Economic Review*, 110, 1177–1205.

Epper, T., E. Fehr, K. B. Hvidberg, C. T. Kreiner, S. Leth-Petersen, and G. N. Rasmussen (2022): "Preferences predict who commits crime among young men," *Proceedings of the National Academy of Sciences*, 119, e2112645119.

Epper, T., H. Fehr-Duda, and A. Bruhin (2011): "Viewing the Future through a Warped Lens: Why Uncertainty Generates Hyperbolic Discounting," *Journal of Risk and Uncertainty*, 43, 169–203.

Epper, T., A. Koch, and J. Nafziger (2018): "Aarhus Survey," SSRN Working Paper No. 3123666.

Eskreis-Winkler, L., A. L. Duckworth, E. P. Shulman, and S. Beal (2014): "The grit effect: Predicting retention in the military, the workplace, school and marriage," *Frontiers in Psychology*, 5, 36.

Farquhar, P. H. (1984): "Utility Assessment Methods," *Management Science*, 30, 1283–1300.

Fe, E., D. Gill, and V. Prowse (2022): "Cognitive skills, strategic sophistication, and life outcomes," *Journal of Political Economy*, 130, 2643–2704.

Fehr, E. and L. Goette (2007): "Do workers work more if wages are high? Evidence from a randomized field experiment," *American Economic Review*, 97, 298–317.

Figlio, D., P. Giuliano, U. Özek, and P. Sapienza (2019): "Long-term orientation and educational performance," *American Economic Journal: Economic Policy*, 11, 272–309.

Fisher, A., C. Rudin, and F. Dominici (2019): "All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously," *Journal of Machine Learning Research*, 20, 1–81.

Frederick, S. (2005): "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 25–42.

Gächter, S., E. J. Johnson, and A. Herrmann (2022): "Individual-level loss aversion in riskless and risky choices," *Theory and Decision*, 92, 599–624.

Gillen, B., E. Snowberg, and L. Yariv (2019): "Experimenting with measurement error: Techniques with applications to the caltech cohort study," *Journal of Political Economy*, 127, 1826–1863.

Golsteyn, B. H., H. Grönqvist, and L. Lindahl (2014): "Adolescent time preferences predict lifetime outcomes," *Economic Journal*, 124, F739–F761.

Goodman, R., H. Meltzer, and V. Bailey (1998): "The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version," *European child & adolescent psychiatry*, 7, 125–130.

Heath, C. and J. B. Soll (1996): "Mental Budgeting and Consumer Decisions," *Journal of Consumer Research*, 23, 40–52.

HECKMAN, J. J., J. E. HUMPHRIES, AND G. VERAMENDI (2018): "Returns to education: The causal effects of education on earnings, health, and smoking," *Journal of Political Economy*, 126, S197–S246.

HECKMAN, J. J., J. STIXRUD, AND S. URZUA (2006): "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labor Economics*, 24, 411–482.

HERSHBEIN, B., M. S. KEARNEY, AND L. W. PARDUE (2020): "College attainment, income inequality, and economic security: A simulation exercise," in *AEA Papers and Proceedings*, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, vol. 110, 352–355.

HORN, J. L. (1965): "A Rationale and Test for the Number of Factors in Factor Analysis," *Psychometrika*, 30, 179–185.

HVIDBERG, K. B., C. T. KREINER, AND S. STANTCHEVA (2023): "Social Positions and Fairness Views on Inequality," *The Review of Economic Studies*, 90, 3083–3118, rdad019.

HVIDMAN, C., A. K. KOCH, J. NAFZIGER, S. A. NIELSEN, AND M. ROSHOLM (2024): "An intensive, school-based learning camp targeting academic and non-cognitive skills evaluated in a randomized trial," *Labour Economics*, 88, 102535.

JOENSEN, J. S. AND E. MATTANA (2021): "Student Aid Design, Academic Achievement, and Labor Market Behavior: Grants or Loans?" working paper, University of Chicago.

JUDGE, T. A., A. EREZ, J. E. BONO, AND C. J. THORESEN (2003): "The core self-evaluations scale: Development of a measure," *Personnel psychology*, 56, 303–331.

JURAJDA, Š. AND D. MÜNICH (2011): "Gender gap in performance under competitive pressure: Admissions to Czech universities," *American Economic Review*, 101, 514–518.

Kahneman, D. and A. Tversky (1984): "Choices, Values, and Frames," *American Psychologist*, 39, 341–350.

Kalamova, M. M., S. Roy, C. S. Sarrico, and T. Weko (2020): *Resourcing Higher Education: Challenges, Choices, and Consequences*, OECD Publishing.

Karle, H., D. Engelmann, and M. Peitz (2022): "Student performance and loss aversion," *Scandinavian Journal of Economics*, 124, 420–456.

Köbberling, V. and P. Wakker (2005): "An Index of Loss Aversion," *Journal of Economic Theory*, 112, 119–131.

Koch, A. K. and J. Nafziger (2016): "Goals and Bracketing under Mental Accounting," *Journal of Economic Theory*, 162, 305–351.

——— (2019): "Correlates of narrow bracketing," *Scandinavian Journal of Economics*, 121, 1441–1472.

Kunz, J. S. and K. E. Staub (2020): "Early subjective completion beliefs and the demand for post-secondary education," *Journal of Economic Behavior & Organization*, 177, 34–55.

Lange, F. and R. Topel (2006): "The social value of education and human capital," *Handbook of the Economics of Education*, 1, 459–509.

LeCun, Y., Y. Bengio, and G. Hinton (2015): "Deep learning," *Nature*, 521, 436–444.

Lira, B., J. M. O'Brien, P. A. Peña, B. M. Galla, S. D'Mello, D. S. Yeager, A. Defnet, T. Kautz, K. Munkacsy, and A. L. Duckworth (2022): "Large studies reveal how reference bias limits policy applications of self-report measures," *Scientific reports*, 12, 19189.

Locke, E. A. and G. P. Latham (1990): *A theory of goal setting & task performance*, Englewood Cliffs, NJ: Prentice-Hall, Inc.

LUNDBERG, S. M., G. ERION, H. CHEN, A. DEGRAVE, J. M. PRUTKIN, B. NAIR, R. KATZ, J. HIMMELFARB, N. BANSAL, AND S.-I. LEE (2020): "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, 2, 56–67.

LUNDBERG, S. M. AND S.-I. LEE (2017): "A unified approach to interpreting model predictions," *Advances in Neural Information processing systems*, 30.

MULLAINATHAN, S. AND J. SPIESS (2017): "Machine learning: an applied econometric approach," *Journal of Economic Perspectives*, 31, 87–106.

NAGEL, R. (1995): "Unraveling in guessing games: An experimental study," *The American Economic Review*, 85, 1313–1326.

NICHOLS, M., J. R. TRIVITT, AND G. ZAMARRO (2018): "College readiness, student expectations and success: The role of non-cognitive skills," EDRE Working Paper No. 2018-04.

NIEDERLE, M. AND L. VESTERLUND (2007): "Do Women Shy Away from Competition? Do Men Compete Too Much?" *Quarterly Journal of Economics*, 122, 1067–1101.

OECD (2022): "Education at a Glance," OECD Publishing, Paris.

OREOPOULOS, P. AND U. PETRONIJEVIC (2013): "Making college worth it: A review of research on the returns to higher education," *National Bureau of Economic Research, Working Paper 19053*.

——— (2019): "The remarkable unresponsiveness of college students to nudging and what we can learn from it," Tech. rep., National Bureau of Economic Research.

OREOPOULOS, P. AND K. G. SALVANES (2011): "Priceless: The nonpecuniary benefits of schooling," *Journal of Economic perspectives*, 25, 159–184.

ORS, E., F. PALOMINO, AND E. PEYRACHE (2013): "Performance gender gap: does competition matter?" *Journal of Labor Economics*, 31, 443–499.

Ost, B., W. Pan, and D. Webber (2018): "The returns to college persistence for marginal students: Regression discontinuity evidence from university dismissal policies," *Journal of Labor Economics*, 36, 779–805.

Patil, V. H., S. N. Singh, S. Mishra, and D. T. Donavan (2008): "Efficient Theory Development and Factor Retention Criteria: Abandon the 'Eigenvalue Greater than One'-criterion," *Journal of Business Research*, 61, 162–170.

Patnaik, A., M. Wiswall, and B. Zafar (2021): "College majors 1," *The Routledge handbook of the economics of education*, 415–457.

Pekkarinen, T. (2014): "Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations," *Journal of Economic Behavior & Organization*, 115, 94–110.

Peña-Ayala, A. (2014): "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert systems with applications*, 41, 1432–1462.

Rabin, M. and G. Weizsäcker (2009): "Narrow bracketing and dominated choices," *American Economic Review*, 99, 1508–43.

Rainio, O., J. Teuho, and R. Klén (2024): "Evaluation metrics and statistical tests for machine learning," *Scientific Reports*, 14, 6086.

Rector's Office (2024): "AU in the World Rankings," https://international.au.dk/about/profile/rankings/.

Reuben, E., M. Wiswall, and B. Zafar (2017): "Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender," *Economic Journal*, 127, 2153–2186.

Rooth, D.-O. (2011): "Work out or out of work—the labor market return to physical fitness and leisure sports activities," *Labour Economics*, 18, 399–409.

Rothstein, J. M. (2004): "College performance predictions and the SAT," *Journal of Econometrics*, 121, 297–317.

SALTIEL, F. (2020): "Gritting it out: The importance of non-cognitive skills in academic mismatch," *Economics of Education Review*, 78, 102033.

SCHOLZ, J. K. AND K. SICINSKI (2015): "Facial attractiveness and lifetime earnings: Evidence from a cohort study," *Review of Economics and Statistics*, 97, 14–28.

SHAPLEY, L. S. (1953): "17. A value for n-person games," *Contributions to the theory of games (AM-28), Volume II*, 307–318.

SHODA, Y., W. MISCHEL, AND P. K. PEAKE (1990): "Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions." *Developmental Psychology*, 26, 978.

STINEBRICKNER, R. AND T. STINEBRICKNER (2014a): "Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model," *Journal of Labor Economics*, 32, 601–644.

STINEBRICKNER, R. AND T. R. STINEBRICKNER (2014b): "A major in science? Initial beliefs and final outcomes for college major and dropout," *Review of Economic Studies*, 81, 426–472.

STONE, M. (1974): "Cross-validatory choice and assessment of statistical predictions," *Journal of the royal statistical society: Series B (Methodological)*, 36, 111–133.

SYDNOR, J. (2010): "(Over) insuring modest risks," *American Economic Journal: Applied Economics*, 2, 177–199.

TANGNEY, J. P., R. F. BAUMEISTER, AND A. L. BOONE (2004): "High self-control predicts good adjustment, less pathology, better grades, and interpersonal success," *Journal of Personality*, 72, 271–324.

TAYLOR, M. P. (2020): "Heterogeneous motivation and cognitive ability in the lab," *Journal of Behavioral and Experimental Economics*, 85, 101523.

TIBSHIRANI, R. (1996): "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288.

TINTO, V. (2012): *Leaving college: Rethinking the causes and cures of student attrition*, University of Chicago press.

TSUKAYAMA, E., A. L. DUCKWORTH, AND B. KIM (2013): "Domain-specific impulsivity in school-age children," *Developmental Science*, 16, 879–893.

TVERSKY, A. AND D. KAHNEMAN (1981): "The Framing of Decisions and the Psychology of Choice," *Science, New Series*, 211, 453–458.

VAN VELDHUIZEN, R. (2022): "Gender Differences in Tournament Choices: Risk Preferences, Overconfidence, or Competitiveness?" *Journal of the European Economic Association*, 20, 1595–1618.

VARMA, S. AND R. SIMON (2006): "Bias in error estimation when using cross-validation for model selection," *BMC bioinformatics*, 7, 1–8.

VISCHER, T., T. DOHMEN, A. FALK, D. HUFFMAN, J. SCHUPP, U. SUNDE, AND G. G. WAGNER (2013): "Validating an ultra-short survey measure of patience," *Economics Letters*, 120, 142–145.

WEBBER, D. A. (2016): "Are college costs worth it? How ability, major, and debt affect the returns to schooling," *Economics of Education Review*, 53, 296–310.

WILLIAMS, S. H. (2008): "Sticky expectations: Responses to persistent over-optimism in marriage, employment contracts, and credit card use," *Notre Dame L. Rev.*, 84, 733.

WISWALL, M. AND B. ZAFAR (2015): "Determinants of college major choice: Identification using an information experiment," *The Review of Economic Studies*, 82, 791–824.

——— (2021): "Human capital investments and expectations about career and family," *Journal of Political Economy*, 129, 1361–1424.

Yu, R., Q. Li, C. Fischer, S. Doroudi, and D. Xu (2020): "Towards accurate and fair prediction of college success: evaluating different sources of student data," in *Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020)*, ERIC, 292–301.

Zimmerman, S. D. (2014): "The returns to college admission for academically marginal students," *Journal of Labor Economics*, 32, 711–754.

Appendix

A.1 Elicitation of risk preferences

We elicited certainty equivalents (ce) for nine binary lotteries of the form $(x_1, p_1; x_2)$: the lottery pays x_1 with probability p_1 and x_2 otherwise.

The lotteries were presented in three different outcome domains (gain, loss, mixed) and grouped so that they imply equivalent *terminal outcomes* $w_1 = w + x_1$ and $w_2 = w + x_2$. We refer to these groupings as "lottery configurations" (see Table A.1).

Table A.1: Lottery Configurations

Lottery Configuration 1							
Domain	ID	w	x_1	x_2	p_1	w_1	w_2
loss	1	80	0	-80	0.50	80	0
mixed	2	40	40	-40	0.50	80	0
gain	3	0	80	0	0.50	80	0
Lottery Configuration 2							
Domain	ID	w	x_1	x_2	p_1	w_1	w_2
loss	4	160	-40	-120	0.50	120	40
mixed	5	80	40	-40	0.50	120	40
gain	6	0	120	40	0.50	120	40
Lottery Configuration 3							
Domain	ID	w	x_1	x_2	p_1	w_1	w_2
loss	7	160	0	-160	0.50	160	0
mixed	8	80	80	-80	0.50	160	0
gain	9	0	160	0	0.50	160	0

Upon entering this part of the survey, students faced nine choice lists. In each choice list, they made a sequence of binary choices between alternatives A and B. Under A, students got an amount of money for sure. Under B, the amount of money they received was uncertain, and it was equally likely that the student received DKK x_1

or DKK x_2, where $x_1 > x_2$. Each choice list contained 21 rows. Alternative B was fixed within a list, while the sure amount under A increased from x_2 to x_1 across rows. We induced a unique switching point by asking students to mark the lowest sure amount of money at which they preferred A over B.[29]

Some lotteries involved losses. Therefore, each question included an endowment of DKK w (shown at the top of the page): any loss was deducted from w and any gain was added.

Incentives followed standard procedures: at the end of the experiment, we drew (i) one list to pay at random and (ii) one of the 21 rows to count. If the student had chosen A in that row, they received the sure amount; if they had chosen B, the computer randomly selected x_1 or x_2. In addition, the student received the endowment w for that list.

We derive a single nonparametric index of loss aversion from the switching behavior in the mixed-domain lotteries (lottery IDs 2, 5, and 8) (for related approaches see Fehr and Goette (2007) and Gächter et al. (2022)).

Let ce denote the certainty equivalent implied by the switching point in a given choice list. We compute the *relative switching point*:

$$\mathrm{sp} = \frac{\mathrm{ce} - x_2}{x_1 - x_2}. \tag{1}$$

This normalization implies that $\mathrm{sp} = 1$ when $\mathrm{ce} = x_1$ and sp=0 when $\mathrm{ce} = x_2$.

We then apply a monotone transformation to obtain a normalized loss-aversion index for each mixed lottery, defined as $\eta = 1 - 2 \cdot \mathrm{sp}$. This transformation maps the switching point onto a symmetric scale such that $\eta = 1$ indicates maximum loss aversion, $\eta = -1$ indicates maximum gain seeking, and $\eta = 0$ indicates neither loss aversion nor gain seeking. Intuitively, switching earlier to the sure option (a lower value of sp implies higher loss aversion).

[29] A screenshot of the task is part of Appendix A.9.

Because we elicit the mixed lottery once in each of the three configurations (lottery IDs 2, 5, and 8), we construct a more robust measure of loss aversion by aggregating across these three observations. Specifically, we define the average index as $\bar{\eta} = \frac{1}{3}(\eta_{id=2} + \eta_{id=5} + \eta_{id=8})$. This aggregation reduces the influence of idiosyncratic reporting and errors at the single-choice-level list, since such errors are partly averaged out across configurations. In our analysis, we use this aggregated index $\bar{\eta}$ as our measure of loss aversion.

We also exploit the full set of nine lotteries to derive parametric (structural) measures of loss aversion by estimating prospect-theory value functions following Köbberling and Wakker (2005), using individual-level maximum likelihood estimation as in Epper et al. (2011). We consider two specifications: a piece-wise linear specification of the value function where we only estimate the loss aversion parameter (specification 1), and a flexible specification which, in addition, estimates value function curvature parameters for both gains and losses (specification 2).

For subjects whose estimation converges, we assess concordance between the estimated loss-aversion parameter λ and our nonparametric index $\bar{\eta}$ using Spearman rank-correlation tests. The resulting rank correlations are 0.676 (p-value of approximately zero) for specification 1, and 0.998 (p-value of approximately zero) for specification 2, respectively.

A.2 Predictive models

We carry out a comprehensive analysis by comparing the performance of five distinct machine learning algorithms. First, we include a **standard logistic regression model**. Due to its simplicity and transparency, this model is widely used in binary classification problems. Despite its linear nature, it often performs well for relatively simple tasks with a limited number of variables, henceforth referred to as *features*—in line with the machine learning lingo. This model serves as a baseline for our comparison, allowing us to assess the performance of more advanced algorithms in relation to this standard method.

Second, we apply **LASSO** (Least Absolute Shrinkage and Selection Operator), an extension of linear regression that incorporates regularization techniques to address overfitting (Tibshirani, 1996). LASSO applies a penalty term that shrinks coefficients, encouraging sparsity and leading to a simpler, more interpretable model. This means that variables with little predictive power may have their coefficients set exactly to zero and be effectively removed from the model. This algorithm is particularly useful for handling high-dimensional data and addressing multicollinearity.

Third, moving towards a more flexible and non-linear model, we examine the **neural network algorithm** (LeCun et al., 2015). Inspired by the structure of the human brain, neural networks consist of layers of interconnected units known as 'nodes' that transform inputs in a hierarchical manner. It is particularly effective for capturing complex, non-linear relationships in high-dimensional data. While the above methods employ linear models, neural networks can be imagined as a vast interconnected system of equations where each of the equations is fine-tuned.

Fourth, we include the **random forest algorithm** (Breiman, 2001). Rather than relying on a single algorithm, it is an ensemble learning method that constructs multiple decision trees to improve the overall prediction. Each tree is trained on a random subset of the data and considers a randomly chosen subset of variables at each split. It is notable for its robustness against overfitting and its capability of handling both categorical and numerical variables well.

Fifth, we include the **XGBoost**, as described in the text.

Suppose we consider three variables: high school GPA, motivation, and risk aversion, each with varying predictive importance. For the logistic regression and LASSO, each variable gets a coefficient based on its influence on dropout prediction. Logistic regression linearly combines these coefficients with the actual variable values for each student. LASSO does the same but with a twist: it penalizes less influential variables and potentially eliminates them from the model. For instance, if risk aversion has no predictive power in terms of dropout risk, the model may assign its coefficient a value of zero and remove it.

Neural networks take a different approach. They consist of multiple layers of nodes, each transforming the data and passing information forward. Instead of explicitly defining interactions, the model learns them automatically through hidden layers. For example, it may implicitly capture how GPA and motivation jointly influence dropout risk, even if we do not specify this interaction in advance. As data flows through these layers, the network gains a comprehensive understanding of how GPA, motivation, and risk aversion, individually and in combination, influence dropout risk.

Random forest constructs multiple decision trees, each trained on a random subset of the data. For instance, some trees might primarily capture the impact of GPA, while others focus on the relationship between motivation and risk aversion. By aggregating the predictions of all trees, the algorithm manages to encapsulate varied patterns within the data. Building on the concept of tree-based models, XGBoost builds decision trees sequentially, with each new tree correcting errors made by the previous ones. For instance, if an initial tree misinterprets the dropout likelihood for students characterized by low GPA but high motivation, subsequent trees in the sequence strive to solve this error.

A.3 Nested cross-validation

Nested cross-validation (CV) involves the use of two levels of cross validation: an outer loop and an inner loop. The outer loop estimates the performance of the model on unseen data, while the inner loop optimizes the model's hyperparameters. This separation ensures that the performance estimation is independent of the tuning process. More specifically, in the outer loop, the data are divided into multiple folds, using k-fold CV. The model is trained on a subset of the data and evaluated on the remaining fold. This process is repeated for each fold, yielding performance metrics for each iteration. In the inner loop, within each fold of the outer loop, a separate CV is performed to fine-tune the model's hyperparameters. This ensures that the model is optimized for each subset of the data, further enhancing its predictive capability. We use five outer as well as inner folds to balance computational efficiency with rigorous model evaluation.

A.4 Hyperparameter tuning

Hyperparameters are tuned using random search within the inner loop of the nested cross-validation procedure described in Appendix A.3. Random search is well suited to high-dimensional tuning problems, where exhaustive grid search quickly becomes computationally infeasible, and has been shown to identify high-performing configurations efficiently (Bergstra and Bengio, 2012). For each model and each outer fold, 20 hyperparameter configurations are randomly sampled and evaluated using five-fold cross-validation within the inner loop. The configuration that maximizes the inner-loop performance metric (AUC) is selected and subsequently used to train the model on the full outer-fold training sample. The fitted model is then applied to the held-out outer-fold test data to generate out-of-sample predictions.

All models are estimated using the off-the-shelf implementations provided by the `caret` package in R, which handles cross-validation, random hyperparameter search, and model fitting in a unified framework.

Table A.2 summarizes the hyperparameters that are tuned for each algorithm. The random search procedure relies on caret's default hyperparameter generators for each algorithm. Accordingly, we report the set of tuned hyperparameters rather than imposing or claiming fixed tuning bounds.

To illustrate the resulting heterogeneity in selected hyperparameters, Table A.3

Table A.2: Tuned hyperparameters by algorithm

Algorithm	Tuned hyperparameters
Logistic regression (glm)	None
LASSO/Elastic net (glmnet)	Mixing parameter α Penalty parameter λ
Neural network (nnet)	Number of hidden units (`size`) Weight decay (`decay`)
Random forest (rf)	Number of variables considered at each split (`mtry`)
XGBoost (xgbTree)	Learning rate (η) Maximum tree depth (`max_depth`) Regularization parameter (γ) Column subsampling ratio (`colsample`) Minimum child weight (`min_child`) Subsampling ratio (`subsample`) Number of boosting rounds (`nrounds`)

Notes: Hyperparameters are tuned via random search within the inner loop of the nested cross-validation procedure using five-fold cross-validation. For each model and outer fold, twenty randomly sampled hyperparameter configurations are evaluated in the inner loop, and the configuration that maximizes inner-loop AUC is selected.

reports the XGBoost hyperparameter configurations selected by inner-loop cross-validation for each information set and outer fold in the program dropout analysis. This table highlights that optimal configurations vary across both folds and information sets, underscoring the importance of fold-specific tuning within the nested cross-validation framework.

A.5 Model performance

A.5.1 Background on Receiver Operating Characteristic and Area Under the Curve

In this section we define a few concepts that are relevant for evaluating model performance in a binary classification problem like ours. A model classifies cases in the data as either positive or negative, in our setting ‘dropout’ or ‘non-dropout’. Such a

Table A.3: XGBoost hyperparameters selected by inner-loop cross-validation (program dropout)

Information set	Fold	η	max_depth	γ	colsample	min_child	subsample	nrounds
	1	0.020	5	6.29	0.443	9	0.492	676
	2	0.101	2	6.44	0.686	13	0.871	87
GPA	3	0.147	10	5.54	0.645	0	0.681	644
	4	0.147	1	5.54	0.645	5	0.787	974
	5	0.147	10	5.54	0.645	0	0.681	644
	1	0.250	8	8.90	0.466	4	0.960	664
	2	0.006	6	6.93	0.484	2	0.405	779
Basic	3	0.118	7	3.43	0.406	6	0.721	300
	4	0.227	8	2.35	0.647	11	0.897	207
	5	0.174	8	4.23	0.642	1	0.518	600
	1	0.031	7	6.93	0.588	5	0.576	847
	2	0.096	10	9.47	0.309	2	0.309	498
Backgr	3	0.482	8	8.04	0.509	2	0.822	721
	4	0.343	5	2.70	0.640	2	0.775	16
	5	0.482	8	8.04	0.509	2	0.822	721
	1	0.031	7	6.93	0.588	5	0.576	847
	2	0.116	8	5.48	0.426	7	0.859	540
Expect	3	0.018	7	7.52	0.369	0	0.766	259
	4	0.295	10	7.50	0.602	3	0.646	520
	5	0.018	7	7.52	0.369	0	0.766	259
	1	0.031	7	6.93	0.588	5	0.576	847
	2	0.046	5	4.73	0.684	1	0.843	618
Noncog	3	0.381	1	3.73	0.458	2	0.882	899
	4	0.272	9	4.75	0.694	1	0.546	786
	5	0.381	1	3.73	0.458	2	0.882	899
	1	0.031	7	6.93	0.588	5	0.576	847
	2	0.029	7	6.70	0.478	0	0.640	766
All	3	0.270	9	9.53	0.675	8	0.835	829
	4	0.029	7	6.70	0.478	0	0.640	766
	5	0.270	9	9.53	0.675	8	0.835	829

Notes: For each information set and outer fold, the table reports the XGBoost hyperparameter configuration selected by inner-loop cross-validation (five-fold CV, random search with 20 candidate configurations). Information-set definitions follow Table 1.

classification can correctly label cases where dropout occurs as true positives or correctly label cases where non-dropout occurs as true negatives. False positives occur when non-dropout students are incorrectly classified as dropouts, while false negatives occur when actual dropout students are incorrectly classified as non-dropouts.

The Receiver Operating Characteristic (ROC) visualizes the tradeoff between true positives and false positives. The x-axis of the ROC represents the false positive rate, which captures the proportion of non-dropout cases incorrectly classified as dropout. The y-axis represents the true positive rate, which is the proportion of actual dropouts correctly identified by the model. Each point thus relates true positives to false positives at a given classification threshold.

The Area Under the Curve (AUC) presents an aggregate measure of performance by aggregating across all possible classification thresholds on the ROC curve. Intuitively, it captures the probability that the model ranks a randomly drawn student *who drops out* higher in terms of dropout risk than a randomly drawn student *who does not drop out*. The range of the AUC is from 0 to 1, where 0.5 corresponds to random ranking, 1 corresponds to perfect ranking, and values below 0.5 indicate that the model ranks dropouts systematically lower than non-dropouts (i.e., worse than random).

While there is a substantial share of students who drop out at the program and institution level, only 10.1 percent of students drop out at the system level. In situations where one class is much more frequently observed than the other, a low false positive rate may reflect the abundance of correctly classified non-dropouts rather than strong dropout detection, which can mask poor performance on the minority class (dropouts).

Precision-recall curves and their associated PRAUC (Precision Recall Area Under the Curve) metric provide a robust alternative criterion when data are imbalanced (e.g., Davis and Goadrich, 2006). Precision is the proportion of correctly predicted dropouts out of all predicted dropouts. Recall measures the proportion of true positives that are correctly identified by the model. By comparing precision and recall, PRAUC emphasizes the model's ability to identify the minority class (dropouts), ensuring that dropout predictions are not overshadowed by the large number of non-dropout cases.

Therefore, as a robustness test, in Appendix A.5.2.1 we use the PRAUC metric to corroborate the conclusions drawn from the AUC analysis in Section 5.2.

A.5.2 Predictive performance

A.5.2.1 Predicting dropout

PRAUC. To further corroborate the conclusions drawn from the AUC analysis, in Section 5.2, we also consider the PRAUC (Precision Recall Area Under the Curve) for the different predictive models as a robustness check—a metric which performs well with imbalanced data (see Appendix A.5.1). Figure A.1 largely mirrors the findings of Figure 2.

For *program dropout* (top panel of Figure A.1), adding expectations and study motivation (Expect) yields the largest improvement relative to Basic. The broader non-cognitive information set (Noncog) also increases PRAUC relative to Basic, but the gain is more modest than for Expect. The full feature set (All) performs best overall.

For *institution dropout* (middle panel of Figure A.1), the main gain comes from moving from the GPA-only baseline to Basic. Adding Expect or All yields small additional improvements beyond Basic, whereas Backgr and Noncog achieve lower PRAUC than Basic, consistent with these variables adding noise or overfitting in this setting.

For *system dropout* (bottom panel of Figure A.1), PRAUC levels are low across all specifications, reflecting the rarity of the outcome. The full feature set (All) performs best. No other information set delivers an improvement relative to Basic.

Comparison of machine learning algorithms. In this section, we compare the predictive performance of the five machine learning algorithms—a standard logistic regression model, LASSO, neural network, random forest, and XGBoost. These models vary in their ability to capture linear and non-linear relationships. Nevertheless, Figures A.2 - A.5 (AUC) and Figures A.7 - A.9 (PRAUC) indicate that the algorithms generally perform similarly. However, Figures A.6 (AUC) and A.11 (PRAUC) reveal that logistic regression is somewhat worse when incorporating many features compared to the performance of the other algorithms.

A.5.3 Identification of at-risk students

Lift curve. Figure A.15 complements the analysis in Section 5.3 by showing the corresponding lift curve. The lift curve plots the cumulative share of all program dropouts captured as one moves through the sample, ranked from highest to lowest predicted dropout risk. A steeper curve indicates better model performance, while a line tracking the upper boundary of the gray box would represent perfect prediction. For example, targeting the 20 percent of students with the highest predicted risk identifies 39 percent of all dropouts. Likewise, if we were to target the 100 students with the highest predicted risk (approximately 16 percent of the sample), 59 of them would be actual dropouts, corresponding to around 33 percent of all observed dropouts.

Comparison of machine learning algorithms. Figures A.12 $-$ A.14 plot the response rates for all machine learning models for the different categories of dropout. Figure A.12 shows that there are similar patterns across models as the ones for XGBoost shown in Figure 3.

Figure A.13 considers *institution dropout*. Overall, 21.6 percent of students drop out of the university. A perfect model therefore would have assigned all actual dropouts to the two highest risk deciles 9 and 10. We actually see that around 45 percent of those assigned to the highest risk decile 10 do drop out of the university, but the different machine learning models vary in terms of their performance. The logistic regression model performs relatively poorly for the highest risk decile 10 and deviates also from the ideal staircase pattern of the response rate.

Figure A.14 considers *system dropout*, i.e., leaving higher education without any bachelor's degree. Overall, 10.1 percent of students do so. A perfect model therefore would have assigned them to the highest risk decile 10. In that case the response rate would have been 75 percent in decile 10. The actual response rate is around 20 percent in that decile. Again, model performance varies. The logistic regression performs poorly for the top decile and it does not produce the staircase pattern in response rates across deciles. LASSO also faces the latter issue.

Appendix Figures A.16 and A.17 show the corresponding lift curves for *institution* and *system* dropout, respectively. The interpretation is analogous to Figure A.15. For institution dropout, targeting the 20 percent of students with the highest predicted risk identifies 39 percent of all dropouts, while targeting the 100 students with the highest predicted risk (about 16 percent of the sample) captures roughly 33 percent of all observed dropouts. For system dropout, the corresponding figures are 35 percent and 32 percent, respectively. In both cases, the models exhibit sizeable lift relative to random assignment, especially for program dropout, indicating that the predicted probabilities meaningfully rank students by their dropout risk.

Overall, these results suggest that machine learning models applied to our data are useful for predicting dropout at the program and institution level in the part of the data not used for training the model. The models perform less satisfactorily when it comes to system level dropout, though one should note that the number of 'positives' here is small as only 62 students leave without a bachelor's degree. Another possible reason is that dropout at the system level might be more likely to occur for very idiosyncratic reasons, such as unexpected illness.

A.5.4 Feature importance

Figure A.18 shows the ten features that contribute most to the prediction of *institution* and system dropout according to *SHAP* feature importance. Across all three dropout definitions, high school GPA is by far the most important feature. Its contribution is similar in magnitude across dropout categories.

For *institution dropout*, degree completion beliefs (*BeliefAUdegree*) is the second most important variable ($|SHAP| = 0.21$), while motivation is still ranked fourth for this outcome ($|SHAP| = 0.12$). This pattern is intuitive: the *BeliefAUdegree* item refers to completion at the university level and therefore relates more closely to institution dropout, whereas *Motivated* targets commitment to the specific study program. As expected, both measures are less important for *system dropout*, though *BeliefAUdegree* still shows up. In addition, *FamilyRecommended*, indicating whether family or friends recommended the student to choose the subject, appears to play a non-trivial role. Demographic variables such as *Age* and *Gap years* between high school and university also emerge as important predictors, and *BMI* contributes as well. This latter finding aligns with empirical studies linking physical attractiveness to labor market success (e.g., Rooth, 2011; Scholz and Sicinski, 2015).

For *system dropout*, *Age* emerges as the second most important feature ($|SHAP| = 0.27$), ahead of the already mentioned study-related self-regulation strategies and overconfidence. This could indicate that older students are generally more deliberate in their educational choices and less likely to re-enroll in another institution if they drop out of their studies. In addition, *BMI* also ranks among the more important features, together with *OnlyProgram*, which captures whether the student reported not knowing what else to study. The latter may reflect a lack of alternative educational pathways, potentially increasing the risk of fully exiting the system rather than switching institutions.

A.5.5 Permutation feature importance

In this section, we consider the other widely applied *Permutation Feature Importance* (PFI) measure.

Background. *PFI* quantifies how much the performance of a model is impacted when the values of a particular feature are randomly shuffled (Fisher et al., 2019). In this process, the original correlation or relationship between the feature and the target variable is effectively broken. The purpose of this shuffling is to create a scenario where the chosen feature essentially provides no informative content to the model. Any subsequent decrease in the performance of the model can be attributed to the importance of that feature in the original prediction. This process is repeated for all features in the dataset, one at a time. By doing this, we create a ranking of feature importance for a model.

A key issue when considering feature importance is how correlated features are taken into account. *SHAP* and *PFI* differ in the way they handle correlated features. In the case of *PFI*, permuting one of the highly correlated features might not decrease model performance much, because the other correlated feature can still provide the same information. This may result in the importance of these correlated features being overestimated: both of these features might be perceived as important even if one could be removed without significantly impacting the performance of the model. *SHAP*, on the other hand, splits importance between two features, rather than attributing all the importance to both features, which would double-count the importance of that information. However, the division of importance can lead to an underestimation of the total effect of a group of highly correlated features. If two features are carrying the same information, each may appear less important than when considered individually.

Results on feature importance. Figure A.19 shows the ten highest ranked variables according to *PFI* for XGBoost models using the full feature set ALL when predicting *program dropout*, *institution dropout*, and *system dropout*, respectively. The patterns are broadly consistent with the *SHAP* rankings reported in Figures 4 and A.18.

For *program dropout*, high school GPA is the most important predictor, followed by *BeliefAUdegree* and *Motivated*. Having worked between high school graduation and university entry (*Worked*) and the number of *Gap years* also contribute meaningfully relative to the remaining variables, although at noticeably lower levels than GPA and degree beliefs. The other predictors add comparatively little incremental explanatory power. Academic performance and expectations about degree completion therefore account for a substantial share of predictive performance at the program level.

For *institution dropout*, the dominance of high school GPA is even more pronounced, with a large gap to the second ranked variable, *BeliefAUdegree*. *Age* and *Motivated* follow at some distance. Institution dropout thus appears to be driven primarily by prior academic achievement, while beliefs and basic demographic characteristics again play secondary roles.

For *system dropout*, the ranking changes. *Age* is the most important predictor, slightly exceeding high school GPA. Commitment strategies, captured by *C4(CommitStrat)*, rank third and are clearly more important than the remaining variables. *BeliefAUdegree* contributes at a lower level of importance but still ranks fourth.

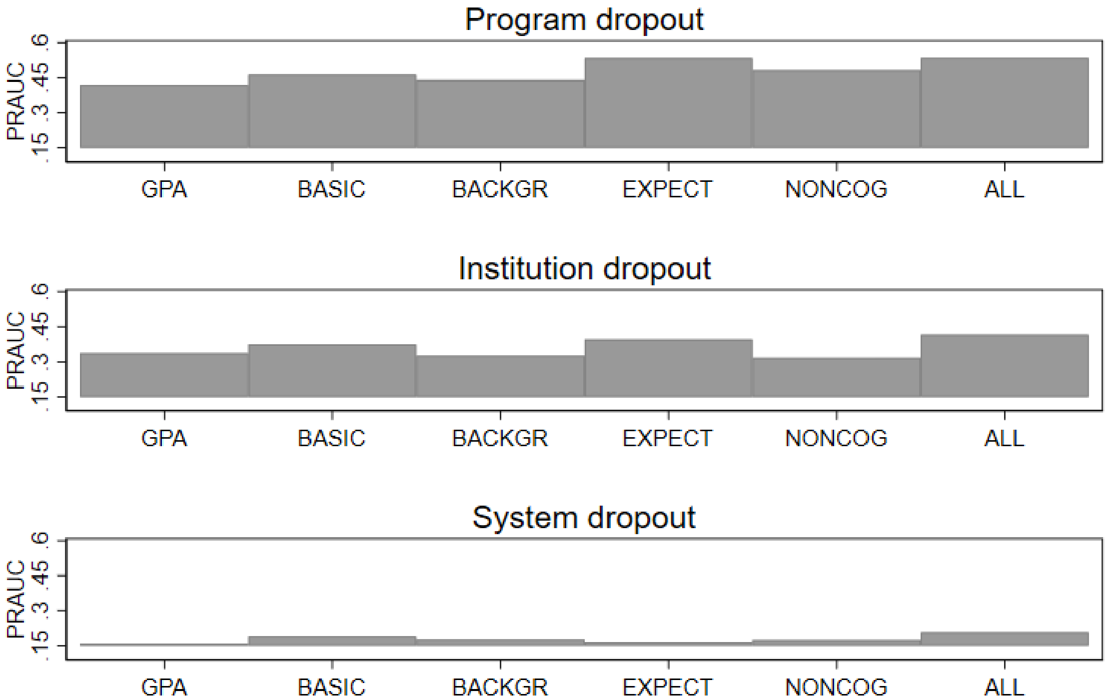
A.6 Selection into the survey

Figure A.20 shows that those who answer the survey have a lower dropout rate compared to the remainder of the cohort of first semester students at the faculty of Business and Social Sciences. Thus, whether or not a student responds to the survey is in itself predictive of dropout from the program that they start their studies in.

Non-respondents are more likely to be male and have slightly fewer gap years on average (see Table A.4). However, they do not differ in terms of share of foreign students or parental education (see Table A.5). In contrast, there are some differences in the composition of study programs across the two samples, which is due to the often small cell sizes (see Table A.4)

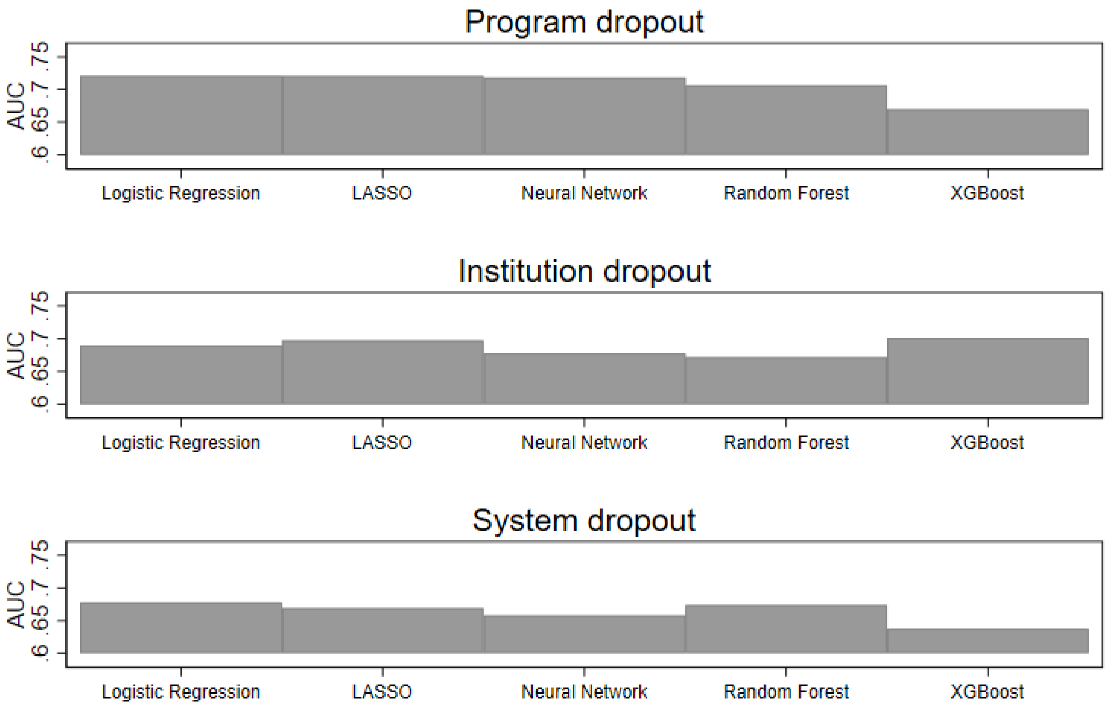
A.7 Additional figures

Figure A.1: PRAUC model performance with different sets of features (XGBoost)



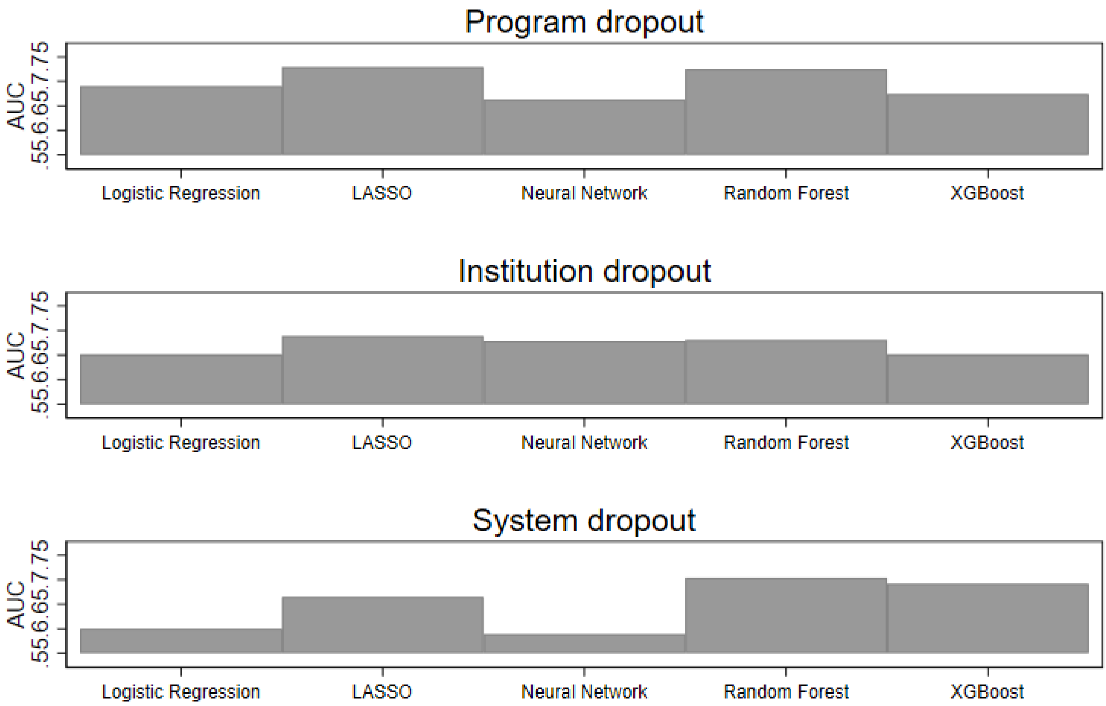
Notes: Model: XGBoost. All sets of features to the right of BASIC include the BASIC features (see Table 1). PRAUC (Precision Recall Area Under the Curve) captures the ability of a model to identify students who drop out. It takes values between 0 (none of the dropouts are correctly classified) and 1 (all of the dropouts are correctly classified).

Figure A.2: Model performance with the Basic information set



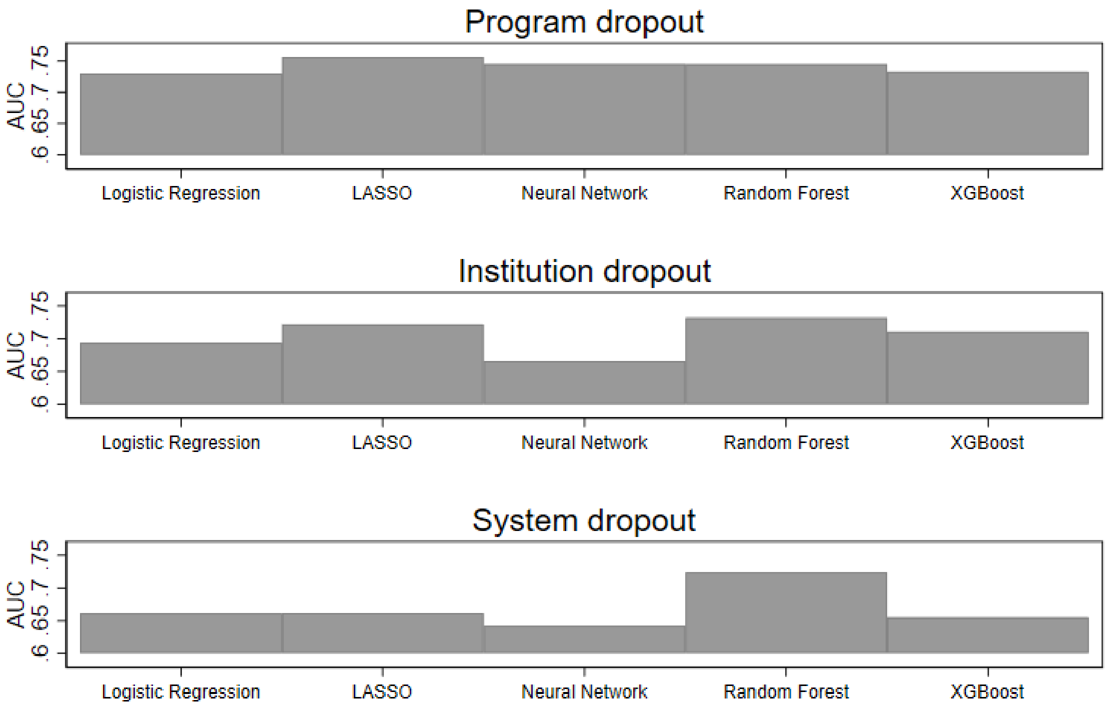
Notes: Models are evaluated using the Basic information set (see Table 1). AUC (area under the Receiver Operating Characteristic) captures the probability that the model ranks a randomly drawn student who drops out more highly in terms of dropout risk than a randomly drawn student who does not drop out. It takes values between 0 (all predictions are wrong) and 1 (all predictions are correct). The 95% confidence intervals are constructed using the percentile bootstrap method with 2000 replications applied to the AUC estimates.

Figure A.3: Model performance with the BACKGR information set



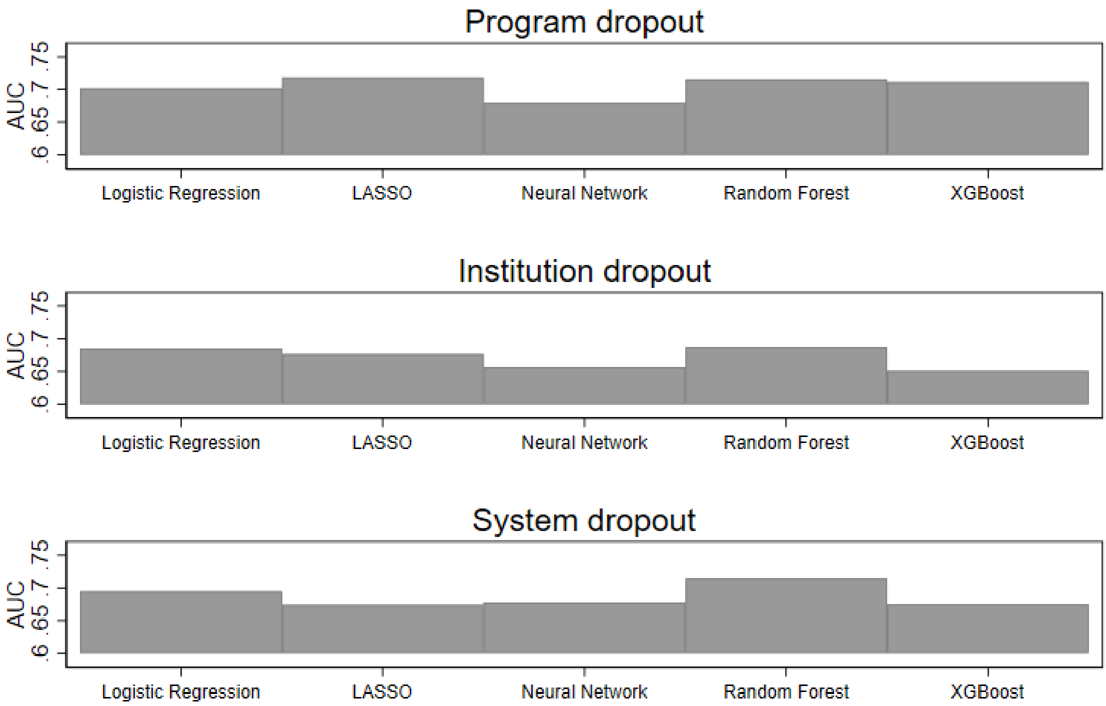
Notes: Models are evaluated using the BACKGR information set (see Table 1). AUC (area under the Receiver Operating Characteristic (ROC) curve) captures the probability that the model ranks a randomly drawn student who drops out more highly in terms of dropout risk than a randomly drawn student who does not drop out. It takes values between 0 (all predictions are wrong) and 1 (all predictions are correct).The 95% confidence intervals are constructed using the percentile bootstrap method with 2000 replications applied to the AUC estimates.

Figure A.4: Model performance with the Expect information set



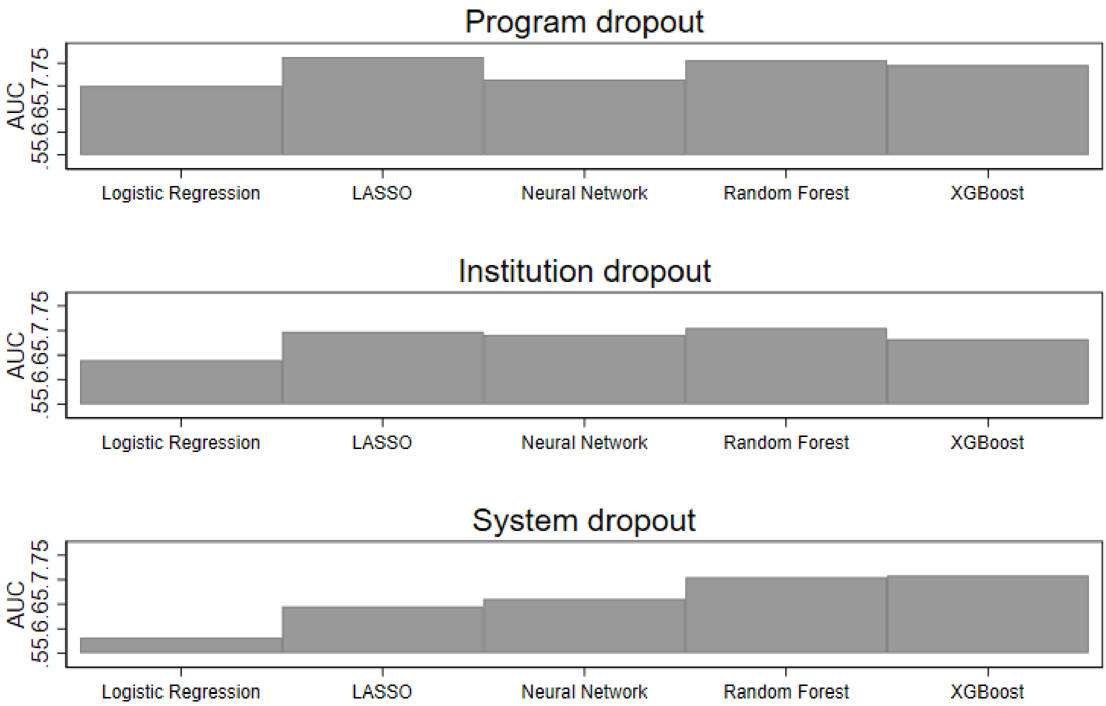
Notes: Models are evaluated using the Expect information set (see Table 1). AUC (area under the Receiver Operating Characteristic (ROC) curve) captures the probability that the model ranks a randomly drawn student who drops out more highly in terms of dropout risk than a randomly drawn student who does not drop out. It takes values between 0 (all predictions are wrong) and 1 (all predictions are correct). The 95% confidence intervals are constructed using the percentile bootstrap method with 2000 replications applied to the AUC estimates.

Figure A.5: Model performance with the NONCOG information set



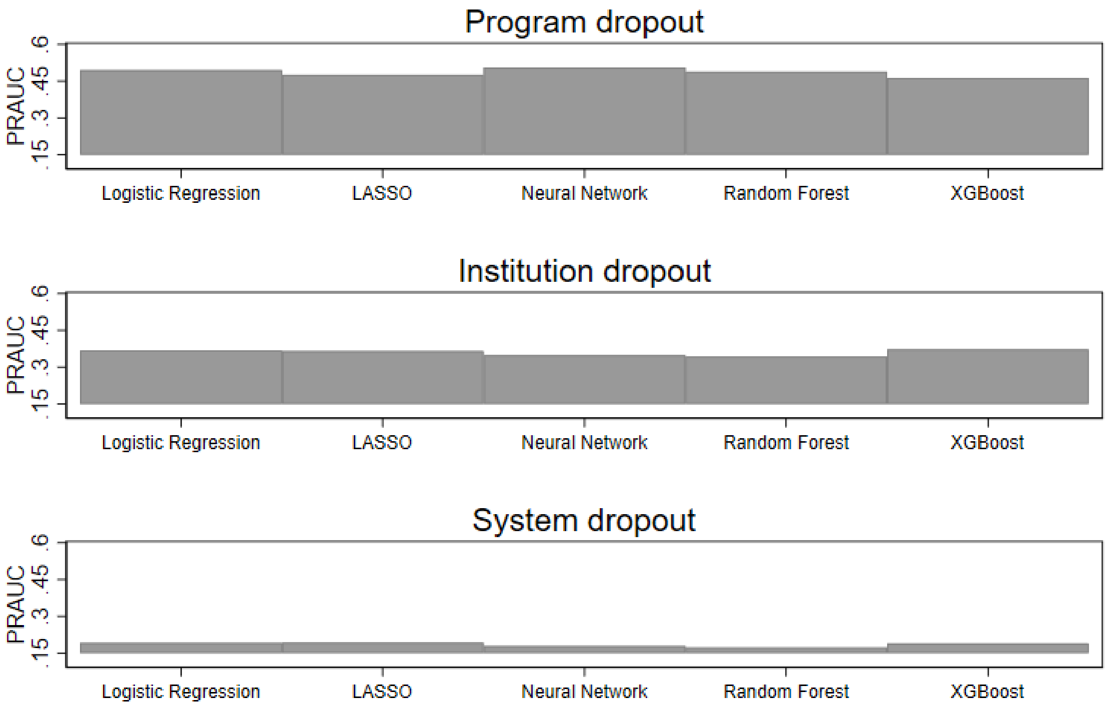
Notes: Models are evaluated using the NONCOG information set (see Table 1). AUC (area under the Receiver Operating Characteristic (ROC) curve) captures the probability that the model ranks a randomly drawn student who drops out more highly in terms of dropout risk than a randomly drawn student who does not drop out. It takes values between 0 (all predictions are wrong) and 1 (all predictions are correct). The 95% confidence intervals are constructed using the percentile bootstrap method with 2000 replications applied to the AUC estimates.

Figure A.6: Model performance using the full feature set ALL



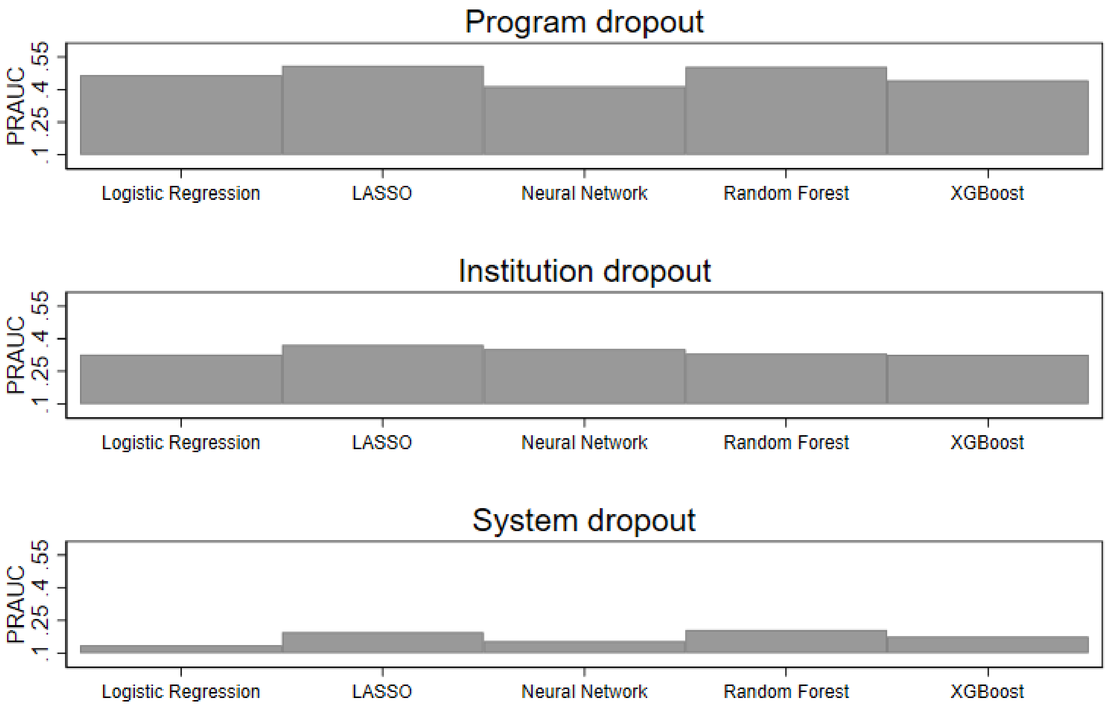
Notes: Models are evaluated using the full feature set ALL (see Table 1). AUC (area under the Receiver Operating Characteristic (ROC) curve) captures the probability that the model ranks a randomly drawn student who drops out more highly in terms of dropout risk than a randomly drawn student who does not drop out. It takes values between 0 (all predictions are wrong) and 1 (all predictions are correct). The 95% confidence intervals are constructed using the percentile bootstrap method with 2000 replications applied to the AUC estimates.

Figure A.7: Model performance with the Basic information set (PRAUC)



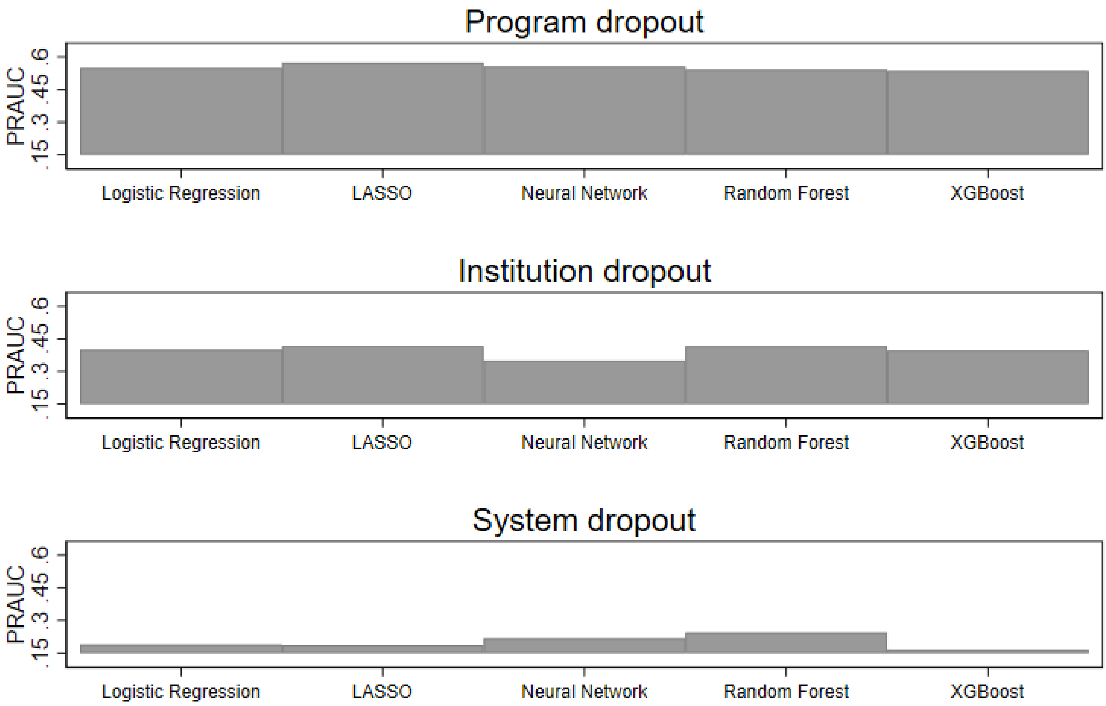
Notes: Models are evaluated using the Basic information set (see Table 1). PRAUC (Precision Recall Area Under the Curve) captures the ability of a model to identify students who drop out. It takes values between 0 (none of the dropouts are correctly classified) and 1 (all of the dropouts are correctly classified).

Figure A.8: Model performance with the Backgr information set (PRAUC)



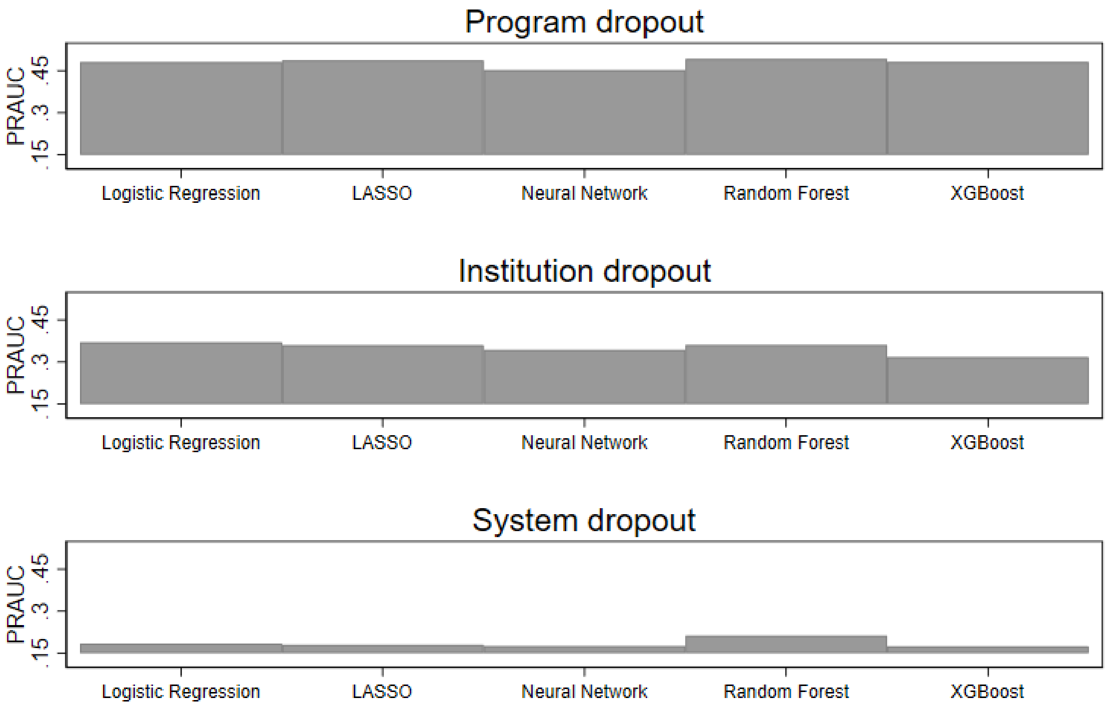
Notes: Models are evaluated using the Backgr information set (see Table 1). PRAUC (Precision Recall Area Under the Curve) captures the ability of a model to identify students who drop out. It takes values between 0 (none of the dropouts are correctly classified) and 1 (all of the dropouts are correctly classified).

Figure A.9: Model performance with the Expect information set (PRAUC)



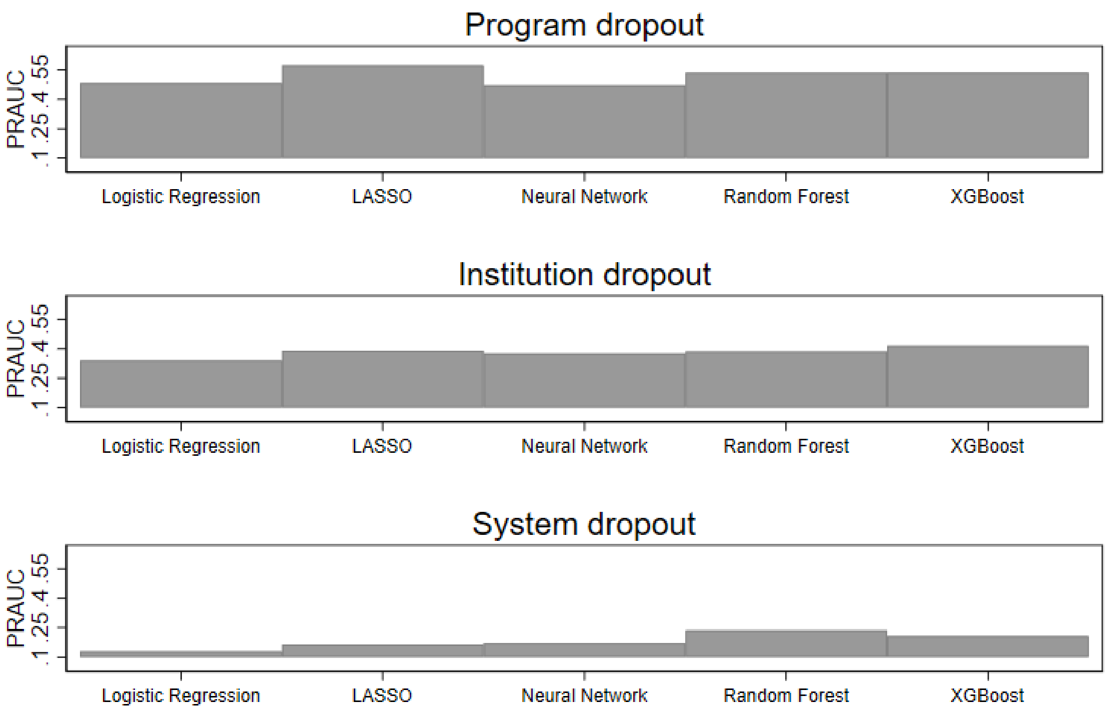
Notes: Models are evaluated using the Expect information set (see Table 1). PRAUC (Precision Recall Area Under the Curve) captures the ability of a model to identify students who drop out. It takes values between 0 (none of the dropouts are correctly classified) and 1 (all of the dropouts are correctly classified).

Figure A.10: Model performance with the NONCOG information set (PRAUC)



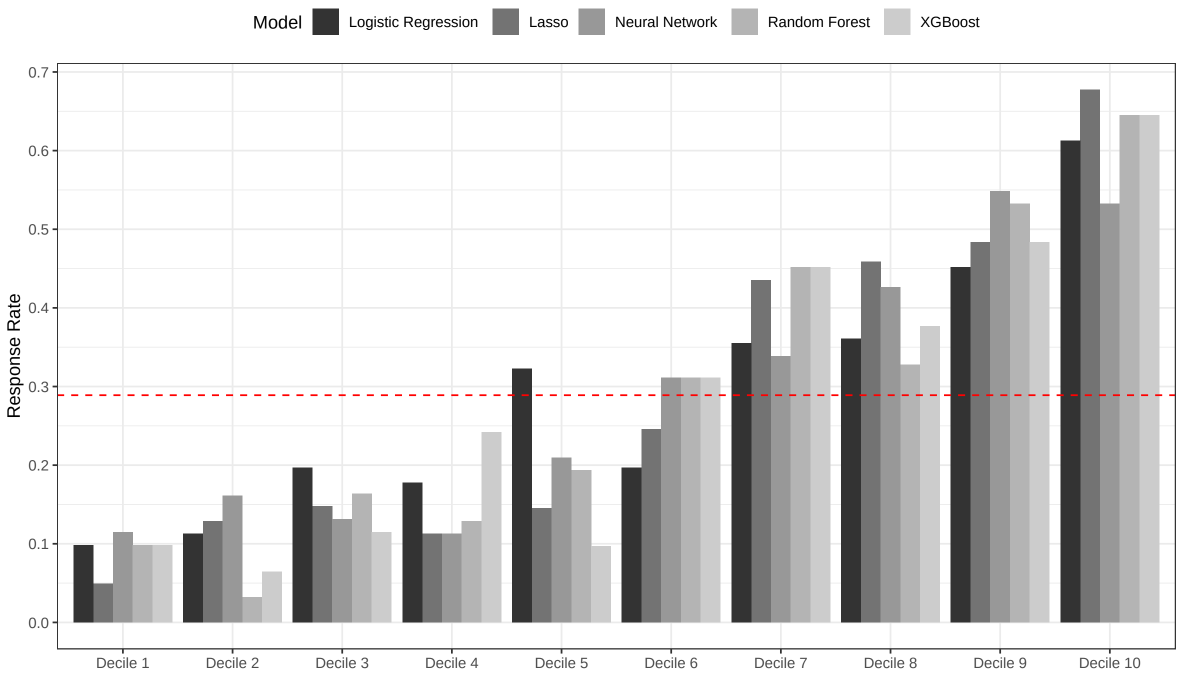
Notes: Models are evaluated using the NONCOG information set (see Table 1). PRAUC (Precision Recall Area Under the Curve) captures the ability of a model to identify students who drop out. It takes values between 0 (none of the dropouts are correctly classified) and 1 (all of the dropouts are correctly classified).

Figure A.11: Model performance using the full feature set ALL (PRAUC)



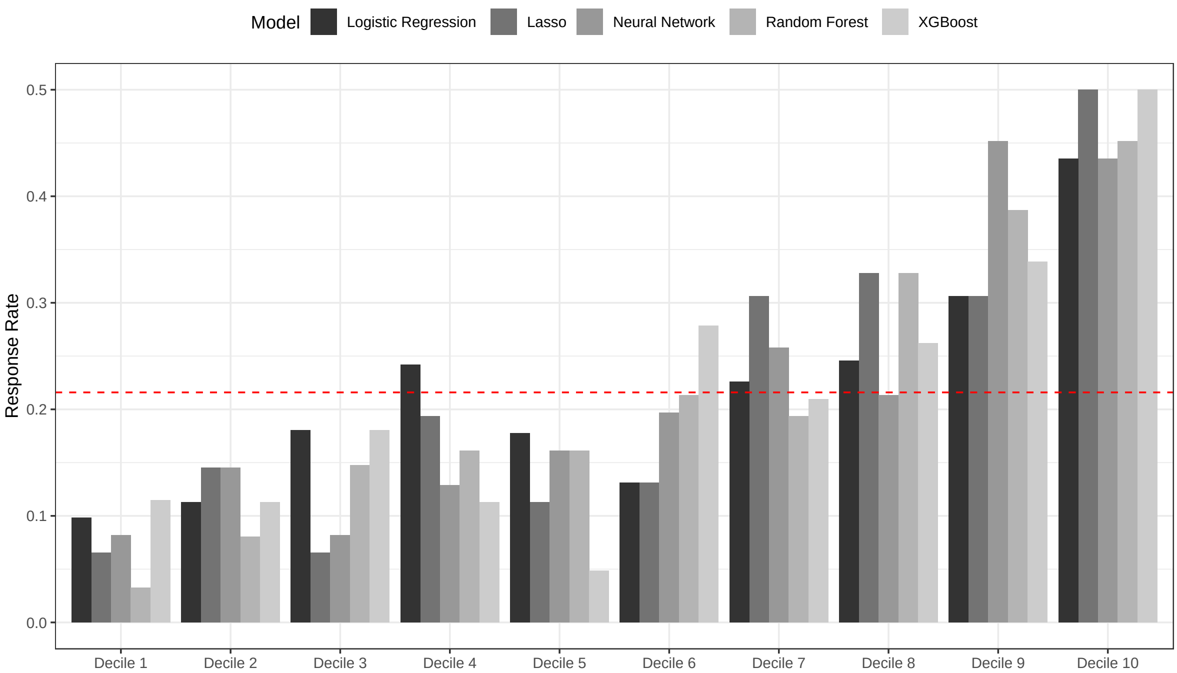
Notes: Models are evaluated using the full feature set ALL (see Table 1). PRAUC (Precision Recall Area Under the Curve) captures the ability of a model to identify students who drop out. It takes values between 0 (none of the dropouts are correctly classified) and 1 (all of the dropouts are correctly classified).

Figure A.12: Response rates for program dropout (All models)



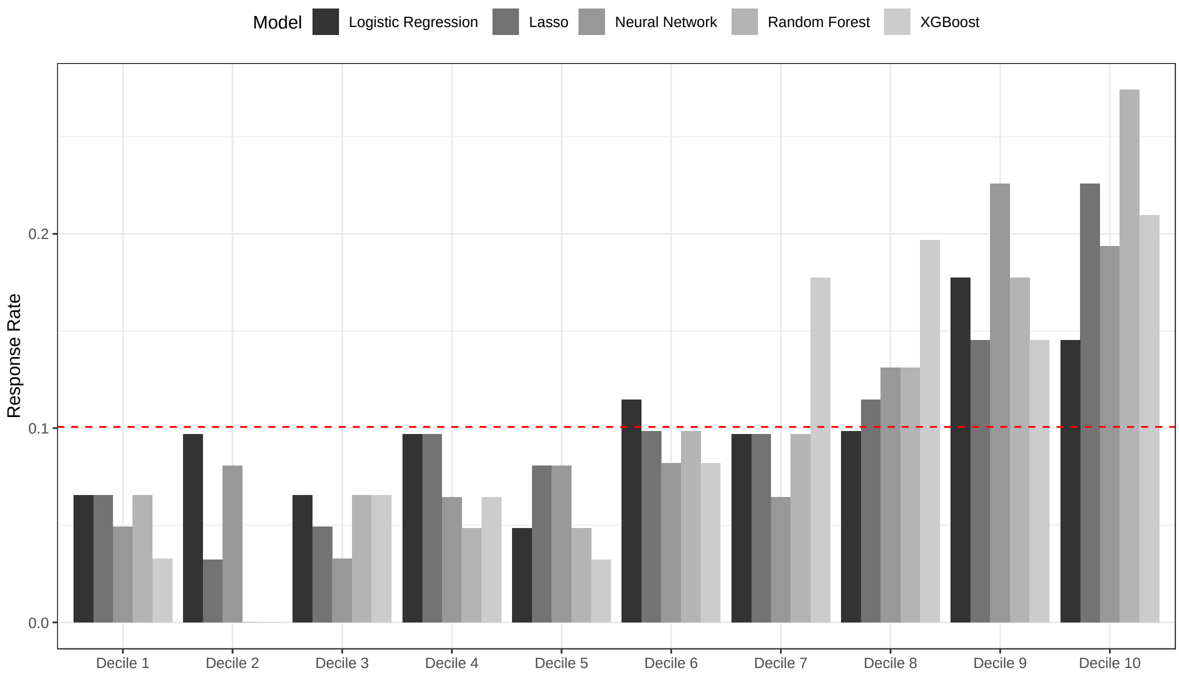
Notes: Predictive models using the full feature set All (see Table 1). Deciles capture the risk ranking assigned by a model: decile 1 contains the 10 percent of students with the lowest predicted dropout risk and decile 10 contains the 10 percent of students with the highest predicted dropout risk. The graph shows the response rate, which measures the proportion of actual cases of dropout (true positives) in a decile. The dotted line shows the 28.9 percent dropout rate at the program level for the entire sample.

Figure A.13: Response rates for institution dropout



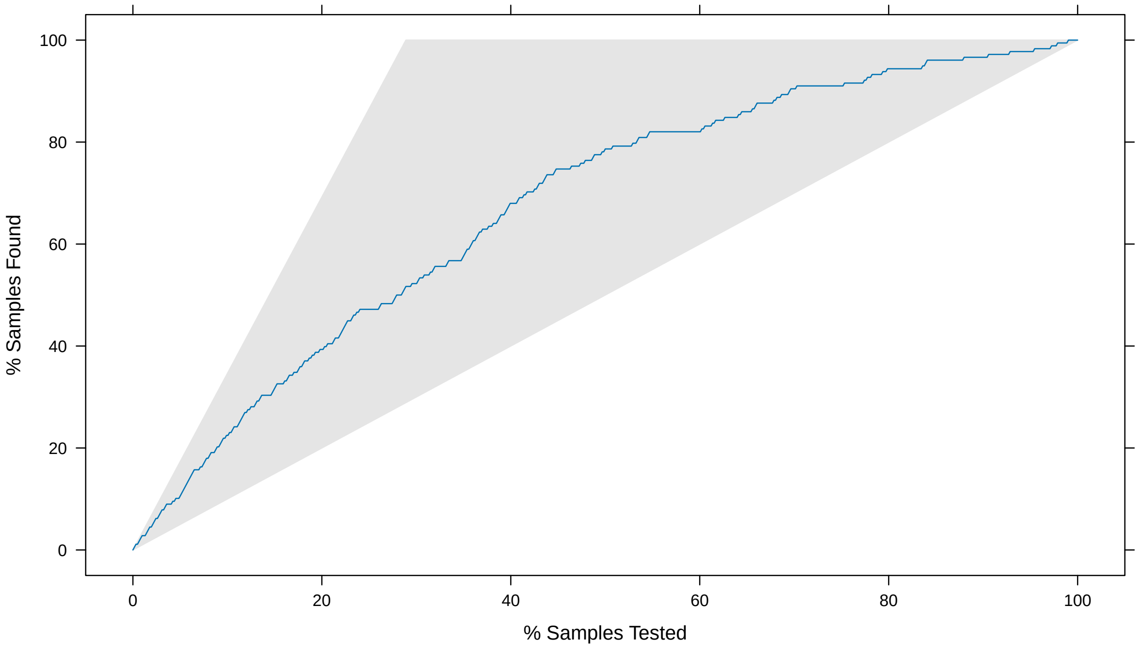
Notes: Predictive models using the full feature set ALL (see Table 1). Deciles capture the risk ranking assigned by a model: decile 1 contains the 10 percent of students with the lowest predicted dropout risk and decile 10 contains the 10 percent of students with the highest predicted dropout risk. The graph shows the response rate, which measures the proportion of actual cases of dropout (true positives) in a decile. The dotted line shows the 21.6 percent dropout rate at the institution level for the entire sample.

Figure A.14: Response rates for system dropout



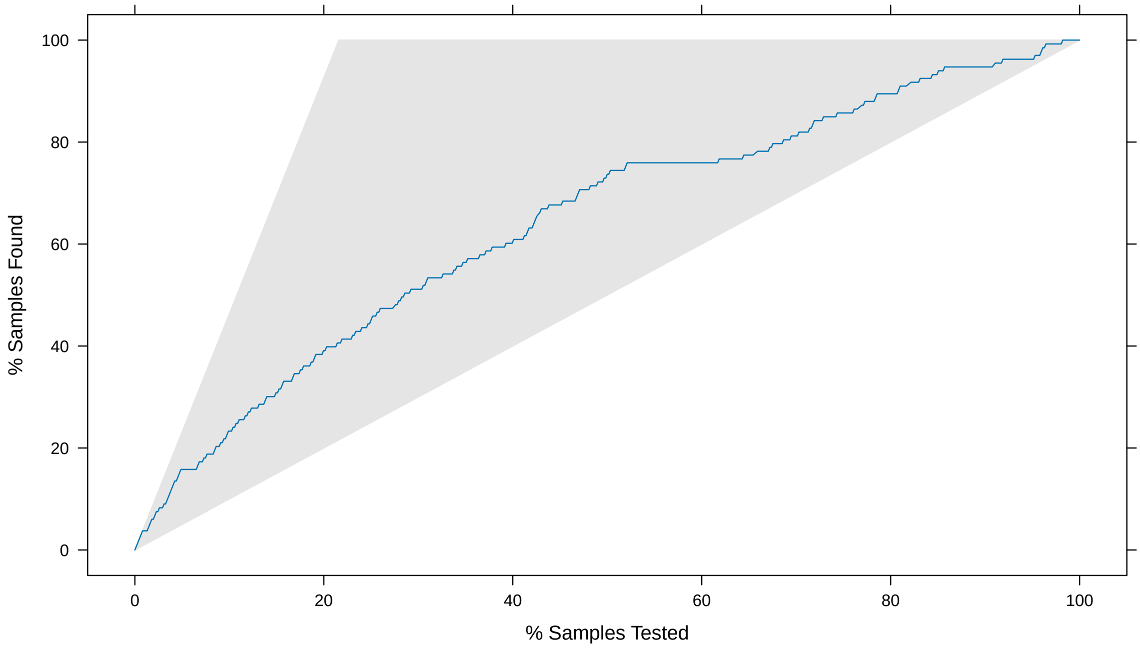
Notes: Predictive models using the full feature set All (see Table 1). Deciles capture the risk ranking assigned by a model: decile 1 contains the 10 percent of students with the lowest predicted dropout risk and decile 10 contains the 10 percent of students with the highest predicted dropout risk. The graph shows the response rate, which measures the proportion of actual cases of dropout (true positives) in a decile. The dotted line shows the 10.1 percent dropout rate at the system level for the entire sample.

Figure A.15: Lift curve for program dropout



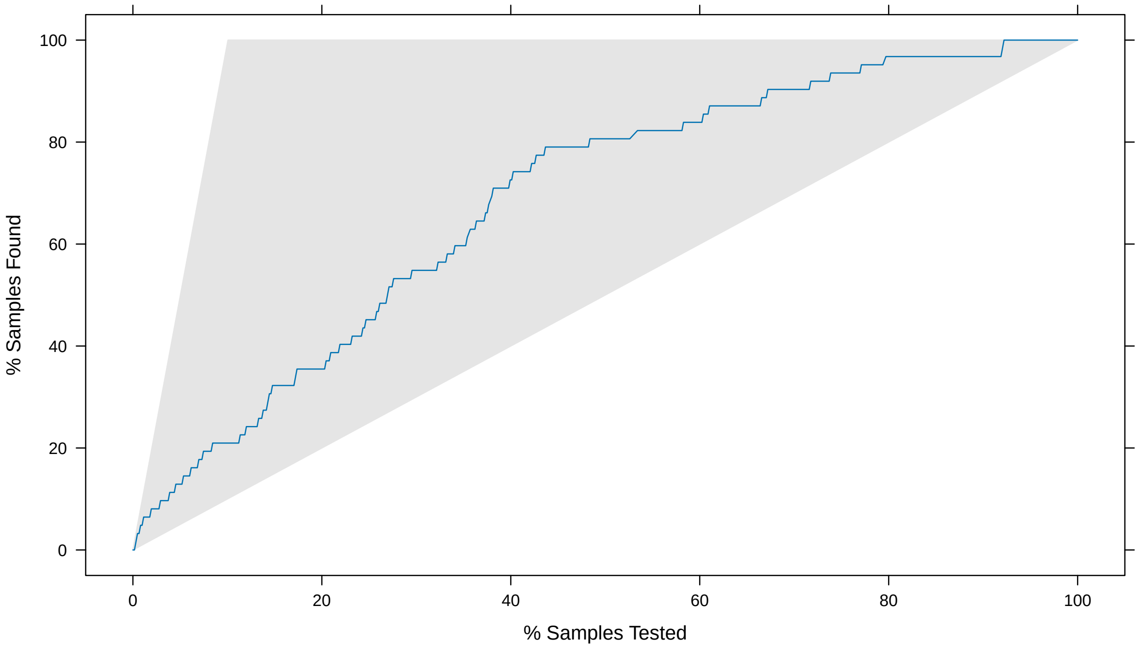
Notes: The figure shows a cumulative lift curve based on predictions from an XGBoost model using the full feature set (ALL; see Table 1). The curve illustrates how much better the model is at identifying program dropouts (*Samples Found*) compared to random targeting across deciles of predicted risk (*Samples Tested*). A steeper curve indicates better model performance, while a line tracking the upper boundary of the gray box would represent perfect prediction.

Figure A.16: Lift curve for institution dropout



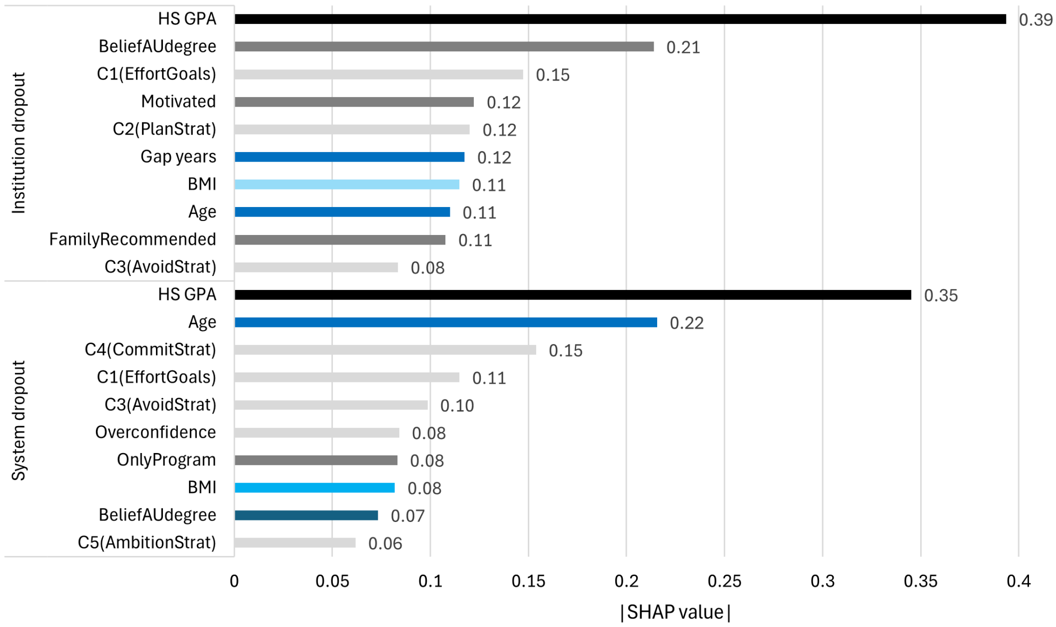
Notes: The figure shows a cumulative lift curve based on predictions from an XGBoost model using the full feature set (ALL; see Table 1). The curve illustrates how much better the model is at identifying institution dropouts (*Samples Found*) compared to random targeting across deciles of predicted risk (*Samples Tested*). A steeper curve indicates better model performance, while a line tracking the upper boundary of the gray box would represent perfect prediction.

Figure A.17: Lift curve for system dropout



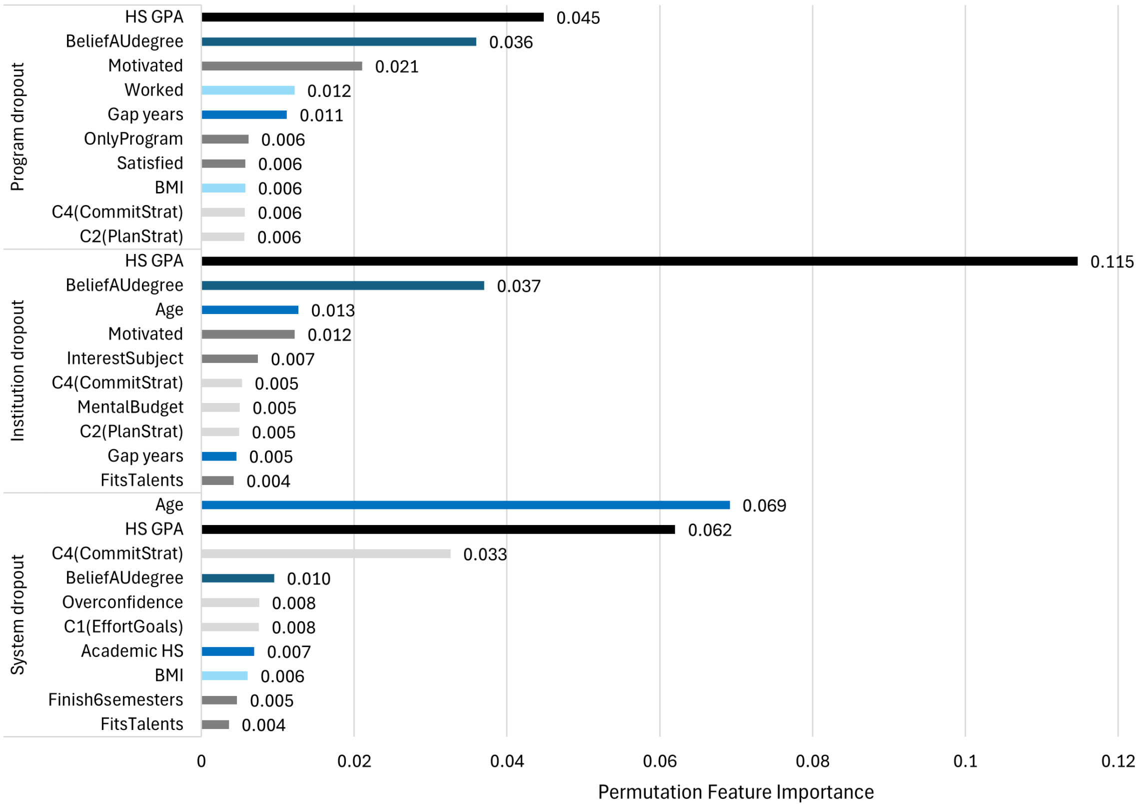
Notes: The figure shows a cumulative lift curve based on predictions from an XGBoost model using the full feature set (ALL; see Table 1). The curve illustrates how much better the model is at identifying system dropouts (*Samples Found*) compared to random targeting across deciles of predicted risk (*Samples Tested*). A steeper curve indicates better model performance, while a line tracking the upper boundary of the gray box would represent perfect prediction.

Figure A.18: Feature importance for institution and system dropout (XGBoost)



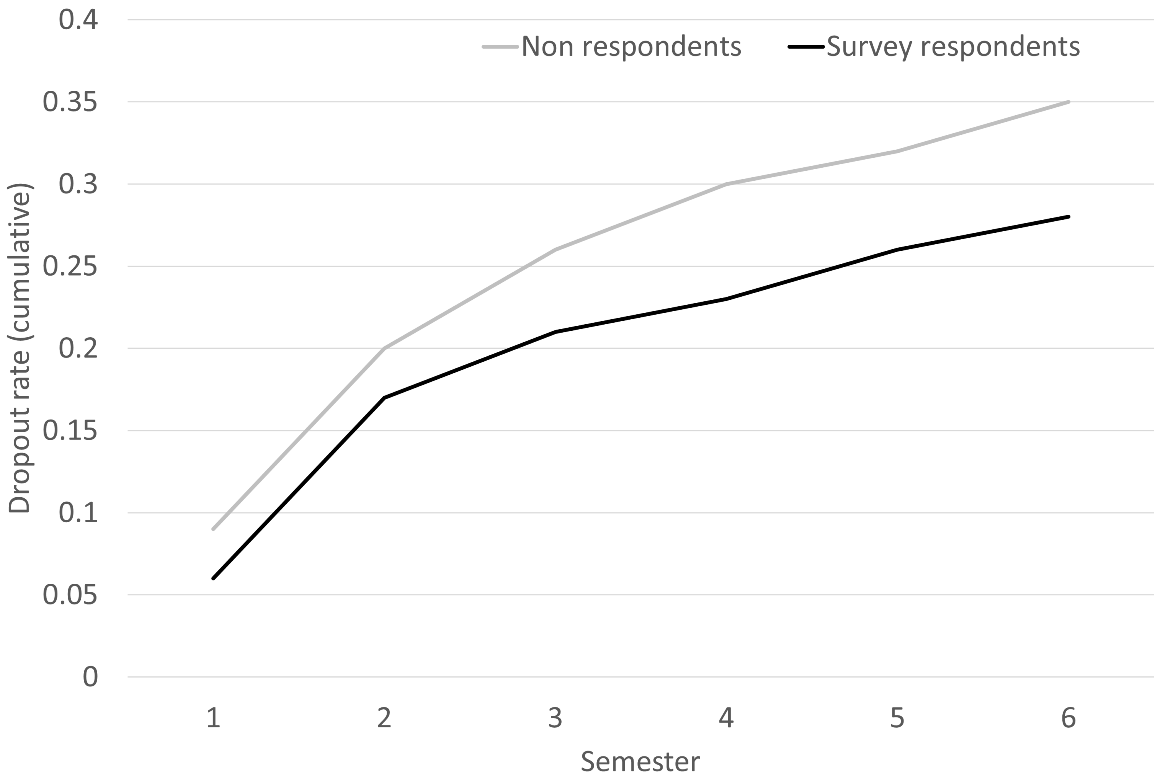
Notes: Model: XGBoost using the full feature set ALL (see Table 1). Top 10 feature importance in terms of absolute *SHAP* value for each type of dropout. Components of the full feature set ALL: Black is included in the GPA set, blue is included in BASIC, light blue is included in BACKGR, dark gray is included in EXPECT, and light gray is included in NONCOG.

Figure A.19: Permutation feature importance (XGBoost)



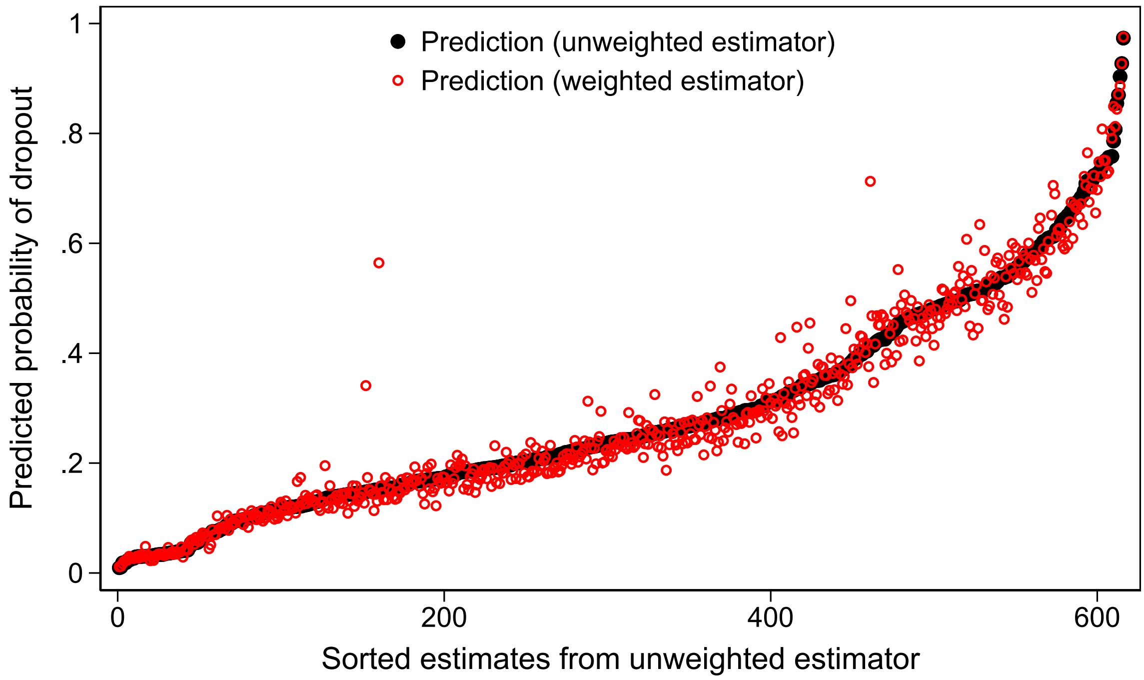
Notes: Model: XGBoost with ALL features (see Table 1). Top 10 permutation feature importance (PFI) for each type of dropout. Components of the full feature set ALL: Black is included in the GPA set, blue is included in BASIC, light blue is included in BACKGR, dark gray is included in EXPECT, and light gray is included in NONCOG.

Figure A.20: Sample selection: Dropout from the program



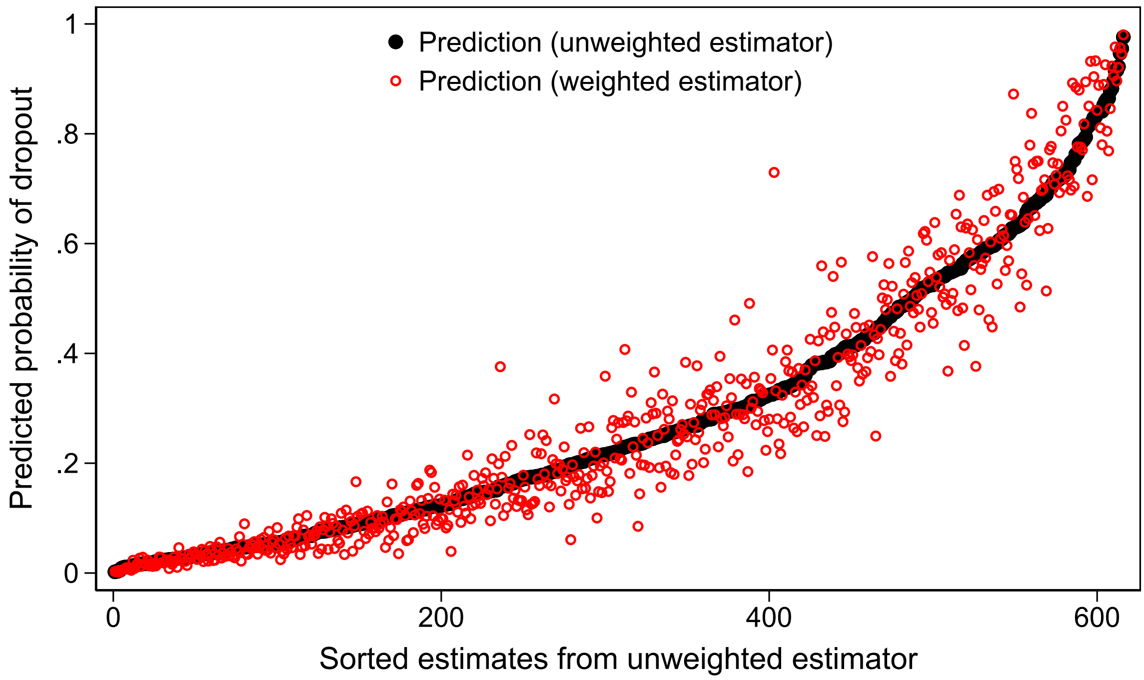
Notes: Cumulative dropout rate for the cohort of first semester students at the faculty of Business and Social Sciences, split into the subset of students who completed the survey (N=616) and the subset who did not complete the survey (N=2,131). The difference is significant at 10 percent level in semesters 1-2 and at 1 percent level for the higher semesters (see Table A.5).

Figure A.21: Predicted probability of program dropout with the BASIC information set, for original vs. weighted sample



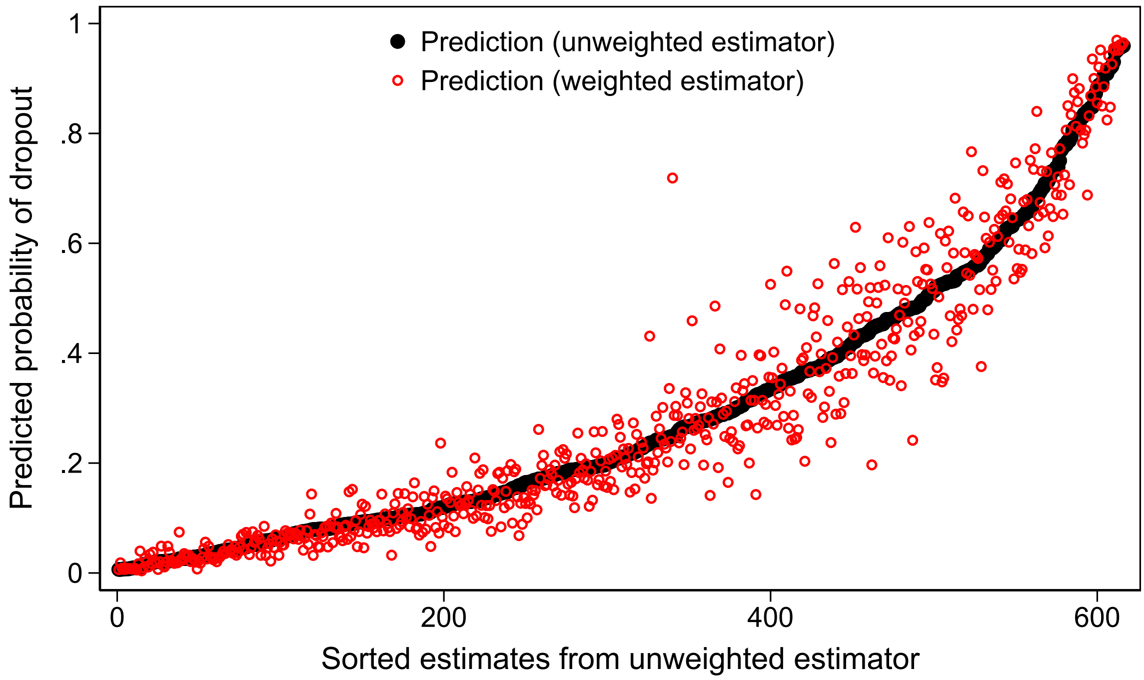
Notes: Predictions are based on a logit model. Weights are estimated with propensity score matching.

Figure A.22: Predicted probability of program dropout with the BACKGR information set, for original vs. weighted sample



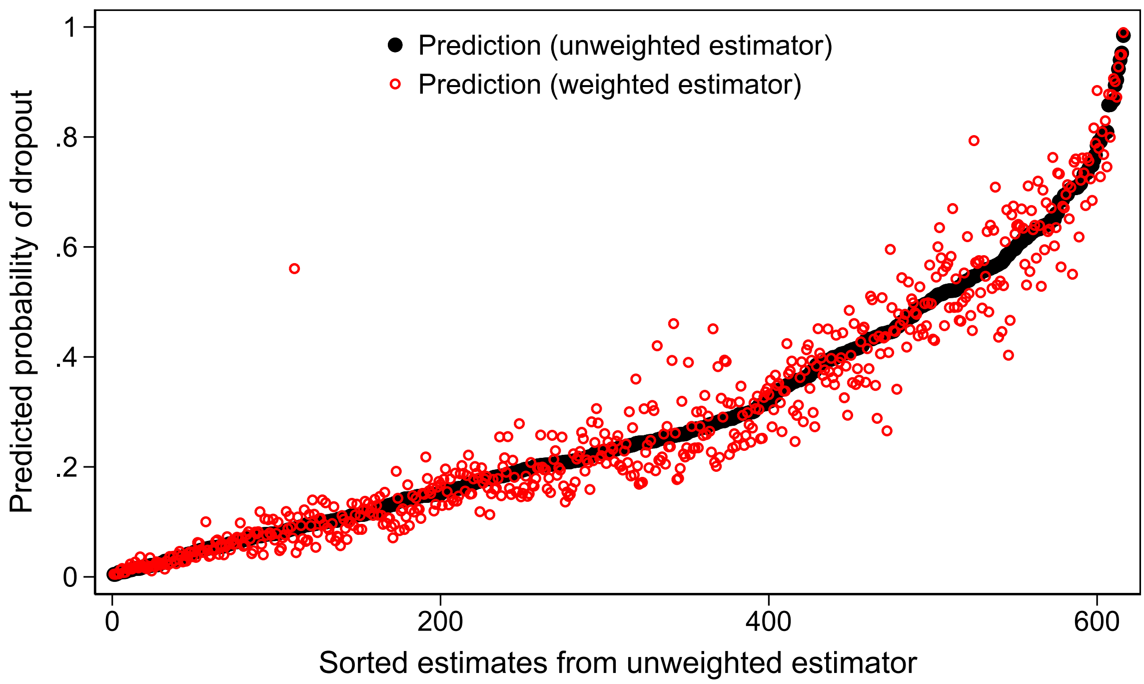
Notes: Predictions are based on a logit model. Weights are estimated with propensity score matching.

Figure A.23: Predicted probability of program dropout with the Expect information set, for original vs. weighted sample



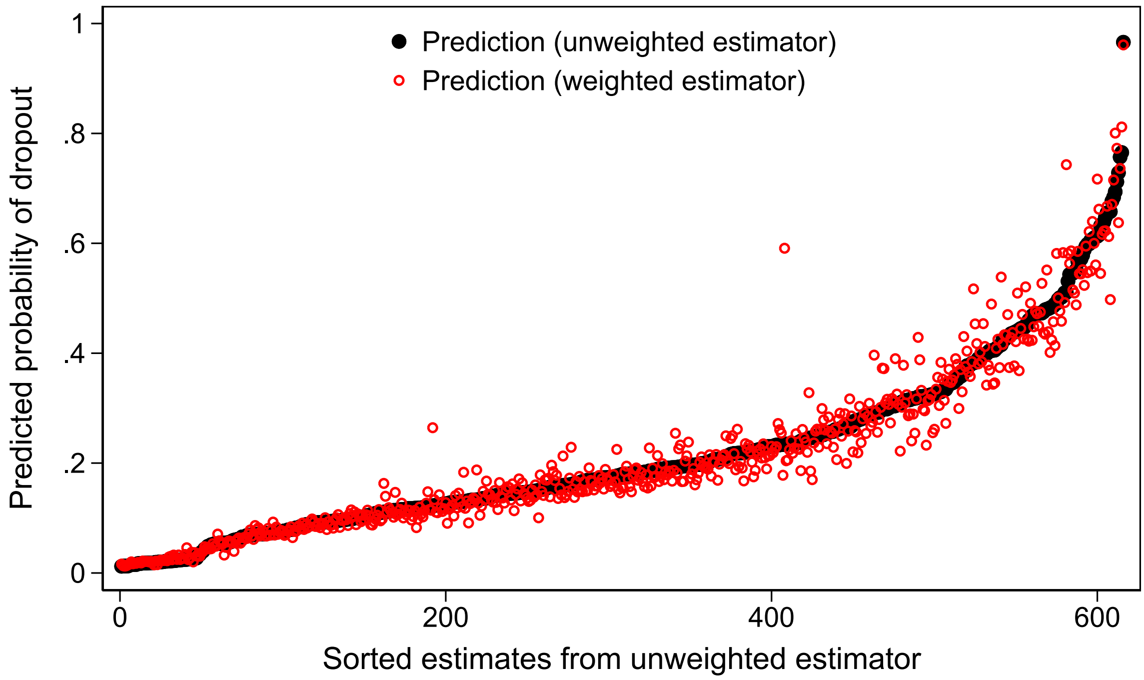
Notes: Predictions are based on a logit model. Weights are estimated with propensity score matching.

Figure A.24: Predicted probability of program dropout with the Noncog information set, for original vs. weighted sample



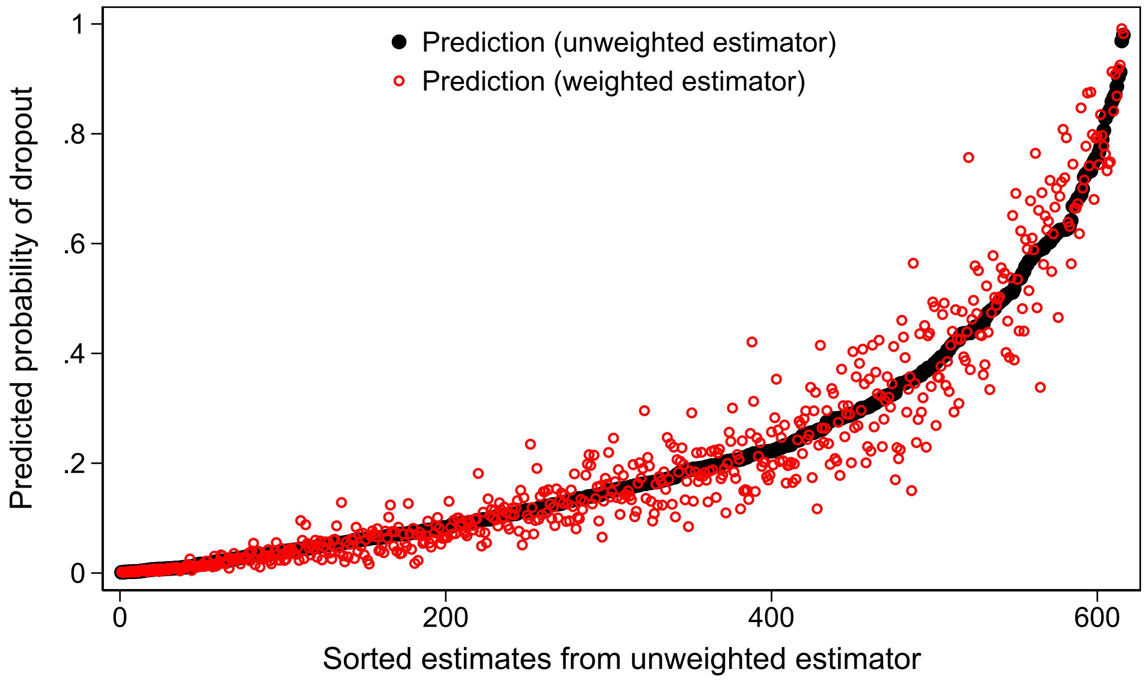
Notes: Predictions are based on a logit model. Weights are estimated with propensity score matching.

Figure A.25: Predicted probability of institution dropout with the Basic information set, for original vs. weighted sample



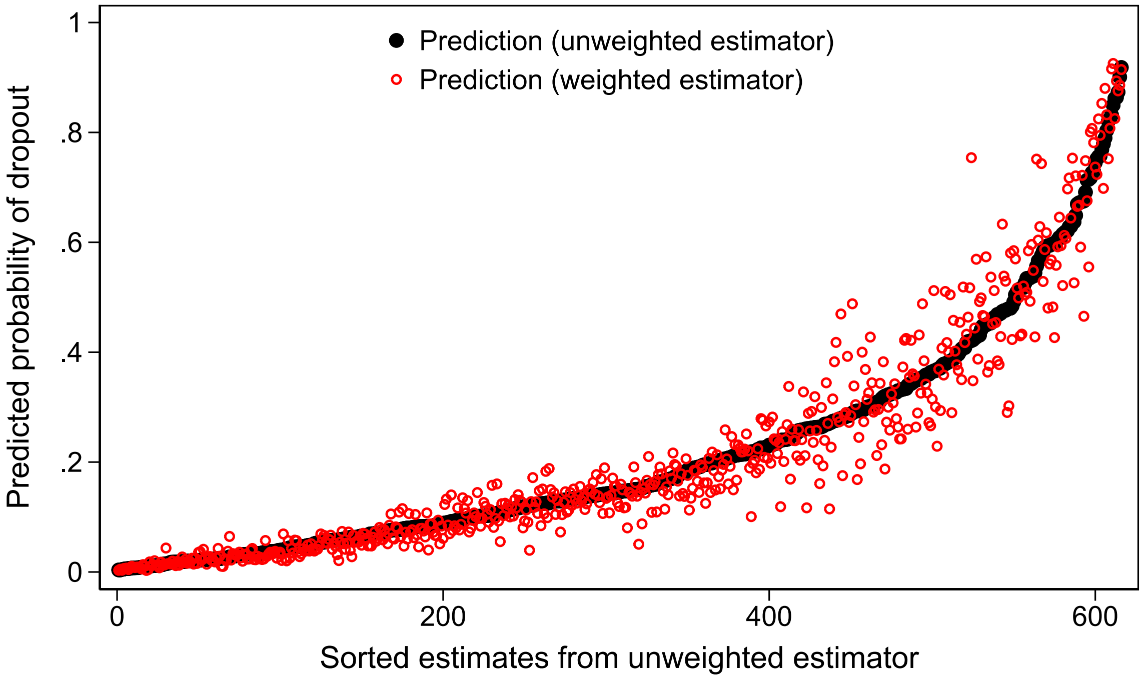
Notes: Predictions are based on a logit model. Weights are estimated with propensity score matching.

Figure A.26: Predicted probability of institution dropout with the Backgr information set, for original vs. weighted sample



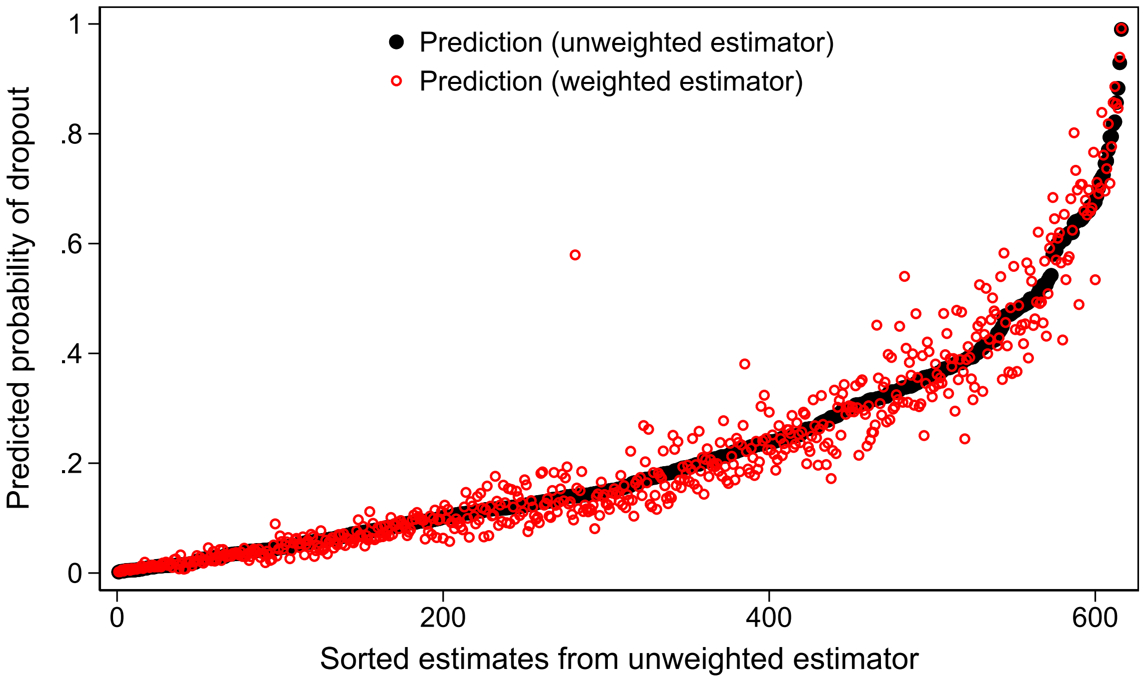
Notes: Predictions are based on a logit model. Weights are estimated with propensity score matching.

Figure A.27: Predicted probability of institution dropout with the EXPECT information set, for original vs. weighted sample



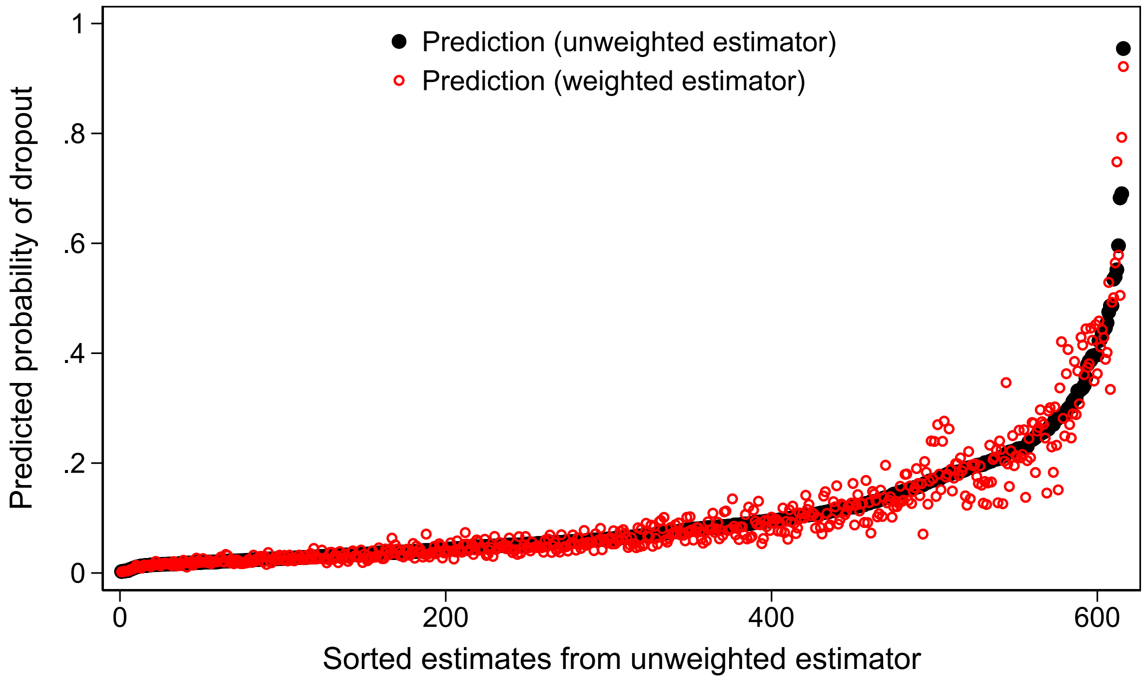
Notes: Predictions are based on a logit model. Weights are estimated with propensity score matching.

Figure A.28: Predicted probability of institution dropout with the NONCOG information set, for original vs. weighted sample



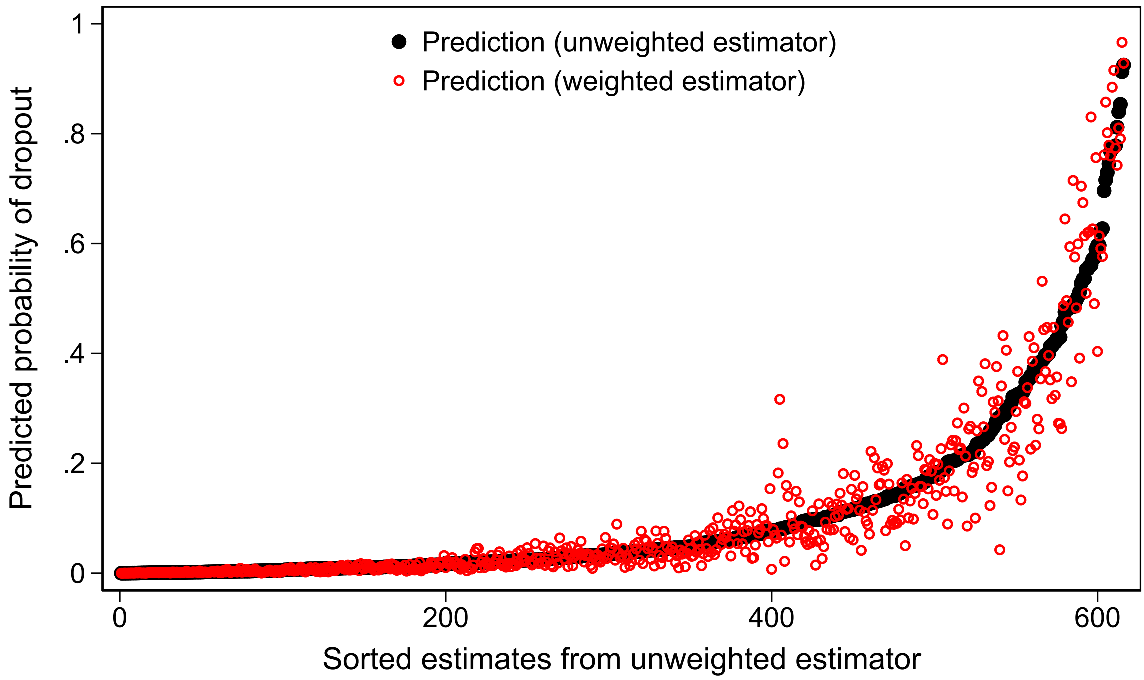
Notes: Predictions are based on a logit model. Weights are estimated with propensity score matching.

Figure A.29: Predicted probability of system dropout with the Basic information set, for original vs. weighted sample



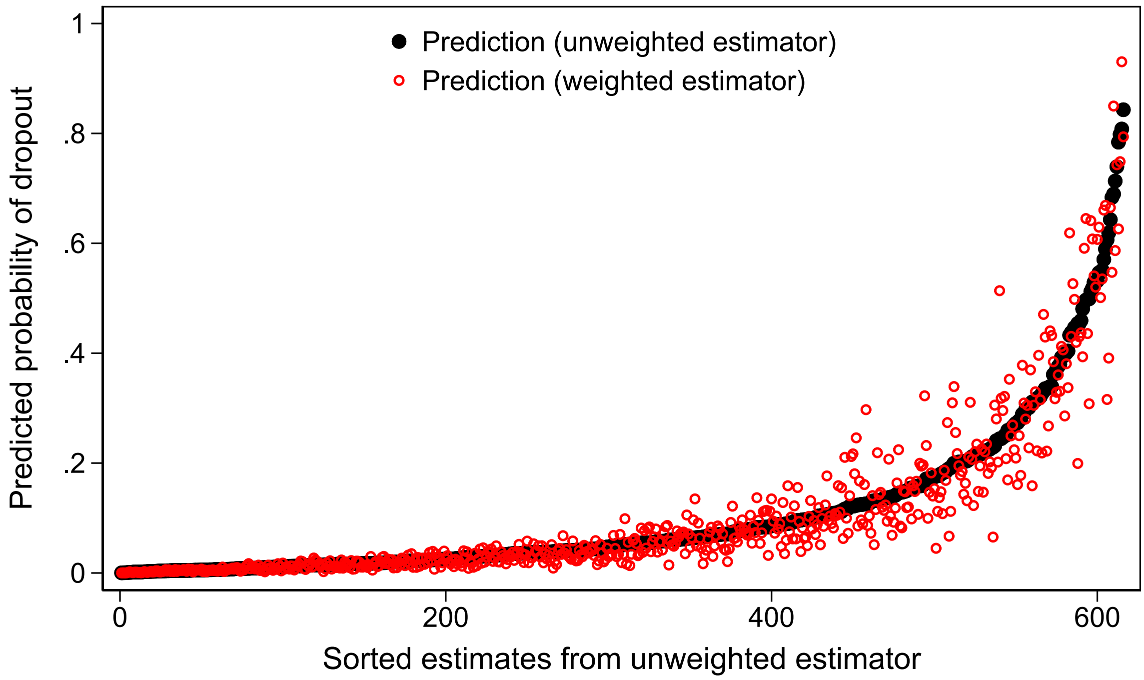
Notes: Predictions are based on a logit model. Weights are estimated with propensity score matching.

Figure A.30: Predicted probability of system dropout with the Backgr information set, for original vs. weighted sample



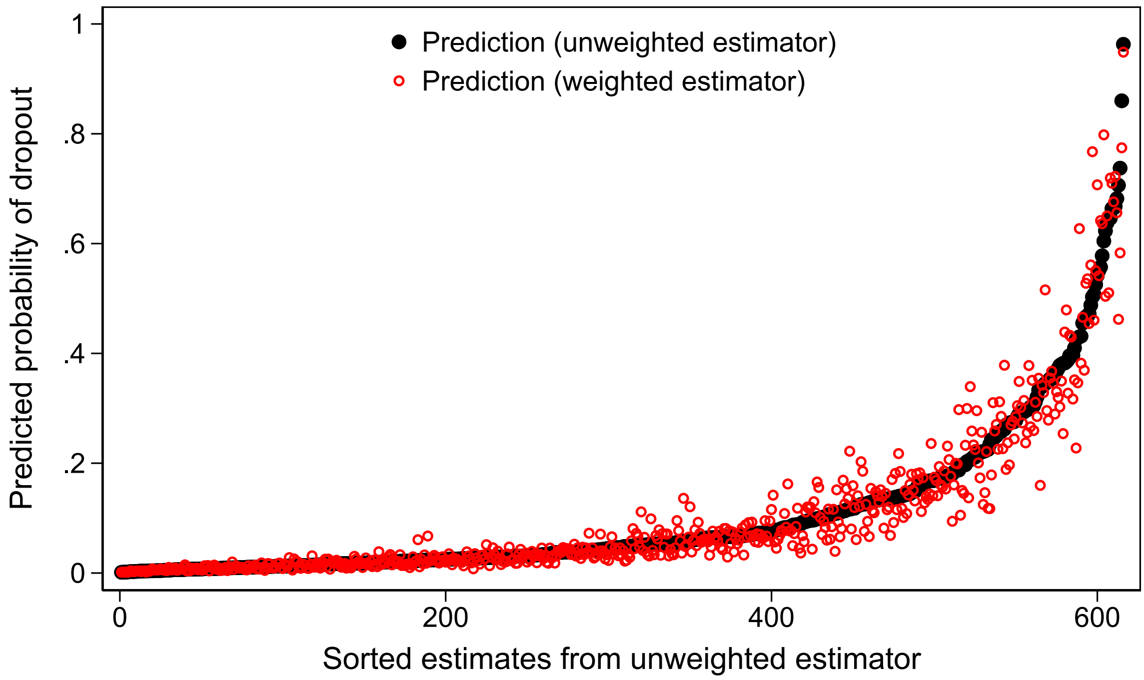
Notes: Predictions are based on a logit model. Weights are estimated with propensity score matching.

Figure A.31: Predicted probability of system dropout with the EXPECT information set, for original vs. weighted sample



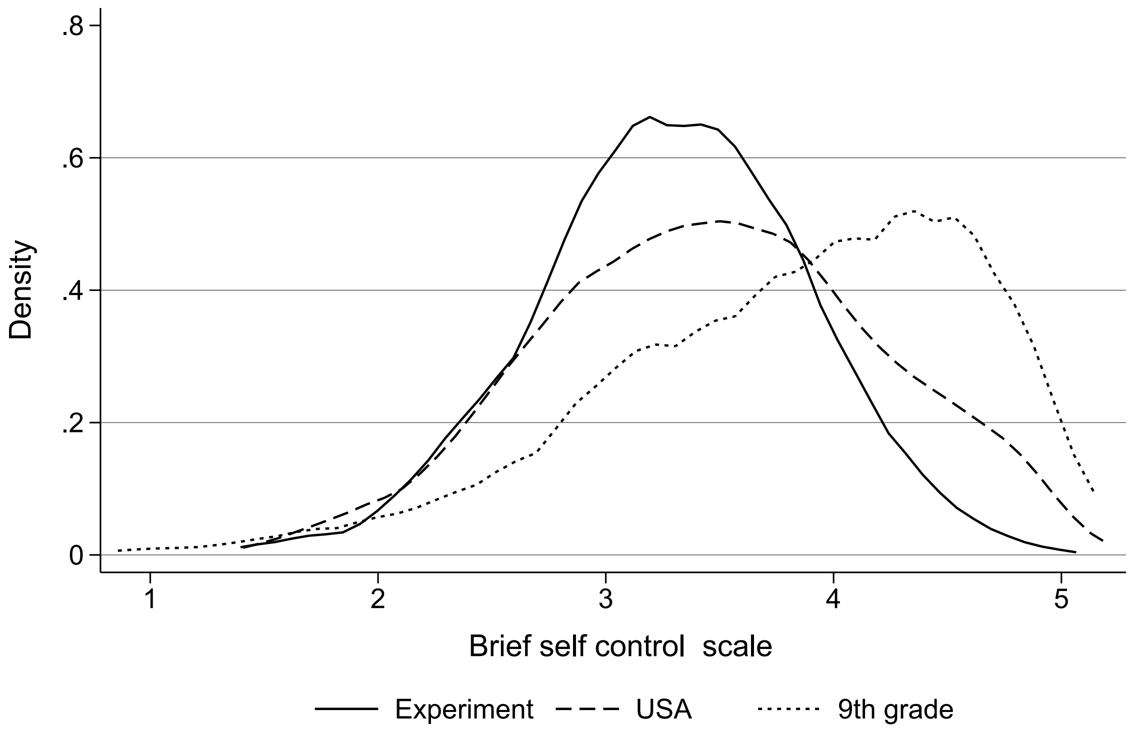
Notes: Predictions are based on a logit model. Weights are estimated with propensity score matching.

Figure A.32: Predicted probability of system dropout with the NONCOG information set, for original vs. weighted sample



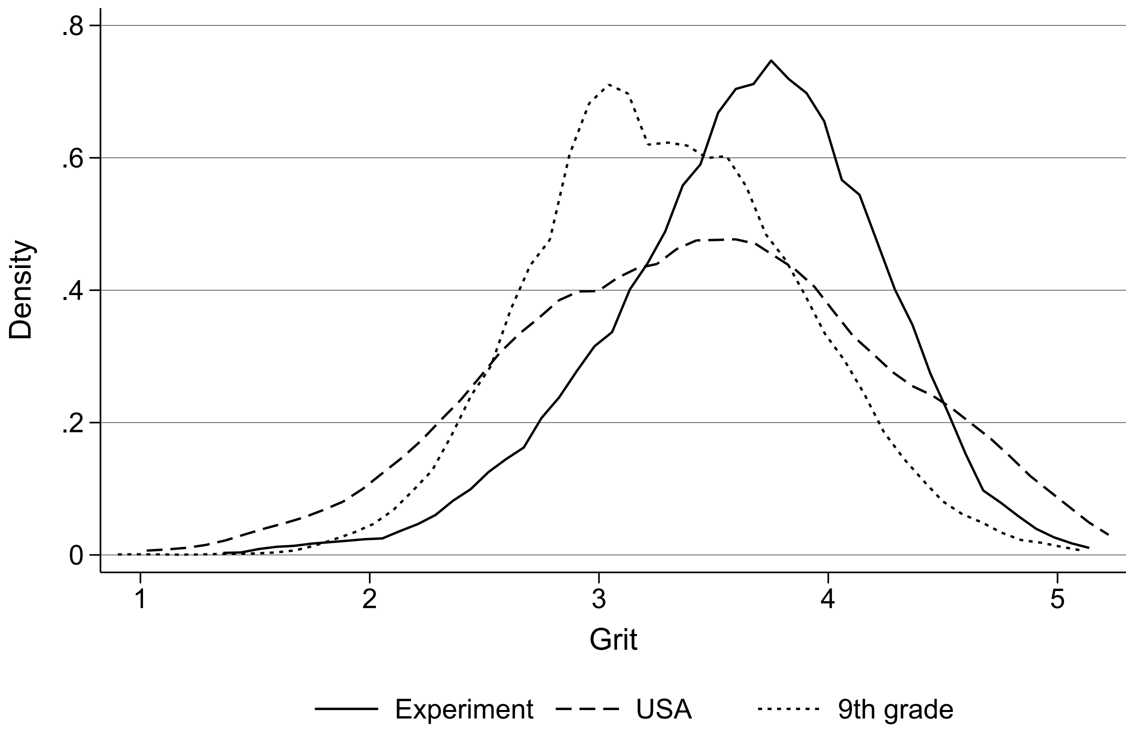
Notes: Predictions are based on a logit model. Weights are estimated with propensity score matching.

Figure A.33: Kernel density plots of self-control



Notes: The figure shows kernel density plots for the 13-item brief self-control scale (Tangney et al., 2004) for three different samples. *Experiment*: the university students who completed our survey experiment; *USA*: a sample of 299 participants, representative of the USA population in terms of sex and age, collected on Prolific Academic in June 2024 (51 percent female; mean age 45.2 years); *9th grade*: sample of 13,698 Danish 9th grade pupils from Hvidman et al. (2024).

Figure A.34: Kernel density plots of grit



Notes: The figure shows kernel density plots for the 8-item grit scale (Duckworth et al., 2007) for three different samples. *Experiment*: the university students who completed our survey experiment; *USA*: a sample of 299 participants, representative of the USA population in terms of sex and age, collected on Prolific Academic in June 2024 (51 percent female; mean age 45.2 years); *9th grade*: sample of 13,698 Danish 9th grade pupils from Hvidman et al. (2024).

A.8 Additional tables

Table A.4: Descriptive statistics for the cohort of first year students

Variable	Non respondents *Mean*	Non respondents *Std.dev.*	Survey respondents *Mean*	Survey respondents *Std.dev.*	*Diff.*
		Feature set GPA			
GPA	7.92	1.93	8.45	1.79	-0.53^{**}
Dummies for the bachelor program					
BusAdmin(HerningCampus)	0.05	0.22	0.04	0.20	0.01
BusAdmin(English)	0.08	0.27	0.03	0.18	0.05^{***}
BusAdmin(Danish)	0.25	0.43	0.19	0.39	0.06^{**}
International Business	0.18	0.38	0.18	0.38	0.00
Political Science	0.10	0.30	0.11	0.32	-0.01
Business & Economics	0.02	0.15	0.04	0.19	-0.01^{+}
Economics	0.06	0.24	0.09	0.29	-0.03^{**}
Business & Law	0.05	0.22	0.05	0.22	0.00
Law	0.14	0.35	0.17	0.38	-0.03^{+}
Psychology	0.06	0.24	0.09	0.28	-0.02^{+}
		Feature set BASIC			
Age	21.47	2.67	21.28	2.58	0.19
Male	0.53	0.50	0.44	0.50	0.09^{***}
Gap years	1.46	2.05	1.29	2.10	0.17^{+}
Priority of the bachelor's program in the application for studying					
Priority1	0.84	0.37	0.87	0.34	-0.03
Priority2	0.09	0.29	0.09	0.28	0.00
Priority3 or greater	0.07	0.25	0.05	0.21	0.02^{*}

Notes: Descriptive statistics for the cohort of first semester students at the faculty of Business and Social Sciences, split into the subset of students who completed the survey (N=616) and the subset who did not complete the survey (N=2,131). Significance levels: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A.5: Descriptive statistics for the cohort of first year students (continued)

	Non respondents		**Survey respondents**		
Variable	*Mean*	*Std.dev.*	*Mean*	*Std.dev.*	*Diff.*
		Administrative data			
Foreign	0.07	0.25	0.06	0.23	0.01
Educational attainment of the father					
No father registered	0.06	0.24	0.06	0.24	0.00
Primary or secondary education only	0.13	0.34	0.14	0.35	−0.01
Vocational education and training	0.37	0.48	0.36	0.48	0.00
Bachelor's, vocational degree or other short cycle higher education	0.24	0.43	0.24	0.43	0.00
Master's degree	0.15	0.36	0.17	0.38	−0.02
Educational attainment of the mother					
No mother registered	0.05	0.21	0.04	0.20	0.00
Primary or secondary education only	0.14	0.35	0.15	0.36	−0.01
Vocational education and training	0.33	0.47	0.31	0.46	0.03
Bachelor's, vocational degree or other short cycle higher education	0.36	0.48	0.35	0.48	0.00
Master's degree	0.09	0.29	0.12	0.32	-0.03^{*}

Notes: Descriptive statistics for the cohort of first semester students at the faculty of Business and Social Sciences, split into the subset of students who completed the survey (N=616) and the subset who did not complete the survey (N=2,131). Significance levels: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A.6: Descriptive statistics for the cohort of first year students (continued)

	Non respondents		**Survey respondents**		
Variable	*Mean*	*Std.dev.*	*Mean*	*Std.dev.*	*Diff.*
		Outcomes			
Cumulative dropout rates for each semester					
1	0.09	0.28	0.06	0.24	0.02*
2	0.2	0.4	0.17	0.38	0.03*
3	0.26	0.44	0.21	0.41	0.05**
4	0.3	0.46	0.23	0.43	0.07**
5	0.32	0.47	0.26	0.44	0.06**
6	0.35	0.48	0.28	0.45	0.07**
Dropout rates					
Program dropout	0.36	0.48	0.29	0.45	0.07**
Institution dropout	0.29	0.46	0.22	0.41	0.08***
System dropout	0.16	0.37	0.10	0.30	0.06***
Completed					
Bachelor's program	0.64	0.48	0.71	0.45	−0.07**
A bachelor's at the university	0.71	0.46	0.78	0.41	−0.08***
A bachelor's in Denmark	0.84	0.37	0.90	0.30	−0.06***
A master's degree	0.65	0.47	0.73	0.44	−0.08***

Notes: Descriptive statistics for the cohort of first semester students at the faculty of Business and Social Sciences, split into the subset of students who completed the survey (N=616) and the subset who did not complete the survey (N=2,131). Significance levels: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A.7: Principal Component Analysis

Variable	Component 1	Component 2	Component 3	Component 4	Component 5
	EffortGoals	PlanStrat	AvoidStrat	CommitStrat	AmbitionStrat
ShowupGoal	**0.49**	-0.05	-0.05	-0.03	0.07
HomeworkGoal	**0.44**	0.05	-0.01	-0.11	0.06
StudygroupGoal	**0.54**	0.00	-0.04	0.10	-0.06
HoursGoal	0.20	**0.24**	-0.02	-0.11	0.15
ForethoughtGoal	0.00	**0.47**	0.10	-0.15	0.07
Deadlines	0.26	**0.28**	0.10	0.02	-0.04
SelfReward	-0.07	**0.37**	-0.15	0.24	-0.08
TellFriend	-0.09	**0.29**	0.01	0.25	0.10
Bet	-0.10	**0.25**	-0.18	0.13	-0.07
Subgoals	0.04	**0.46**	0.03	-0.15	-0.03
Avoidgoals(R)	-0.03	-0.16	**0.50**	-0.01	0.13
SpontaneousGoal	-0.09	0.21	**0.51**	0.04	-0.26
Giveup(R)	0.02	0.01	**0.60**	0.03	0.05
GradeReward	-0.05	0.06	-0.14	**0.35**	-0.05
AvoidTempt	-0.14	0.16	-0.03	**0.26**	0.07
MandAssign	0.02	-0.15	0.03	**0.47**	0.17
ProjectDeadlines	0.02	-0.04	0.10	**0.46**	0.03
Studygroup	0.31	-0.07	0.10	**0.37**	-0.26
GradeGoal	0.04	0.03	0.08	0.02	**0.49**
AmbitiousGoal	-0.02	-0.02	-0.02	0.02	**0.63**
AngerGoal	0.01	0.15	0.00	0.08	**0.32**

Notes: Principal component analysis after orthogonal varimax rotation (eigenvalues averaged over 1,000 replications, parallel analysis suggests five components with components marked in bold face).

Table A.8: Correlation matrix with outcomes (selected features)

	HS GPA	Program Dropout	Institution Dropout	System Dropout	GPA Finished
			Feature set GPA		
HS GPA	-	-0.28***	-0.26***	-0.20***	0.41***
			Feature set BASIC		
Age	-0.09*	0.07+	0.11*	0.18***	-0.03
Gender (male = 1)	-0.08+	0.05	0.05	0.08+	-0.04
Gap years	-0.14***	0.02	0.04	0.11*	0.06
First priority	0.23***	-0.15***	-0.11*	-0.08+	0.10*
Second priority	-0.13***	0.07+	0.02	0.01	-0.09+
Third or lower priority	-0.20***	0.15***	0.14***	0.10*	-0.04
Foreign student	-0.03	0.00	0.02	0.03	-0.03
			Feature set BACKGR		
EduMom	-0.01	-0.08*	-0.05	-0.09*	-0.07
EduDad	-0.02	-0.05	-0.06	-0.04	-0.08
BMI	-0.10*	0.11*	0.11*	0.05	-0.03
Math A-level	0.04	-0.11*	-0.07+	-0.05	0.16***
			Feature set NONCOG		
Grit	0.15***	-0.04	-0.06	-0.04	0.08
Self-control	0.18***	-0.04	-0.06	-0.04	0.11*
CRT	0.09*	-0.08+	-0.06	-0.02	0.10*
Narrow Goal	0.10*	-0.06	-0.05	-0.03	0.03
Overconfidence	-0.04	0.08+	0.06	0.07+	-0.03
C1(EffortGoals)	0.06	-0.14***	-0.14***	-0.11***	-0.02
C2(PlanStrat)	-0.02	0.00	-0.01	-0.08+	-0.04
C3(AvoidStrat)	0.15***	0.02	0.03	0.03	0.06
C4(CommitStrat)	-0.01	-0.04	-0.05	-0.12***	-0.11*
C5(AmbitionStrat)	0.07+	-0.01	-0.03	0.03	0.08
Loss aversion	0.00	0.03	0.04	-0.04	0.01
			Feature set EXPECT		
BeliefAUdegree	0.04	-0.25***	-0.21***	-0.14***	-0.02
Motivated	0.13***	-0.26***	-0.20***	-0.12***	0.02
Most Desired	0.20***	-0.24***	-0.19***	-0.15***	0.02
Satisfied	0.10*	-0.27***	-0.21***	-0.11*	0.00
N	616	616	616	616	438

Notes: Pearson correlation coefficients. Significance levels: $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. *HS GPA*: high school grade point average. *GPA Finished*: grade point average for the bachelor's degree (for those who graduated in their original degree program).

Table A.9: Test of pairwise AUC differences - with NONCOG without BASIC features, except age and gender

	AUC	Age+Gender	BASIC	NONCOG	NONCOG w/o BASIC
			Program dropout		
Age+Gender	0.6196	–	0.050^{+}	**0.092****	-0.043
BASIC	0.6698		–	**0.041***	**-0.094****
NONCOG	0.7112			–	**-0.135*****
NONCOG w/o BASIC	0.5761				–
			Institution dropout		
	AUC	Age+Gender	BASIC	NONCOG	NONCOG w/o BASIC
Age+Gender	0.5327	–	**0.168*****	**0.118*****	0.044
BASIC	0.7008		–	**-0.050****	**-0.125*****
NONCOG	0.6510			–	**-0.075****
NONCOG w/o BASIC	0.5762				–
			System dropout		
	AUC	Age+Gender	BASIC	NONCOG	NONCOG w/o BASIC
Age+Gender	0.5792	–	0.059	**0.095***	0.052
BASIC	0.6377		–	0.037	-0.007
NONCOG	0.6747			–	-0.044
NONCOG w/o BASIC	0.6310				–

Notes: Each panel reports pairwise differences in the area under the ROC curve (AUC) between XGBoost models, based on DeLong's test for two correlated ROC curves. Cells above the diagonal show AUC(column) – AUC(row); positive values indicate superior predictive performance of the column model. "w/o BASIC " denotes the NONCOG model estimated without BASIC features, but including age and gender. Significance levels: $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table A.10: Test of pairwise AUC differences – completion of higher secondary school within 4 years of graduating from lower secondary school

	AUC	Age+Gender	Basic	Backgr	Noncog	All
Age+Gender	0.6434	–	**0.209*****	**0.212*****	**0.213*****	**0.217*****
Basic	0.8520		–	**0.003*****	**0.004*****	**0.009*****
Backgr	0.8551			–	0.001	**0.006*****
Noncog	0.8561				–	**0.005*****
All	0.8606					–

Notes: Column 2 reports the Area Under the ROC Curve (AUC) for XGBoost models predicting completion of higher secondary school within 4 years of graduating from lower secondary school, using data from Hvidman et al. (2024). Columns 3–7 report pairwise differences in AUC between predictive models, based on DeLong's test for two correlated ROC curves. Cells above the diagonal show AUC(column) − AUC(row); positive values therefore indicate that the model in the column achieves higher predictive performance than the model in the row. Basic, Backgr, Noncog, and All refer to the information sets described in Section 6.3.3. Significance levels: $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

A.9 Survey instructions

.

Overview

The study was conducted online using the Qualtrics survey software. It could be taken either in Danish or English. We reproduce below the English version. Potential participants were invited in separate waves spread over September to October 2013.

Overview and order of elements in the online study

1. Introduction and consent
2. Competition
3. Time preferences task
4. Risk preferences
5. Beauty contest, survey questions, cognitive reflection test, concluding remarks

Introduction and Consent

Page 1

Introduction

Welcome to the scientific study on Aarhus University students' traits, behaviors and study outcomes conducted by Alexander Koch(Department of Economics and Business, Aarhus University). Thanks for your help with this study! The study is funded by the Aarhus University Research Foundation (AUFF) and the Danish Council for Independent Research | Social Sciences (FSE). The study is conducted online and consists of a **survey** and **several tasks**.

If you complete all parts you can earn in total up to 480 kr.

First part of study: You now start with the first part of the study, which runs this week. If you complete it you will for sure receive **50 kr. for your participation.** You will perform several tasks that allow you to earn more than these 50 kr. The exact amount depends on your and others' decisions and chance. **All in all you can earn up to 280 kr.** This part will take about **45-60 minutes**.

To be eligible for these payments you need to complete the entire first part.

Second part of study: By completing the first part you qualify for the second part of the study. This part requires making five decisions today and working on tasks in week 37 (9.-15.9.2013) and week 38 (16.-22.9.2013). In total, working on the tasks takes about **60 minutes** spread over 2 weeks. If you complete the second part, you receive **an additional 200kr**. We will ask you later whether or not you want to participate in the second part of the study.

Navigation:

- You do not need to do everything in one go. Your completed answers will be automatically saved and you can use your personalized link from the email to return as often as you like to complete the remaining parts before 23:59h on Sunday, 8.9.2013.
- Use the >> button to move to the next page. Note that once you pressed the >> button, in most cases you can't access that page anymore.
- You can choose the language (Danish or English) in the box at the upper right corner.
- Sometimes you might have to scroll down to reach the end of a page.
- Closing help boxes: you find the "close"-button at the bottom of the help-page. If you open a help box you might need to scroll up or down to find it.

Page 2

Eligibility for this study: To participate in this study you need to have a **Danish bank account** and will need to enter your **CPR number**, which will be transmitted by a secure internet connection. The CPR number is needed to pay you for participating in this study.
Payments: Aarhus University will automatically transfer the amount you earn into your NemKonto. This is simply your existing bank account, into which all payments from the public sector flow (e.g. tax refunds or SU student grants). Alexander Koch and his team will start registering the payments with the administration of Aarhus University in week 39 (23.-29.9.2013). Then the administrative process might take between 2-6 weeks. You can contact Alexander Koch by email (akoch@econ.au.dk) if you want information on the payment process. Please write this email address down, so that you have his contact details in case you later have any questions!
Taxes: According to Danish law, Aarhus University reports payments to the tax authorities. Please note that taxes might be deducted from the amount of money you earn. That is, the amount you will receive might be lower than the one stated.
Data protection: The data from this study will only be used for the purpose of scientific investigations. All the information will be analyzed and reported anonymously. CPR numbers are used to anonymously link the data with data from studies in which you may choose to participate in the future, student registers and public registers. The project is notified to the Danish Data Protection Agency (Datatilsynet) and the Ethics Commission (Videnskabsetisk Komité), and complies with their terms for protecting your privacy. By participating you agree that your data is used in the described way. Your participation in this study is voluntary and you are free to withdraw from it at any time.

Contact information: You can contact Alexander Koch (akoch@econ.au.dk) if you have further questions.

Page 3

I have read this information, accept the terms and conditions, and would like to participate in this study.

Yes No

Page 4

Please enter your CPR-Number (or your temporary CPR-number), which will be transmitted by a secure internet connection. Write it in without spaces or hyphen (e.g. 0112401234): We cannot pay you for your participation in the study without a correct and complete CPR-number!

Please confirm your CPR-Number:

Page 5

You start with several tasks that allow you to earn an additional amount of money beyond the 50 kr. paid for participating in this week's part of the study. The exact amount will depend on your and others' decisions and chance. After you performed these tasks, you move on to some survey questions. Remember that you need to complete the entire first part of the study this week to be eligible for the payments from the tasks and the 50 kr. for participation.

Instructions competition

Page 1

Your task is to count zeros in a series of tables. Such a table looks like follows and once you have counted the number of zeros in a table, you should enter the number of zeros in that table into a field below the table.

0	1	1	0	0
0	0	0	1	0
0	1	0	0	1
1	0	0	0	0
1	1	0	1	0
1	0	1	0	0

How many zeros are in the table?

(11 is the correct answer for this table)

On the next page you will have 3 minutes to count zeros in up to 40 tables. You earn 50 ører for each table where you counted the number of zeros correctly.

Once you finished a table, please scroll down to access the next table. Use the tab key to jump to the next data entry field, or select the field with a mouse click. The remaining time will be displayed on the right-hand side of the screen. After the 3 minutes have elapsed, all your entered answers will be saved and you will automatically be redirected to the next screen.

Do not use the back/forward/reload screen, etc. buttons on your browser toolbar. Do not close the browser. Doing so may invalidate results, in which case you will not receive payments for this task.

When you are ready to start, press the >> button.

Page 2

You have 3 minutes to count the number of zeros in up to 40 tables.

[Tables]

Page 3

Your answers have been registered. Continue now with the task.

Page 4

<u>Round 2</u>

You will again have 3 minutes to count the number of zeros in up to 40 tables. But now you can choose whether you want to be paid based just on your own performance or whether you want to compete with the performance of other participants in this study. Please select how you want to be paid for round 2:

- No competition: 50 ører per correctly counted table.
- Competition: 1 kr. per correctly counted table if you correctly count more tables than one randomly chosen participant did in round 1. If you count the same number, you get paid 1 kr. per correct table with probability 50 percent. If you count fewer tables correctly than the randomly chosen participant did in round 1, you earn nothing.

Page 5

Before you start counting for round 2, we ask you to rank your performance in round 1 relative to the performance of the other participants in round 1. Drag the slider to indicate your belief about your rank. For example, positioning the slider at 30, means that you think that 30 percent of all participants have fewer correct tables than you in round 1, and that 70 percent have more.
We add 5 kr. to your earnings if your answer hits your true rank plus / minus 5 percentage points. For example, suppose 30 percent of all participants had fewer correct tables in round 1 than you had in round 1. Then you get 5kr. if your slider was positioned somewhere between 25 and 35 percent.

What percent of participants had fewer correct tables than you in round 1?

[slider]

Page 6

When you are ready to start round 2 of counting zeros, move to the next page.

Do not use the back/forward/reload screen, etc. buttons on your browser toolbar. Do not close the browser. Doing so may invalidate results, in which case you will not receive payments for this task.

Page 7

You have 3 minutes to count the number of zeros in up to 40 tables.

[Tables]

Page 8

Your answers for this task have been registered. Please continue now with the next task. If you complete the entire first part, then you will receive an email when we initiate the payments to your bank account with feedback about the number of tables you correctly counted and a summary of your earnings from this task.

Instructions time preferences

In the survey experiment, before the risk task (Week 0)

Notes:

- If participants did not complete this part, they could do it again at the end of the survey

- The current weekday (Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, or Sunday) and date (referred to below as t_0) is stored by the survey and used to display dates.

Page 1
Preparation for the second part of the study

Before continuing with the first part of the study, you now can decide whether to participate in the second part of the study. Here you can earn an additional 200 kr. If you participate, then you **make five decisions now; and on [weekday] of next week ([date for t_0+7 days]) as well as on [weekday] of the week after ([date for t_0+14 days]), you will have to count tables** – just like the ones you counted now. Counting will take approximately 60 minutes in total.

If you complete all these tasks you will receive 200kr. in addition to your other earnings from the first part of study. Please note: you need to complete the entire first part this week to be eligible for the second part of the study. You can opt out (at any time) of the second part without losing your earnings from the first part.

Would you like to participate?

- Yes, I want to participate in the first and second part of the study.
- No, I only want to participate in the first part of the study.[> continue with risk question]

Page 2
Today's five decisions for the second part: Schedule your work!

Next week on [weekday] ([date for t_0+7 days]) and in two weeks on [weekday] ([date for t_0+14 days]) you will have to count zeros in a number of tables – just like the ones you counted before. A table is only completed if you counted the number of zeros in it correctly. If you miscount a table, you will be asked to count it again.

In each week, you first have to complete 40 tables. In addition to these 40 tables, you have to complete a certain number of tables. You choose how many of these tables to complete in each week by making work schedules**. In a work schedule, you state how many tables you want to complete one week from today ([weekday, date for t_0+7 days]), and how many you want to complete two weeks from today ([weekday, date for t_0+14 days]).**

Page 3
Work schedules

You choose a work schedule from a list. Look at the example for such a list below. A work schedule states how many tables you want to complete next week and how many in two weeks. For example, the row "60 tables next week - 60 tables two weeks from now" means "I want to complete 60 tables on [weekday] of next week ([date for t_0+7 days]) and 60 tables on [weekday] in the week after ([date for t_0+14 days])." From the 31 possible work schedules in the list, you select the work schedule that you like best.

In the example, every table you complete in next week reduces the number of tables you have to complete in two weeks by one. We refer to this as a **1:1** "**exchange rate"**. On the next screen we explain exchange rates further.

Work schedule example - exchange rate 1:1

[drop down list – see table below; one needs to choose away from default text “*Work schedule example - exchange rate 1:1****”]***

Page 4

Work schedules and exchange rates

There are 5 different exchange rates. For each of these exchange rates you choose a work schedule. That is, you make 5 work schedules. For example, the exchange rate may be 1:1.5, such that every table you complete next week reduces the number of tables you have to complete in two weeks by 1.5. Or, the exchange rate may be 1:0.5, such that every table you complete next week reduces the number of tables you have to complete in two weeks by 0.5.

One of the 5 work schedules may become the "work schedule that counts". If a work schedule is the "work schedule that counts", you have to complete the number of tables that you specified in this work schedule to be eligible for payments. Next week, we will inform you which work schedule is the "work schedule that counts" and give more details about the process.

You receive 200kr. if you complete all the tables as specified in the "work schedule that counts" and the additional 40 tables each week.

Page 5

Choose work schedules

Choose your work schedules for the 5 different exchange rates below. There are no right or wrong choices!

Remember:

- **Each work schedule could be chosen to be the "work schedule that counts". Thus, you should make each work schedule as if it were the "work schedule that counts".**
- **The tables in the work schedule are in addition to the 40 tables you have to complete each week.**

Help ***[see below for help text that appears when clicking here]***

[Decisions]

Work schedule 1: exchange rate 1:1.5	[dropdown list]
Work schedule 2: exchange rate 1:1.25	[dropdown list]
Work schedule 3: exchange rate 1:1	[dropdown list]
Work schedule 4: exchange rate 1:0.75	[dropdown list]
Work schedule 5: exchange rate 1:0.5	[dropdown list]

Help text (pop-up window):

Click on each of the dropdown lists to select the 5 work plans

For example, the exchange rate may be 1:1.5, such that every table you complete next week reduces the number of tables you have to complete two weeks from now by 1.5. Or, the exchange rate may be 1:0.5, such that every table you complete next week reduces the number of tables you have to complete two weeks from now by 0.5.

Dropdown menu items (next page)

Work schedule 1: exchange rate 1:1.5	**Work schedule 2: exchange rate 1:1.25**	**Work schedule 3: exchange rate 1:1**	**Work schedule 4: exchange rate 1:0.75**	**Work schedule 5: exchange rate 1:0.5**
Choose work schedule 1	Choose work schedule 2	Choose work schedule 3	Choose work schedule 4	Choose work schedule 5
120 tables next week - 0 tables two weeks from now	120 tables next week - 0 tables two weeks from now	120 tables next week - 0 tables two weeks from now	120 tables next week - 0 tables two weeks from now	120 tables next week - 0 tables two weeks from now
116 tables next week - 6 tables two weeks from now	116 tables next week - 5 tables two weeks from now	116 tables next week - 4 tables two weeks from now	116 tables next week - 3 tables two weeks from now	116 tables next week - 2 tables two weeks from now
112 tables next week - 12 tables two weeks from now	112 tables next week - 10 tables two weeks from now	112 tables next week - 8 tables two weeks from now	112 tables next week - 6 tables two weeks from now	112 tables next week - 4 tables two weeks from now
108 tables next week - 18 tables two weeks from now	108 tables next week - 15 tables two weeks from now	108 tables next week - 12 tables two weeks from now	108 tables next week - 9 tables two weeks from now	108 tables next week - 6 tables two weeks from now
104 tables next week - 24 tables two weeks from now	104 tables next week - 20 tables two weeks from now	104 tables next week - 16 tables two weeks from now	104 tables next week - 12 tables two weeks from now	104 tables next week - 8 tables two weeks from now
100 tables next week - 30 tables two weeks from now	100 tables next week - 25 tables two weeks from now	100 tables next week - 20 tables two weeks from now	100 tables next week - 15 tables two weeks from now	100 tables next week - 10 tables two weeks from now
96 tables next week - 36 tables two weeks from now	96 tables next week - 30 tables two weeks from now	96 tables next week - 24 tables two weeks from now	96 tables next week - 18 tables two weeks from now	96 tables next week - 12 tables two weeks from now
92 tables next week - 42 tables two weeks from now	92 tables next week - 35 tables two weeks from now	92 tables next week - 28 tables two weeks from now	92 tables next week - 21 tables two weeks from now	92 tables next week - 14 tables two weeks from now
88 tables next week - 48 tables two weeks from now	88 tables next week - 40 tables two weeks from now	88 tables next week - 32 tables two weeks from now	88 tables next week - 24 tables two weeks from now	88 tables next week - 16 tables two weeks from now
84 tables next week - 54 tables two weeks from now	84 tables next week - 45 tables two weeks from now	84 tables next week - 36 tables two weeks from now	84 tables next week - 27 tables two weeks from now	84 tables next week - 18 tables two weeks from now
80 tables next week - 60 tables two weeks from now	80 tables next week - 50 tables two weeks from now	80 tables next week - 40 tables two weeks from now	80 tables next week - 30 tables two weeks from now	80 tables next week - 20 tables two weeks from now
76 tables next week - 66 tables two weeks from now	76 tables next week - 55 tables two weeks from now	76 tables next week - 44 tables two weeks from now	76 tables next week - 33 tables two weeks from now	76 tables next week - 22 tables two weeks from now
72 tables next week - 72 tables two weeks from now	72 tables next week - 60 tables two weeks from now	72 tables next week - 48 tables two weeks from now	72 tables next week - 36 tables two weeks from now	72 tables next week - 24 tables two weeks from now
68 tables next week - 78 tables two weeks from now	68 tables next week - 65 tables two weeks from now	68 tables next week - 52 tables two weeks from now	68 tables next week - 39 tables two weeks from now	68 tables next week - 26 tables two weeks from now
64 tables next week - 84 tables two weeks from now	64 tables next week - 70 tables two weeks from now	64 tables next week - 56 tables two weeks from now	64 tables next week - 42 tables two weeks from now	64 tables next week - 28 tables two weeks from now
60 tables next week - 90 tables two weeks from now	60 tables next week - 75 tables two weeks from now	60 tables next week - 60 tables two weeks from now	60 tables next week - 45 tables two weeks from now	60 tables next week - 30 tables two weeks from now
56 tables next week - 96 tables two weeks from now	56 tables next week - 80 tables two weeks from now	56 tables next week - 64 tables two weeks from now	56 tables next week - 48 tables two weeks from now	56 tables next week - 32 tables two weeks from now
52 tables next week - 102 tables two weeks from now	52 tables next week - 85 tables two weeks from now	52 tables next week - 68 tables two weeks from now	52 tables next week - 51 tables two weeks from now	52 tables next week - 34 tables two weeks from now
48 tables next week - 108 tables two weeks from now	48 tables next week - 90 tables two weeks from now	48 tables next week - 72 tables two weeks from now	48 tables next week - 54 tables two weeks from now	48 tables next week - 36 tables two weeks from now
44 tables next week - 114 tables two weeks from now	44 tables next week - 95 tables two weeks from now	44 tables next week - 76 tables two weeks from now	44 tables next week - 57 tables two weeks from now	44 tables next week - 38 tables two weeks from now
40 tables next week - 120 tables two weeks from now	40 tables next week - 100 tables two weeks from now	40 tables next week - 80 tables two weeks from now	40 tables next week - 60 tables two weeks from now	40 tables next week - 40 tables two weeks from now
36 tables next week - 126 tables two weeks from now	36 tables next week - 105 tables two weeks from now	36 tables next week - 84 tables two weeks from now	36 tables next week - 63 tables two weeks from now	36 tables next week - 42 tables two weeks from now
32 tables next week - 132 tables two weeks from now	32 tables next week - 110 tables two weeks from now	32 tables next week - 88 tables two weeks from now	32 tables next week - 66 tables two weeks from now	32 tables next week - 44 tables two weeks from now
28 tables next week - 138 tables two weeks from now	28 tables next week - 115 tables two weeks from now	28 tables next week - 92 tables two weeks from now	28 tables next week - 69 tables two weeks from now	28 tables next week - 46 tables two weeks from now
24 tables next week - 144 tables two weeks from now	24 tables next week - 120 tables two weeks from now	24 tables next week - 96 tables two weeks from now	24 tables next week - 72 tables two weeks from now	24 tables next week - 48 tables two weeks from now
20 tables next week - 150 tables two weeks from now	20 tables next week - 125 tables two weeks from now	20 tables next week - 100 tables two weeks from now	20 tables next week - 75 tables two weeks from now	20 tables next week - 50 tables two weeks from now
16 tables next week - 156 tables two weeks from now	16 tables next week - 130 tables two weeks from now	16 tables next week - 104 tables two weeks from now	16 tables next week - 78 tables two weeks from now	16 tables next week - 52 tables two weeks from now
12 tables next week - 162 tables two weeks from now	12 tables next week - 135 tables two weeks from now	12 tables next week - 108 tables two weeks from now	12 tables next week - 81 tables two weeks from now	12 tables next week - 54 tables two weeks from now
8 tables next week - 168 tables two weeks from now	8 tables next week - 140 tables two weeks from now	8 tables next week - 112 tables two weeks from now	8 tables next week - 84 tables two weeks from now	8 tables next week - 56 tables two weeks from now
4 tables next week - 174 tables two weeks from now	4 tables next week - 145 tables two weeks from now	4 tables next week - 116 tables two weeks from now	4 tables next week - 87 tables two weeks from now	4 tables next week - 58 tables two weeks from now
0 tables next week - 180 tables two weeks from now	0 tables next week - 150 tables two weeks from now	0 tables next week - 120 tables two weeks from now	0 tables next week - 90 tables two weeks from now	0 tables next week - 60 tables two weeks from now

Page 6

Your work schedules have been registered. Remember that you need to complete the first part of the study this week to be eligible for the second part of the study. If you complete the entire first part, you will in 6 days, on [date for t_0+6 days] at 20:00h, receive an email with further instructions and a link allowing you to log in and complete the second part of the study.

Please continue now with the first part of the study.

One week after the survey experiment (Week 1)

Page 1

Complete 40 tables

Welcome to today's tasks. First, please count the number of zeros in the following 40 tables. If you miscount a table, you will be asked to count it again. Thereafter, we will give you information on the work schedules.

40 pages with tables like this one

Table 1/ 40

0	1	1	0	0
0	0	0	1	0
0	1	0	0	1
1	0	0	0	0
1	1	0	1	0
1	0	1	0	0

How many zeros are in the table?

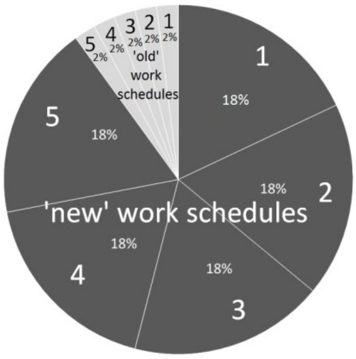
Page 42

Make 5 new work schedules

Remember that last week you made 5 work schedules for how many tables you wanted to complete this week and how many you wanted to complete one week from today. **You can now revise your work schedules and make 5 "new" work schedules**. A work schedule states how many tables you want to complete today and how many you want to complete one week from today (on [weekday], [date for t_0+14 days]).

The computer picks one work schedule to be the "work schedule that counts". Each of the "new" work schedules has an 18 percent chance of being picked as the "work schedule that counts". Each of the "old" work schedules has a 2 percent chance of being picked as the "work schedule that counts" (see figure).

That is, overall, **there is a 90 percent probability that one of the 5 "new" work schedules will be the "work schedule that counts", and there is a 10 percent probability that one of 5 the "old" work schedules from last week will be the "work schedule that counts"**. You will be informed about the "work schedule that counts" before you start counting tables. Remember: you have to complete the exact number of tables that is specified in the "work schedule that counts".



Page 43

Choose work schedules

Choose your work schedules for the 5 different exchange rates below. There are no right or wrong choices!

Remember:

- **Each work schedule could be chosen to be the "work schedule that counts". Thus, you should make each work schedule as if it was the "work schedule that counts".**
- **To complete the task and receive the 200 kr. you have to complete today's tables from the "work schedule that counts" by 23:59h on [weekday], [date for t_0+7 days], and you have to complete next week's tables on [weekday], [date for t_0+14 days], by 23:59h.**
- **The tables in the work schedule are in addition to the 40 tables you have to complete next week.**
- **Next week on at 20:00h you will receive an email with a link allowing you to log in to complete next week's tables.**

Help

[Decisions]

Work schedule 1: exchange rate 1:1.5	[dropdown list]
Work schedule 2: exchange rate 1:1.25	[dropdown list]
Work schedule 3: exchange rate 1:1	[dropdown list]
Work schedule 4: exchange rate 1:0.75	[dropdown list]
Work schedule 5: exchange rate 1:0.5	[dropdown list]

Help text and dropdown menu items (see week 0 instructions above)

Page 44

The following work schedule from this/last week ('new/old work schedule') has been chosen and thus is the "work schedule that counts":

[X] tables now and [Y] tables next week on [weekday], [date for t_0+14 days]

That is, **to complete the task and receive the 200 kr. you have to complete [X] tables by 23:59h on [weekday], [date for t_0+7 days] and you have to complete [Y] + 40 tables next week [weekday], [date for t_0+14 days].** Next week on [date for t_0+13 days] at 20:00h you will receive an email with a link allowing you to log in to complete next week's tables.

Page 45

Complete the tables from the binding work schedule

Now you have to **complete the [X] tables that were specified for this week in the "work schedule that counts".** If you miscount a table, you will be asked to count it again.

X pages with tables
[Tables from Work schedule]

Final page
You have now completed the tables for this week. Next week on [date for t_0+13 days] at 20:00h you will receive an email with a link allowing you to log in to complete next week's tables. Remember, **to complete the task and receive the 200 kr. you have to complete [Y] + 40 tables next week (on [weekday], [date for t_0+14 days]).**

Two weeks after the survey experiment (Week 2)

Page 1
Complete 40 tables

Welcome to today's tasks. First, please count the number of zeros in the following 40 tables. If you miscount a table, you will be asked to count it again. Thereafter, we will give you information on the work schedules.

40 pages with tables
[40 Tables]

Page 42
Complete the tables from the binding work schedule

Now you have to **complete the [Y] tables that were specified in the "work schedule that counts"**. If you miscount a table, you will be asked to count it again.

Remember, to complete the task and receive the 200 kr. you have to complete the tables **by 23:59h on [weekday], [date for t_0+14 days]**.

Y pages with tables
[Tables from Work schedule]

Final page
Thank you for participating. **You completed the second part of the study and you will thus receive 200 kr. in addition to your other earnings from the first part of the study.**

Alexander Koch and his team will start registering the payments with the administration of Aarhus University in week [week number]. Then the administrative process might take 2-6 weeks. You can contact Alexander Koch by email (akoch@econ.au.dk) if you want information on the payment process.

Instructions risk preferences

Page 1
Task 2

In this task, there are 9 questions. In each question you make choices between two alternatives - Alternative A and Alternative B. There are no right or wrong answers. Here is an example.

Alternative A: you get an amount of money for sure.

Alternative B: the amount of money you get is uncertain. That is, you win 0 kr. with probability 50 percent and you win 100 kr. with probability 50 percent.

Click image to enlarge

Each question gives you a list with different sure amounts of money. Each amount corresponds to a possible Alternative A. For each amount you decide whether you like Alternative A or Alternative B better.

To make your life easier, we implement a simple procedure such that you do not have to enter an answer for each amount. Look at the table below. **Consider the first row "Alternative A: win 0 kr. for sure":**

- You might prefer the 50 percent chance of winning 100kr. (Alternative B) over getting nothing for sure (Alternative A).

Now consider the row at the bottom "Alternative A: win 100 kr. for sure"

- Here you might prefer to win 100kr. for sure (Alternative A) over taking the risk of getting nothing in 50 percent of all cases (Alternative B).

And somewhere in between these two rows, there is a point where the sure Alternative A becomes more attractive to you than the risky Alternative B.

Click on the box for this Alternative A. Based on this answer the computer automatically fills in the rest of the table:

- The computer ticks Alternative A for the amount you selected and for all larger amounts.
- The computer ticks Alternative B for all smaller amounts.

After you clicked a box, you can change your choices by clicking on a different box.

Try this now! On the next screen we explain how you get paid.

Do you like Alternative A or Alternative B better?
(Remember: you only have to click on the first row where Alternative A becomes more attractive to you than the risky Alternative B)

	Alternative A You get the sure amount	**Alternative B**  **Click image to enlarge**
Alternative A: win **0kr.** for sure	○	○
Alternative A: win **5kr.** for sure	○	○
Alternative A: win **10kr.** for sure	○	○
Alternative A: win **15kr.** for sure	○	○
Alternative A: win **20kr.** for sure	○	○
Alternative A: win **25kr.** for sure	○	○
Alternative A: win **30kr.** for sure	○	○
Alternative A: win **35kr.** for sure	○	○
Alternative A: win **40kr.** for sure	○	○
Alternative A: win **45kr.** for sure	○	○
Alternative A: win **50kr.** for sure	○	○
Alternative A: win **55kr.** for sure	○	○
Alternative A: win **60kr.** for sure	○	○
Alternative A: win **65kr.** for sure	○	○
Alternative A: win **70kr.** for sure	○	○
Alternative A: win **75kr.** for sure	○	○
Alternative A: win **80kr.** for sure	○	○
Alternative A: win **85kr.** for sure	○	○
Alternative A: win **90kr.** for sure	○	○
Alternative A: win **95kr.** for sure	○	○
Alternative A: win **100kr.** for sure	○	○

Page 2

Here is how you will be paid.

After you have answered all 9 questions, the computer randomly selects one of the 9 questions as the **'question that is paid'**. Each question is equally likely to be selected.

For the 'question that is paid' the computer randomly selects one of the rows from the list in that question as the **'row that counts'**. Each row is equally likely to be selected.

For the row that counts the computer checks whether you liked Alternative A or Alternative B better. If you liked Alternative A better, then you get the sure amount that is listed in that row. If you liked Alternative B better, then the computer randomly selects the outcome for this alternative.

Let's consider the example from the previous screen. Suppose 20 kr. is the sure amount where Alternative A becomes more attractive to you than the Alternative B (row 5).

- If, for example, row 7 was selected as the 'row that counts': For that row your choice is Alternative A. Hence, you would get paid according to Alternative A. That is, you win 30 kr.
- If, for example, row 3 was selected as the 'row that counts': For that row your choice is Alternative B. Hence, you would get paid according to Alternative B. That is, you win 0 kr. with 50 percent chance and win 100 kr. with 50 percent chance.

Start with the questions on the next screen.

	Alternative A You get the sure amount	Alternative B  Click image to enlarge
1. Alternative A: win **0kr.** for sure	○	◉
2. Alternative A: win **5kr.** for sure	○	◉
3. Alternative A: win 10kr. for sure	○	◉
4. Alternative A: win **15kr.** for sure	○	◉
5. Alternative A: win 20kr. for sure	◉	○
6. Alternative A: win **25kr.** for sure	◉	○
7. Alternative A: win 30kr. for sure	◉	○
8. Alternative A: win **35kr.** for sure	◉	○
9. Alternative A: win **40kr.** for sure	◉	○
10. Alternative A: win **45kr.** for sure	◉	○
11. Alternative A: win **50kr.** for sure	◉	○
12. Alternative A: win **55kr.** for sure	◉	○
13. Alternative A: win **60kr.** for sure	◉	○
14. Alternative A: win **65kr.** for sure	◉	○
15. Alternative A: win **70kr.** for sure	◉	○
16. Alternative A: win **75kr.** for sure	◉	○
17. Alternative A: win **80kr.** for sure	◉	○
18. Alternative A: win **85kr.** for sure	◉	○
19. Alternative A: win **90kr.** for sure	◉	○
20. Alternative A: win **95kr.** for sure	◉	○
21. Alternative A: win **100kr.** for sure	◉	○

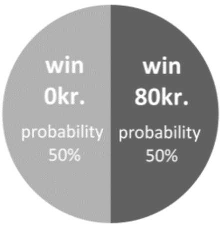
Pages 3-5

Question Block I: gain questions g40_120, g0_80, g0_160 (randomized order). They all have the same structure as below. Sure amounts are summarized in a table at the end for all risk questions.

<u>Question nr. /9</u>

Consider the following alternatives.

- Alternative A: you win an amount of money for sure.
- Alternative B: the amount of money you receive is uncertain. That is, you win 0 kr. with probability 50 percent and you win 80 kr. with probability 50 percent.



Click image to enlarge

Help

Do you like Alternative A or Alternative B better?
(Remember: you only have to click on the first row where Alternative A becomes more attractive to you than the risky Alternative B)

	Alternative A You win the sure amount	**Alternative B**  **Click image to enlarge**
Alternative A: win **0kr.** for sure	○	○
Alternative A: win **4kr.** for sure	○	○
Alternative A: win **8kr.** for sure	○	○
Alternative A: win **12kr.** for sure	○	○
Alternative A: win **16kr.** for sure	○	○
Alternative A: win **20kr.** for sure	○	○
Alternative A: win **24kr.** for sure	○	○
Alternative A: win **28kr.** for sure	○	○
Alternative A: win **32kr.** for sure	○	○
Alternative A: win **36kr.** for sure	○	○
Alternative A: win **40kr.** for sure	○	○
Alternative A: win **44kr.** for sure	○	○
Alternative A: win **48kr.** for sure	○	○
Alternative A: win **52kr.** for sure	○	○
Alternative A: win **56kr.** for sure	○	○
Alternative A: win **60kr.** for sure	○	○
Alternative A: win **64kr.** for sure	○	○
Alternative A: win **68kr.** for sure	○	○
Alternative A: win **72kr.** for sure	○	○
Alternative A: win **76kr.** for sure	○	○
Alternative A: win **80kr.** for sure	○	○

Help text (pop-up window)

You only have to click on the first row where Option A becomes more attractive to you than the risky Option B.

Based on this answer the computer automatically fills in the rest of the table:

- The computer ticks Alternative A for the amount you selected and for all larger amounts.
- The computer ticks Alternative B for all smaller amounts.

After you clicked a box, you can change your choices by clicking on a different box.

Reminder of how you get paid:

After you have answered all 9 questions, the computer randomly selects one of them as the **'question that is paid'**. Each question is equally likely to be selected.

For the 'question that is paid' the computer randomly selects one of the rows from the list in that question as the **'row that counts'**. Each row is equally likely to be selected.

For the row that counts the computer checks whether you liked Alternative A or Alternative B better. If you liked Alternative A better, then you get the sure amount that is listed in that row. If you liked Alternative B better, then the computer randomly selects the outcome for this alternative.

Page 6

Introduction to questions with possible losses

If one of the next 6 questions is selected for payment you will be given an extra amount on top of your other earnings. Each question has a different extra amount. You can see the exact extra amount when you answer the question.

You will be asked to make choices, which may involve losing money. If your choice involves losing money, these losses will be taken out of the extra amount you receive for the question.

Page 7

Question Block II: mixed gain-loss question m40_40 with endowment 40 or 80 (randomized; the version with the other endowment is then shown as question 9, i.e. in block IV).

<u>Question nr. 4/9</u>

If this question is selected for payment you will be given 40 kr. extra on top of your other earnings.

Consider the following alternatives.

- Alternative A: you lose or win an amount of money for sure.
- Alternative B: the amount of money you receive is uncertain. That is, you lose 40kr. with probability 50 percent and you gain 40kr. with probability 50 percent.

Click image to enlarge

Help

Do you like Alternative A or Alternative B better?
(Remember: you only have to click on the first row where Alternative A becomes more attractive to you than the risky Alternative B)

	Alternative A You lose/gain the sure amount	Alternative B  Click image to enlarge
Alternative A: lose **40kr.** for sure	○	○
Alternative A: lose **36kr.** for sure	○	○
Alternative A: lose **32kr.** for sure	○	○
Alternative A: lose **28kr.** for sure	○	○
Alternative A: lose **24kr.** for sure	○	○
Alternative A: lose **20kr.** for sure	○	○
Alternative A: lose **16kr.** for sure	○	○
Alternative A: lose **12kr.** for sure	○	○
Alternative A: lose **8kr.** for sure	○	○
Alternative A: lose **4kr.** for sure	○	○
Alternative A: lose **0kr.** for sure	○	○
Alternative A: win **4kr.** for sure	○	○

…

Help text (pop-up window)

You only have to click on the first row where Option A becomes more attractive to you than the risky Option B.

Based on this answer the computer automatically fills in the rest of the table:

- The computer ticks Alternative A for the amount you selected and for all larger amounts.
- The computer ticks Alternative B for all smaller amounts.

After you clicked a box, you can change your choices by clicking on a different box.

Reminder of how you get paid:

If this question is selected for payment you you will be given 40kr. extra on top of your other earnings. If your choice involves losing money, these losses will be taken out of these 40 kr.

After you have answered all 9 questions, the computer randomly selects one of them as the **'question that is paid'**. Each question is equally likely to be selected.

For the 'question that is paid' the computer randomly selects one of the rows from the list in that question as the **'row that counts'**. Each row is equally likely to be selected.

For the row that counts the computer checks whether you liked Alternative A or Alternative B better. If you liked Alternative A better, then you get the sure amount that is listed in that row. If you liked Alternative B better, then the computer randomly selects the outcome for this alternative.

Pages 8-10

Question Block III: mixed gain-loss question m80_80 with endowment 80 and loss questions l0_160 (endowment 160), l0_80 (endowment 80), l40_120 (endowment 160), (randomized). All questions have the same structure as the questions above. Sure amounts are summarized in a table at the end for all risk questions.

Pages 11

Question Block IV: mixed gain-loss question m40_40 with endowment 40 or 80 (the one not shown in block II).

Overview of certain amounts in the tables shown:

Row	g40_120	g0_80	g0_160	m40_40	m80_80	l0_160	l0_80	l40_120
1	40	0	0	-40	-80	-160	-80	-120
2	44	4	8	-36	-72	-152	-76	-116
3	48	8	16	-32	-64	-144	-72	-112
4	52	12	24	-28	-56	-136	-68	-108
5	56	16	32	-24	-48	-128	-64	-104
6	60	20	40	-20	-40	-120	-60	-100
7	64	24	48	-16	-32	-112	-56	-96
8	68	28	56	-12	-24	-104	-52	-92
9	72	32	64	-8	-16	-96	-48	-88
10	76	36	72	-4	-8	-88	-44	-84
11	80	40	80	0	0	-80	-40	-80
12	84	44	88	4	8	-72	-36	-76
13	88	48	96	8	16	-64	-32	-72
14	92	52	104	12	24	-56	-28	-68
15	96	56	112	16	32	-48	-24	-64
16	100	60	120	20	40	-40	-20	-60
17	104	64	128	24	48	-32	-16	-56
18	108	68	136	28	56	-24	-12	-52
19	112	72	144	32	64	-16	-8	-48
20	116	76	152	36	72	-8	-4	-44
21	120	80	160	40	80	0	0	-40

Note: certainty equivalents are calculated as the average between the first certain amount preferred over lottery (CE row) and the certain amount in the row before. Exceptions: the very first row (CE= lowest certain amount), or if always the lottery is preferred (CE=highest certain amount). See next table.

	sure amount	CE	sure amount	CE	sure amount	CE	sure amount	CE	sure amount	CE	sure amount	CE	sure amount	CE	sure amount	CE
Row	g40_120		g0_80		g0_160		m40_40		m80_80		l0_160		l0_80		l40_120	
1	40	40	0	0	0	0	-40	-40	-80	-80	-160	-160	-80	-80	-120	-120
2	44	42	4	2	8	4	-36	-38	-72	-76	-152	-156	-76	-78	-116	-118
3	48	46	8	6	16	12	-32	-34	-64	-68	-144	-148	-72	-74	-112	-114
4	52	50	12	10	24	20	-28	-30	-56	-60	-136	-140	-68	-70	-108	-110
5	56	54	16	14	32	28	-24	-26	-48	-52	-128	-132	-64	-66	-104	-106
6	60	58	20	18	40	36	-20	-22	-40	-44	-120	-124	-60	-62	-100	-102
7	64	62	24	22	48	44	-16	-18	-32	-36	-112	-116	-56	-58	-96	-98
8	68	66	28	26	56	52	-12	-14	-24	-28	-104	-108	-52	-54	-92	-94
9	72	70	32	30	64	60	-8	-10	-16	-20	-96	-100	-48	-50	-88	-90
10	76	74	36	34	72	68	-4	-6	-8	-12	-88	-92	-44	-46	-84	-86
11	80	78	40	38	80	76	0	-2	0	-4	-80	-84	-40	-42	-80	-82
12	84	82	44	42	88	84	4	2	8	4	-72	-76	-36	-38	-76	-78
13	88	86	48	46	96	92	8	6	16	12	-64	-68	-32	-34	-72	-74
14	92	90	52	50	104	100	12	10	24	20	-56	-60	-28	-30	-68	-70
15	96	94	56	54	112	108	16	14	32	28	-48	-52	-24	-26	-64	-66
16	100	98	60	58	120	116	20	18	40	36	-40	-44	-20	-22	-60	-62
17	104	102	64	62	128	124	24	22	48	44	-32	-36	-16	-18	-56	-58
18	108	106	68	66	136	132	28	26	56	52	-24	-28	-12	-14	-52	-54
19	112	110	72	70	144	140	32	30	64	60	-16	-20	-8	-10	-48	-50
20	116	114	76	74	152	148	36	34	72	68	-8	-12	-4	-6	-44	-46
21	120	118	80	78	160	156	40	38	80	76	0	-4	0	-2	-40	-42
22		120		80		160		40		80		0		0		-40

Survey questions, beauty contest and Cognitive Reflection Test

Beauty Contest

You have to write down a number between 0 and 100 (the number can have decimals). All the other participants of this survey do the same.

The average of all these numbers will be computed, and then this average is multiplied by two thirds. Call this number X. The winner is the participant who chose the number which is closest to X. If there are several participants who chose this number, the winner will be selected at random among them.

The winner will receive 200 kr. All other participants will receive 0 kr. for this task. You will be notified in week [Dates for PaymentWeek] whether you won.

Please enter your number here (enter decimals after a point):

Survey questions

We now would like to ask you several questions. Be honest – there are no right or wrong answers!

Remember: to be eligible for payments from the previous tasks and for the 50kr. you need to complete the entire survey.

Please read the following sentences and state how well they describe you. I decided to follow the study program I am currently enrolled in, because …

	Not like me at all	Not much like me	Somewhat like me	Mostly like me	Very much like me
I am very interested in the subject area, and I would like to know more about it:	❍	❍	❍	❍	❍
the study program fits my talents:	❍	❍	❍	❍	❍
I believe that as a graduate in this program I will have very good job opportunities and income prospects:	❍	❍	❍	❍	❍
I did not know what I should do otherwise:	❍	❍	❍	❍	❍
my family/friends recommended me to study this subject:	❍	❍	❍	❍	❍

Please read the following sentences and state how well they describe you.

	Not like me at all	Not much like me	Somewhat like me	Mostly like me	Very much like me
The study program I am enrolled in is my most desired study program:	❍	❍	❍	❍	❍
I was very certain about choosing my study program:	❍	❍	❍	❍	❍
I am very satisfied now with my chosen study program:	❍	❍	❍	❍	❍
I am very motivated for my studies:	❍	❍	❍	❍	❍
I am very certain that I will finish my studies at Aarhus University with a bachelor's or master's degree:	❍	❍	❍	❍	❍
I believe that my future income depends on my final average grade in my studies:	❍	❍	❍	❍	❍

How do you finance your studies? (you can name more than one option)

- ❑ My parents support me financially
- ❑ I get SU (Danish student grant and loan scheme)
- ❑ I have a job at the university
- ❑ I have a job outside of the university
- ❑ Other funding

What is the highest amount of money you could pay out of your own pocket within the next 3 days?

- ❍ less than 350 kr.
- ❍ 350 kr.
- ❍ 700 kr.
- ❍ 1500 kr.
- ❍ 2000 kr.
- ❍ 3500 kr.
- ❍ 7000 kr.
- ❍ more than 7000 kr.

Please state how well this sentence describes you: I divide my monthly budget into several separate budgets (such as budgets for housing, clothes, leisure expenditures, study related expenditures and the like).

- ❍ Not like me at all
- ❍ Not much like me
- ❍ Somewhat like me
- ❍ Mostly like me
- ❍ Very much like me

How many semesters do you think you will actually need to obtain the following degree:

______ a bachelor's degree in your current studies:
______ a master's degree in your current studies (exclusive semesters for bachelor's degree):

Suppose you will obtain a bachelor’s degree in your subject. What do you think will be your monthly gross income in your first year of employment (in kr.)?

- ❍ Select from the list
- ❍ less than 15 000 kr.
- ❍ 15 000 - 20 000 kr.
- ❍ 20 000 - 25 000 kr.
- ❍ 25 000 - 30 000 kr.
- ❍ 30 000 - 35 000 kr.
- ❍ 35 000 - 40 000 kr.
- ❍ 40 000 - 45 000 kr.
- ❍ 45 000 - 50 000 kr.
- ❍ 50 000 - 55 000 kr.
- ❍ 55 000 - 60 000 kr.
- ❍ more than 60 000 kr.

Suppose you will obtain a master's degree in your subject. What do you think will be your monthly gross income in your first year of employment (in kr.)?

- ❍ Select from the list
- ❍ less than 15 000 kr.
- ❍ 15 000 - 20 000 kr.
- ❍ 20 000 - 25 000 kr.
- ❍ 25 000 - 30 000 kr.
- ❍ 30 000 - 35 000 kr.
- ❍ 35 000 - 40 000 kr.
- ❍ 40 000 - 45 000 kr.
- ❍ 45 000 - 50 000 kr.
- ❍ 50 000 - 55 000 kr.
- ❍ 55 000 - 60 000 kr.
- ❍ more than 60 000 kr.

Which university qualifying exam do you have?

- ❍ Studentereksamen (stx)
- ❍ Højere forberedelseseksamen (hf)
- ❍ Højere handelseksamen (hhx)
- ❍ Højere teknisk eksamen (htx)
- ❍ International high school degree
- ❍ Another university qualifying exam

Which grade did you obtain in your university qualifying exam in the following subjects (if you have several qualifying exams, write down the best grade at the highest level)?

	Grade (Danish 7-point-scale)								Subject level (Danish classification)			
	-3	00	02	4	7	10	12	Did not have subject	A	B	C	I do not know
Danish	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
Mathematics	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
English	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
Physics	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍

Which grade did you obtain in your university qualifying exam in the following subjects (if you have several qualifying exams, write down the best grade at the highest level)?

	Grade							Subject level (Danish classification)			
	A	B	C	D	E	F	Did not have subject	A	B	C	I do not know
Danish	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
Mathematics	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
English	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍
Physics	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍	❍

When did you obtain your university qualifying exam?

❍ 2013
❍ 2012
❍ 2011
❍ 2010
❍ 2009
❍ 2008
❍ 2007
❍ 2006
❍ 2005
❍ before 2005

What did you do between your university qualifying exam and now? (You can give more than one answer)

❑ Travel
❑ Work
❑ Voluntary social work
❑ Højskole
❑ Second university qualifying exam
❑ Vocational training
❑ Completed university degree
❑ Started studying, but dropped out
❑ Other

What is the highest completed education of your parents?

Highest completed education of your mother

- 9-10 years of secondary school
- Higher secondary school (University entrance exam)
- Vocational education
- Short higher education, less than 3 years
- Long higher education, more than 3 years
- No completed education
- Other
- I do not know

Highest completed education of your father

- 9-10 years of secondary school
- Higher secondary school (University entrance exam)
- Vocational education
- Short higher education, less than 3 years
- Long higher education, more than 3 years
- No completed education
- Other
- I do not know

Which language do you speak at home with your parents?

- Danish
- Another language
- Danish and another language

Please read the following sentences and state how well they describe you.

	Not like me at all	Not much like me	Somewhat like me	Mostly like me	Very much like me
New ideas and projects sometimes distract me from previous ones:	❍	❍	❍	❍	❍
Setbacks don't discourage me:	❍	❍	❍	❍	❍
I have been obsessed with a certain idea or project for a short time but later lost interest:	❍	❍	❍	❍	❍
I am a hard worker:	❍	❍	❍	❍	❍
I often set a goal but later choose to pursue a different one:	❍	❍	❍	❍	❍
I have difficulty maintaining my focus on projects that take more than a few months to complete:	❍	❍	❍	❍	❍
I finish whatever I begin:	❍	❍	❍	❍	❍
I am diligent:	❍	❍	❍	❍	❍

Brief Self-Control Scale

Please read the following sentences and state how well they describe you.

	Not like me at all	Not much like me	Somewhat like me	Mostly like me	Very much like me
I am good at resisting temptation:	❍	❍	❍	❍	❍
I do certain things that are bad for me, if they are fun:	❍	❍	❍	❍	❍
I have a hard time breaking bad habits:	❍	❍	❍	❍	❍
I wish I had more self-discipline:	❍	❍	❍	❍	❍
I am lazy:	❍	❍	❍	❍	❍
I say inappropriate things:	❍	❍	❍	❍	❍
Pleasure and fun sometimes keep me from getting work done:	❍	❍	❍	❍	❍
I have trouble concentrating:	❍	❍	❍	❍	❍
I am able to work effectively toward long-term goals:	❍	❍	❍	❍	❍

Sometimes I can't stop myself from doing something, even if I know it is wrong:	❍	❍	❍	❍	❍
I often act without thinking through all the alternatives:	❍	❍	❍	❍	❍
People would say that I have iron self-discipline:	❍	❍	❍	❍	❍
I refuse things that are bad for me:	❍	❍	❍	❍	❍
I know that I often cannot resist temptations and thus try to avoid these temptations:	❍	❍	❍	❍	❍

Small-scale insurance

Have you ever have bought one of the following types of insurance:

	Yes	No
mobile phone theft/damage insurance:	❍	❍
bicycle insurance:	❍	❍
insurance of computer/laptop:	❍	❍
baggage insurance:	❍	❍
travel cancelation insurance:	❍	❍

Lost ticket - lost money questions (topical mental accounts)

[order randomized]

Imagine that you decided to see a play and that you paid the admission price of 200 kr. for the ticket. As you enter the theatre you notice that you have lost the ticket. Would you pay 200 kr. for another ticket?

- ❍ Very likely
- ❍ Likely
- ❍ Neither likely nor unlikely
- ❍ Unlikely
- ❍ Very unlikely

Imagine that you decided to see a play where the admission price is 200 kr. for a ticket. As you enter the theatre you notice that you have lost 200 kr. Would you still pay 200 kr. for a ticket for the play?

- ❍ Very likely
- ❍ Likely
- ❍ Neither likely nor unlikely
- ❍ Unlikely
- ❍ Very unlikely

Exam vignette (narrow goals)

Imagine that two weeks before an exam the professor hands out 30 practice exams. Furthermore, the professor tells you that all questions for the actual exam will be drawn from these practice exams. It takes you 4 hours to work through a practice exam. How would you plan your workload? Pick the one answer that describes you best:

- ❍ I set a daily study goal that specifies for each day between now and the exam date how many practice exams I want to work on.
- ❍ I set a weekly study goal that specifies for each of the two weeks between now and the exam date how many practice exams I want to work on.
- ❍ I set an overall goal that specifies how many practice exams I want to work on between now and the exam date.
- ❍ I set no goal and just see how many practice exams I manage to work on between now and the exam date.

Please read the following sentences and state how well they describe you. Related to my studies, I set...

	Not like me at all	Not much like me	Somewhat like me	Mostly like me	Very much like me
Goals for course grades:	❍	❍	❍	❍	❍
Goals for the number of study hours per day/week:	❍	❍	❍	❍	❍
Goals for regularly attending lectures and seminars:	❍	❍	❍	❍	❍
Goals for doing course work (e.g. problem sets):	❍	❍	❍	❍	❍
Goals for preparing work in study groups:	❍	❍	❍	❍	❍
Deadlines for when to complete different steps in project work:	❍	❍	❍	❍	❍

Please read the following sentences and state how well they describe you.

	Not like me at all	Not much like me	Somewhat like me	Mostly like me	Very much like me
I divide a goal into subgoals, to keep track of how I am doing:	❍	❍	❍	❍	❍
When setting a goal, I carefully think about what I want to achieve and when to achieve it:	❍	❍	❍	❍	❍
I sometimes do not set goals because I am afraid that I will not be able to achieve them:	❍	❍	❍	❍	❍
I feel angry with myself when I give up a goal:	❍	❍	❍	❍	❍
When I reach a goal I sometimes reward myself by buying something nice:	❍	❍	❍	❍	❍
I tell friends or family about my goals, to increase my motivation to achieve these goals:	❍	❍	❍	❍	❍
I set goals, but then often give them up:	❍	❍	❍	❍	❍
I set goals spontaneously:	❍	❍	❍	❍	❍
The goals I set for myself are very ambitious:	❍	❍	❍	❍	❍

Please read the following sentences and state to what extent you agree with the statement.

	Strongly disagree	Disagree	Neither/nor	Agree	Strongly agree
Mandatory course assignments are better than course assignments with a voluntary hand-in option:	❍	❍	❍	❍	❍
Project work should come with evenly spaced, strict deadlines rather than only being due at the end of term:	❍	❍	❍	❍	❍
If someone paid me money for good exam grades, I would study more:	❍	❍	❍	❍	❍
A study group motivates me to get more work done:	❍	❍	❍	❍	❍
To increase my motivation, I sometimes bet with friends or family for money, that I will reach a certain goal:	❍	❍	❍	❍	❍

What is your height in cm? (If you do not know your exact height, please make an estimate)

What is your weight in kg? (If you do not know your exact weight, please make an estimate)

How strong are you? Please rate your physical strength compared to the average of people of your age and gender:

- ❍ Much below the average
- ❍ Somewhat below the average
- ❍ Average
- ❍ Somewhat above the average
- ❍ Much above the average

How attractive are you? Please rate your physical attractiveness compared to the average of people of your age and gender:

- ❍ Much below the average
- ❍ Somewhat below the average
- ❍ Average
- ❍ Somewhat above the average
- ❍ Much above the average

ABCD question testing for viewing lotteries in isolation

For this question the computer will randomly select one participant as the 'participant who is paid'. If you are the 'participant who is paid':

- you will be given an extra 100 kr. on top of your other earnings. If your choice involves making losses, these losses will be taken out of these 100 kr.

-you will be paid for your Decision 1 and for your Decision 2 below.

You face the following pair of concurrent decisions. First examine both decisions, then indicate your choices, by ticking one of the two boxes in each decision.

Decision 1: Choose between (before answering, read Decision 2):

- ❍ winning 24 kr.
- ❍ a 25% chance of winning 100 kr. and a 75% chance of not winning or losing any money.

Decision 2: Choose between:

- ❍ losing 75 kr.
- ❍ a 75% chance of losing 100 kr., and a 25% chance of not winning or losing any money.

Cognitive reflection test

For the final 3 questions you earn 2kr. for each question that you answer correctly.

A bat and a ball cost 110kr. in total. The bat costs 100 kr. more than the ball. How much does the ball cost (in kr.)?

If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets (in minutes)?

In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake (in days)?

Concluding remarks

Thank you for participating in this study.

You now completed the first part of the study. In week [dates], you will receive an email which summarizes all your earnings and which gives you feedback on the tasks. Alexander Koch and his team will then also start registering the payments with the administration of Aarhus University. The administrative process might take up 2-6 weeks. You can contact Alexander Koch by email (akoch@econ.au.dk) if you want information on the payment process.

[*If time preference part skipped before:*
So far, you skipped the second part of the study, where you can earn an additional 200kr. Would you nevertheless like to participate in the second part of the study?

- ❍ Yes
- ❍ No]

[If participated in time preference part:
Next week [weekday, date], 20:00h you will receive an email with further instructions for the second part of the study, where you can earn an additional 200kr.]

Thank you for participating in this study.

Do you want to receive invitations to other studies in the Aarhus Cognition and Behavior Lab in which you can earn money?

- ❍ Yes
- ❍ No