Economics of Education Review 62 (2018) 170–182

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Thrivers and divers: Using non-academic measures to predict college success and failure☆

Graham Beattie[a], Jean-William P. Laliberté[b], Philip Oreopoulos[b,c,d,⁎]

[a] University of Pittsburgh, United States
[b] University of Toronto, Canada
[c] National Bureau of Economic Research, United States
[d] Canadian Institute for Advanced Research, Canada

ABSTRACT

We introduce a novel method for collecting a comprehensive set of non-academic characteristics for a representative sample of incoming freshman to explore which measures best predict the wide variance in first-year college performance unaccounted for by past grades. We focus our attention on student outliers. Students whose first-year college average is far below expectations (divers) have a high propensity for procrastination – they self-report cramming for exams and wait longer before starting assignments. They are also considerably less conscientious than their peers. Divers are more likely to express superficial goals, hoping to 'get rich' quickly. In contrast, students who exceed expectations (thrivers) express more philanthropic goals, are purpose-driven, and are willing to study more hours per week to obtain the higher GPA they expect. A simple seven-variable average of these key non-academic variables does well in predicting college achievement relative to adding more variables or letting a machine-algorithm choose, and improves our ability to predict at-risk students when used jointly with past grades.

1. Introduction

In recent decades, college enrollment has increased and both policy makers and parents have continued to emphasize the importance of post-secondary education as a worthy investment. In parallel, more attention is now directed towards helping entrants actually complete their degrees and exit with valuable experience and skills. But despite efforts to increase college support – additional tutoring, counseling, stress management workshops, time management assistance, and other resources – the fraction of students completing a degree remains alarmingly low. Only about half of students who begin a bachelors' degree in the United States complete it within six years (Symonds, Schwartz, & Ferguson, 2011). In Canada, three-quarters complete but many do so with minimum requirements and questionable skill improvement (Arum & Roksa, 2011).[1]

Understanding what factors can improve college performance predictions would allow administrators to better target students at risk of struggling and identify incoming skills particularly helpful for academic success. Previous research shows that past performance strongly predicts college achievement, which explains why institutions rely on past grades or standardized tests for admission.[2] But even for students with similar past grades, a high variance exists in subsequent performance. Similarly, there is considerable variance in high school grades among freshmen at the bottom of the college grade distribution, those most at risk of failing to graduate. Of the students who perform well enough in high school to make it to selective postsecondary institutions, a substantial fraction end up struggling and eventually drop out. Transitioning from high school to college can be challenging and success in one level of education does not guarantee success in another.[3]

☆ We are very grateful to Aloysius Siow, Nathalie Bau, Yosh Halberstam, Uros Petronijevic for helpful discussions and to seminar participants at the University of Toronto and at the NBER Education meeting for comments. We also wish to thank Aaron de Mello for spectacular web development and research assistance, and Matthew Hendrickson for helping us compile the administrative data. Financial support for this research was provided by the Ontario Human Capital Research and Innovation Fund, a Social Sciences and Humanities Research Council Insight Grant (#435-2015-0180), and a JPAL Pilot Grant. Any omissions or errors are our own responsibility.

⁎ Corresponding author.
E-mail address: philip.oreopoulos@utoronto.ca (P. Oreopoulos).

[1] Bound and Turner (2011) and Bound, Lovenheim, and Turner (2010) discuss recent trends in the US and Finnie, Childs, Finnie, and Martinello (2016) provide an analysis of Canadian trends.

[2] Bettinger, Evans, and Pope (2013), Dooley, Payne, and Robb (2012), Cyrenne and Chan (2012), Rothstein (2004).

[3] For example, Scott-Clayton, Crosta, and Belfield (2014) discuss how the difficulties associated with identifying at-risk students generate substantial mis-assignment of students to remedial classes.

https://doi.org/10.1016/j.econedurev.2017.09.008
Received 11 January 2017; Received in revised form 19 July 2017; Accepted 30 September 2017
Available online 21 November 2017
0272-7757/

Navigating this new environment with ease may require more than strong academic capabilities – students may have to “become new kinds of learners” (Farrington et al., 2012). Hence, when it comes to predicting who among admitted students with similar grades will eventually ‘thrive’ and who will ‘dive’, we need to look beyond their high school academic performance. College students arrive from an increasing variety of backgrounds with different initial abilities, hopes, goals, and expectations, all of which may influence the degree of ease with which they transition from high-school to college. For example, Bound, Lovenheim, and Turner (2010) show that a third of the decline in completion rates in recent decades can be explained by a surge in the fraction of students with weaker preparation.

Recent research on non-academic factors suggests that variables aside from past grades may help identify students who are at risk of floundering in college and those who are likely to succeed. There is ample evidence that these skills, particularly personality traits and social background, exhibit substantial predictive power for a variety of life outcomes such as educational attainment, earnings, and health.[4] Conscientiousness – a personality trait associated with staying organized, working hard, and persistence – is positively associated with educational achievement independent of intelligence.[5] Gritty students, who persevere towards achieving particular goals, tend to have higher college GPAs than their peers even after conditioning on SAT scores (Duckworth, Peterson, Matthews, & Kelly, 2007). Also, work by Mischel, Shoda, and Rodriguez (1989) and Kirby, Winston, and Santiesteban (2005) suggests that the ability to delay gratification also predicts future achievement.

In this paper, we collect a comprehensive set of non-academic characteristics for a large sample of incoming college freshmen from various backgrounds to explore which measures best predict success and failure in first-year that could not have been foreseen on the basis of past grades.[6] More specifically, we focus on the *incremental* predictive power of these measures by first absorbing the variation in college grades explained by high school grades, thereby accounting for any correlation between non-academic variables and past grades. We depart from the previous literature in focusing our attention on the highly informative, yet understudied groups of student outliers – those who end up in the bottom and top deciles in our sample, in terms of the difference between actual performance and predicted performance based on high school grades. We call students in the top decile thrivers, and those in the bottom, divers. Thrivers and divers are opposite extremes, making it easier to examine key differences in initial characteristics relative to the rest of the student population. Examining them in isolation helps avoid measuring small linear relationships from the majority of students ‘in the middle’ and allows for asymmetries between outliers. That a typical B student obtains a GPA of B+, or that a former high school valedictorian receives a GPA of A-, is not out of the ordinary. The search for non-academic predictors of a successful or failed transition to college is concentrated by isolating outlier groups that, on their own, are of particular interest. In particular, our data indicates that divers are four times more likely to drop out after a year than the average, and therefore fall in the category of students often implicitly targeted by support interventions. Focusing on them may help administrators better understand how to avoid pitfalls and promote environments for helping incoming college students.

Our approach offers a new, innovative and low-cost way of collecting both quantitative and qualitative data from large samples of students with near-perfect consent rates. Our data come from partnering with all first-year economics instructors at the three campuses of the University of Toronto and asking students to complete an online ‘warm-up exercise’ for 2% of their final grade, as part of a broader research program (Oreopoulos & Petronijevic, 2016). Over 45–90 min during the first weeks of school, the participating students completed survey questions about procrastination, study habits, social identity, academic expectations, and agreed to link their responses to the university's administrative database of background characteristics and future academic performance.[7] A subset of our data allows us to explore a wide variety of non-academic characteristics, including grit, risk aversion, time preferences, locus of control, as well as the Big Five personality traits; agreeableness, conscientiousness, extraversion, openness to experience, and emotional stability. Our sample is also very large, allowing sufficient statistical power to detect even small differences between performance groups. We explore what variables best predict first-year performance, both unconditionally as well as when conditioning on all other predictors.[8] The exercise does not attempt to uncover causal estimates, but rather document the independent and incremental predictive properties of a large number of characteristics, above and beyond what could be expected on the basis of high school grades. All relationships between these non-academic variables and college grades are estimated on the same sample, ensuring that the set of controls is consistent and that coefficients’ magnitudes are directly comparable.

We find that objective and subjective measures of procrastination and impatience are the best predictors of failing to keep up with grade expectations. Whether conditioning on other traits or characteristics or not, students that self-report tending to cram for exams, wait until the last minute in general to complete deadlines, or even wait last minute to complete the survey we collected data from them are much more likely to end up in the lowest or second lowest grade decile relative to expectations. Poor performers also tend to work many more hours for pay than their peers and are less conscientious on average. These patterns are not the same for thrivers. The best predictors for far exceeding grade expectations are self-reported intended hours of study and expected grades. Students who expect higher grades tend to get them, and thrivers plan to study over three hours a week more than divers do, on average.

Another subset of students was asked to write freely about their future goals, anticipated setbacks, and mindset. Examining their answers offers the opportunity to vastly expand the set of potential predictors beyond those explicitly measured by questionnaires reflecting researchers’ priors. We find that thrivers and divers answered these open-ended written questions differently. Thrivers write longer answers and use better spelling than divers, and are also are more likely to identify self-discipline as a trait they admire in themselves. In addition, when asked to identify future goals, thrivers are more likely to discuss the impact they want to make on society, while divers are more likely to emphasize wanting to ‘get rich’.

These findings have both theoretical and practical implications. A better understanding of the characteristics of student outliers informs us about the shape of the college education production function. Even among those who were admitted to the University of Toronto, several noncognitive skills sharply distinguish divers from other students. Accounting for the skills we measure increases the explanatory power in predicting performance over the full distribution compared to using only past performance alone, but high school grades remain the single best predictor of college grades. Also, skills that characterize students who are the most successful in their transition to college are not necessarily the ones that divers lack. On practical grounds, this paper

[4] Kautz et al. (2014), Almlund et al. (2011), Borghans et al. (2008), Roberts et al. (2007), Heckman, Stixrud, and Urzua (2006).

[5] Burks et al. (2015), Almlund et al. (2011), Komarraju, Karau, and Schmeck (2009), Poropat (2009), O'Connor and Paunonen (2007), De Fruyt and Mervielde (1996).

[6] We use the expression ‘non-academic’ to refer to any variable that is not an explicit measure of academic performance, such as grades or test scores. This broad category therefore includes measures often labeled as noncognitive or soft skills, but also demographic information.

[7] The consent rate was 97%.

[8] Access to administrative files will allow us to consider other outcomes such as persistence and academic performance in future work.

highlights some specific skills that educational policies might target to improve. The abilities to persist, to self-regulate and to set high expectations for oneself all contribute to reducing the risk of struggling in higher education. Our findings also motivate further research on possible policies likely to restrict the scope for the negative effects of behaviors shared by most divers, such as increasing the frequency of deadlines to mitigate procrastination. By helping characterize the profile of students exceptionally poor or great at transition to college, this research may also prove useful for catching students before they run into difficulty, and advising students about how to excel in school.

The paper proceeds as follows. First, we briefly review the existing literature on predictors of college success in Section 2. Section 3 explains the data collection process and the institutional environment and provides an overview of our estimation samples. The methodology is presented in Section 4 and results are displayed in Section 5. In Section 6, we combine the best predictors into a uni-dimensional "at-risk" factor, and document its predictive power over the full distribution of grades, as well as for more policy-relevant extreme negative outcomes. We then benchmark the predictive properties of our simple summary measure against machine learning results that let a computer algorithm choose the best predictors and find the weights on them that maximizes the predictive power. Section 7 concludes with a discussion of the policy implications of this research.

2. Background

Social scientists increasingly stress the importance of noncognitive abilities for a host of socioeconomic outcomes. Both in the labor market and in school, the explanatory power of personality traits and personal preferences is comparable to or greater than that of cognitive abilities (Almlund, Duckworth, Heckman, & Kautz, 2011). In a similar vein, successful childhood interventions that have long-term impacts on adult outcomes often show no persistent effect on cognitive skills while significantly improving children's non-academic skills (Chetty et al., 2011; Kautz, Heckman, Diris, Weel, & Borghans, 2014). Grades in high school as well as in college partly reflect both the cognitive and non-cognitive abilities of students.

The emphasis on personality traits and other non-academic measures as determinants of educational success has a long tradition in the fields of education and psychology.[9] In recent decades, the emergence of the Big Five dimensions of personality as a broadly accepted general taxonomy (John, Naumann, & Soto, 2008), along with an increasing interest in motivational theories (Robbins et al., 2004), generated a substantial amount of research on the incremental effects of personality and individual goals on college success over that of standard predictors such as standardized tests (Conard, 2006). The number of noncognitive measures that have been found to correlate significantly with college GPA is large. Yet, it remains unclear which of them or which set constitute the best predictors of success in college, since few studies consider a broad selection of predictors simultaneously and many distinct measures considerably overlap conceptually and empirically.

The lack of a thorough evaluation of how different measures used in separate literatures are related has rendered integration of independent findings difficult. For example, conscientiousness[10] and grit,[11] which have been the focus of most personality research, are both strong predictors of postsecondary education performance, but recent evidence suggests that the latter might be a facet of the former (Credé, Tynan, & Harms, 2017; Dumfart & Neubauer, 2016). In parallel, the literature on motivational theories has emphasized the importance of goals and beliefs about performance. The most comprehensive meta-analytic reviews in psychology and education research generally find that academic self-efficacy – the belief in one's capability to succeed academically – and grade goals – exhibit the strongest correlations with college GPA (Richardson, Abraham, & Bond, 2012; Robbins et al., 2004).[12] More recently, researchers in economics of education have emphasized the role of time preferences as important inputs in schooling decision and in the educational production function.[13]

These separate branches of research in education have yet to integrate findings from one another. Our paper casts a wider net by considering multiple predictors from all three fields simultaneously, notably including standard personality constructs, measures of motivational factors previously found to be good predictors of college GPA such as locus of control and grade expectations, as well as economic preference parameters. We further broaden the set of predictors by moving beyond traditional questionnaire-based measures through text analysis, and complement our examination with machine learning techniques.

3. Data

Our data comes from an online exercise completed by first year economics students in all three campuses of the University of Toronto. While more than half of the university's student population attend the main campus, over 25,000 students are registered at two smaller satellite campuses to the West and East of downtown, both about 20 miles away. These campuses receive more commuter students than the main campus and have different admission requirements. The downtown campus is perceived as more elite, whereas the satellite campuses resemble other smaller institutions across Ontario. As a result, the university's student population comes from a very diverse set of academic backgrounds.

Early in the 2015 Fall semester, all undergraduate students enrolled in an introduction to economics course (approx. 6000) across all three campuses were asked to participate in an online 'warm-up' exercise. The nature of the exercise varied randomly across students – some were asked to complete a comprehensive personality test while others were assigned a goal-setting program which asks them to write freely about their future goals. Each group was shown a short video created to introduce the purpose of the program and key take-away points. Beforehand, students were required to fill in a brief survey and were asked for consent to work with their administrative data (97% agreed). Completion of this one- to two-hour exercise counted for 2 percentage points of their overall grade in the course.[14]

The group of students who took part in the program represents about a third of all first year students enrolled at this university, and almost 10% of the entire undergraduate student population.[15] Linked administrative variables include gender, citizenship, registration status, GPA, all courses taken and grades received at this postsecondary institution and, for the majority of students, the high school performance measure used for admission to Canadian universities (the *admission grade*).[16] In the analyses below, we restrict our estimation sample to full-time students for which we have this measure of high school

[9] Willingham (1985) provides an excellent overview of the early work on the topic.

[10] Burks et al. (2015), Komarraju, Karau, and Schmeck (2009), Poropat (2009).

[11] Duckworth et al. (2007).

[12] While these meta-analyses consider many characteristics as predictors, the underlying studies rarely do, plausibly introducing bias. Our setup overcomes this methodological drawback.

[13] Lavecchia, Liu, and Oreopoulos (2016), Cadena and Keys (2015), Burks et al. (2015).

[14] The warm-up exercise was setup, in part, to test the effectiveness of new online and text-based approaches for providing student support. For more information about the experimental design, we refer readers to Oreopoulos and Petronijevic (2016).

[15] Introduction to Economics is an extremely popular course. Many students in fields other than business or economics take this course as an elective. The sample also includes students who enrolled but dropped the course later in the semester.

[16] This corresponds to the student's average of her best six grades for a standardized set of high school courses taken by all students in the province of Ontario. Admission to postsecondary education in Ontario is based solely on academic performances. There is no admission criterion, implicit or explicit, based on personal characteristics such as race, ancestry, ethnic origin, sex or age.

achievement (77% of the sample).

The set of variables that was collected as part of the survey from all students contains detailed background characteristics such as international student status and parental education, as well as a large set of other measures of noncognitive skills, in particular reports of study habits and subjective expectations. Survey questions are presented in Online Appendix A.

For a 30% random subsample of students (henceforth the *personality sample*), we collected additional data on a large array of traditional personality traits and economic preference measures as part of the online exercise. These include self-assessed propensity to procrastinate and summary measures of *perseverance of effort* and *consistency of interest*, two latent factors loading onto the construct of *grit* (Duckworth & Quinn, 2009). Two complementary measures of each Big Five trait were also constructed: an absolute measure obtained by implementing the Likert-scale Mini-IPIP questionnaire (Donnellan, Oswald, Baird, & Lucas, 2006), and a relative-scored ipsative measure. The ipsative measure indicates the extent to which a given trait is dominant in one's personality profile *relative* to other traits. This relative-scored method is known to be more resistant to biased responding (Hirsh & Peterson, 2008).[17]

We also assess students' level of tolerance for risk using both a simple survey question as well as a series of hypothetical choices between a lottery and a certain amount of money (Dohmen et al., 2011; Dohmen, Falk, Huffman, & Sunde, 2010). Finally, we elicit time preferences using lists of hypothetical choices between an amount of money paid at some early point in time and a larger amount received later (Andersen, Harrison, Lau, & Rutstrom, 2008; Dohmen et al., 2010).[18]

The first column of Table 1 shows descriptive statistics for all students included in the personality sample for whom the admission grade is non-missing.[19]

The average admission grade is 87% with the majority of students scoring above 80.[20] The summary statistics for demographic variables underline the sample's diversity. Roughly half the students have a mother tongue other than English and a citizenship other than Canadian, and a third self-report as international students.[21] Approximately 53% are women, and 81% started their first year of university in the Fall of 2015. More than 40% of our sample intends to major in a field other than economics or business (the two programs for which the introduction to economics course is required). Only 25% are first-generation college students (i.e. neither of their parents is college-educated).

There is substantial variation in average first-year college grades. The mean is 66% with a standard deviation of 13% age points, almost three times larger than the standard deviation of admission grades.[22] Of particular interest is the fact that students with the lowest college grades are not systematically the ones with the lowest high school grades. The large variance of college grades around high school grades is shown in Fig. A1 in the Online Appendix.

Table 1
Summary statistics.

Variables	Mean	Standard deviation
Age at entry	18.07	[0.959]
Mother tongue: English	0.48	[0.500]
Citizenship: Canadian	0.52	[0.500]
Women	0.52	[0.500]
First-year student in 2015	0.82	[0.388]
International student	0.34	[0.473]
Economics is a required course	0.59	[0.491]
Living in Residence	0.30	[0.459]
Mother has BA or more	0.50	[0.500]
Father has BA or more	0.59	[0.491]
First-generation student	0.25	[0.430]
Hours expected to study	18.18	[10.816]
Hours expected to work for pay	7.46	[9.807]
Expects to get more than undergraduate degree	0.63	[0.482]
Expected college GPA	3.61	[0.434]
Day started the survey (relative to first day of class)	3.84	[5.257]
Admission grade	87.38	[5.121]
Average college grade	66.33	[13.467]
Observations	1317	

Notes: Sample is restricted to students in the personality sample whose admission grade is not missing, and who finished at least one university course in their first year. First-year and international student status, gender, parental education, study habits and expectations are self-reported. We infer that economics is a required course if a student intends to major in either Economics or Business. Age, mother tongue, citizenship and grades are from administrative records. The average college grade is calculated over all courses for which a valid grade is reported in the administrative file and weighted by number of credits.

In terms of study habits, students expect to study for approximately 18 hours per week on average and work at a paid job for less than 8 hours per week. Students come in with high expectations: approximately 63% intend to eventually pursue graduate studies,[23] and the average expected GPA is 3.6, more than one grade point above the actual first-year mean GPA (2.3) – a difference greater than a full standard deviation. In addition to these subjective expectations, we also consider an objective measure of procrastination, which is the number of days between the first day of class and the time a student started the online survey for this study. Students were encouraged to complete the task early before being burdened with other homework, and given a two-week deadline. On average, four days passed between the beginning of classes and the moment students started the survey, with about half the sample registering within 2 days, but a fifth of students waiting more than a week.

In complementary analyses, we focus on a separate 50% random subsample of students (henceforth the *text sample*) who were asked to answer open-ended questions such as “describe what kind of person you want to become later on in life”. The qualitative answers to each of these questions provide sufficient information to analyze whether outliers tend to discuss different topics than other students when they are allowed to choose what to write about.[24] Students were prompted to take their time and take the exercise seriously because it was intended for their benefit. Some questions contained word count and time constraints, with a friendly message of encouragement to students that

[17] The relative-scored measure combines rank-order and forced-choice approaches. The main drawback to this approach is that relative-scored traits are negatively correlated with each other by construction.

[18] It must be noted that skipping questions was not permitted. Interested readers will find the personality test questions in the Online Appendix.

[19] Admission grades are more likely missing for transfer, non-traditional and international students.

[20] In terms of high school performance, our sample is reasonably close to the provincial average for those enrolling in university. The most recent application data from the Council of Ontario Universities (2014) indicates that the secondary school average of Full-Time, First Year students at the University of Toronto is 85.9%. The average across Ontario universities is 83.4% with some institutions with entering average grades above 86%.

[21] In practice, domestic student are those with either a Canadian citizenship or a Permanent Resident status.

[22] By construction, the distribution of admission grades we observe is truncated at the bottom. It does not reflect the full distribution of potential applicants as it only includes enrollees. This restriction of range raises methodological issues if one tries to extrapolate the relationship between past grades and college grades to non-enrolled students (Rothstein, 2004). Our objective in this paper is not to inform admission policy and the interpretation of our results is independent of restriction of range issues.

[23] In comparison, only 20% of the university's student population is enrolled in a graduate program.

[24] 50% of first-year students and 70% of upper-year students were randomly assigned to a goal-setting exercise. The proportion of first-year/upper-year students was unknown prior to assignment. Overall, about 53% of students who took part in the warm-up exercise were assigned to the goal-setting exercise. By construction, the personality sample and the text sample are mutually exclusive.

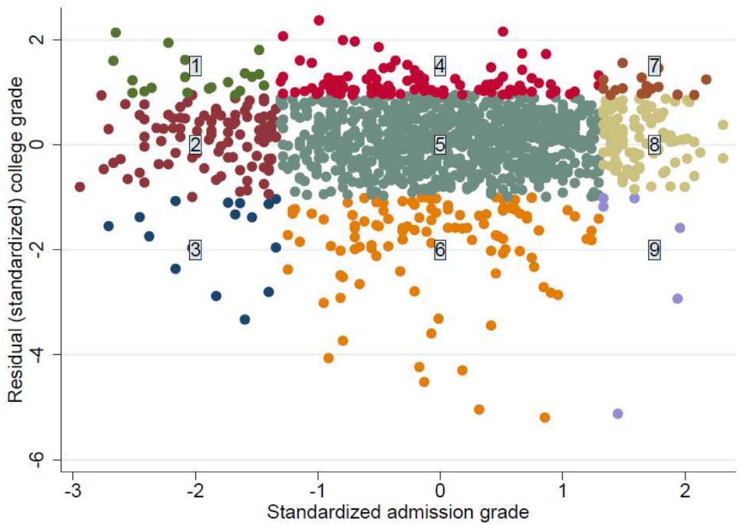


Fig. 1. Distribution of grade residuals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Notes: Colors indicate whether students are in the top or bottom decile of the distribution on each dimension. College grade residuals are obtained from specification (2), Table A1. The sample is restricted to the personality sample.

tried to complete a question before removing these constraints. The large majority of students wrote in detail, with emotion, clarity and personal insight.

4. Methodology

4.1. Defining outliers

Admission to college generally relies on standardized tests or high school grades. Yet, substantial variation in freshman performance around past grades remains. High school GPA alone is not sufficient to predict which students are the most likely to struggle and eventually drop out of college. The methodology developed below aims at exploring whether adding more variables is useful for improving predictions of these extreme outcomes. To emphasize the incremental predictive power of non-academic characteristic, we focus on the part of college grades that cannot be expected on the basis of past grades. It must be noted that high school grades partly reflect both cognitive and noncognitive skills. As a result, controlling for past academic performance absorbs part of the total contribution of our non-academic measures in explaining the raw variation in college grades. This approach aims at improving our predictions of successful and failed transitions, and helps us understand what makes divers and thrivers different than other students.[25]

To identify students who perform unusually above or below expectations, we first residualize college grades on past performance. More specifically, we extract the portion of college grades that is linearly predicted by past grades and a set of background characteristics (κ_{ics}) by estimating the following equation:

$$CollegeGrade_{ics} = \alpha_0 + \alpha_1 HSGRADE_{ics} + \alpha_2 \kappa_{ics} + \delta_c + \delta_s + \epsilon_{ics} \quad (1)$$

where $CollegeGrade_{ics}$ is the credit-weighted first-year average college grade of student i who started college in semester s and at campus c, and $HSGRADE_{ics}$ is her high school average used for admission. Campus fixed-effects (δ_c) are included to take into account differences in admission criteria across campuses, as well as any discrepancy in grading practices. Upper year students included in our sample are more likely to be enrolled in STEM programs and to take introduction to economics as an elective than are first-year students. Therefore, cohort fixed-effects (δ_s) are added to the model. We estimate the model separately for the personality sample and the text sample.

Fig. 1 plots residualized college grades against admission grades for the personality sample. In both dimensions, we highlight students who belong to either the top or the bottom decile of the distribution. The vast majority of students who perform significantly above expectations (groups 1, 4 and 7 on the figure) or below expectations (groups 3, 6 and 9) come from the middle of the admission grade distribution. Put differently, students who thrive are not simply students who were already expected to do well and did even better, and students who dive are not merely students who were expected to have relatively low grades and did even worse, nor students expected to do exceptionally well but who instead regressed towards the mean. In fact, the performance gap between the two outlier groups is colossal: divers' average first-year college grade is 40, and thrivers' is 81.

In our main specifications we define the two groups of students who rank in the top and bottom deciles of the distribution of ϵ_{ics} as thrivers and divers, respectively. We explore the robustness of our results with respect to the definition of divers and thrivers in Section 5.2.

4.2. Differences in quantitative non-academic measures

The main exercise we undertake compares the distributions of a large set of non-academic measures for the two outlier groups relative to the full sample. Unconditional mean differences for each characteristic $x \in \boldsymbol{X}$ are obtained from the following regression:

$$x_i = \gamma_1 D_i + \gamma_2 T_i + u_i \quad (2)$$

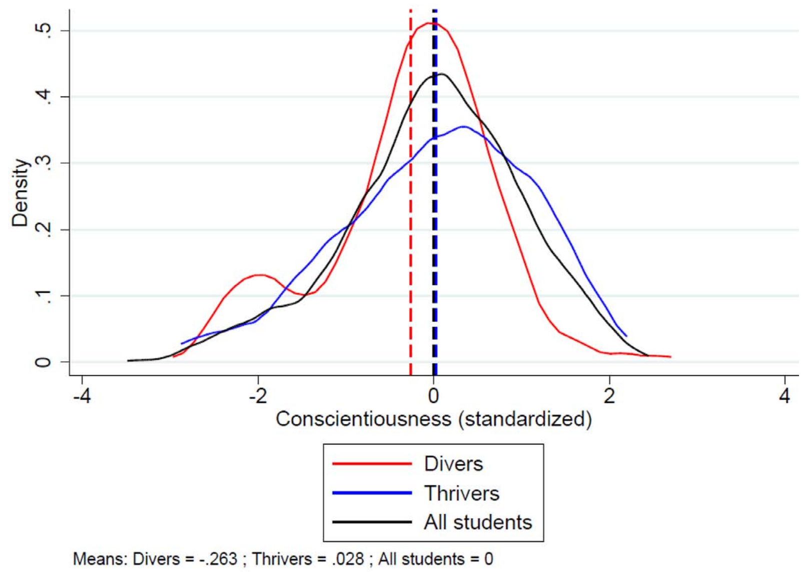
where x_i is a non-academic measure, D_i is a dummy for diver status and T_i is a dummy for thriver status. To ease the interpretation of the results, each non-binary individual characteristic of interest is standardized with mean zero and unit variance. For continuous predictors, the coefficients of interest, γ_1 and γ_2, indicate the difference in mean for each outlier group relative to the main distribution, in standard deviations units.[26] Correspondingly, binary measures are centered such that their mean is zero and the estimated coefficients reflect the percentage point difference in the fraction of thrivers or divers who exhibit the characteristic of interest relative to the main sample.

[25] See Farrington et al. (2012) for an extensive discussion of transition points in education.

[26] The coefficients are relative to the full distribution since the model does not include a constant.

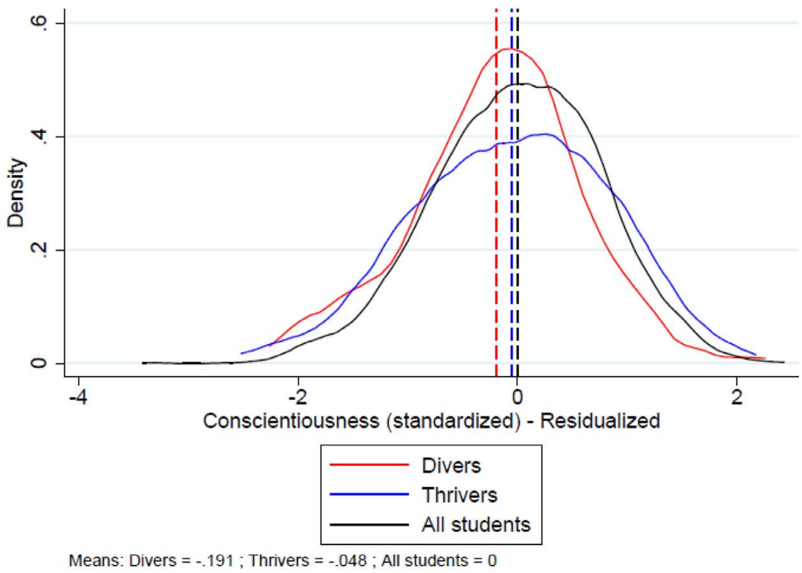
Fig. 2. Differences in distributions of conscientiousness.

Panel A: Unconditional distributions



Conscientiousness is relative-scored and unadjusted. Divers are defined as students with residual college grade below the 10th percentile. Thrivers have residual college grades above the 90th percentile. The full distribution corresponds to the personality sample.

Panel B: Conditional distributions



Conscientiousness is relative-scored and residualized from a regression on all other non-academic characteristics. Divers are defined as students with residual college grade below the 10th percentile. Thrivers have residual college grades above the 90th percentile. The full distribution corresponds to the personality sample.

As discussed in Section 2, there is substantial conceptual overlap between different non-academic constructs. To find which of these measures are the best predictors of success and failure in transitioning to college, we assess whether the mean differences remain significant when using only variation in the distribution of a given characteristic that is unexplained by other predictors. These conditional differences are calculated in two-steps. First, we residualize each characteristic x:

$$x_i = a + \boldsymbol{b}\boldsymbol{X}_{-x,i} + v_{x,i} \quad (3)$$

where $\boldsymbol{X}_{-x}$ is the subset of $\boldsymbol{X}$ that excludes characteristic x. Then, differences in means of residualized characteristics are obtained by substituting $v_{x,\ i}$ for x_i in Eq. (2). This strategy amounts to comparing the outlier distributions with the main distribution using only the fraction of the variation in a given construct that is orthogonal to other non-academic measures.[27]

Fig. 2 illustrates the nature of the comparison exercise. In panel A we show the unconditional distributions of (relative-scored)

[27] This is similar but not numerically equivalent to including all other characteristics as controls in Eq. (2). Results are not sensitive to this choice of methodology. See Online Appendix Table A5.

conscientiousness for thrivers, divers and the full personality sample. Divers are considerably less conscientious than average (0.26 standard deviations below the sample mean). This pattern is not symmetric – on average, thrivers are just as conscientious as others. Conditional differences are presented in panel B, where each density plot shows the distribution of residual conscientiousness that is unaccounted for by variation in other non-academic measures. The mass of divers with very low conscientiousness (around −2 s.d.) observed in the unconditional distribution is explained by other predictors, but the conditional mean difference between divers and the full sample remain substantial. Figs. A2 and A3 show similar density plots for other non-academic measures.

4.3. Text analysis

Students in the text sample were asked a series of open-ended questions, such as "Name at least one thing that you admire about yourself". This type of question allows students much more freedom to answer, so individual answers are often very informative. For instance, answers are not restricted to a set of goals pre-selected by the researcher, but rather include any goals students may have. However, aggregating the results over all students in a meaningful way is a challenge. We use two techniques to quantify the writing, one evaluating effort and writing quality, the other analyzing which topics students choose to write about.

There are three measures of effort and writing quality. Firstly, the programming of the survey website allows us to measure how many seconds each student takes to answer a given question. Secondly, we count the number of words each student uses for each answer, where a word is defined as one or more characters separated by one or more spaces. Finally, we run each of these words through the Microsoft Word Canadian English spellchecker, and calculate the proportion of words which are spelled correctly.[28] These variables are taken as measures of conscientiousness and language ability and analyzed using the method described in Section 4.2.

We also compare the topics that divers and thrivers discuss in their answers using a simplified topic modeling text analysis approach. In topic modeling, it is assumed that an author makes a series of decisions about which topics to discuss. Each topic then maps to a series of words (Blei, Ng, & Jordan, 2003; Hofmann, 2000). For example, discussion of procrastination might use words like ``procrastinate'', ``cram'', or ``all-nighter''. The researcher measures the amount of space devoted to a topic by comparing the frequency of words across documents. If one author, or group of authors, use a particular set of words more often, it is assumed that they devote a higher proportion of their documents to topics related to those words.

Given the sample size and the fact that students often give brief responses, we adopt a very simple method to apply this approach. Firstly, we clean the students' answers to generate more meaningful results with the following rules: If a word was spelled incorrectly according to the Word spellchecker, we replace it with Word's top suggestion for a replacement. These words are then stemmed, to remove grammatical constructions such as pluralisation and verb tenses. This ensures that words such as "class" and "classes" are treated as identical. Finally, we remove stopwords, which are short, common words such as "and" or "the".

For each word in the cleaned text, we calculate the proportion of students who use the word to answer a given question among divers, thrivers, and in the entire sample. A chi-squared test comparing the share of divers who use a word with the share of the entire sample shows if low performing students are more likely than others to use a given word. If many of the words used more often by divers are related to a given topic, the intuition of the topic modeling approach suggests that divers are more likely to spend more space discussing that topic.

5. Results

5.1. Predicting college grades using past academic achievement

Estimates of the relationship between past academic performance and college grades (Eq. (1)) are shown in Table A1 of the Online Appendix.

A one standard deviation higher admissions grade is associated with a 0.41–0.43 standard deviation higher first-year average college grade. Older students and non-domestic students receive lower grades in college than do younger and domestic students with equivalent admission grades. While past measures of academic performance do predict success in college, the explanatory power of this model is modest. When no demographics are included, less than 20% of the observed variation in college grades is explained by admission grades, in line with previous findings (Bettinger, Evans, & Pope, 2013; Richardson et al., 2012; Stephan, Davis, Lindsay, & Miller, 2015). The inclusion of age at entry and non-domestic student status adds some explanatory power, but more than three quarters of the variation in college grades remain unexplained.[29] We next explore which non-academic characteristics best characterize outliers relative to the main distribution.

5.2. Predicting student outliers with non-academic outcomes

Columns (1) and (3) of Table 2 report average unconditional deviations from the sample mean for divers and thrivers, respectively. For each possible predictor, columns (2) and (4) report deviations from the mean conditional on all other predictors listed in the table. In the last two columns, we test whether the difference between the top and bottom outliers for each non-academic measure is significantly different from zero.

Relative to the full distribution, students who perform largely below expectations are much more likely to self report they cram for exam (0.30 s.d. above the mean), much more likely to start the online survey later (0.29 s.d. above the mean) and tend to work much more hours at paid jobs (0.22 s.d. above the mean). They are also significantly less conscientious (0.26 s.d. below the mean) and more impatient than their peers (0.2 s.d. above the mean), consistent with prior evidence (Burks et al., 2015). Even conditional on other predictors, most of these patterns remain strong and statistically significant. Being sure about one's major and intending to pursue graduate studies has little explanatory power, and, if anything, divers are more likely to say they often think about the future. We interpret these results as evidence that students who perform significantly below expectations are neither lacking ambition nor vision, but tend to put themselves in situations that hinder their academic success.

Thrivers are not the mirror image of divers; they are no less likely to cram for exams or to work many hours for pay than the average student. However, they tend to study for relatively more hours (0.22 s.d. above the mean), and expect a higher GPA than divers (difference of 0.23 s.d.). We find that thrivers are more introverted than divers (unconditional difference of −0.27 s.d.), but that the conditional difference is not statistically significant.[30] Relative to the full distribution,

[28] Note that this is a noisy measure of spelling quality. If a student's misspelling of a word is a correct spelling of another word – for example, "coarse" for "course" – it will count as a correct spelling. On the other hand, some widely acceptable abbreviations, such as GPA, are not recognized by the spellchecker and counted as incorrect spellings.

[29] Adding polynomials of admission grades does not affect subsequent results. Similarly, including high school fixed effects (unreported) does not qualitatively affect most of our conclusions, but comes at the high cost of precision since high school identifiers are only observed for domestic students, and the sample contains multiple high schools from which only one student is observed. Also, our results are robust to further including high school grades in mathematics and English as additional covariates.

[30] While less common in the literature, this result is not entirely new (O'Connor & Paunonen, 2007; Noftle & Robins, 2007). Chamorro-Premuzic and Furnham (2005) discuss how introverts may have a greater ability to consolidate learning and have better study habits (e.g. spend more time studying than socializing).

Table 2
Differences between outliers and full distribution - Personality sample.

	Bottom decile		Top decile		Difference between outliers	
	Unconditional Mean diff. [s.e.] (1)	Conditional Mean diff. [s.e.] (2)	Unconditional Mean diff. [s.e.] (3)	Conditional Mean diff. [s.e.] (4)	Unconditional (3) − (1) [*p*-value test (3) = (1)] (5)	Conditional (4) − (2) [*p*-value test (4) = (2)] (6)
Study hours per week (z-score)	−0.079	−0.019	.224**	.226***	.303**	.245**
	[0.087]	[0.083]	[0.087]	[0.083]	[0.014]	[0.037]
Sure about program of study (z-score)	−0.074	−0.085	−0.067	−0.08	0.007	0.005
	[0.087]	[0.082]	[0.087]	[0.082]	[0.954]	[0.966]
Think about future goals (z-score)	0.134	.150**	−0.119	−0.095	−.253**	−.244**
	[0.087]	[0.075]	[0.087]	[0.075]	[0.040]	[0.022]
Identify with university (z-score)	0.067	0.04	−0.011	0.04	−0.078	0.001
	[0.087]	[0.080]	[0.087]	[0.080]	[0.529]	[0.995]
Transition has been challenging (z-score)	0.061	−0.024	−0.059	−0.091	−0.12	−0.067
	[0.087]	[0.078]	[0.087]	[0.079]	[0.330]	[0.548]
Cram for exams (z-score)	.297***	.209***	−0.043	−0.058	−.34***	−.268**
	[0.087]	[0.076]	[0.087]	[0.076]	[0.006]	[0.013]
Work hours per week (z-score)	.216**	.140*	0.049	0.076	−0.168	−0.064
	[0.087]	[0.083]	[0.087]	[0.084]	[0.173]	[0.588]
Expected GPA (z-score)	−0.097	−0.106	0.137	0.126	.233*	.232**
	[0.087]	[0.080]	[0.087]	[0.080]	[0.058]	[0.040]
Day started exercise (z-score)	.288***	.199**	−0.041	−0.03	−.329***	−.229*
	[0.087]	[0.083]	[0.087]	[0.083]	[0.007]	[0.051]
Expects more than undergraduate	0.003	−0.009	0.016	0.033	0.012	0.043
	[0.042]	[0.040]	[0.042]	[0.040]	[0.834]	[0.448]
Agreeableness (z-score)	−0.023	0.045	−0.03	−0.056	−0.007	−0.101
	[0.087]	[0.079]	[0.087]	[0.080]	[0.956]	[0.368]
Conscientiousness (z-score)	−.263***	−.191***	0.028	−0.048	.292**	0.143
	[0.087]	[0.066]	[0.087]	[0.067]	[0.018]	[0.128]
Extraversion (z-score)	.170*	0.1	−0.102	−0.038	−.272**	−0.138
	[0.087]	[0.078]	[0.087]	[0.078]	[0.027]	[0.212]
Openness (z-score)	0.087	0.035	0.094	0.011	0.007	−0.024
	[0.087]	[0.077]	[0.087]	[0.077]	[0.953]	[0.829]
Emotional stability (z-score)	0.05	−0.014	0.024	−0.102	−0.026	−0.088
	[0.087]	[0.076]	[0.087]	[0.077]	[0.834]	[0.414]
Risk tolerance (z-score)	0.098	0.042	−.231***	−0.109	−.328***	−0.151
	[0.087]	[0.078]	[0.087]	[0.078]	[0.008]	[0.171]
Impatience (z-score)	.199**	.180**	−0.134	−0.112	−.333***	−.292**
	[0.087]	[0.085]	[0.087]	[0.085]	[0.007]	[0.016]
Procrastination (z-score)	0.102	0.067	0.043	−0.039	−0.059	−0.106
	[0.087]	[0.078]	[0.087]	[0.078]	[0.633]	[0.335]
Locus of Control (z-score)	0.112	0.094	−0.034	0.02	−0.146	−0.074
	[0.087]	[0.081]	[0.087]	[0.081]	[0.238]	[0.519]
Perseverance of effort (z-score)	−0.135	−0.089	−0.04	0.038	0.095	0.127
	[0.087]	[0.081]	[0.087]	[0.081]	[0.439]	[0.265]
Consistency of interest (z-score)	0.053	0.051	−0.132	−0.086	−0.185	−0.137
	[0.087]	[0.078]	[0.087]	[0.078]	[0.133]	[0.214]
Women	−.105**	−.082**	−.094**	−.090**	0.011	−0.008
	[0.043]	[0.041]	[0.044]	[0.041]	[0.860]	[0.888]
English mother tongue	0.013	0.017	0.009	0.041	−0.004	0.024
	[0.044]	[0.031]	[0.044]	[0.031]	[0.950]	[0.594]
Canadian citizenship	−0.017	−0.029	−0.036	−0.034	−0.019	−0.005
	[0.044]	[0.026]	[0.044]	[0.026]	[0.757]	[0.899]
International student	−0.005	−0.014	−0.01	−0.015	−0.005	−0.001
	[0.041]	[0.028]	[0.041]	[0.028]	[0.931]	[0.976]
Economics is required	0.021	0.011	−0.059	−0.022	−0.079	−0.033
	[0.043]	[0.041]	[0.043]	[0.041]	[0.191]	[0.570]
Mother has at least bachelor degree	−0.022	−0.049	−0.003	0.004	0.019	0.053
	[0.044]	[0.034]	[0.044]	[0.034]	[0.759]	[0.270]
Father has at least bachelor degree	0.043	0.033	0.018	0.034	−0.026	0.001
	[0.043]	[0.029]	[0.043]	[0.029]	[0.672]	[0.983]
First-generation student	−0.026	−0.011	0.03	0.041	0.055	0.052
	[0.037]	[0.026]	[0.038]	[0.026]	[0.299]	[0.151]

Notes: Diver and thrivers status is defined using residuals from the specification reported in column (2) of Table A1. All non-z-score predictors are binary. In columns (1) through (4), coefficients represent the difference in means between outlier groups and the full sample. For conditional differences (columns (2) and (4)), each characteristic is first regressed on the set of other characteristics reported in this table. Big Five traits are relative-scored. Likert-scale Big Five traits are used as controls in lieu of relative-scored traits in the residualization process for columns conditional differences. *** $p < .01$, ** $p < .05$, * $p < .1$

thrivers are more risk averse, but this difference is mostly accounted for by variation in other characteristics. Students who excel above expectations do not report finding the transition to university any less challenging than the average student does, and intend to pursue graduate studies in the same proportions as average students and divers do.

We find no statistically significant differences between outliers in terms of agreeableness, openness to experience or emotional stability. Similarly, grit (perseverance of effort and consistency of effort) and locus of control do not help predict extreme outcomes. The point estimates for our subjective measure of procrastination indicate that

Table 3
Differences between outliers and full distribution - Text sample.

	Bottom decile Unconditional Mean diff. [s.e.] (1)	Conditional Mean diff. [s.e.] (2)	Top decile Unconditional Mean diff. [s.e.] (3)	Conditional Mean diff. [s.e.] (4)	Difference between outliers Unconditional (3) - (1) [p-value test (3) = (1)] (5)	Conditional (4) - (2) [p-value test (4) = (2)] (6)
Total number of words used (z-score)	−.263***	−.134**	0.046	.130**	.309***	.264***
	[0.065]	[0.061]	[0.065]	[0.061]	[0.001]	[0.002]
Proportion spelled correctly (z-score)	−.135**	−0.058	0.046	0.077	.170*	0.135
	[0.066]	[0.064]	[0.065]	[0.064]	[0.066]	[0.136]
Time taken on written questions (z-score)	−0.043	−0.066	0.098	0.066	0.141	0.132
	[0.065]	[0.062]	[0.065]	[0.062]	[0.126]	[0.135]

Notes: In columns (1) through (4), coefficients represent the difference in means between outlier groups and the full sample. For conditional differences (columns (2) and (4)), each characteristic is first regressed on the set of controls (variables from survey and administrative data. *** $p < .01$, ** $p < .05$, * $p < .1$

thrivers are less likely to procrastinate than divers, but we cannot reject the null hypothesis of no difference.

Men are overrepresented in both tails of the distribution of college grade residuals: the proportion of women is approximately 10 percentage point lower among divers and thrivers than in the full sample. Previous research has also found that boys exhibit higher variance in test scores than girls (Hedges & Nowell, 1995; Machin & Pekkarinen, 2008). Other demographic characteristics have little or no predictive power.

To evaluate the robustness of these findings, we consider two alternative definitions of outliers in the Online Appendix. In Table A2, divers (thrivers) are defined as students who fall in the bottom (top) 20% of the distribution of college grade residuals. Broadening the groups' composition improves precision, but may dilute results by including students with less extreme outcomes in the outlier groups. We continue to find that divers are less conscientious and more impatient, more likely to cram for exam and to start the exercise later, and that thrivers study more hours on average. Under this specification, the difference between divers and the full sample in terms of procrastination does reach statistical significance at conventional levels, but differences in gender composition do not. In Table A3, we verify that our results are not driven by students who came in with extraordinarily high or low admission grades by restricting the sample to those with past grades in the middle 80% of the distribution.[31]

Overall, students who perform markedly lower than expected given their past academic achievement are more prone to procrastination, more impatient and less conscientious than the average. In contrast, students who perform significantly better than expected exhibit very few differences with the full sample, with the exception of the number of hours spent studying. This suggests that most measured characteristics conductive to success in college are already reflected through high school grades, but that non-academic measures do help predict negative outcomes that were unexpected on the basis of past performance.

We note that it is unlikely that these results reflect purely transitory phenomena. For instance, while personality traits are not fixed, they particularly stable over time (Almlund et al., 2011; Cobb-Clark & Schurer, 2012; John et al., 2008). Also, the students we identified as divers in first-year do not appear to catch-up with other students in later years – their second-year grades are still more than a full standard deviation below the mean (Fig. A4). In addition, divers were more than four times more likely to have dropped out after the first-year than the average student.[32]

For completeness, Table A4 shows differences between students in the top and bottom distribution of college grades that are not adjusted for high school grades, so that students in the right tail also include students who do very well and were expected to do so.[33] While results for students in the bottom 10% are essentially the same as for divers, the picture for top students is noticeably different than for thrivers: relative to the main distribution, students with the best college grades are more conscientious and less extroverted, less likely to cram for exams, expect higher GPAs and are significantly less tolerant of risk. Our interpretation is that these characteristics contribute to success both in college and in high school, but cannot explain why some students thrive beyond expectations. However, students who obtain low college grades unconditionally or relative to expectations share the same harmful traits of impatience and lack of conscientiousness.

5.3. *Text analysis of student outliers*

The analysis of open ended questions yields results that are consistent with the main results. Table 3 replicates the methodology in Table 2, and shows that divers use fewer words when answering questions and thrivers use more, which suggests that thrivers are providing more detailed and careful answers. Thrivers spend more time answering these questions, although this difference is not significant. Finally, in the unconditional comparisons, thrivers have stronger spelling than the average and divers have weaker spelling. The significance of this result dissipates in the conditional comparisons, which suggests that most of this difference is a function of other non-academic variables. Overall, thrivers appear to put in more effort when answering these open-ended questions, consistent with the finding that they expect to study for more hours than others.

Topic analysis results are shown in Table 4. For a selected set of questions, the table shows a word if the difference between the share of thrivers (or divers) who use it and the share of the whole sample is significant at 5%, and if at least 5 thrivers (or divers) use it. These results reinforce the point that conscientiousness is a crucial trait. When asked to identify traits they admired about themselves, thrivers were more likely to use words such as “discipline”, “practice”, or “responsibility”, which are indicative of conscientiousness. Sample phrases using these words include “I admire the fact that I have discipline”, and “One of the qualities I admire most about myself is responsibility”.[34]

Thrivers and divers can also be differentiated when they are asked to list their goals or hopes for the future. Divers are significantly likely to use words which highlight wealth. Examples include “rich”, in contexts such as “be a rich man” and “business” in contexts such as “being successful, having so many successful businesses.” Thrivers, on the other hand, are more likely to highlight how they plan to contribute to

[31] We restrict the sample to students in groups 4, 5 and 6 on Fig. 1.

[32] The dropout rate for the full personality sample is 8%. For divers, it is 36%.

[33] We here use the distribution of grades adjusted only for cohort and campus fixed effect, as well as age at entry and non-domestic status.

[34] A list of sample phrases for each of the words in Table 4 is shown in the Online Appendix.

Table 4
Words used more frequently by outliers.

Question	Top decile words	Bottom decile words
Name two goals	Build	Rich business own actuary
Qualities admire in self	Discipline specific word practice responsibility smart confident game	Cause communicate receive friendly trust
Your future self	Human meet people deal god trustworthy computer whole provide famous love mature determine helpful wise	Tough man book father moment rich
Qualities admire in others	Weakness avoid challenge overcome read mistake creativity word people Steve area general initiative understand	Power concentration waste

Notes: The words listed are used with higher frequency by thrivers (or divers) in response to the questions listed in the first column. Words are included if the frequencies are different at the 5% significance level, according to a Pearson's Chi-square test.

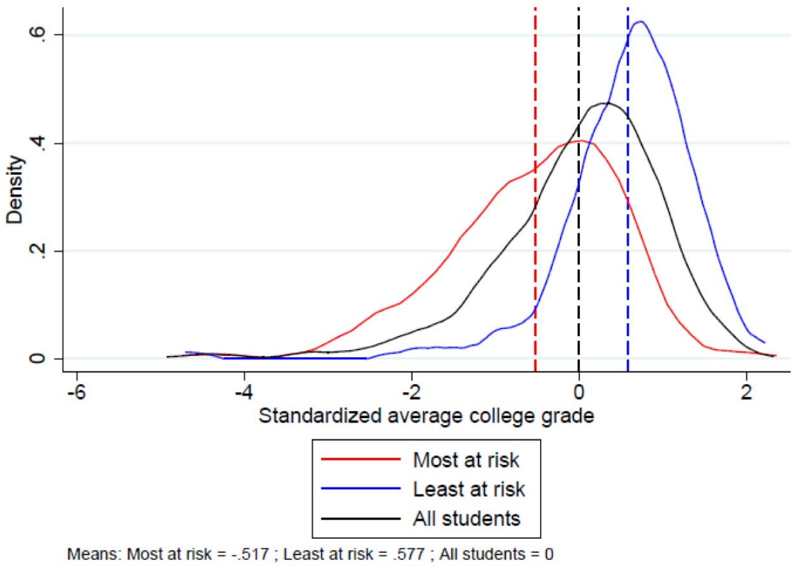


Fig. 3. College grades by at-risk status.

Notes: College grades are unadjusted. Most at-risk students are defined as students below the 10th percentile in the distribution of the seven-variable unweighted average of key characteristics. Least at-risk students rank above the 90th percentile. The full distribution corresponds to the personality sample.

society, using words such as “human” and “people”. Previous work has emphasized the importance for educational success of pursuing long-term goals. Our text analysis stresses the importance of the nature and content of these goals.

6. Summary measures of non-academic characteristics

We combine our key non-academic predictors of college success and failure into an overall predictor, to examine how well it performs in forecasting outcomes over the full distribution as well as in the tails. Our most robust predictors of outliers are: propensity to cram for exams, number of hours studying, number of hours of paid work, expected GPA, time started the exercise, conscientiousness and impatience. We remain agnostic about the exact relative importance of each of these seven constructs and take the unweighted average of these standardized variables for each student. We later explore whether we can improve our predictions by using weights obtained from machine learning techniques.

Fig. 3 shows the distributions of unadjusted college grades for students in the top and bottom 10% of the distribution of our relatively simple summary measure of non-academic characteristics. Students deemed the most at-risk under this metric have first-year grades on average more than a full standard deviation below students considered least at risk of struggling during the transition to college.

The one-dimensional measure performs relatively well in terms of predicting freshman performance, but its incremental explanatory power after accounting for past performance is modest, as shown in Table 5. The table displays estimates of a modified version of Eq. (1) in which one-dimensional at-risk measures are substituted for admission grades in panel A, and added as regressors in panel B. Our preferred metric, the simple unweighted average of 7 large predictors of extreme unexpected performance, correlates strongly with college grades (column (3)). In terms of adjusted R-squared, the explanatory power of this measure alone (0.163) is not as high as that of high school grades (0.218), but adding the at-risk factor to past grades increases the model's fit by almost 4 percentage points.[35]

We then benchmark the predictive abilities of our summary at-risk factor against measures computed with more sophisticated but less transparent approaches to constructing indices. In column (4), the seven best predictors are summarized by their principal component. The model's fit is actually lower with this method than under our preferred approach, suggesting that only using the variance common to all 7 variables is too restrictive.

In columns (5)–(9) we use least angle regressions (LARS) and let the algorithm pick the best predictors and put optimal weights on these (Efron, Hastie, Johnstone, & Tibshirani, 2004). A summary measure of characteristics can then be defined as the fitted values associated with the LARS estimates. The dependent variable used in the process is average college grade adjusted for our conditioning variables, but not for high school grades. Since we chose our 7 best predictors by examining outliers, we first

[35] The adjusted R-square is 0.096 when only conditioning variables (cohort and campus fixed effects, age, non-domestic status) are included.

Table 5
Predictive properties of summary measures of non-academic characteristics.

	Dependent variable: Standardized first-year college grades								
Summary measure	–	–	Unweighted average	Principal component	LARS on outliers	LARS on outliers	LARS on outliers	LARS on outliers	LARS on full sample
Predictors included	–	–	7 best	7 best	7 best	7 best + polynomials	7 best + polynomials + interactions	All	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Separate predictive power									
Admission grade	.433***	.446***							
	[0.030]	[0.030]							
Admission grade^2		.046**							
		[0.019]							
Summary measure			.653***	.249***	.560***	.410***	.701***	.499***	1.355***
			[0.061]	[0.027]	[0.054]	[0.045]	[0.076]	[0.045]	[0.110]
Observations	1317	1317	1317	1317	1317	1317	1317	1317	1317
Adjusted R^2	0.218	0.221	0.163	0.145	0.160	0.145	0.146	0.169	0.184
Panel B: Incremental predictive power									
Admission grade	.433***	.446***	.387***	.401***	.392***	.408***	.405***	.390***	.370***
	[0.030]	[0.030]	[0.030]	[0.030]	[0.030]	[0.030]	[0.030]	[0.030]	[0.030]
Admission grade^2		.046**	.037**	.038**	.038**	.042**	.041**	.037**	.035*
		[0.019]	[0.018]	[0.019]	[0.018]	[0.019]	[0.019]	[0.018]	[0.018]
Summary measure			.466***	.171***	.406***	.308***	.515***	.381***	.992***
			[0.059]	[0.026]	[0.052]	[0.043]	[0.072]	[0.043]	[0.109]
Observations	1317	1317	1317	1317	1317	1317	1317	1317	1317
Adjusted R^2	0.218	0.221	0.255	0.245	0.255	0.250	0.249	0.265	0.267

Notes: All regressions include campus and cohort fixed-effects, as well as non-domestic status and age at entry. Standard errors are in brackets. In column (6), quadratic and cubic terms for each of the 7 best predictors are used in the LARS algorithm. In column (7) all pairwise interactions between the 7 best predictors are further added. In columns (8) and (9), the set of potential predictors used in the algorithm is all variables encompass all variables listed in Table 2. *** $p < .01$, ** $p < .05$, * $p < .1$

Table 6
Proportion of 'at-risk' students in bottom tail of distribution of college grades.

	Distribution of unadjusted first-year college grades				
	Bottom 10%	Bottom 20%	Bottom 30%	Bottom 40%	Bottom 50%
Most at-risk:					
(a) Bottom decile of summary measure	23%	40%	52%	61%	71%
(b) Bottom decile of admission grades	25%	43%	60%	76%	87%
(c) Bottom decile in both metrics	38%	55%	76%	86%	100%
Least at-risk:					
(d) Top decile of summary measure	3%	6%	8%	14%	24%
(e) Top decile of admission grades	2%	2%	5%	9%	15%
(f) Top decile in both metrics	0%	0%	0%	0%	7%

Notes: Each cell indicates the fraction of our personality sample who fall into the categories in rows and columns. By construction 10% of our sample is defined as most/least at-risk in rows (a), (b), (d) and (e). Only 2.2% of students in our sample satisfy the 'at-risk' criterion in row (c), and 2.3% satisfy the criterion in (f).

run the LARS algorithm on the subsample of divers and thrivers, as they are defined in Section 4.1. Comparing the adjusted R-squared in columns (5), (6) and (7) with column (3) indicates that the weights put on the seven selected predictors that maximize the share of the variance in grades *among* outliers do not necessarily generalize to the full distribution since the unweighted average has better predictive power over the full personality sample. This observation underscores the importance of non-linearities in the education production function. Column (8) demonstrates that the fit of the model is minimally improved by letting the algorithm pick more predictors than the ones we selected.[36] The summary measure used in column (9) is obtained using LARS on the full distribution of students in the personality sample and therefore puts an upper bound on the joint predictive power of all the non-academic characteristics over the full distribution. We find that using information from the full sample (column (9)) rather than from outliers only (column (8)) to calibrate the weights increases the R-squared by 0.015 (from 0.169 to 0.184) if admission grades are omitted, but only by 0.002 if past grades are account for. Our simple summary measure raises the adjusted R-squared almost as high as these upper-bound measures do, but without compromising on transparency.

Results on outliers highlight important asymmetries in the distribution of non-academic characteristics across the grade distribution, suggesting that the at-risk factor may have more predictive power for extreme outcomes than over the entire distribution of grades. Table 6 shows the proportion of students considered most or least at-risk under different criteria who fall in the bottom of the distribution of raw first-year college grades. About a quarter of all students below the 10th percentile of admission grades end up below the 10th percentile of college grades (row (a)). The proportion of students deemed 'at-risk' by our simple measure (row (b)) that ends up with such dramatic outcomes is very similar (23%).[37] Yet, there is little overlap in the tails of the distributions of admission grades and of our at-risk measure – only

[36] Reassuringly, the algorithm tends to select the same predictors we picked.

[37] This result suggests that this measure can be a useful substitute for admission grades when these are missing. For instance, the relationship between the at-risk measure and grades observed in the personality sample also holds in the smaller sample of students for which non-academic measures are observed but past grades are missing (non-reported).

2.2% of students in our sample fall below the 10th percentile in both distributions (row (c)). Among students in this situation, 38% will fall in the bottom decile of college grades, and all of them will end up below the median. Importantly, falling below the 10th percentile of college grades may have serious consequences: In our sample, no student in the bottom decile of college grades has an average grade above 50 and all are therefore put on probation at the end of their first year in college, which substantially reduce the probability of graduating (Lindo, Sanders, & Oreopoulos, 2010). When used jointly with high school grades, the at-risk factor can substantially improve the prediction of extreme outcomes, with potentially important benefits for school administrators and students alike.

7. Conclusion

A vast array of personality traits and other noncognitive constructs are used in education research in order to predict performance in college, with substantial overlap across distinct measures. Samples are often based only on a select group of volunteers. In this paper, we were able to gather a more comprehensive set of non-academic measures for virtually all students taking a large first-year college course by assigning a small grade requirement to the survey. We investigated which variables, unconditional and conditional on other predictors, best explain the variation in college grades that could not have been expected on the basis of variables known upon admission, notably past academic performance. Our results suggest that a few non-academic measures have reasonable predictive power and that linear assumptions often implicit in prior research mask interesting asymmetries.

Students whose first-year college average is far below expectations (divers) have a high propensity for procrastination – they self-report cramming for exams and wait longer before starting a short exercise worth 2% of their overall grade in a first-year economics course. They are also considerably less conscientious than their peers. Divers are generally more impatient for positive experiences. For instance, qualitative analyses of short texts written by students suggest divers are more likely to express superficial goals, hoping to 'get rich' quickly. In contrast, students who exceed expectations (thrivers) express more philanthropic goals, are purpose-driven, and are willing to study more hours per week to obtain the higher GPA they expect. The only background characteristic that help predict outlier status is gender, with men being more likely to both thrive and dive.

Divers are considerably more likely to drop out after first year than other students, and even those who remain in school continue lagging behind. Our results, which indicate that divers are more impatient and tend to wait to the last minute before getting to work, suggest some possible ways to help early on. For example, interventions emphasizing efforts around staying organized and structured to avoid wasting time may be fruitful. In a follow-up paper (Beattie, Laliberté, Michaud-Leclerc, & Oreopoulos, 2017), we notably find that students at the bottom of the grade distribution severely lack time management skills, are aware of these issues, but are lost when it comes to finding solutions. Proactive guidance on how to improve these skills, for instance how to design a study schedule and how to respect it, are promising avenues.

Consistent with the extensive literature on the correlates of college GPA, we found that high school grades remain the best predictor of college grades in general. However, non-academic constructs are especially useful for predicting extreme outcomes that cannot be explained by prior educational achievement. Importantly, the characteristics that best predict successful transitions to college are not necessarily the ones that struggling students lack. Our results, descriptive in nature, warrant further research on the importance of non-linearities, notably at the bottom of grade distributions, for the design and targeting of successful interventions in higher-education.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.econedurev.2017.09.008.

References

Chapter 1Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Vol. Eds.), *Handbook of the economics of education: Vol. 4*, (pp. 1–181). Elsevier.

Andersen, S., Harrison, G. W., Lau, M. I., & Rutstrom, E. E. (2008). Eliciting risk and time preferences. *Econometrica, 76*(3), 583–618.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses.* University of Chicago Press.

Beattie, G., Laliberté, J.-W. P., Michaud-Leclerc, C., & Oreopoulos, P. (2017). *What sets college thrivers and divers apart? A contrast in study habits, attitudes, and mental health.* National Bureau of Economic Research *Working Paper No. 23588*.

Bettinger, E. P., Evans, B. J., & Pope, D. G. (2013). Improving college performance and retention the easy way: Unpacking the ACT exam. *American Economic Journal: Economic Policy, 5*(2), 26–52.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(January), 993–1022.

Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources, 43*(4), 972–1059.

Bound, J., & Turner, S. (2011). Chapter 8 - Dropouts and diplomas: The divergence in collegiate outcomes. In E. A. Hanushek, S. Machin, & L. Woessmann (Vol. Eds.), *Handbook of the economics of education: Vol. 4*, (pp. 573–613). Elsevier.

Bound, J., Lovenheim, M. F., & Turner, S. (2010). Why have college completion rates declined? An analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics, 2*(3), 129–157.

Burks, S. V., Lewis, C., Kivi, P. A., Wiener, A., Anderson, J. E., Gotte, L., & Rustichini, A. (2015). Cognitive skills, personality, and economic preferences in collegiate success. *Journal of Economic Behavior & Organization, 115*, 30–44.

Cadena, B. C., & Keys, B. J. (2015). Human capital and the lifetime costs of impatience. *American Economic Journal: Economic Policy, 7*(3), 126–153.

Chamorro-Premuzic, T., & Furnham, A. (2005). *Personality and intellectual competence.* Lawrence Erlbaum Associates Publishers.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from project star. *The Quarterly Journal of Economics, 126*(4), 1593–1660.

Childs, S. E., Finnie, R., & Martinello, F. (2016). Postsecondary student persistence and pathways: Evidence from the YITS-A in Canada. *Research in Higher Education, 58*(3), 270–294.

Cobb-Clark, D. A., & Schurer, S. (2012). The stability of big-five personality traits. *Economics Letters, 115*(1), 11–15.

Conard, M. A. (2006). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality, 40*(3), 339–346.

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology, 113*(3), 492–511.

Cyrenne, P., & Chan, A. (2012). High school grades and university performance: A case study. *Economics of Education Review, 31*(5), 524–542.

De Fruyt, F., & Mervielde, I. (1996). Personality and interests as predictors of educational streaming and achievement. *European Journal of Personality, 10*(5), 405–425.

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *The American Economic Review, 100*(3), 1238–1260.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association, 9*(3), 522–550.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18*(2), 192.

Dooley, M. D., Payne, A. A., & Robb, A. L. (2012). Persistence and academic success in university. *Canadian Public Policy, 38*(3), 315–339.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of Personality Assessment, 91*(2), 166–174.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087.

Dumfart, B., & Neubauer, A. C. (2016). Conscientiousness is the most powerful non-cognitive predictor of school achievement in adolescents. *Journal of Individual Differences, 37*(1), 8–15.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of statistics, 32*(2), 407–499.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., et al. (2012). *Teaching adolescents to become learners: The role of noncognitive factors in shaping school Performance–a critical literature review.* University of Chicago Consortium on Chicago School Research.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics, 24*(3), 411–482.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*(5220), 41.

Hirsh, J. B., & Peterson, J. B. (2008). Predicting creativity and academic success with a

"fake-proof" measure of the Big Five. *Journal of Research in Personality, 42*(5), 1323–1333.
Hofmann, T. (2000). Learning the similarity of documents: An information-geometric approach to document retrieval and categorization. *Advances in neural information processing systems* (pp. 914–920). .
John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research, Vol. 3*, 114–158.
Kautz, T., Heckman, J. J., Diris, R., Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success.* OECD Publishing OECD Education Working Papers.
Kirby, K. N., Winston, G. C., & Santiesteban, M. (2005). Impatience and grades: Delay-discount rates correlate negatively with college GPA. *Learning and Individual Differences, 15*(3), 213–222.
Komarraju, M., Karau, S. J., & Schmeck, R. R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences, 19*(1), 47–52.
Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2016). Chapter 1 - Behavioral economics of education: Progress and possibilities. Chapter 1 In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.). Elsevier.
Lindo, J. M., Sanders, N. J., & Oreopoulos, P. (2010). Ability, gender, and performance standards: Evidence from academic probation. *American Economic Journal: Applied Economics, 2*(2), 95–117.
Machin, S., & Pekkarinen, T. (2008). Global sex differences in test score variability. *Science, 322*(5906), 1331–1332.
Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science, 244*(4907), 933–938.
Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology, 93*(1), 116.
O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual differences, 43*(5), 971–990.
Oreopoulos, P., & Petronijevic, U. (2016). *Student coaching: How far can technology go? National Bureau of Economic Research Working Paper No. 22630*.
Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*(2), 322–338.
Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387.
Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2), 261–288.
Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science, 2*(4), 313–345.
Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics, 121*(1), 297–317.
Scott-Clayton, J., Crosta, P. M., & Belfield, C. R. (2014). Improving the targeting of treatment evidence from college remediation. *Educational Evaluation and Policy Analysis, 36*(3), 371–393.
Stephan, J. L., Davis, E., Lindsay, J., & Miller, S. (2015). *Who will succeed and who will struggle? Predicting early college success with Indiana's Student Information System*U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest Technical report.
Symonds, W. C., Schwartz, R., & Ferguson, R. F. (2011). *Pathways to prosperity: Meeting the challenge of preparing young Americans.* Cambridge, MA: Pathways to Prosperity Project at Harvard Graduate School of Education.
Willingham, W. W. (1985). *Success in college.* New York: College Entrance Examination Board.