Using Machine Learning to Identify Personal Qualities in College Applications at Scale

Benjamin Lira[1], Angela L. Duckworth[1], Margo Gardner[2], Abigail Quirk[1], Cathlyn Stone[2], Arjun Rao[2], Stephen Hutt[1], and Sidney K. D'Mello[2]

[1]University of Pennsylvania

[2]University of Colorado, Boulder

Author Note

This research was supported by the Walton Family Foundation, the Student Experience Research Network, the Bill & Melinda Gates Foundation, the Joyce Foundation, the Overdeck Family Foundation, and the Raikes Foundation. Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

Correspondence concerning this article should be addressed to Benjamin Lira, University of Pennsylvania. Email: blira@upenn.edu

Abstract

To what extent can machine learning identify personal qualities from college application essays? We instructed human raters to code 3,131 essays for seven different personal qualities and then used these codes to train a computer algorithm to detect personal qualities in the same essays. Next, we applied the trained algorithm to a separate sample of 43,667 essays and analyzed whether computer-coded personal qualities predicted college graduation 4 and 6 years later. We found that computer-coded personal qualities of leadership, mastery orientation, and prosocial purpose predicted graduation, but teamwork, goal pursuit, self-concordant motivation, and perseverance—which human raters found difficult to code reliably—did not. These findings highlight both the future potential and current limitations of machine learning in college admissions.

Keywords: educational data mining, college admissions, social-emotional learning, motivation

Using Machine Learning to Identify Personal Qualities in College Applications at Scale

The college admissions process is ripe for innovation. For more than a century, standardized tests and report card grades have been the twin pillars of college admissions (Gaertner & Roberts, 2018). A more holistic approach would include an array of personal qualities not captured by these traditional academic metrics. A school may be keen to admit students who are kind, for example, or goal-directed or who excel as leaders. All these qualities could be predictive of graduation—one of the main goals behind admissions decisions—but assessing personal qualities in a high-stakes situation like college admissions is far from straightforward (Willingham, 1985).

Where might one look for indications of an applicant's personal qualities? One window into a student's character is what they write about themselves. But not all schools can afford an army of admissions officers to read essays one by one, searching each for evidence of personal qualities of interest. What's more, human judgment is influenced by a variety of unconscious biases and can be surprisingly unreliable across identical cases (Kahneman et al., 2021). In this exploratory investigation, we used machine learning— an artificial intelligence technique—to identify personal qualities in a large national sample (*N* = 43,667) of college applicants. We then examined the predictive power of computer-coded personal qualities for college graduation up to 6 years later.

Personal Qualities in College Admissions

Personal qualities are of growing interest in education (Authors et al., 2015). Whether referred to as character, social-emotional skills, noncognitive skills, soft skills, 21st century skills, or personality traits, personal qualities encompass an array of capabilities not fully captured by test scores or grades. Recently, more than 140 deans of admissions signed a pledge stating that admissions should take personal qualities into account when selecting applicants (Weissbourd et al., 2019). Why? Personal qualities are of intrinsic value in a vital and diverse student body. The admissions office at Stanford University, for example, states that “each year we aim to enroll a class of diverse backgrounds and experiences, talents, academic interests, and ways of viewing the world.” Practically speaking, the college admissions process should prioritize students who possess the varied capabilities needed to perform, persist, and graduate. Finally, insofar as personal qualities are independent of socioeconomic opportunity, broadening the scope of admissions may make college access more equitable.

Dozens of studies have shown that personal qualities are relevant to performance and persistence in college (Kyllonen et al., 2014; Robbins et al., 2004), and a handful have followed students long enough to assess their relevance to college success. In one study of 1,364 high school seniors, self-reported prosocial purpose—the desire to contribute to the world—predicted continued college enrollment (Authors et al., 2014). More recently, a study of 2,135 high school seniors found that self-reported grit predicted on-time graduation, but only at colleges with strong institutional support (Goyer et al., 2021). One cross-sectional analysis of the National Longitudinal Study of Adolescent Health found that self-reported agreeableness and emotional stability were consistent correlates of college graduation, whereas relationships with openness, conscientiousness, and introversion were moderated by sex and maternal education (Lundberg, 2013). Although self-report measures of personal qualities can be useful in research settings, “fakeability” and other limitations circumscribe their utility in high-stakes settings like college admissions (Authors et al., 2015).

A logical place to look for evidence of an applicant's personal qualities is in the words they use to describe themselves.[1] For example, a student may write about how getting cut from the basketball team and then practicing enough to make the team the following year taught them to persevere in the face of challenge. Another student might reflect on how they now want to pursue a major in social work after developing a deep commitment to the less privileged after volunteering at a soup kitchen. Recently, the National Association for College Admission Counseling (NACAC) asked 447 admissions officers where they look for evidence of personal qualities. The most common response—endorsed by 87% of admissions officers—was "the content of essay/personal statement" (National Association for College Admission Counseling, 2020).

Nevertheless, surprisingly little research has been published on personal qualities in the admissions essay. In the landmark Personal Qualities Project, researchers rated personal statements of about 4,000 applicants for writing quality and writing content, but there was no attempt to infer particular personal qualities from the essays (Willingham, 1985). In fact, we were able to locate only one peer-reviewed study of admissions essays and college graduation. In a sample of 236 nursing students, graduates were more likely to have written essays about helping and caring for others, whereas those who did not finish were more likely to have written about "nursing as external to themselves, as something to 'do' rather than to 'be'" (Sadler, 2003, p. 625).

[1] The historical origin of measuring personal qualities through essays can be traced to the early 20th century. Before the 1920s, admissions to college were based solely on academic performance. The shift towards considering personal qualities and character began with the establishment of the first office of admissions at Columbia University in 1910. Ironically, its objective was to devise ways of limiting the number of "undesirable students" and "uncultured Jews" who were increasingly gaining admission under the old system. To this end, the admissions office at Columbia incorporated qualitative features they could use to justify excluding Jewish applicants. Other elite universities soon followed, and the modern college application process—entailing personal essays, interviews, and letters of recommendation—was born (Karabel, 2005).

A classic study in psychology suggests that autobiographical narratives can indeed reveal personal qualities. In 1930, nuns were asked to “write a short sketch of [their] lives… including … influences that led to the convent, religious life, and outstanding events” (Danner et al., 2001, p. 806). More than 60 years later, researchers coded these essays for positive, neutral, and negative emotions, and then identified the age at which each nun had died. They found no link between neutral or negative words and longevity, but nuns whose essays contained more positive words (e.g., “happy,” “eager”) lived longer. More recently, another group of researchers coded the same essays and discovered that the more the nuns wrote about how they autonomously chose to join the convent, the longer the nuns lived, even after controlling for positive emotions (Weinstein et al., 2019).

Machine Learning and College Admissions

Even at well-resourced universities, admissions officers have minutes, not hours, to evaluate each application (Hoover, 2017). What’s more, regardless of the fact that officers might have precise criteria for identifying personal qualities, human beings tend to be unreliable in their judgments—applying harsher standards when hungry or tired, for example—and differing in idiosyncratic ways from one another (Kahneman et al., 2021).

How might colleges analyze student essays for evidence of personal qualities at scale? One possibility is to employ artificial intelligence, broadly defined as the use of computers to generate rational or humanlike behavior and decision making (Russell & Norvig, 2010). In particular, essays might be coded for personal qualities using machine learning, a subtype of artificial intelligence in which an algorithm is trained to find patterns in a sample of data and then generalize them for use in other data sets. In the rapidly evolving field of educational data mining (Baker, 2010), machine learning has been applied to grading, the identification of at-risk students, and other applications (Kucak et al., 2018; W. Wang et al., 2020). Although machine learning has not been used to capture personal qualities from written text in the context of education, it has been used to do so in social media, including posts on Facebook (Park et al., 2015) and Twitter (Pang et al., 2020), as well as in online personal blogs (Yarkoni, 2010).

Current Investigation

In this study, we used machine learning to identify personal qualities from college applicants' short (150-word limit) essays describing a significant extracurricular activity or work experience.[2] On average, American high school students spend around 3.5 hours per week in sports, student government, community service, paid work, and other extracurricular activities (U.S. Bureau of Labor Statistics, 2016). Experience sampling data reveals that teenagers experience such out-of-school pursuits as both intrinsically motivating and challenging (Larson, 2000). In contrast, academic work is typically experienced as challenging but not intrinsically rewarding to students, while socializing with friends is intrinsically motivated but not challenging. Thus, extracurricular activities are a primary avenue for high school students to express their autonomy (Leversen et al., 2012) and therefore may be prognostic of how they will behave in college, where they are afforded more autonomy and independence.

As shown in **Figure 1,** we began by instructing human raters to code essays for the presence or absence of seven different personal qualities: leadership, mastery orientation, prosocial purpose, teamwork, goal pursuit, self-concordant motivation, and perseverance. After establishing inter-rater reliability, these raters then coded a *development sample* of 3,131 essays. Next, in the same sample, we used these manually assigned codes to train a computer algorithm

[2] Note that as part of the Common Application, all applicants answered a prompt about activities or work experiences. Our sample is not limited to students who participated in planned school extracurriculars. Note that the Common Application includes a longer essay component, which we could not access due to confidentiality agreements.

to identify the likelihood that each personal quality is present in a given essay. Within hours, the trained computer algorithm was able to generate estimated likelihoods for each personal quality in a *holdout sample* of 43,667 essays. Finally, we used longitudinal data from the National Student Clearinghouse to establish the predictive validity of these computer-generated likelihoods for college graduation.

Method

Participants

The data for this study came from a legally and IRB-reviewed collaboration with the Common Application and the National Student Clearinghouse Research Center, in which de-identified data were made available for research on predictors of college graduation. The Common Application is a national nonprofit membership organization representing hundreds of colleges and universities that provides a standardized college application platform. The National Student Clearinghouse Research Center is a nonprofit organization that works with colleges and universities as well as states, districts, high schools, and educational organizations across the U.S. "to inform practitioners and policymakers about student educational pathways and enable informed decision making."

As shown in **Figure 2**, 413,675 students completed the Common Application (www.commonapp.org) during the 2008-2009 college admissions cycle. We were able to link 362,205 of these applications to graduation data from the National Student Clearinghouse data set. After removing 50,894 students who enrolled in college prior to 2008 and an additional 3 students due to data integrity issues, we were left with 311,308 students. That year, high school counselors had the option to submit report card grades either online or by uploading hard-copy transcripts. This was consequential for our data access because transcripts included student names and were not possible to de-identify. This yielded 47,303 applications with high school GPA (HSGPA) data and 264,005 without it.

To provide labeled data for the machine learning algorithm, we set aside a *development sample* of 4,000 essays for manual coding. We used stratified random sampling to ensure representation across demographic groups and levels of involvement in extracurricular activities (see **Supplementary Online Materials** for details). After excluding missing data, invalid responses, and essays coded by one rater who ultimately failed to achieve agreement with other raters, the development sample consisted of 3,131 students.

Our main analytic sample—which we refer to as the *holdout sample*—comprised a subset of n = 43,667 students for whom we had access to de-identified high-school GPA data, and who were not part of the development sample. As shown in **Table S1**, the holdout sample was representative of the full sample of 311,308 students.

To test the robustness of our results on a larger sample, we included all the students not in the development sample, regardless of whether or not they had HSGPA data in a *robustness sample.* After exclusions and data-entry errors, the robustness sample consisted of 307,251 students. See **Figure 2** for a visual representation of the composition of each sample.

Measures

Essay on Extracurriculars and Work Experiences

All applicants who completed the Common Application responded in up to 150 words to the following essay prompt: “Please briefly elaborate on one of your activities or work experiences.”

High School GPA (HSGPA)

For students in the *holdout sample*, school counselors reported cumulative HSGPA and their school's GPA scale. To achieve a common metric, we proportionalized students' HSGPAs by dividing their raw scores by the maximum value on the school's scale, following Authors et al. (2019). Because some schools used weighted scales that allowed individual HSGPAs to exceed the maximum HSGPA, some scores exceeded 1.00. Additionally, extremely out-of-range GPA proportions (e.g., proportions many times greater than 1) suggested reporting errors. To address these outliers, GPA proportions were capped at the 1st and 99th percentiles.

College Admissions Test Scores

Over half (55%) of the robustness check sample completed the SAT, 14% completed the ACT, and 25% completed both. Using published guidelines, we converted ACT scores to SAT scores. For students who submitted scores from both tests, we selected the higher score.

Extracurricular Activities

Applicants listed up to seven extracurricular activities and for each, indicated participation intensity in terms of years, the type of activity, and hours per week they participated. For each applicant, we computed the total number of extracurricular activities, mean years per activity, and the proportion of activities that were sports.[3]

Demographics

We obtained the following demographic information from the Common Application: gender, race/ethnicity, type of high school (i.e., Title I public school, non-Title I public school, private school, or homeschool), English language learner status, and parents' education level and marital status.

---

[3] Sports participation often predicts positive academic outcomes, but it also predicts risk behaviors like alcohol use that may derail college progress (see Farb & Matjasko, 2012, for a review). Note that applicants indicated whether each of the activities reported were sports in the Common Application.

College Graduation

We obtained data from the 2015 National Student Clearinghouse database (www.studentclearinghouse.org) to create two binary graduation measures: 4-year and 6-year graduation (0 = *did not earn a bachelor's degree within 4 [6] years of initial enrollment*; 1 = *earned a bachelor's within 4 [6] years*).

Institutional Graduation Rates

We obtained institutional rates of graduation within 4 and 6 years from the National Center for Educational Statistics. We controlled for any potential effects of baseline institutional effects on the odds of graduation in the **Supplementary Online Materials.**

Procedure

To begin, the second and third authors read random batches of 50 applicant essays to identify common personal qualities. After reading and discussing nine batches (i.e., 450 essays), these authors developed criteria for seven personal qualities: leadership, mastery orientation, prosocial purpose, teamwork, goal pursuit, self-concordant motivation, and perseverance. See **Table 1** for coding rules and fictionalized examples.

Next, we trained five research assistants to apply these criteria until each coder achieved adequate inter-rater reliability with either the second or third author across all seven attributes (Krippendorff's alpha > .80). Then, raters coded all essays in the development sample. Most of the essays were coded by a single rater (n = 2,925; 93% of the development sample). To assess inter-rater reliability, pairs of raters independently coded a subset of essays (n = 206; 7% of the development sample) before discussing disagreements until they reached consensus.

In the development sample, we used manually assigned codes to train a computer algorithm to estimate the probability of each personal quality. Specifically, we employed

Bidirectional Encoder Representations from Transformers (BERT), an advanced machine learning approach considered a significant innovation upon prior algorithms in the field of natural language processing (Devlin et al., 2019). BERT is a language representation model—a deep neural network that is pre-trained on extremely large volumes of generic text (e.g., 2.5 billion words in English Wikipedia)—and is then fine-tuned (by adjusting its parameters) using sample-specific training data. We used 10-fold cross validation, dividing the 3,131 coded essays into 10 random subsets. We trained BERT models on nine subsets and generated predictions on the held-out subset. We repeated this process until each subset was used for testing once. We then pooled the computer-generated likelihoods over the 10 iterations to estimate reliability. Rather than modeling all seven codes simultaneously, we opted for individual models per code.

Next, we applied the 10 models for each code to generate predictions for the remaining 307,308 essays. Each BERT model estimated the likelihood that an essay would earn a score of 1 on a given personal quality, which we averaged across the 10 models, resulting in seven different computer-generated, continuous (0 to 1) codes per essay. The computer failed to generate likelihoods for 57 essays (0.02%), suggesting data errors, yielding 307,251 and 43,667 computer-generated likelihoods in the robustness and holdout samples, respectively.

In the holdout sample, 4.2% cases were missing admissions test scores, 12.1% were missing institutional graduation rates, and 5.2% were missing data on high school Title I status. In the robustness sample, 5.7%, 12.2%, and 7.1% of students were missing data on admissions test scores, 4-year and 6-year institutional graduation rates, and high school Title I status, respectively. To handle missing data, we used multiple imputation (m = 25), employing the mice package in R. We used predictive mean matching for graduation rates and college admissions test scores and polytomous regression for school type.

Finally, we entered all seven computer-generated likelihoods of personal qualities into binary logistic regression models predicting 4-year and 6-year graduation, respectively. These predictors, as well as all other continuous variables, were standardized to facilitate interpretation of odds ratios. All models controlled for a suite of covariates, including demographics, high school GPA, and college admissions test scores. Our BERT settings python code and R code for analysis are available here on OSF.

Results

Descriptive Statistics

Human raters assigned an average of two (out of seven) personal qualities to essays in the development sample. Some personal qualities were more commonly observed than others. For instance, the personal qualities of mastery orientation and self-concordant motivation were coded for 42% of essays, whereas the personal qualities of leadership and perseverance were coded for only 18% and 19% of essays, respectively.

Our human raters were able to code some personal qualities more reliably than others. According to rules of thumb in Krippendorff (2004), prosocial purpose was reliably coded (Krippendorff's $\alpha > .80$), and reliabilities for leadership and mastery orientation merited "drawing tentative conclusions" (Krippendorff's αs > .67) (p. 241). In contrast, estimated reliabilities for perseverance, self-concordance, teamwork, and goal pursuit were lower (Krippendorff's αs < .67). Not surprisingly, the correlation between the computer-generated likelihoods and manually assigned codes were stronger for leadership, mastery orientation, and prosocial purpose (point-biserial *rs* ranged from .75 to .83), compared to those for teamwork, goal pursuit, self-concordant motivation and perseverance (point-biserial *rs* ranged from .54 to

.70). See **Table 2** for correlations, descriptive statistics, Krippendorff's alphas for the human raters, and human-computer correlations**.**

Whether coded by human raters or the computer algorithm, each of the seven personal qualities appeared to capture unique information. As shown in **Table 2**, tetrachoric correlations among manually assigned codes ranged from r = -.22 to r = .27 (average r = .01), suggesting relatively orthogonal dimensions. Likewise, bivariate associations among computer-generated likelihoods ranged from r = -.21 to r = .26 (average r = -.01). As further evidence of discriminant validity, bivariate associations with covariates in the college admissions application varied by personal quality. For instance, computer-generated likelihoods for leadership correlated positively with both high school GPA (r = .08, p < .001) and college entrance test scores (r = .06, p < .001). Mastery orientation, however, was not (r = .01, p = .11 and r = -.01, p = .08, respectively). Prosocial purpose showed yet another pattern, correlating positively with high school GPA (r = .03, p < .001) and negatively with college entrance test scores (r = -.01, p = .02).

Predicting Graduation

In the holdout sample, the average graduation rate was 40% and 78% for 4-year and 6-year graduation, respectively. Three of seven personal qualities reliably predicted both 4-year and 6- year college graduation. As shown in **Table 3**, computer-generated likelihoods for leadership (4-year OR = 1.03, p = .002; 6-year OR = 1.06, p < .001), mastery orientation (4-year OR = 1.03, p = .005; 6-year OR = 1.03, p = .009), and prosocial purpose (4-year OR = 1.02, p = .023; 6-year OR = 1.07, p < .001) each demonstrated incremental predictive validity for both 4-year and 6-year college graduation over and above traditional predictors and demographics. It is worth noting that predictive validities for these personal qualities, while reliable, were modest in size. For example, students whose essays were one standard deviation higher in leadership had only 3% higher odds of graduating from college within 4 years and 6% higher odds of graduating within 6 years. In contrast, a 1 standard deviation increase in high school GPA was associated with 25% higher odds of graduation within 4 years and 43% higher odds within six years. Even the total number of extracurricular activities a student participated in—the coarsest of measures of extracurricular involvement—more accurately predicted college graduation: A 1 standard deviation increase in that indicator resulted in a 9% and 20% higher odds of graduation in 4 and 6 years, respectively.

As detailed in the **Supplementary Online Materials**, a series of robustness checks confirmed the predictive validity of leadership, mastery orientation, and prosocial purpose. These alternative model specifications varied on the following dimensions: whether or not missing data was imputed, use of the holdout versus robustness samples, inclusion of all seven personal qualities in the same model versus entering one at a time, using 4-year versus 6-year graduation as the outcome, and whether or not institutional graduation rates were included as a covariate. These variations resulted in a set of 32 estimated coefficients for each personal quality. The predictive validities of leadership, mastery orientation, and prosocial purpose reached statistical significance in 100%, 87.5%, and 87.5% of model specifications, respectively. In contrast, the predictive validities of teamwork, goal pursuit, self-concordant motivation, and perseverance reached positive statistical significance in 50%, 12.5%, 0%, and 6.25% of model specifications, respectively. As a final check of robustness, we calculated multilevel models with random intercepts to account for the possibility that our results were confounded by the nested structure of the data. Results were largely consistent (See **Table S8**).

Discussion

We used machine learning to identify personal qualities from college essays in a national longitudinal dataset of 307,251 students. We found that students whose written descriptions of extracurricular activities were coded by a trained computer algorithm as indicating leadership, mastery orientation, or prosocial purpose were more likely to earn their college diplomas 4 or 6 years later—even when controlling for each other and a rich set of covariates, including grades and college admissions test scores. In contrast, computer-generated likelihoods for teamwork, goal pursuit, self-concordant motivation, and perseverance did not reliably predict college graduation.

Although our application of machine learning to the assessment of personal qualities in the college application is novel, the relevance of leadership, mastery orientation, and prosocial purpose to college success is consistent with prior research. For example, in the National Education Longitudinal Study, holding a leadership role during high school (e.g., athletic team captain, club officer) predicts college graduation, as well as enrollment in and total years of postsecondary education (Rouse, 2012). Likewise, although mastery goals are not always predictive of academic performance in college, mastery orientation (i.e., the motivation to learn and improve) has been shown in experimental and correlational research to predict continued interest and motivation, positive coping, self-regulation, and deep learning strategies (Kaplan & Maehr, 2007). Finally, prosocial purpose (i.e., the desire to contribute to the world) has been shown in longitudinal and experimental research to improve academic effort, performance, and persistence (Authors et al., 2014).

Why weren’t goal pursuit, teamwork, self-concordance, or perseverance prognostic of college graduation? Each of these personal qualities has been associated with academic achievement in prior research (see Black & Deci, 2000; Thorsen et al., 2021; L. Wang et al., 2009; Zimmerman et al., 1992). Moreover, these four personal qualities were as commonly observed in personal essays (mean likelihood = 39%) as the three that were predictive (mean likelihood = 35%). The most likely explanation, in our view, is that personal qualities that failed to predict graduation were more difficult for human raters to code reliably. There were stark differences between the reliabilities of predictive (range of alphas = .73 - .83) versus non-predictive personal qualities (range of alphas = .57 - .66). Moreover, estimates of inter-rater reliability correlated positively with estimates of human-computer correlations (r = .93), as well as predictive validities for 4-year and 6-year graduation (r = .63; r = .91).[4] In other words, disagreement between the human raters corresponded with lower correspondence with computer-estimated likelihoods as well as lower predictive validities.

It is worth noting that observed effect sizes for leadership, mastery orientation, and prosocial purpose were quite small when compared to the predictive validities of high school grades and college admissions test scores. Why? The simplest explanation is that these personal qualities are less important to earning a college diploma than academic self-regulation or cognitive skills indexed by these traditional admissions criteria (Authors et al., 2019a). Of course, we might have obtained larger effects had we increased reliability by applying the algorithm to a variety of writing samples for each applicant and then aggregating estimates. Perhaps, but there is a limit to the expected prognostic power of any personal quality, no matter how precisely measured. After all, personal qualities change over the life course, including during adolescence and emerging adulthood (Roberts et al., 2006). In addition, personal qualities are both domain-general and domain-specific. A student who expresses mastery orientation for

[4]We used linear coefficients rather than odds ratios to calculate correlations with Krippendorf alpha reliability estimates.

dancing, ceteris paribus, likely enjoys learning in other activities as well, but is not guaranteed to express mastery orientation in their college classes. Finally, whether or not a student earns their college diploma also depends on financial circumstances and other non-psychological factors.

Several limitations of this study are worth highlighting. By necessity, we relied on 150-word descriptions of extracurricular activities. The Common Application includes longer essays, but in the cohort to which we had access, these longer essays included student identification information and were not shared with our team. Now that the Common Application process is entirely electronic, there is the possibility of replicating this investigation with longer and more varied written components within the college application. Second, graduation is one among many consequential outcomes we could have examined. For instance, when defining success in college, Willingham (1985) included academic honors, leadership, and community involvement. Third, the field of machine learning algorithms is continually evolving. There are now language models that are arguably superior to BERT (e.g., XLNet; Yang et al., 2020) that could be tested in future research. Fourth, it is possible we could have improved prediction by omitting the intermediate step of manually coding personal qualities and instead directly linking essay text to graduation data. However, if that did improve prediction, such improvement would come at a cost in terms of theoretical interpretability. Finally, the essay is not the only element of the college application that may reveal an applicant’s personal qualities. In a recent survey, 80% of college admissions officers said they also relied on written recommendations of teachers and counselors, and 73% said they look at the list of extracurricular activities applicants report (NACAC, 2020). Future research is needed to assess the potential of machine learning for assessing personal qualities from other elements in the college application.

This inaugural application of machine learning to the assessment of personal qualities in the college application underscores the importance of the labeled data on which the computer algorithms are trained. The adage “garbage in, garbage out” has often been used to sum up the major limitation of this approach (Geiger et al., 2021). The challenge of establishing a high-fidelity dataset on which to train the computer algorithm makes adding machine learning to the admissions process premature. Nevertheless, we are optimistic about the long-term promise of artificial intelligence to revolutionize college admissions. Compared to the intuitive judgments of human raters, algorithms are dramatically less noisy (Kahneman et al., 2021), presumably more easily tested for bias and fairness across gender, ethnicity, or other subgroups (Authors et al., 2019b), and are dramatically more cost-effective. At the same time, their use for high-stakes decision making raises several foundational issues pertaining to privacy, ethics, equity, and justice (https://www.ajl.org/). Contingent upon advances in the conventional assessment of personal qualities, if done responsibly, machine learning has the potential to make the college admissions process both more efficient and perhaps even more equitable.

References

Authors et al. (2014)

Authors et al. (2015)

Authors et al. (2019a)

Authors et al. (2019b)

Baker, R. S. (2010). Data mining for education. In *International encyclopedia of education* (3rd ed., pp. 112–118). Elsevier.

Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motivation on learning organic chemistry: A self-determination theory perspective. *Science Education*, *84*(6), 740–756. https://doi.org/10.1002/1098-237X(200011)84:6<740::AID-SCE4>3.0.CO;2-3

Danner, D. D., Snowdon, D. A., & Friesen, W. V. (2001). Positive emotions in early life and longevity: Findings from the nun study. *Journal of Personality and Social Psychology*, *80*(5), 804–813. https://doi.org/10.1037/0022-3514.80.5.804

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Arxiv Preprints*, *1810.04805v2*.

Farb, A. F., & Matjasko, J. L. (2012). Recent advances in research on school-based extracurricular activities and adolescent development. *Developmental Review*, *32*(1), 1–48. https://doi.org/10.1016/j.dr.2011.10.001

Gaertner, M. N., & Roberts, R. D. (2018). More than a test score: Defining and measuring personal qualities. In K. L. McClarty, K. D. Mattern, & M. N. Gaertner (Eds.), *Preparing students for college and careers: Theory, measurement, and educational practice*. Routledge, Taylor & Francis Group.

Geiger, R. S., Cope, D., Ip, J., Lotosh, M., Shah, A., Weng, J., & Tang, R. (2021). “Garbage in, garbage out” revisited: What do machine learning application papers report about human-labeled training data? *Quantitative Science Studies,* 1–32. https://doi.org/10.1162/qss_a_00144

Goyer, J. P., Walton, G. M., & Yeager, D. S. (2021). The role of psychological factors and institutional channels in predicting the attainment of postsecondary goals. *Developmental Psychology, 57*(1), 73–86. https://doi.org/10.1037/dev0001142

Hoover, E. (2017, March 12). Working smarter, not harder, in admissions. *The Chronicle of Higher Education.* https://www.chronicle.com/article/working-smarter-not-harder-in-admissions/

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Harper Collins.

Kaplan, A., & Maehr, M. L. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review*, *19*(2), 141–184. https://doi.org/10.1007/s10648-006-9012-5

Karabel, J. (2005). *The chosen: The hidden history of admission and exclusion at Harvard, Yale, and Princeton.* Houghton Mifflin Harcourt.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). SAGE Publications.

Kucak, D., Juricic, V., & Dambic, G. (2018). Machine learning in education—A survey of current research trends. In B. Katalinic (Ed.)*, DAAAM Proceedings* (pp. 0406–0410). https://doi.org/10.2507/29th.daaam.proceedings.059

Kyllonen, P. C., Lipnevich, A. A., Burrus, J., & Roberts, R. D. (2014). *Personality, motivation,* *and college readiness: A prospectus for assessment and development.* (Report No. ETS RR–14-06). Educational Testing Service. https://doi.org/10.1002/ets2.12004

Larson, R. W. (2000). Toward a psychology of positive youth development. *American Psychologist*, *55*(1), 170. https://doi.org/10.1037/0003-066X.55.1.170

Leversen, I., Danielsen, A. G., Birkeland, M. S., & Samdal, O. (2012). Basic psychological need satisfaction in leisure activities and adolescents' life satisfaction. *Journal of Youth and Adolescence*, *41*(12), 1588–1599. https://doi.org/10.1007/s10964-012-9776-5

Lundberg, S. (2013). The college type: Personality and educational inequality. *Journal of Labor Economics*, *31*(3), 421–441. https://doi.org/10.1086/671056

National Association for College Admission Counseling. (2020). *Character and the college admission process*. https://www.nacacnet.org/globalassets/documents/publications/research/character-brief/nacac_brief_character-0120_2.pdf

Pang, D., Eichstaedt, J. C., Buffone, A., Slaff, B., Ruch, W., & Ungar, L. H. (2020). The language of character strengths: Predicting morally valued traits on social media. *Journal of Personality*, *88*(2), 287–306. https://doi.org/10.1111/jopy.12491

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, *108*(6), 934–952. https://doi.org/10.1037/pspp0000020

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. https://doi.org/10.1037/0033-2909.130.2.261

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*(1), 1–25. https://doi.org/10.1037/0033-2909.132.1.1

Rouse, K. E. (2012). The impact of high school leadership on subsequent educational attainment: The impact of high school leadership. *Social Science Quarterly*, *93*(1), 110–129. https://doi.org/10.1111/j.1540-6237.2011.00836.x

Russell, S., & Norvig, P. (2010). *Artificial intelligence: A modern approach*. Prentice Hall.

Sadler, J. (2003). Effectiveness of student admission essays in identifying attrition. *Nurse Education Today*, *23*(8), 620–627. https://doi.org/10.1016/S0260-6917(03)00112-6

Thorsen, C., Hansen, K. Y., & Johansson, S. (2021). The mechanisms of interest and perseverance in predicting achievement among academically resilient and non-resilient students: Evidence from Swedish longitudinal data. *British Journal of Educational Psychology*, e12431. https://doi.org/10.1111/bjep.12431

U.S. Bureau of Labor Statistics. (2016). *American time use survey: Charts by topic: Students*. https://www.bls.gov/tus/charts/students.htm

Wang, L., MacCann, C., Zhuang, X., Liu, O. L., & Roberts, R. D. (2009). Assessing teamwork and collaboration in high school students: A multimethod approach. *Canadian Journal of School Psychology*, *24*(2), 108–124. https://doi.org/10.1177/0829573509335470

Wang, W., Zhuang, H., Zhou, M., Liu, H., & Li, B. (2020). What makes a star teacher? A hierarchical BERT model for evaluating teacher’s performance in online education. *ArXiv:2012.01633 [Cs]*. http://arxiv.org/abs/2012.01633

Weinstein, N., Legate, N., Ryan, W. S., & Hemmy, L. (2019). Autonomous orientation predicts longevity: New findings from the Nun Study. *Journal of Personality, 87*(2), 181–193.

https://doi.org/10.1111/jopy.12379

Weissbourd, R., Anderson, T. R., Barnard, B., Cashin, A., & Ditkowsky, A. (2019). *Turning the tide II: How parents and high schools can cultivate ethical character and reduce distress in the college admissions process*. https://mcc.gse.harvard.edu/

Willingham, W. W. (1985). *Success in college: The role of personal qualities and academic ability*. College Board Publications.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2020). XLNet: Generalized autoregressive pretraining for language understanding. *Arxiv Preprints*, *1906.08237v2*.

Yarkoni, T. (2010). Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363–373. https://doi.org/10.1016/j.jrp.2010.04.001

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, *29*(3), 663–676. https://doi.org/10.3102/00028312029003663

Table 1

Personal Qualities, Coding Rules and Fictionalized Example Essays

Personal quality	Criteria for coding by human raters	Fictionalized example essay with relevant phrases in **bold**
Prosocial purpose	Helping others, wanting to help others, consideration of the benefits to others, mention of reasons for helping others, or reflection on how enjoyable or rewarding it is to help others	Every summer for the last three years, **I worked as camp counselor at a camp for young children from underprivileged families. Helping children realize their hidden talents is one of the most rewarding experiences I have ever had. I've been so fulfilled by watching these children develop confidence in their abilities. This experience has been so important to me,** and it showed me that a career in education is where I belong.
Leadership	Serving in a leadership role, commenting on what he or she did in his or her capacity as a leader, or discuss the value, meaning, or importance of leadership	I was chosen to be cheerleading **captain** during my senior year. My freshman year captain had a huge impact on my life, and I felt like it was my time to pay it forward. **I am so proud of everything I did for the girls: creating a mentorship system, organizing events and fundraisers, and encouraging everyone to work as hard as they could.** At the end of the year, a few girls thanked me. I was completely overcome with emotion. **I've never felt so gratified in my life.**
Mastery orientation	Improving, learning, or developing knowledge, skills, or abilities	I played softball in high school. When I started, **I was not a very strong player.** When I finally made the varsity team my senior year, I was determined to have a better season. **I worked constantly to improve my game – during practice and on my own time. My skills grew so much.** Because of my hard work, I finished the year with the best record on my team!
Goal pursuit	Having a goal and/or a plan	I have been playing soccer since I was six years old. Unfortunately, last year I injured my knee, and it has been a struggle to get back to the level I was playing at before my injury. It has been really challenging, but **I've been doing physical therapy and practicing everyday so that I can be a varsity starter this year.**
Self-concordant motivation	Describing the activity as enjoyable or interesting. Liking the activity or identifying with it.	**Running track is so much more than a sport to me.** It's a challenge and an adventure, and I put everything I have into it. **I love every aspect of it, even the afternoons I spend drenched in sweat in the scorching heat.**
Teamwork	Working with or learning from others. Valuing what fellow participants bring to the activity.	I've been on my school's debate team since my freshman year, and was elected co-captain **because of my commitment to the team's success. My fellow co-captains and I worked together to get our team ready for competitions.** We knew that **a strong team performance was more important than the successes of a few individuals. We stressed teamwork and cooperation between our teammates.** Because **we focused on team effort**, we earned first place at the state meet.
Perseverance	Persisting in the face of challenge	I've learned to become a gracious victor and to **grow from defeat.** Track has **helped me overcome my fear of losing,** and even helped me put my life in perspective. I've learned to **keep working and fighting even when the odds seem impossible to beat. There were many times that I found myself lagging, but I pulled ahead at the end because I never gave up.** The most important thing I've learned is **to never let anything stand in my way.**

Table 2

Descriptive Statistics and Bivariate Associations of Manually-Assigned Codes and Computer-Generated Likelihoods of Personal Qualities

Variable	1	2	3	4	5	6	7
Personal qualities							
1. Leadership	–	-.06†	-.01	.27***	.08*	-.13***	-.02
2. Mastery orientation	-.02***	–	-.13***	.09*	.01	-.04	.15***
3. Prosocial purpose	-.01	-.13***	–	-.07*	.09**	-.07*	-.22***
4. Teamwork	.26***	.05***	-.11***	–	.05†	-.04	.13***
5. Goal pursuit	.07***	-.08***	.12***	-.01	–	.03	.06†
6. Self-concordant motivation	-.16***	-.03***	-.10***	-.06***	.00	–	.06*
7. Perseverance	-.04***	.11***	-.21***	.06***	.00	.08***	–
College graduation							
4-year graduation	.03***	.01**	.02***	.02***	.01†	.00	.00
6-year graduation	.05***	.02***	.03***	.03***	.02***	.01*	.02***
Traditional admissions criteria							
HSGPA	.08***	.01	.03***	.04***	.01*	-.01*	.03***
College admissions test scores	.06***	-.01†	-.01*	.05***	.02***	.00	.08***
Extracurricular activities							
Number of activities	.09***	.03***	.10***	.07***	.07***	.01	.04***
Years per activity	.05***	-.01†	-.04***	.01*	-.01**	.06***	.01**
Proportion sports	-.03***	.01*	-.10***	.03***	-.03***	.04***	.07***
Race/ethnicity							
White	.03***	-.01	-.03***	.03***	-.01*	.03***	.00
Latino	-.01*	.00	.04***	-.02***	.00	-.01*	-.02***
Black	-.01†	-.01**	.01	-.02***	-.01†	-.03***	-.03***
Asian	.01	.02***	.04***	.00	.03***	-.03***	.02***
Other	.00	.00	-.01†	.00	.00	-.01	-.01†
No race/ethnicity reported	-.03***	.00	-.01*	-.01	.00	.01*	.02***
Parental education							
No parent with college degree	-.02***	-.02***	.00	-.04***	-.03***	-.03***	-.05***
One parent with college degree	.01†	.00	-.01	.00	.01†	.00	-.01

Two parents with college degrees	.01*	.02***	.00	.03***	.02***	.03***	.05***
High school type							
Non-Title I public school	.03***	.00	-.02***	.02***	-.01*	-.02***	-.02***
Private school	-.04***	.00	.03***	-.02***	.02***	.04***	.04***
Public school Title I	.02***	.00	-.01†	.00	-.01*	-.03***	-.02**
Homeschool	-.02***	-.01**	.00	-.02***	-.01**	.00	-.01
Other demographics							
Female	.02***	.03***	.10***	.00	.02***	.06***	.00
English language learner	-.02***	.01**	.04***	-.02***	.02***	-.03***	.00
Parents not married	-.02***	-.01**	-.02***	-.03***	.00	.00	-.02***
Summary statistics							
Manually-coded frequency	.18	.42	.34	.26	.31	.42	.19
Computer-generated likelihood	.22	.49	.33	.34	.42	.56	.24
Reliability of human raters	.78	.73	.83	.61	.57	.63	.66
Human-computer correlation	.79***	.75***	.83***	.54***	.55***	.70***	.63***

Note. Correlations above the diagonal are tetrachoric coefficients for manually coded essays in the development sample (n = 3,131). Correlations below the diagonal are Pearson coefficients for the computer-generated likelihoods (n = 43,667). Correlations between manually assigned codes and computer-generated likelihoods, and the frequency of manually assigned codes were calculated in the development sample (n = 3,131). Mean of computer-generated likelihoods (n = 43,667). Reliability estimates are Krippendorff's alpha of the manually coded essays (n = 206).

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3

Binary Logistic Regressions Predicting 4-Year and 6-Year Graduation

Term	4-year graduation		6-year graduation	
	OR	[95% CI]	*OR*	[95% CI]
Computer-generated likelihoods of personal qualities				
Leadership	1.03**	[1.01-1.05]	1.06***	[1.03-1.09]
Mastery orientation	1.03**	[1.01-1.05]	1.03**	[1.01-1.06]
Prosocial purpose	1.02*	[1.00-1.05]	1.07***	[1.04-1.10]
Teamwork	1.01	[0.99-1.03]	1.00	[0.98-1.03]
Goal pursuit	1.00	[0.98-1.02]	1.02	[0.99-1.04]
Self-concordant motivation	1.00	[0.98-1.02]	1.01	[0.99-1.03]
Perseverance	0.98	[0.97-1.01]	1.02†	[1.00-1.05]
Traditional admissions criteria				
HSGPA	1.25***	[1.22-1.28]	1.43***	[1.39-1.47]
College admissions test scores	1.10***	[1.07-1.13]	1.24***	[1.21-1.28]
Extracurricular activities				
Number of activities	1.09***	[1.06-1.11]	1.20***	[1.17-1.23]
Years per activities	1.01	[0.99-1.03]	1.08***	[1.06-1.11]
Proportion sports	1.03**	[1.01-1.06]	1.06***	[1.04-1.09]
Parents' education$_a$				
No parents with college degrees	0.84***	[0.80-0.89]	0.68***	[0.64-0.73]
One parent with college degree	0.98	[0.93-1.03]	0.86***	[0.81-0.92]
Race/ethnicity$_b$				
Latino	0.83***	[0.76-0.90]	0.94	[0.86-1.04]
Black	0.78***	[0.71-0.86]	0.82***	[0.75-0.91]
Asian	0.85***	[0.79-0.92]	0.76***	[0.69-0.83]
Other	0.86***	[0.80-0.93]	0.76***	[0.70-0.83]
No race reported	0.97	[0.92-1.02]	0.84***	[0.79-0.90]
Type of high school$_c$				
Private school	1.05†	[1.00-1.10]	0.95	[0.90-1.01]
Public school Title I	0.94*	[0.89-0.99]	0.89***	[0.83-0.95]
Homeschool	0.79†	[0.61-1.02]	0.67**	[0.50-0.90]
Other demographics				
Female$_d$	1.22***	[1.17-1.27]	1.39***	[1.32-1.46]
Parents not married$_e$	0.87***	[0.83-0.92]	0.74***	[0.70-0.79]
English language learner$_f$	0.90**	[0.84-0.97]	0.71***	[0.66-0.77]
(Intercept)	0.68***	[0.65-0.72]	4.88***	[4.60-5.18]
AUC	.61		.70	
Scaled Brier Score	.03		.10	
Nagelkerke's R^2	.05		.14	

Note. To facilitate interpretation of odds ratios, all variables were standardized prior to the analysis. Results are pooled estimates from m = 25 multiply imputed datasets.

$_a$Reference group is two parents with college degrees. $_b$Reference group is White. $_c$Reference group is non-Title I high school. $_d$Reference group is male. $_e$Reference group is married parents. $_f$Reference group is English as a first language.

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 1

Machine Learning of Open-Ended Text Responses in the Common Application

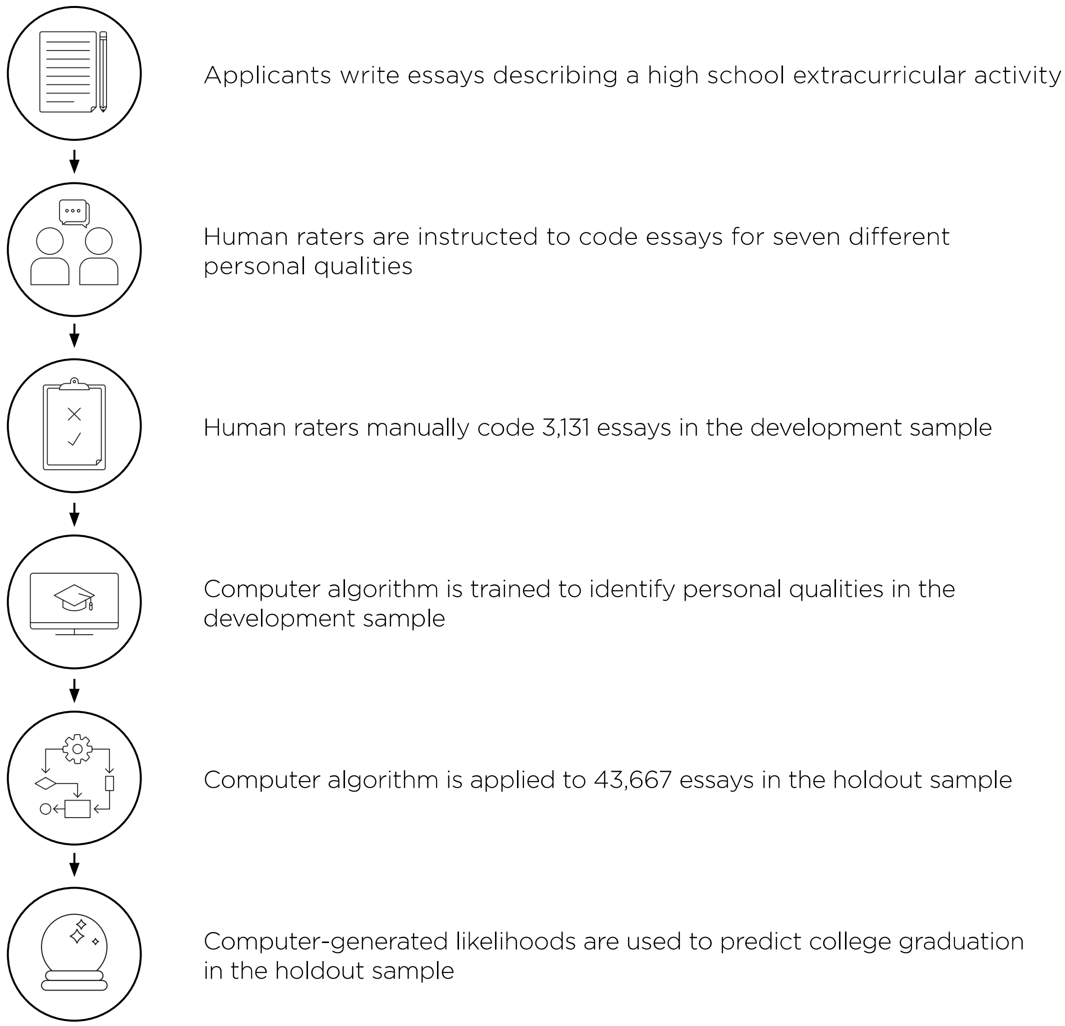
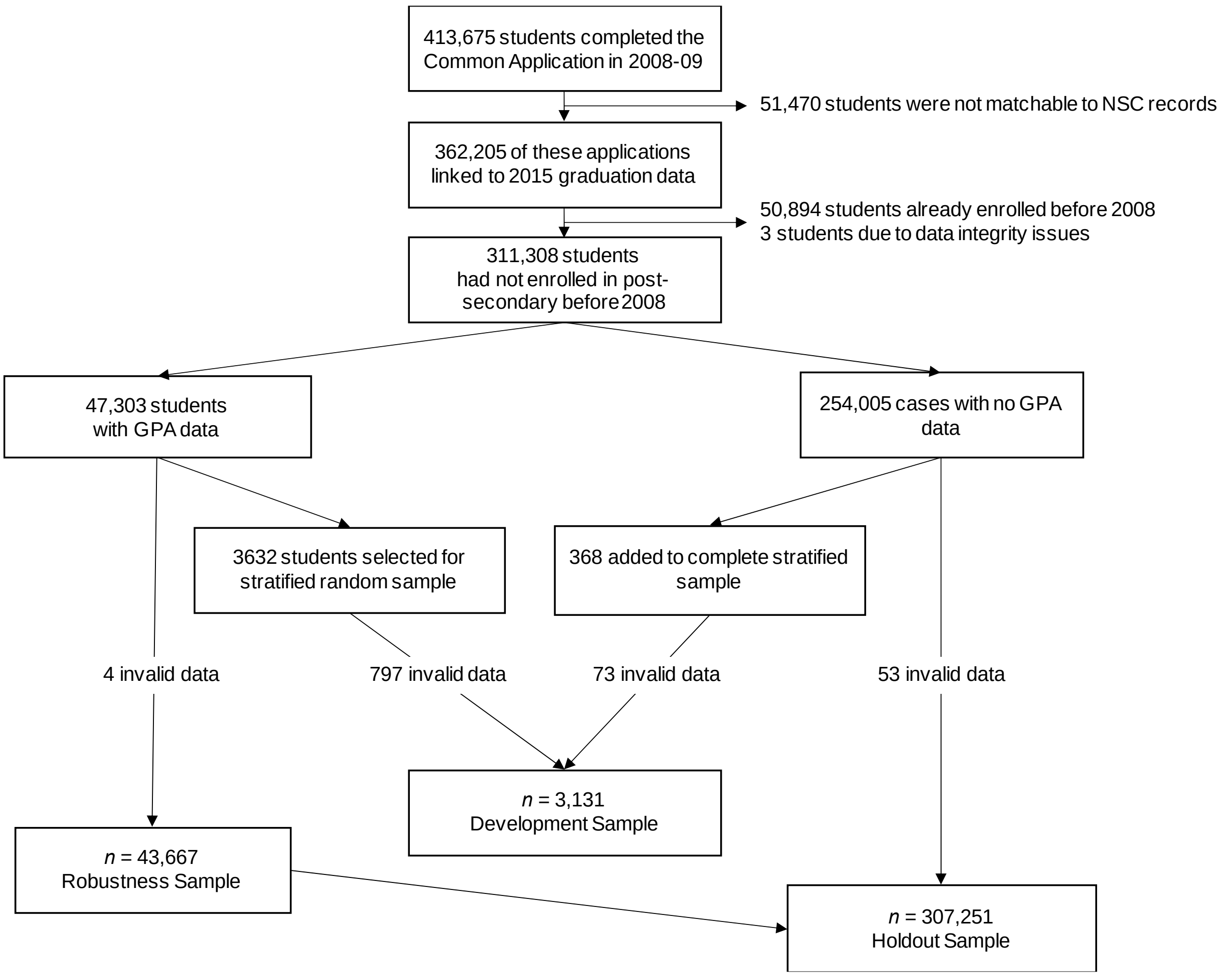


Figure 2

Sampling and Participants



Supplementary Online Materials

Using Machine Learning to Identify Personal Qualities in College Applications at Scale

Additional Sampling Information

The development sample was a stratified sample of *n* = 4,000 students from the full sample. As reported in Authors et al. (2019), we defined sampling strata based on the number of extracurricular activities reported on the Common Application as well as membership in one of five multi-dimensional demographic groups identified using latent class analysis (LCA). Specifically, our LCA model classified students according to profiles across race/ethnicity, parental education, parents' marital status, English language learner status (ELL), attended a Title I high school, and high school race/ethnic composition. The LCA was performed in MPlus 7 on all public school students (*n* = 213,091) in the larger sample of 311,308. We excluded private and homeschooled students from this analysis because their school-level demographic data were not available. This procedure ensured adequate representation throughout our development sample.

Forty-eight cases had to be removed from the development sample due to data entry errors or invalid responses to the essay prompt (e.g., applicant wrote "response will be mailed separately"). To maximize reliability, we excluded ratings from one research assistant who failed to achieve reliability with other coders. Ultimately, our development sample of 3,131 students for manual coding comprised 2,835 students with available HSGPA data and 296 students without it. To establish interrater reliability, raters coded 206 essays in the development sample twice.

Additional Methods Information

We used the BERT-base-uncased model, which we obtained from huggingface's 'transformers' python library. See this link to the model hosted on the huggingface website. We used 4 training epochs, with 32 examples used to predict on before updating the weights in each iteration. See our settings in **Appendix A.**

We used the mice library in R to impute missing data (van Buuren & Groothuis-Oudshoorn, 2011). Multiple imputation was a better approach than full-information maximum likelihood estimation in this data because it produces complete datasets that can be analyzed using normal methods and then pooled. This makes it more flexible, as there is no need to use complex estimators that might be unavailable in certain software packages.

Descriptive Statistics

Table S1 shows means and standard deviations for the development sample, the holdout sample, and the robustness check sample.

Table S1

Descriptive Statistics for the Development, Holdout, and Robustness Check Samples

Variable	Development sample		Holdout sample		Robustness sample	
	M	*SD*	*M*	*SD*	*M*	*SD*
Personal qualities						
Leadership	0.18	0.38	0.22	0.35	0.21	0.34
Mastery orientation	0.42	0.49	0.49	0.43	0.49	0.43
Prosocial purpose	0.34	0.47	0.33	0.43	0.34	0.43
Teamwork	0.26	0.44	0.34	0.30	0.33	0.30
Goal pursuit	0.31	0.46	0.42	0.33	0.42	0.33
Self-concordant motivation	0.42	0.49	0.56	0.42	0.55	0.42
Perseverance	0.19	0.39	0.24	0.32	0.24	0.32
College graduation						
Four-year graduation	0.66	0.47	0.78	0.42	0.78	0.42
Six-year graduation	0.32	0.47	0.40	0.49	0.42	0.49
Traditional admissions criteria						
HSGPA	0.89	0.14	0.92	0.13	0.92	0.13
College admissions test scores	1693.08	306.58	1837.99	268.91	1826.41	267.90
Extracurricular activities						
Number of activities	3.46	2.29	5.37	1.86	5.16	1.98
Years per activity	2.11	1.13	2.57	0.72	2.53	0.76
Proportion sports	0.23	0.32	0.26	0.27	0.27	0.29
Race/ethnicity						
White	0.28	0.45	0.51	0.50	0.52	0.50
Latino	0.16	0.37	0.08	0.28	0.06	0.24
Black	0.16	0.36	0.05	0.22	0.05	0.23
Asian	0.19	0.39	0.10	0.30	0.10	0.30
Other	0.09	0.29	0.09	0.28	0.08	0.27
No race/ethnicity reported	0.12	0.32	0.17	0.37	0.19	0.39
Parental education						
No parent with college degree	0.52	0.50	0.27	0.45	0.26	0.44
One parent with college degree	0.21	0.41	0.24	0.43	0.24	0.43
Two parents with college degrees	0.27	0.45	0.48	0.50	0.50	0.50
High school type						
Non-Title I public school	0.42	0.49	0.19	0.39	0.16	0.36
Private school	0.00	0.00	0.34	0.47	0.33	0.47
Public school Title I	0.58	0.49	0.46	0.50	0.51	0.50
Homeschool	0.00	0.00	0.01	0.08	0.00	0.05
Other demographics						
Female	0.55	0.50	0.56	0.50	0.55	0.50
English language learner	0.26	0.44	0.12	0.33	0.12	0.32
Parents not married	0.34	0.53	0.23	0.58	0.22	0.59

Correlations in the Development, Holdout and Robustness Samples

Table S2, Table S3, and **Table S4** show correlations between computer generated likelihoods, traditional predictors of graduation, and 4-year and 6-year graduation from college in the holdout and robustness samples, respectively. In each table, zero-order bivariate Pearson correlations are below the diagonal. Above the diagonal, we present partial correlations controlling for demographic factors, including gender, parental education and marital status, English as a first language, ethnicity, and type of high school.

Table S2

Correlations in the Development Sample

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Leadership		-.04	-.01	.21***	.09**	-.15***	-.04	.07*	.07*	.12***	.09***	.00	.03	.04
2. Mastery orientation	-.03		-.09***	.02	-.04	-.03	.09**	.02	-.01	.05	.01	.04	.03	.05
3. Prosocial purpose	.00	-.09***		-.07†	.08*	-.08*	-.18***	.02	.02	.08**	-.03	-.10***	-.04	.03
4. Teamwork	.20***	.05**	-.06**		.06	-.02	.05	-.01	.03	.07*	.02	.06	.01	.01
5. Goal pursuit	.08***	-.03†	.09***	.07***		.02	.03	.02	.01	.07*	.04	.02	-.03	.03
6. Self-concordant motivation	-.13***	-.02	-.06***	.00	.04*		.08*	.01	.02	-.03	.03	.08*	.02	.04
7. Perseverance	-.03	.10***	-.15***	.04*	.04*	.07***		.06	.07†	.04	.05	.07†	.01	.00
8. HSGPA	.09***	.05*	.02	.03†	.02	.02	.07***		.45***	.27***	.11***	-.08*	.16***	.24***
9. College admissions test scores	.08***	.04†	-.01	.07***	.02	.02	.10***	.47***		.33***	.14***	-.17***	.14***	.24***
10. Number of activities	.14***	.08***	.08***	.12***	.09***	-.01	.06***	.32***	.36***		.18***	-.14***	.07†	.14***
11. Years per activity	.13***	.04*	-.01	.06***	.06**	.03†	.05**	.16***	.26***	.40***		.21***	.07†	.09**
12. Proportion sports	.02	.02	-.13***	.05**	.02	.06***	.05**	-.10***	-.11***	-.06**	.28***		.02	.01
13. 4-year graduation	.05**	.04*	-.03	.04*	.00	.02	.02	.20***	.22***	.14***	.11***	.00		.45***
14. 6-year graduation	.07***	.06***	.03	.04*	.04*	.03	.03	.28***	.30***	.22***	.17***	.00	.49***	

Note. Correlations above the diagonal control for demographic factors. Correlations below the diagonal are pairwise bivariate correlations.

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table S3

Correlations in the Holdout Sample

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Leadership		-.02**	-.01	.26***	.07***	-.16***	-.03***	.06***	.06***	.08***	.04***	-.02***	.02***	.04***
2. Mastery orientation	-.02***		-.14***	.05***	-.09***	-.03***	.11***	.00	-.02**	.02***	-.01	.02**	.01	.01
3. Prosocial purpose	-.01	-.13***		-.12***	.12***	-.12***	-.22***	.02***	.00	.09***	-.04***	-.07***	.01†	.03***
4. Teamwork	.26***	.05***	-.11***		-.01	-.06***	.06***	.03***	.03***	.06***	.00	.04***	.01	.02*
5. Goal pursuit	.07***	-.08***	.12***	-.01		.00	.00	.01	.01	.06***	-.01†	-.02***	.00	.02*
6. Self-concordant motivation	-.16***	-.03***	-.10***	-.06***	.00		.08***	-.01†	-.01	-.01	.05***	.04***	.00	.00
7. Perseverance	-.04***	.11***	-.21***	.06***	.00	.08***		.02***	.06***	.02***	.01	.08***	.00	.02*
8. HSGPA	.08***	.01	.03***	.04***	.01*	-.01*	.03***		.48***	.22***	.08***	-.15***	.12***	.19***
9. College admissions test scores	.06***	-.01†	-.01*	.05***	.02***	.00	.08***	.48***		.27***	.11***	-.24***	.09***	.16***
10. Number of activities	.09***	.03***	.10***	.07***	.07***	.01	.04***	.25***	.31***		-.09***	-.28***	.07***	.12***
11. Years per activity	.05***	-.01†	-.04***	.01*	-.01**	.06***	.01**	.10***	.16***	-.05***		.07***	.02*	.05***
12. Proportion sports	-.03***	.01*	-.10***	.03***	-.03***	.04***	.07***	-.18***	-.21***	-.30***	.08***		-.02***	-.04***
13. 4-year graduation	.03***	.01**	.02***	.02***	.01†	.00	.00	.15***	.12***	.09***	.04***	-.03***		.42***
14. 6-year graduation	.05***	.02***	.03***	.03***	.02***	.01*	.02***	.23***	.21***	.16***	.08***	-.05***	.44***	

Note. Correlations above the diagonal control for demographic factors. Correlations below the diagonal are pairwise bivariate correlations.

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table S4

Correlations in the Robustness Sample

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Leadership		-.02***	-.03***	.26***	.07***	-.14***	-.04***	.07***	.06***	.08***	.06***	-.02***	.03***	.04***
2. Mastery orientation	-.02***		-.13***	.05***	-.08***	-.03***	.10***	.00	-.02***	.03***	-.01**	.02***	.01***	.02***
3. Prosocial purpose	-.02***	-.12***		-.12***	.11***	-.09***	-.21***	.02***	.00	.09***	-.03***	-.07***	.02***	.03***
4. Teamwork	.26***	.05***	-.12***		.00	-.04***	.06***	.03***	.04***	.07***	.01***	.03***	.03***	.03***
5. Goal pursuit	.07***	-.08***	.11***	.00		.01***	.01*	.01***	.01***	.07***	-.01***	-.02***	.01***	.01***
6. Self-concordant motivation	-.14***	-.03***	-.09***	-.04***	.01***		.07***	-.02***	.00	.01**	.05***	.05***	.00	.00
7. Perseverance	-.04***	.11***	-.20***	.06***	.01***	.08***		.02***	.06***	.03***	.01***	.09***	.01**	.02***
8. HSGPA	.08***	.01	.03***	.04***	.01*	-.01*	.03***		.48***	.23***	.07***	-.16***	.13***	.20***
9. College admissions test scores	.07***	.00	-.01***	.06***	.02***	.01***	.08***	.48***		.28***	.10***	-.22***	.10***	.17***
10. Number of activities	.09***	.04***	.10***	.08***	.08***	.02***	.05***	.25***	.34***		.00	-.25***	.07***	.14***
11. Years per activity	.07***	.00	-.03***	.02***	-.01***	.06***	.02***	.10***	.14***	.03***		.09***	.02***	.05***
12. Proportion sports	-.03***	.01***	-.10***	.03***	-.03***	.04***	.08***	-.18***	-.21***	-.27***	.09***		-.02***	-.04***
13. 4-year graduation	.03***	.02***	.03***	.03***	.01***	.01**	.01***	.15***	.13***	.10***	.04***	-.03***		.45***
14. 6-year graduation	.05***	.02***	.04***	.05***	.02***	.01***	.03***	.23***	.22***	.18***	.09***	-.05***	.46***	

Note. Correlations above the diagonal control for demographic factors. Correlations below the diagonal are pairwise bivariate correlations.

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Do Manually Assigned Codes for Personal Qualities Predict Graduation?

As shown in **Table S5** and **S6,** the predictive validities of manually assigned codes for college graduation are slightly smaller in magnitude than those in the main model specification. This suggests that the effect sizes of computer-generated likelihoods to predict college graduation are not caused by a failure of the machine learning algorithm to adequately capture the predictive variance in the manually coded samples.

Table S5

Binary Logistic Regression Models Predicting 4-Year College Graduation from Manually Coded and Computer-Estimated Personal Qualities in the Development Sample

Term	Computer-estimated		Manually coded	
	OR	[95% CI]	*OR*	[95% CI]
Personal qualities				
Leadership	1.01	[0.99-1.03]	1.01	[0.99-1.03]
Mastery orientation	1.01	[0.99-1.03]	1.01	[1.00-1.03]
Prosocial purpose	0.98†	[0.96-1.00]	0.98	[0.97-1.00]
Teamwork	1.00	[0.98-1.03]	1.00	[0.98-1.02]
Goal pursuit	0.98†	[0.96-1.00]	0.97**	[0.96-0.99]
Self-concordant motivation	1.01	[0.99-1.03]	1.00	[0.98-1.02]
Perseverance	0.99	[0.97-1.01]	0.99	[0.97-1.01]
Traditional admissions criteria				
HSGPA	1.06***	[1.03-1.08]	1.06***	[1.03-1.08]
College admissions test scores	1.05***	[1.02-1.08]	1.05***	[1.02-1.08]
Extracurricular activities				
Number of activities	1.01	[0.98-1.03]	1.01	[0.98-1.03]
Years per activities	1.02	[0.99-1.04]	1.02	[0.99-1.04]
Proportion sports	1.02	[1.00-1.04]	1.02	[1.00-1.04]
Parents' education[a]				
Two parents with college degrees	1.02	[0.97-1.08]	1.02	[0.97-1.08]
One parent with college degree	1.00	[0.95-1.05]	1.00	[0.95-1.06]
Race/ethnicity[b]				
Latino	0.91**	[0.85-0.97]	0.91**	[0.85-0.97]
Black	0.92*	[0.87-0.98]	0.92*	[0.86-0.98]
Asian	0.96	[0.90-1.02]	0.96	[0.90-1.02]
Other	0.98	[0.92-1.05]	0.98	[0.92-1.05]
No race reported	0.99	[0.93-1.05]	0.99	[0.93-1.05]
Type of high school[c]				
Public school Title I	0.98	[0.94-1.02]	0.98	[0.94-1.02]
Other demographics				
Female[d]	1.11***	[1.06-1.15]	1.10***	[1.06-1.15]
Parents not married[e]	0.98	[0.94-1.02]	0.98	[0.94-1.02]
English language learner[f]	1.04	[0.98-1.10]	1.04	[0.98-1.10]
(Intercept)	1.36***	[1.29-1.44]	1.36***	[1.29-1.44]
AUC	.67		.67	
Scaled Brier Score	.07		.08	
Nagelkerke's R^2	.02		.02	

Note. To facilitate interpretation of odds ratios, all variables were standardized prior to the analysis. Since the development sample had complete data, these results are not based on imputed datasets.
[a]Reference group is two parents with college degrees. [b]Reference group is White. [c]Reference group is non-Title I high school. [d]Reference group is male. [e]Reference group is married parents. [f]Reference group is English as a first language.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table S6

Binary Logistic Regression Models Predicting 6-Year College Graduation From Manually Coded and Computer-Estimated Personal Qualities in the Development Sample

Term	Computer-estimated		Manually coded	
	OR	[95% CI]	*OR*	[95% CI]
Personal qualities				
Leadership	1.01	[0.99-1.03]	1.01	[0.99-1.03]
Mastery orientation	1.02**	[1.01-1.04]	1.02*	[1.00-1.04]
Prosocial purpose	1.01	[1.00-1.03]	1.01	[0.99-1.03]
Teamwork	1.00	[0.98-1.02]	0.99	[0.97-1.00]
Goal pursuit	1.01	[0.99-1.03]	1.00	[0.99-1.02]
Self-concordant motivation	1.02*	[1.00-1.04]	1.01	[0.99-1.03]
Perseverance	0.99	[0.97-1.01]	1.00	[0.98-1.01]
Traditional admissions criteria				
HSGPA	1.07***	[1.05-1.09]	1.07***	[1.05-1.09]
College admissions test scores	1.09***	[1.07-1.12]	1.09***	[1.07-1.12]
Extracurricular activities				
Number of activities	1.02*	[1.00-1.05]	1.02*	[1.00-1.05]
Years per activities	1.01	[0.99-1.04]	1.01	[0.99-1.04]
Proportion sports	1.02*	[1.00-1.04]	1.02*	[1.00-1.04]
Parents' education$_a$				
Two parents with college degrees	1.09***	[1.04-1.14]	1.09***	[1.04-1.14]
One parent with college degree	1.06*	[1.01-1.11]	1.06*	[1.01-1.11]
Race/ethnicity$_b$				
Latino	0.97	[0.92-1.04]	0.97	[0.91-1.04]
Black	1.04	[0.98-1.10]	1.04	[0.98-1.10]
Asian	1.02	[0.95-1.08]	1.02	[0.96-1.08]
Other	1.00	[0.94-1.07]	1.00	[0.94-1.07]
No race reported	0.96	[0.90-1.02]	0.96	[0.91-1.02]
Type of high school$_c$				
Public school Title I	1.03	[0.99-1.07]	1.03	[0.99-1.07]
Other demographics				
Female$_d$	1.06**	[1.02-1.10]	1.06**	[1.02-1.10]
Parents not married$_e$	0.99	[0.95-1.03]	0.99	[0.95-1.03]
English language learner$_f$	1.00	[0.95-1.06]	1.01	[0.96-1.06]
(Intercept)	1.85***	[1.75-1.94]	1.85***	[1.76-1.95]
AUC	.73		.73	
Scaled Brier Score	.14		.14	
Nagelkerke's R^2	.13		.13	

Note. To facilitate interpretation of odds ratios, all variables were standardized prior to the analysis. Since the development sample had complete data, these results are not based on imputed datasets.
$_a$Reference group is two parents with college degrees. $_b$Reference group is White. $_c$Reference group is non-Title I high school. $_d$Reference group is male. $_e$Reference group is married parents. $_f$Reference group is English as a first language.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

College Selectivity

To account for the possibility that college selectivity might affect our results, we ran logistic regression models that controlled for institutional graduation rates. As shown in **Table S7,** the effect sizes for personal qualities are virtually identical to those presented in the main text, suggesting that our results are not driven by college-level factors.

Table S7

Binary Logistic Regression Models Predicting 4- and 6-Year College Graduation From Computer-Estimated Personal Qualities in the Holdout and Robustness Samples Controlling for Institutional Graduation Rates

	4-year graduation holdout sample		4-year graduation robustness sample		6-year graduation holdout sample		6-year graduation robustness sample	
Term	*OR*	[95% CI]	*OR*	[95% CI]	*OR*	[95% CI]	*OR*	[95% CI]
Personal qualities								
Leadership	1.03**	[1.01-1.05]	1.03***	[1.02-1.04]	1.06***	[1.03-1.08]	1.05***	[1.04-1.07]
Mastery orientation	1.03**	[1.01-1.05]	1.03***	[1.02-1.03]	1.03*	[1.01-1.06]	1.04***	[1.03-1.05]
Prosocial purpose	1.02	[1.00-1.04]	1.04***	[1.03-1.04]	1.06***	[1.03-1.09]	1.07***	[1.06-1.08]
Teamwork	1.01	[0.99-1.03]	1.03***	[1.02-1.04]	0.99	[0.97-1.02]	1.03***	[1.02-1.04]
Goal pursuit	1.00	[0.98-1.02]	0.99	[0.99-1.00]	1.01	[0.99-1.04]	1.00	[0.99-1.01]
Self-concordant motivation	1.00	[0.98-1.02]	1.00	[0.99-1.01]	1.01	[0.99-1.04]	1.01	[1.00-1.02]
Perseverance	0.98*	[0.96-1.00]	0.99	[0.99-1.00]	1.01	[0.99-1.04]	1.00	[0.99-1.01]
Traditional admissions criteria								
HSGPA	1.23***	[1.20-1.26]			1.37***	[1.34-1.41]		
College admissions test scores	0.96*	[0.94-0.99]	1.05***	[1.04-1.06]	1.04*	[1.00-1.08]	1.19***	[1.17-1.20]
Extracurricular activities								
Number of activities	1.03*	[1.00-1.05]	1.04***	[1.03-1.05]	1.13***	[1.10-1.16]	1.17***	[1.16-1.18]
Years per activities	0.99	[0.97-1.01]	0.99*	[0.98-1.00]	1.07***	[1.04-1.09]	1.07***	[1.06-1.08]
Proportion sports	1.02	[0.99-1.04]	1.00	[0.99-1.01]	1.05***	[1.02-1.08]	1.03***	[1.02-1.04]
Parents' education$_a$								
No parents with college degrees	0.88***	[0.84-0.93]	0.99	[0.97-1.01]	0.73***	[0.69-0.78]	0.80***	[0.78-0.82]
One parent with college degree	1.02	[0.97-1.07]	1.09***	[1.07-1.11]	0.91**	[0.86-0.97]	0.95***	[0.92-0.97]
Race/ethnicity$_b$								

Term	4-year graduation holdout sample		4-year graduation robustness sample		6-year graduation holdout sample		6-year graduation robustness sample	
	OR	[95% CI]	*OR*	[95% CI]	*OR*	[95% CI]	*OR*	[95% CI]
Latino	0.79***	[0.73-0.86]	0.75***	[0.73-0.78]	0.90*	[0.82-1.00]	0.88***	[0.84-0.91]
Black	0.72***	[0.65-0.79]	0.67***	[0.65-0.70]	0.77***	[0.70-0.86]	0.77***	[0.74-0.80]
Asian	0.84***	[0.78-0.91]	0.81***	[0.79-0.84]	0.71***	[0.65-0.78]	0.69***	[0.67-0.72]
Other	0.84***	[0.78-0.91]	0.80***	[0.77-0.82]	0.73***	[0.67-0.80]	0.75***	[0.72-0.77]
No race reported	0.94*	[0.89-0.99]	0.98*	[0.96-1.00]	0.82***	[0.76-0.88]	0.84***	[0.82-0.86]
Type of high school[c]								
Private school	0.95*	[0.90-0.99]	0.92***	[0.90-0.93]	0.86***	[0.81-0.91]	0.75***	[0.73-0.76]
Public school Title 1	0.93*	[0.88-0.99]	0.87***	[0.85-0.89]	0.87***	[0.81-0.93]	0.86***	[0.84-0.89]
Homeschool	0.79†	[0.60-1.02]	0.61***	[0.52-0.71]	0.69*	[0.51-0.92]	0.49***	[0.42-0.57]
Other demographics								
Female[d]	1.22***	[1.17-1.27]	1.28***	[1.26-1.30]	1.42***	[1.35-1.50]	1.46***	[1.43-1.48]
Parents not married[e]	0.87***	[0.83-0.92]	0.87***	[0.85-0.88]	0.75***	[0.71-0.80]	0.77***	[0.76-0.79]
English language learner[f]	0.85***	[0.79-0.92]	0.92***	[0.89-0.94]	0.64***	[0.59-0.69]	0.70***	[0.67-0.72]
Institutional graduation rates	1.56***	[1.52-1.60]	1.61***	[1.60-1.63]	1.74***	[1.69-1.79]	1.80***	[1.78-1.83]
(Intercept)	0.70***	[0.66-0.73]	0.73***	[0.72-0.74]	5.29***	[4.98-5.63]	5.19***	[5.07-5.31]
AUC	.65		.65		.74		.74	
Scaled Brier Score	.07		.07		.14		.13	
Nagelkerke's R^2	.09		.09		.19		.19	

Note. To facilitate interpretation of odds ratios, all variables were standardized prior to the analysis. Results are pooled estimates from *m* = 25 multiply imputed datasets.

[a]Reference group is two parents with college degrees. [b]Reference group is White. [c]Reference group is non-Title I high school. [d]Reference group is male. [e]Reference group is married parents. [f]Reference group is English as a first language.

† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Clustering

As a robustness check, we fit random-intercept multilevel models predicting 4-year and 6-year graduation, respectively, nesting students within college and estimating a free intercept for each college. Because models failed to converge at lower numbers of students per college, these models were limited to the subsample of colleges for which there were 40 enrolled students, reducing the sample by about 41% to n = 25,835 students nested within k = 235 universities. As shown in **Table S8**, parameter estimates in these robustness models were largely consistent with our main models but, in this smaller sample, confidence intervals were wider.

Table S8

Multilevel Logistic Regression Models Predicting 4- and 6-Year College Graduation From Computer-Estimated Personal Qualities in the Holdout Sample

Term	4-year graduation		6-year graduation	
	OR	[95% CI]	*OR*	[95% CI]
Personal qualities				
Leadership	1.02	[0.99 - 1.05]	1.04*	[1.00 - 1.08]
Mastery orientation	1.03*	[1.00 - 1.05]	1.05**	[1.01 - 1.09]
Prosocial purpose	1.00	[0.98 - 1.03]	1.08***	[1.04 - 1.12]
Teamwork	1.01	[0.98 - 1.04]	1.02	[0.98 - 1.06]
Goal pursuit	1.00	[0.97 - 1.02]	1.00	[0.96 - 1.03]
Self-concordant motivation	0.99	[0.97 - 1.02]	1.02	[0.98 - 1.06]
Perseverance	0.96**	[0.94 - 0.99]	1.00	[0.96 - 1.04]
Traditional admissions criteria				
HSGPA	1.25***	[1.21 - 1.29]	1.32***	[1.26 - 1.38]
College admissions test scores	1.01	[0.97 - 1.05]	1.17***	[1.10 - 1.24]
Extracurricular activities				
Number of activities	1.03†	[1.00 - 1.06]	1.13***	[1.09 - 1.18]
Years per activities	0.99	[0.96 - 1.01]	1.08***	[1.04 - 1.12]
Proportion sports	1.03	[0.99 - 1.06]	1.08***	[1.03 - 1.13]
Parents' education[a]				
No parents with college degrees	0.99	[0.92 - 1.07]	0.76***	[0.69 - 0.84]
One parent with college degree	1.05	[0.98 - 1.12]	0.92	[0.84 - 1.02]
Race/ethnicity[b]				
Latino	0.88*	[0.78 - 0.98]	0.99	[0.84 - 1.16]
Black	0.75***	[0.65 - 0.86]	0.73***	[0.60 - 0.88]
Asian	0.86**	[0.79 - 0.95]	0.69***	[0.60 - 0.79]
Other	0.85***	[0.77 - 0.93]	0.73***	[0.64 - 0.84]
No race reported	0.87***	[0.81 - 0.94]	0.73***	[0.66 - 0.81]
Type of high school[c]				
Private school	0.83***	[0.78 - 0.89]	0.71***	[0.64 - 0.78]
Public school Title 1	0.95	[0.88 - 1.03]	0.96	[0.84 - 1.08]
Homeschool	1.06	[0.70 - 1.61]	0.96	[0.51 - 1.79]
Other demographics				
Female[d]	1.18***	[1.12 - 1.24]	1.54***	[1.43 - 1.67]
Parents not married[e]	0.92**	[0.86 - 0.98]	0.74***	[0.68 - 0.82]
English language learner[f]	0.82***	[0.75 - 0.90]	0.54***	[0.48 - 0.60]
(Intercept)	1.01	[0.91 - 1.12]	9.63***	[8.30 - 11.18]

Note. To facilitate interpretation of odds ratios, all variables were standardized prior to the analysis. Results are pooled from m = 25 imputed datasets.

[a]Reference group is two parents with college degrees. [b]Reference group is White. [c]Reference group is non-Title I high school. [d]Reference group is male. [e]Reference group is married parents. [f]Reference group is English as a first language.

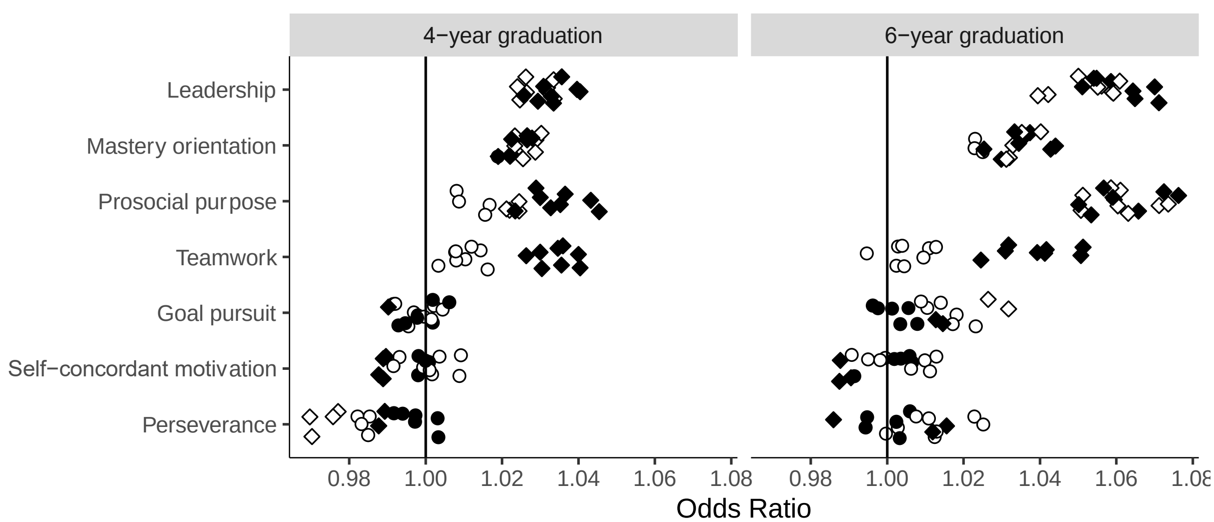
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Robustness Analyses

As a robustness check, we fit a series of binary logistic regression models with different specifications: entering personal quality separately or simultaneously (8 models), with either 4- or 6-year graduation as the outcome, either controlling for institutional graduation rates or not, in either the holdout sample or the robustness check sample, and with or without imputation for missing data. This resulted in the estimation of 128 independent models, which produced 32 sets of estimates for each of the seven personal qualities. As shown in **Figure S1** and consistent with results shown in the main text, leadership, mastery orientation, and prosocial purpose robustly predicted college graduation.

Figure S1

Distributions of Odds Ratios for Personal Qualities in 32 Model Specifications



Note. Filled points correspond to the robustness sample, open points correspond to the holdout sample. Circles indicate non-significant estimates, diamonds represent significant estimates.

References

Authors et al. (2019)

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*(3), 1 - 67. doi:http://dx.doi.org/10.18637/jss.v045.i03

Appendix A: BERT settings python code

```
# Modify this file as needed
import os

from sklearn.ensemble import RandomForestClassifier
from sklearn.feature_selection import SelectKBest
from sklearn.naive_bayes import MultinomialNB
from imblearn.combine import SMOTEENN
from imblearn.over_sampling import RandomOverSampler, ADASYN
from imblearn.under_sampling import RandomUnderSampler

from src.configuration.settings_template import Settings, SettingsEnumOptions
from src.models.nets.windows_of_context_rnn import WindowRnn
from src.pipeline.resampling import DatasetSampler
from src.pipeline.transformers.empty_transformer import EmptyTransformer
from src.pipeline.transformers.standard_scaler_3d import StandardScaler3D
from src.common.meta import MetaUtils

from src.metrics.custom_scorers import PearsonCorrelationScorer



# Path to the data file for input
Settings.IO.DATA_INPUT_FILE = "common_app_new.csv"


# Folder to output results in, helpful to change if you don't want to overwrite previous results
Settings.IO.RESULTS_OUTPUT_FOLDER = "bert_results_multilabel"

Settings.SAVE_MODELS = True

# Names of columns in spreadsheet to identify what should be the input data, and what should be the
predicted labels
Settings.COLUMNS.IDENTIFIER = "applicantprofileid"

Settings.COLUMNS.MAKE_ALL_LABELS_BINARY = True
Settings.COLUMNS.Y_LABELS_TO_PREDICT = [
  "multilabel"
]

Settings.FEATURE_INPUT_SOURCES_TO_RUN = [
  SettingsEnumOptions.LanguageFeatureInput.with_language_from_column(
    "response")
]

Settings.BERT_FEATURES.sentence_column_name = "response"

# Same team from same university cannot be in both train and test folds.
# Group ID is created using team and school
# Settings.COLUMNS.GROUP_BY_COLUMN = "GROUPID"
Settings.PREDICTION = SettingsEnumOptions.Prediction.CLASSIFICATION
```

```
label_list = ["Type_bin", "Accolades_bin", "Connection_bin", "Goal_bin",
          "Goal_r_bin", "Leadership_bin", "Learning_bin", "Persevere_bin", "Selftrans_bin", "Team_bin"]

bert_labels = {
  "multilabel": {
    "is_multilabel": True,
    "num_labels": 10,
    "label_list": label_list,
  },
}

for label in label_list:
  bert_labels[label] = {
    "is_multilabel": False,
    "num_labels": 1,
    "convert_to_onehot": True,
  }

Settings.COLUMNS.BERT_LABELS = bert_labels



# Settings.COLUMNS.ORDER_IN_GROUPS_BY_COLUMN = 'ObsID'
# Settings.COLUMNS.ORDER_IN_GROUPS_SORT_BY_COLUMN = 'UtteranceID'


#USE_ONE_VS_ALL_CLF_FOR_MULTICLASS = True

# ------------------------------------------------------------------------------------------------------------


# Settings relating to the models to be run and the parameters to be cross validated
# ------------------------------------------------------------------------------------------------------------
# Classes of the models to be run

# c = MetaUtils.get_dynamic_class(WindowRnn, {'window_size': 12})
# WindowRnn] # [HierarchicalClassifier |with_base| RandomForestClassifier] # models # models
#WindowRnn] #, rnn.Gru] #rnn.Gru] #RandomForestClassifier] #gru12, gru15, gru31, gru44, gru67] #,
rnn.Gru, rnn.Lstm]
Settings.MODELS_TO_RUN = [RandomForestClassifier]

# Number of folds to use in outer loop to split all data into train / test
Settings.CROSS_VALIDATION.NUM_TRAIN_TEST_FOLDS = 10
# Number of folds to use in nested cross validation to split train data into train / validation
Settings.CROSS_VALIDATION.NUM_CV_TRAIN_VAL_FOLDS = 5

Settings.CROSS_VALIDATION.SCORING_FUNCTION = 'roc_auc'


# Configure BERT parameters here
# epochs: number of training epochs
# batch_size: Number of examples to predict on before updating weights in each iteration.
```

```
# max_seq_len: The number of tokens to truncate/pad all sentences to.
Settings.CROSS_VALIDATION.HYPER_PARAMS.BERT = {
    'epochs': 4,
    'batch_size': 32,
    'max_seq_len': 256
}




Settings.RANDOM_STATE = 42
```