Research Report
2024-12

Predicting Academic Success in College: The Comparative Strength of High School GPA, ACT® Score, and Demographic Factors

EDGAR I. SANCHEZ, PHD

Conclusions

This study offers a novel approach to evaluating the predictors of first-year grade point average (FYGPA) in college by using multiple imputation and dominance analysis to compare high school GPA (HSGPA), ACT® test score, and demographic factors. Unlike previous research that typically relied on single regression models, this study reveals more nuanced relationships between predictors. A key differentiator is the finding that neither HSGPA nor ACT Composite score consistently dominated across all models, challenging prior conclusions about their relative importance. The analysis also highlights the dominant predictive roles of English and math performance in high school. Additionally, this study confirms that academic achievement measures are far stronger predictors of FYGPA than demographic variables, with family income and gender being the least important.

So What?

These findings are important because they provide a more nuanced understanding of the factors that predict college success, which can have significant implications for how students are admitted and how they are supported once enrolled. By demonstrating that high HSGPA and ACT scores are stronger predictors of FYGPA than demographic factors, the study reinforces the importance of academic readiness for student outcomes. Also, the emphasis on English and math performance as key predictors of success underscores the value of strong foundational skills, which can inform curriculum design and high school preparation programs.

Now What?

This study has important implications for educational institutions, students, and policymakers. For these institutions, it suggests that admissions processes should prioritize academic achievement measures such as HSGPA and ACT scores over demographic factors, highlighting the stronger predictive power of academic readiness: The nuanced findings can help institutions refine admissions strategies and develop more data-driven student support programs, especially for students with discrepancies between their ACT scores and HSGPAs. For students, the study emphasizes the importance of strong preparation in key subjects such as English and math, which emerged as the dominant predictors of first-year success. For policymakers, these insights can help them advocate for a balanced, holistic admissions approach that considers both standardized testing and continuous GPA assessment, ensuring more equitable outcomes for students and better alignment with their needs in higher education.

About the Author

Edgar I. Sanchez, PhD

Dr. Sanchez is a lead research scientist at ACT where he studies postsecondary admissions, national testing programs, test preparation efficacy, and intervention effectiveness. Throughout his career, Dr. Sanchez has focused on studying the transition between high school and college and supporting the decision-making capacity of college administrators, students, and their families. His research has been widely cited in academic literature and by the media, including *The Wall Street Journal*, *The Washington Post*, *USA Today*, and the education trade press.

Acknowledgements

The author would like to thank Jeff Allen, Joann Moore, Nola Daley, and Meg Hillie for their comments on previous drafts of this report.

Introduction

High school students transitioning to college have reached a turning point in their academic journey, a point often marked by significant changes in the expectations for their educational performance. Accurately predicting the first-year grade point average (FYGPA) of students entering college is crucial for admissions decisions, academic advising, and student support services. Colleges commonly predict FYGPA from both high school GPA (HSGPA) and standardized test scores, such as the ACT® and SAT. However, the relative strength and utility of these predictors has been the subject of extensive research and debate.

Recent research about the relationship between HSGPA and ACT Composite (ACTC) scores and degree completion with FYGPA as a mediator indicate that both achievement measures are direct and indirect predictors of graduation by the fourth or sixth year after college enrollment. Importantly, Loria and Sanchez (in press) found that HSGPA was a stronger predictor of FYGPA than ACT score at predicting whether students will graduate from college by their fourth or sixth years. Sanchez (2024) also found that HSGPA was a stronger predictor of FYGPA than ACT Composite score when both measures were used together to predict FYGPA. This same study further highlighted that, while the predictive validity of HSGPA has shifted over time, the predictive validity of ACT Composite score has remained relatively stable. The instability of HSGPA is potentially a result of grade inflation that has been observed in the past decade (Sanchez & Moore, 2022).

Consistently, research has demonstrated that HSGPA is a robust predictor of FYGPA. A comprehensive meta-analysis by Westrick et al. (2015) concluded that HSGPA and ACT scores were both highly correlated with FYGPA, with HSGPA showing slightly higher predictive validity. Along the same line of research, Noble and Sawyer (2004) found that HSGPA was a better predictor for moderate levels of FYGPA, whereas ACT scores were more effective at predicting higher levels of academic performance.

Kobrin et al. (2002) demonstrated that SAT scores add significant incremental validity to the prediction of FYGPA, in particular for students with discrepant HSGPAs and SAT scores, which indicates the compensatory nature of both achievement measures. Research making use of the ACT has found similar results to those found by Kobrin et al. (Sanchez & Mattern, 2018).

Zwick and Sklar (2005) found that, while both HSGPA and SAT scores predict FYGPA, the effectiveness of these measures as predictors of FYGPA differed across racial/ethnic and language groups, with HSGPA generally providing stronger predictions. Furthermore, Sawyer (2013) highlighted that HSGPA is particularly useful in less selective institutions where admissions criteria may be broader, whereas ACT test scores offered greater incremental predictive value for highly selective institutions.

Prior research has indicated that early success in college as measured by indicators such as FYGPA is strongly linked to successful and timely degree completion (Demeter et al., 2022; Gershenfeld et al., 2016). This link between FYGPA and long-term college success highlights the importance of being able to accurately predict FYGPA. Together, HSGPA and standardized test scores have historically been found to be significant predictors of FYGPA (Marini et al., 2019; Beard & Marini, 2018; Curabay, 2016; Westrick et al., 2015; Warren & Goins, 2019). Newer research has further indicated that considering test scores and HSGPA together results in greater accuracy in predicting FYGPA than when using either measure alone (Sanchez, 2024).

Consistent in all previously mentioned research is the use of regression models that involve predicting an outcome such as FYGPA from a set of predictors (e.g., gender, HSGPA, ACT Composite score) entered in a specific order. This type of methodology focuses on model fit, coefficient estimates, and predictive power. In this methodology, beta weights are typically used to identify the relative strength of the predictors and whether HSGPA or ACT Composite score is the stronger predictor of the model outcome.

In the current study, I make use of dominance analysis, which is a technique for comparing the relative importance of predictors in the regression model by evaluating their contribution to the model's predictive power in terms of variance explained across all possible subset models. There are three key characteristics to this type of analysis: The first is using subset models in which the outcome of the model is predicted with every possible subset of the predictors. The second is examining the incremental contribution of each predictor by assessing how much additional variance each predictor explains when added to subsets of the other predictors. The third is applying dominance hierarchy in which the predictors are ranked according to their overall contribution to the model's predictive power across all possible subset models.

There are important advantages to choosing dominance analysis over a traditional regression methodology. The first is being able to assess the comprehensive importance of the predictors: The traditional regression model assesses the importance of the predictors within the context of one specific model, whereas dominance analysis is more comprehensive because it considers all possible combinations of predictors, which helps to identify the true importance of each predictor. The second advantage of dominance analysis is how it inherently handles multicollinearity: In the traditional model, highly correlated predictors may cause multicollinearity, which may lead to unreliable coefficient estimates, but dominance analysis instead evaluates predictors across multiple subset models, therefore mitigating the effect of multicollinearity and offering a more robust measure of predictor importance. The third advantage includes the robustness of model specification: In the traditional model, the importance of predictors may vary significantly depending upon which other specific predictors are included or omitted, whereas dominance analysis provides a stable measure of importance that is less sensitive to the specific model used because dominance analysis aggregates information across many model subsets. The fourth advantage is that the dominance analysis model allows for enhanced interpretability: While the traditional regression model may offer limited interpretability, especially in complex models, dominance analysis provides a clear and comprehensive ranking of the predictors that makes it easier to interpret their relative importance. Although the traditional model may offer a more straightforward approach to assessing predictor importance, the collective advantages of dominance analysis show that this method provides a more robust, comprehensive, and reliable evaluation because it considers the contribution of predictors across all possible models.

In addition to dominance analysis, I used the multiple imputation method to address the limitations of missing data. Because sets of educational data typically are missing some data, multiple imputation offers a valuable way to use the existing data to address any gaps. This methodology is described in the Methods section of this paper.

In the present study, I conducted two dominance analyses to examine the relative importance of high school achievement on FYGPA in conjunction with student demographic information. In the first analysis, HSGPA and ACT Composite score in conjunction with student demographics (family income, race/ethnicity, and gender) were evaluated for their relative importance in predicting FYGPA. In the second analysis, high school subject GPAs (English, math, social studies, and natural science), ACT section scores (English, mathematics, reading, and science), and student demographics (family income, race/ethnicity, and gender) were evaluated for their relative importance in predicting FYGPA. Based on previous research about these relationships, I expected that HSGPA would emerge as a moderately more dominant predictor of FYGPA than ACT Composite score and that both together would be more dominant predictors of FYGPA than student demographics. This study provides a first look at the relative importance of high school subject GPAs and ACT section test scores on FYGPA. I considered two research questions:

1. What are the dominant predictors of FYGPA when comparing HSGPA, ACT Composite score, family income, race/ethnicity, and gender?

2. What are the dominant predictors of FYGPA when comparing high school subject GPAs, ACT section test scores, family income, race/ethnicity, and gender?

Methods

Analytical Sample

For this study, the analytical sample was drawn from a southern state in the United States that administers the ACT test to all its public high school 11th graders. In such statewide ACT adoption contexts, nearly all public high school graduates in the state take the ACT. The sample was limited to ACT-tested public high school graduates of 2021 who enrolled in a public 4-year college in the same state in the fall immediately after high school graduation. The sample includes 7,924 students across 10 institutions. These students represented 25% of all public high school ACT-tested students in the state in 2021 and 70% of ACT-tested graduates who enrolled at in-state public colleges in the fall immediately after high school graduation.

Measures

ACT Composite and Section Scores

The official ACT Composite and the English, math, reading, and science section scores were obtained from the last ACT test that students took before high school graduation. These scores may have been obtained either during statewide school-day testing or during a National test administration.

Cumulative HSGPA

Self-reported grades in up to 23 courses in English, mathematics, social studies, and natural science were averaged to calculate each student's HSGPA. Sanchez and Buddin (2016) showed that students' self-reported HSGPA is highly correlated with students' transcript GPA, and other research supports the use of self-reported data for research purposes (Camara et al., 2003; Kuncel et al., 2005; Shaw & Mattern, 2009).

English, Math, Natural Science, and Social Studies High School Subject GPA

Self-reported high school grades in up to 4 English, 7 math, 6 natural science, and 7 social studies courses were averaged to calculate each student's cumulative GPA for each subject.

Official FYGPA

Official FYGPA was obtained from student transcripts at the colleges where students enrolled immediately after high school.

Demographic Variables

The study considered three demographic variables: gender, race/ethnicity, and family income. All demographic information is provided in Table 2.

Students self-reported their gender as male, female, another gender, or prefer not to respond; or they gave no response. For this analysis, we combined the data of the students who identified as another gender (0.2%), preferred not to respond (0.5%), or did not provide a response (5.7%) into a single group.

Students selected their racial/ethnic identity as Asian, Black, Hispanic, American Indian/Alaska Native, Native Hawaiian/Pacific Islander, White, two or more races, or prefer not to respond; or they gave no response. Because some groups had low group numbers overall, for this analysis we combined the data of the students who identified as American Indian/Alaska Native (0.4%), Native Hawaiian/Pacific Islander (< 0.1%), and two or more races (4.7%).

Students self-reported their family income in four groups: below $36,000, $36,000–$60,000, $60,000–$100,000, and above $100,000. I also include a missing response category for income (25.2%).

Data Analysis

This study uses both the multiple imputation methodology and the dominance analysis methodology. Multiple imputation helps to impute missing cumulative and high school subject GPA values. As shown in Table 1, HSGPA values were missing from approximately 17–18% of the sample.

Table 1. Missing Percentages by GPA

GPA	% Missing
Cumulative GPA (HSGPA)	17.71
English GPA	17.37
Math GPA	17.62
Natural science GPA	18.07
Social studies GPA	17.83

In this study, I used both the MICE package for imputation and the domir package in R for dominance analysis, following this process to use them together:

First, to handle missing data, I used the MICE, or Multiple Imputation by Chained Equations, method (van Buuren & Groothuis-Oudshoorn, 2011). Using the MICE package, I generated 20 imputed datasets, which accounts for the uncertainty associated with missing data. I used predictive mean matching as the imputation method, which helps to ensure that realistic values are imputed based on all observed data. When conducting the multiple imputation, I used the following predictors: test type (national or school day); FYGPA; HSGPA; English, math, natural science, and social studies GPAs; ACT Composite, English, math, reading, and science scores; and student demographics (gender, race/ethnicity, and family income).

Subsequently, I used the domir package to perform dominance analysis with each of the 20 imputed datasets to evaluate the relative importance of the predictors (Luchman, 2024). The predictors analyzed in the dominance analysis included HSGPA (overall and high school subject specific), ACT score (Composite, English, math, reading, and science), family income, race/ethnicity, and gender. Then I fit regression models for each of the imputed datasets and conducted the dominance analysis to determine the contribution of each predictor to FYGPA, the outcome of interest.

After conducting dominance analysis on each of the 20 imputed datasets, I combined the results to obtain comprehensive dominance statistics and metrics across all imputations. I examined three types of dominance. In the hierarchy of dominance, (a) general dominance is the weakest form, followed by (b) conditional dominance, and finally by (c) complete dominance, the strongest form. Stronger forms of dominance imply all weaker forms of dominance.

Consistent with Rubin's rules for combining statistics across imputations (Rubin, 1987), I calculated the mean statistic provided across the 20 datasets. For general dominance, the overall contribution of each predictor across all imputed datasets was averaged to provide the overall general dominance (the weakest of the three forms of dominance). The general dominance statistics are the average of all conditional dominance statistics for a predictor.

For conditional dominance, I used a similar process in which conditional dominance for each imputed dataset was averaged to provide the overall conditional dominance. Conditional dominance statistics compare R^2 contributions between predictors for various combinations of models with differing numbers of predictors. For conditional dominance to be established, all R^2 values for a given predictor should be larger than all R^2 values for another predictor across subset models.

Complete dominance compares each predictor across all possible subset models, resulting in a categorical True/False determination of whether a given predictor is dominant. Complete dominance is established if a given predictor is dominant (i.e., has the largest R^2 value) across all possible subset models and, as such, is the strongest form of dominance. Rubin's rules do not explain how to combine categorical data across imputations, so in this study I used both the mode and the number of results instances across all 20 imputations to assess the frequency of categorical results. For the complete dominance results, I report the mode result (e.g., see Table 6). For the strongest level of dominance attained, I report the frequency of the dominance measures across the 20 imputed datasets (e.g., see Table 7).

Results

Descriptive Statistics

As shown in Table 2, the analyzed ACT records were predominantly from National testing (66%). The sample was also predominantly White (61%). The largest group of students (23%) came from families with a family income greater than $100,000. The sample was also predominantly female (55%). Table 3 shows that there were only minor differences between HSGPA and high school subject GPAs as well as between the ACT Composite and four section test scores among students whose test administration was from National testing or school-day testing. The FYGPA for students who tested during National testing was higher than the FYGPA for students from school-day testing (2.96 vs. 2.59, respectively).

Table 2. Descriptive Statistics for the Analytical Sample

Characteristics	Level	National testing (%)	School-day testing (%)	Overall (%)
Sample size		4,795 (66)	2,489 (34)	7,284 (100)
Race/ethnicity	African American	638 (13.3)	391 (15.7)	1029 (14.1)
	American Indian/ Alaska Native	17 (0.4)	10 (0.4)	27 (0.4)
	Asian	165 (3.4)	48 (1.9)	213 (2.9)
	Hispanic	482 (10.1)	202 (8.1)	684 (9.4)
	Native Hawaiian/ Pacific Islander	1 (< 0.1)	2 (0.1)	3 (< 0.1)
	Prefer not to respond/ missing	103 (2.1)	475 (19.1)	578 (7.9)
	Two or more races	229 (4.8)	114 (4.6)	343 (4.7)
	White	3,160 (65.9)	1,247 (50.1)	4,407 (60.5)
Family income	< $36K	983 (20.5)	321 (12.9)	1,304 (17.9)
	$36K–$60K	823 (17.2)	226 (9.1)	1,049 (14.4)
	$60K–$100K	1,115 (23.3)	292 (11.7)	1,407 (19.3)
	> $100K	1,448 (30.2)	240 (9.6)	1,688 (23.2)
	Missing	426 (8.9)	1,410 (56.6)	1,836 (25.2)
Gender	Female	2,874 (59.9)	1,106 (44.4)	3,980 (54.6)
	Male	1,897 (39.6)	946 (38.0)	2,843 (39.0)
	Another gender	5 (0.1)	7 (0.3)	12 (0.2)
	Prefer not to respond	16 (0.3)	18 (0.7)	34 (0.5)
	Missing	3 (0.1)	412 (16.6)	415 (5.7)

Table 3. Means and Standard Deviations for the Analytical Sample

Achievement Indicators	National testing mean (SD)	School-day testing mean (SD)	Overall mean (SD)
HSGPA	3.61 (0.4)	3.53 (0.5)	3.59 (0.4)
English GPA	3.63 (0.5)	3.52 (0.6)	3.61 (0.5)
Math GPA	3.52 (0.5)	3.44 (0.6)	3.50 (0.6)
Natural science GPA	3.57 (0.5)	3.51 (0.6)	3.55 (0.5)
Social studies GPA	3.70 (0.4)	3.64 (0.5)	3.69 (0.4)
ACT Composite	21.91 (5.0)	21.89 (5.4)	21.90 (5.1)
ACT English	22.20 (6.1)	22.01 (6.6)	22.14 (6.3)
ACT math	20.62 (5.0)	20.57 (5.1)	20.60 (5.0)
ACT reading	22.50 (6.3)	22.47 (6.6)	22.49 (6.4)
ACT science	21.80 (4.8)	21.97 (5.4)	21.86 (5.0)
FYGPA	2.96 (1.0)	2.59 (1.2)	2.84 (1.1)

What are the dominant predictors of FYGPA when comparing HSGPA, ACT Composite score, family income, race/ethnicity, and gender?

General Dominance Results

Table 4 displays the general dominance statistics among predictors. While general dominance is the weakest of the three forms of dominance because it does not require a greater contribution to R^2 in all subset models, its statistics provide a useful ranking of predictor strength. The general dominance statistics are the average of all conditional dominance statistics for a predictor. In this case with multiple imputations, it is the average R^2 contribution across 20 imputations. For example, ACT Composite (ACTC) score and HSGPA have the highest R^2 contribution among all predictors and tie for being the most dominant predictor in this table. Following ACTC and HSGPA, in order of R^2 contribution and rank, are race/ethnicity, family income, and gender. Both ACTC and HSGPA have notably higher R^2 contributions to the prediction of FYGPA than any of the demographic characteristics. Race/ethnicity contributed slightly more than either family income or gender to the percentage of variance explained in FYGPA. Finally, family income contributed slightly more than gender to the percentage of variance explained in FYGPA.

Table 4. General Dominance Statistics

Predictor	R^2 contribution	Rank
ACTC	0.11	1
HSGPA	0.11	1
Race/ethnicity	0.03	3
Family income	0.02	4
Gender	0.01	5

Conditional Dominance Results

Table 5 shows the conditional dominance statistics among predictors for various combinations of models with differing numbers of predictors. In this table, each cell represents the average (across 20 imputations) of the incremental R^2—or the difference between each model containing the row predictor and a comparable model not containing the row predictor—by the number of predictors in the model. For example, the column labeled 1 shows the average R^2 contribution for each predictor if it were the only predictor in the model. Similarly, the column labeled 5 shows the average R^2 contribution for each row predictor if all 5 predictors were included in the model (versus a model without that predictor). To have conditional dominance, all row values for a predictor must be larger than all row values for another predictor. Once again, with imputation, the R^2 contributions in the table are the average across 20 imputations. For example, ACTC dominates HSGPA in models containing 1, 2, 3, or 5 predictors but not in the models containing 4 predictors; therefore, conditional dominance was not established between ACTC and HSGPA. These statistics show that ACTC and HSGPA conditionally dominate family income, race/ethnicity, and gender as predictors of FYGPA. No clear conditional dominance was determined between family income and gender. Race/ethnicity conditionally dominated family income.

Table 5. Conditional Dominance Statistics

Predictor	Number of predictors in the model				
	1	2	3	4	5
HSGPA	0.171	0.133	0.103	0.080	0.060
ACTC	0.175	0.136	0.104	0.080	0.061
Family income	0.040	0.023	0.013	0.008	0.005
Race/ethnicity	0.062	0.037	0.021	0.011	0.007
Gender	0.015	0.016	0.015	0.014	0.012

Complete Dominance Results

Table 6 contains results of the analysis of complete dominance between predictors. A predictor demonstrates complete dominance over another predictor if it contributes more explained variance in the dependent variable in every possible subset of regression models in which both predictors are included. This means that, for a predictor to have complete dominance over another predictor, its inclusion must result in a higher R^2 value in every model comparison. This table should be read first by row and then by column. For example, it shows that HSGPA and ACTC both completely dominated family income, race/ethnicity, and gender. N/A indicates where complete dominance could not be determined, which may result from several scenarios. One such example is an ambiguity in complete dominance where the data may show some level of dominance but not a total or consistent dominance or where the relationship is not strong enough to classify as complete dominance. Among the predictors of family income, race/ethnicity, and gender, there were no cases of complete dominance. Similarly, between HSGPA and ACTC, neither variable completely dominated the other.

Table 6. Complete Dominance Statistics

	Predictor	Dominated by				
		HSGPA	ACTC	Family income	Race/ethnicity	Gender
Dominates	HSGPA	—	N/A	True	True	True
	ACTC	N/A	—	True	True	True
	Family income	False	False	—	N/A	N/A
	Race/ethnicity	False	False	N/A	—	N/A
	Gender	False	False	N/A	N/A	—

Note. N/A indicates that complete dominance could not be determined.

Strongest Level of Dominance Results

Table 7 contains the strongest level of dominance found for each pair of predictors and the number of imputations for which the relationship held. This table shows that both ACTC and HSGPA completely dominated the race/ethnicity, gender, and family income predictors. Additionally, the relationship between HSGPA and ACTC was mixed across the 20 imputations. According to the modal strongest level of dominance achieved, HSGPA completely dominated ACTC in six of the 20 imputations. As shown in the previous dominance tables, however, no clear dominance was determined between ACTC and HSGPA. Table 7 presents the uncertainty that such a conclusion implies: HSGPA dominated ACTC to some degree in 11 instances, and ACTC dominated HSGPA to some degree in nine instances. Furthermore, race/ethnicity generally dominated gender in all 20 imputations. Family income was conditionally dominated by race/ethnicity in 14 imputations, and family income generally dominated gender in 18 imputations. Table 7 and the previous tables do not make it clear whether ACTC or HSGPA was the most important predictor of FYGPA, but it is clear that ACTC and HSGPA were the two most important predictors, followed distantly by student demographics.

Table 7. Strongest Level of Dominance Achieved

Predictor1	Relationship	Predictor2	No. of imputations
ACTC	completely dominates	race/ethnicity	20
	completely dominates	gender	20
	completely dominates	family income	20
Race/ethnicity	generally dominates	gender	20
HSGPA	completely dominates	ACTC	6
	conditionally dominates		4
	is completely dominated by		4
	is conditionally dominated by		4
	generally dominates		1
	is generally dominated by		1
	completely dominates	race/ethnicity	20
	completely dominates	gender	20
	completely dominates	family income	20
Family income	is conditionally dominated by	race/ethnicity	14
	is generally dominated by		6
	generally dominates	gender	18
	conditionally dominates		1
	is generally dominated by		1

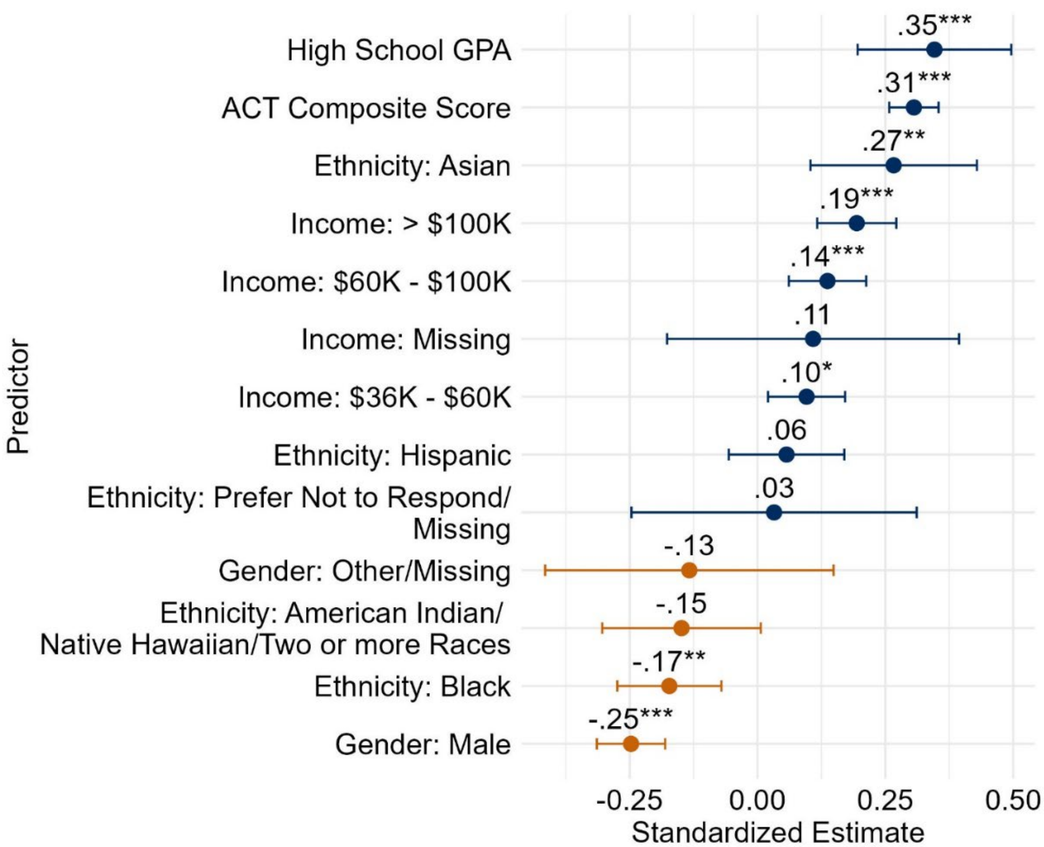
Comparison of Multiply Imputed Dominance Results and Regression With Multiply Imputed Data

Contrary to what has been generally found in studies using regression analysis to predict FYGPA from HSGPA and ACTC, in this study I found that ACTC was as similarly predictive of FYGPA as HSGPA. This raises the question of how these results compare to a linear regression analysis with multiply imputed data. To explore this further, I conducted a linear regression using the lme4 package in R (Bates et. al., 2015). For this linear regression analysis, I used the same model I used for the dominance analysis (see Appendix for regression results):

$$FYGPA = ACTC + HSGPA + \frac{Race}{Ethnicity} + Family\ Income\ + Gender + (1 \mid College\ Code)$$

As Figure 1 shows, the beta coefficient for HSGPA averaged across imputations is slightly larger than that for ACTC (although the confidence interval for ACTC encompasses the coefficient for HSGPA), and both are larger than the averaged beta coefficients for the demographic variables across imputations. If linear regression with multiple imputation suggests that HSGPA is a slightly stronger predictor of FYGPA than ACTC, then why does HSGPA not emerge as the clearly more dominant predictor in a dominance analysis? This is most likely due to the different goals and approaches to understanding predictor importance between the two methods. Linear regression uses beta weights (standardized coefficients) to compare the direct effect of a given predictor on an outcome of interest while controlling for all other predictors in the model; this means that the influence of a given predictor is dependent on the specific model chosen. Dominance analysis, on the other hand, assesses the unique contribution of each predictor to the explained variance (R^2) across all possible subset models; dominance analysis is therefore a model-independent measure of the relative importance of a predictor.

Figure 1. Standardized Coefficients for the Linear Regression With Multiple Imputation



Note. The error bars represent the pooled standard error across imputations and are calculated with the formula $SE = \sqrt{\bar{U} + \left(1 + \frac{1}{M}\right)B}$, where $\bar{U}$ represents the within-imputation variance, M represents the number of imputed datasets, and B represents the between-imputation variance.
**p* < 0.05; ***p* < 0.01; ****p* < 0.001.

There are several reasons why there may be a slight discrepancy in the findings of importance between a predictor and its outcome. These reasons include multicollinearity and model dependence (previously mentioned in the Introduction section). Dominance analysis considers the combined contribution of all predictors in all subset models. If ACTC operates in combination with other predictors in a way that bolsters its overall contribution to the R^2, this would be captured by the iteration through subset models conducted in a dominance analysis, which may make ACTC appear more dominant than HSGPA at times. In linear regression, multicollinearity can affect a predictor's beta weight. If ACTC and HSGPA are highly correlated, which prior research suggests, then standardized beta weights may not fully capture the unique contribution of a single predictor. Through the iteration of all subset models, dominance analysis may reveal a truer reflection of the importance of ACTC. Again, beta weights reflect the predictor's importance in the context of a given model, whereas dominance analysis provides a more flexible and global measure of its importance by considering all possible subset models.

As mentioned before, this discrepancy may arise because linear regression and dominance analysis each measures predictor importance differently. While beta weights provide insight into the predictor's effect within a single specified model, dominance analysis offers a more holistic perspective of the predictor's effect across many models. The ambiguity in the importance of ACTC and HSGPA may be due to ACTC's combined contribution with the other predictors or to how it contributes to the overall variance explained across model subsets.

What are the dominant predictors of FYGPA when comparing high school subject GPAs, ACT section scores, family income, race/ethnicity, and gender?

General Dominance Results

Table 8 contains the general dominance statistics among predictors. Examining the general dominance statistics with high school subject GPAs shows that English GPA was the most dominant predictor of FYGPA. Closely following were ACT English score, ACT math score, math GPA, and ACT science score. Note that ACT math score, math GPA, and ACT science score had equal R^2 contributions. These rankings were followed closely by ACT reading score, natural science GPA, social studies GPA, and race/ethnicity. It is interesting that ACT reading score, natural science GPA, and social studies GPA were ranked as equally important as race/ethnicity. Gender and family income were ranked as the least important predictors.

Table 8. ACT Section and High School Subject GPA General Dominance Statistics

Predictor	R^2 contribution	Rank
English GPA	0.05	1
ACT English	0.04	2
ACT math	0.03	3
Math GPA	0.03	3
ACT science	0.03	3
ACT reading	0.02	4
Natural science GPA	0.02	4
Social studies GPA	0.02	4
Race/ethnicity	0.02	4
Gender	0.01	5
Family income	0.01	5

Conditional Dominance Results

Table 9 shows that English GPA and ACT English score were the two most conditionally dominant predictors of FYGPA. Neither was conditionally dominant over the other, but both were conditionally dominant over eight of the 10 other predictors in the analysis. English GPA conditionally dominated math GPA, natural science GPA, social studies GPA, ACT reading score, ACT science score, family income, race/ethnicity, and gender (all the areas except ACT English and math scores). Math GPA conditionally dominated natural science GPA, social studies GPA, and family income. Table 9 also shows that ACT English score conditionally dominated math GPA, natural science GPA, social studies GPA, ACT math score, ACT reading score, ACT science score, family income, and race/ethnicity (all except English GPA and gender). ACT math score also conditionally dominated natural science GPA, social studies GPA, ACT reading score, ACT science score, and family income. ACT science score conditionally dominated ACT reading score. Race/ethnicity conditionally dominated family income.

Table 9. ACT Section and High School Subject GPA Conditional Dominance Statistics

Predictor	Number of predictors										
	1	2	3	4	5	6	7	8	9	10	11
ACT English	0.172	0.102	0.063	0.042	0.028	0.020	0.015	0.011	0.009	0.007	0.006
ACT math	0.148	0.085	0.051	0.032	0.021	0.014	0.010	0.008	0.006	0.005	0.004
ACT reading	0.125	0.066	0.036	0.020	0.011	0.006	0.003	0.002	0.001	0.000	0.000
ACT science	0.129	0.070	0.038	0.022	0.012	0.007	0.004	0.002	0.001	0.000	0.000
English GPA	0.147	0.096	0.069	0.053	0.043	0.036	0.031	0.027	0.024	0.022	0.020
Math GPA	0.111	0.065	0.042	0.029	0.021	0.016	0.012	0.010	0.008	0.006	0.005
Nat. sci. GPA	0.102	0.057	0.035	0.023	0.016	0.011	0.008	0.006	0.004	0.003	0.003
Soc. studies GPA	0.083	0.047	0.029	0.020	0.014	0.010	0.007	0.005	0.004	0.003	0.002
Family income	0.040	0.022	0.014	0.010	0.008	0.006	0.005	0.005	0.004	0.004	0.004
Race/ ethnicity	0.062	0.035	0.022	0.015	0.011	0.009	0.008	0.007	0.007	0.006	0.006
Gender	0.015	0.017	0.017	0.015	0.014	0.013	0.012	0.011	0.010	0.009	0.009

Complete Dominance Results

According to the complete dominance analysis using high school subject GPAs and ACT section scores, English GPA completely dominated math GPA, natural science GPA, social studies GPA, family income, race/ethnicity, and gender (see Table 10; this table should be read first by row and then by column.). Additionally, ACT English and math scores completely dominated ACT reading and science scores. There was no complete dominance in either direction between any of the high school subject GPAs and the ACT section scores.

Table 10. ACT Section and High School Subject GPA Complete Dominance Statistics

	Predictor	Dominated by										
		Eng. GPA	Math GPA	Nat. sci. GPA	Soc. stud. GPA	ACT Eng.	ACT math	ACT reading	ACT science	Family income	Race/ ethnicity	Gender
Dominates	Eng. GPA	—	True	True	True	N/A	N/A	N/A	N/A	True	True	True
	Math GPA	False	—	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Nat. sci. GPA	False	N/A	—	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Soc. stud. GPA	False	N/A	N/A	—	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	ACT Eng.	N/A	N/A	N/A	N/A	—	N/A	True	True	N/A	N/A	N/A
	ACT math	N/A	N/A	N/A	N/A	N/A	—	True	True	N/A	N/A	N/A
	ACT reading	N/A	N/A	N/A	N/A	False	False	—	N/A	N/A	N/A	N/A
	ACT science	N/A	N/A	N/A	N/A	False	False	N/A	—	N/A	N/A	N/A
	Family income	False	N/A	N/A	N/A	N/A	N/A	N/A	N/A	—	N/A	N/A
	Race/ethnicity	False	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	—	N/A
	Gender	False	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	—

Note. N/A indicates complete dominance could not be determined.

Strongest Level of Dominance Results

Table 11 contains the strongest level of dominance achieved between the ACT sections and each predictor as well as the number of imputations for which the relationship held: ACT English score completely dominated ACT reading and science scores in all 20 imputations, indicating that ACT English score is consistently a stronger predictor than ACT reading and science scores are.[1] ACT English score conditionally or generally dominated ACT math score, race/ethnicity, gender, and family income, though the dominance is not absolute for these predictors. ACT math score completely dominated ACT science score in all 20 imputations, indicating that ACT math score is a stronger predictor than ACT science score. ACT math score also has a complex relationship with both ACT reading score and family income, dominating these predictors to some degree. ACT reading score is dominated by ACT science, English, and math scores, suggesting that relative to the other sections, ACT reading score is a weaker predictor. At the same time, ACT reading score consistently generally dominated race/ethnicity, gender, and family income, indicating that ACT reading score is a better predictor than these demographic variables. Finally, ACT science score generally dominated race/ethnicity, gender, and family income, indicating that ACT science score is a better predictor than these demographic variables.

Table 11. Strongest Level of Dominance Achieved by ACT Section Score

Predictor1	Relationship	Predictor2	No. of imputations
ACT English	conditionally dominates	ACT math	20
	completely dominates	ACT reading	20
	completely dominates	ACT science	20
	generally dominates	race/ethnicity	9
	conditionally dominates		8
	completely dominates		3
	generally dominates	gender	19
	conditionally dominates		1
	conditionally dominates	family income	14
	completely dominates		5
	generally dominates		1
ACT math	completely dominates	ACT reading	13
	conditionally dominates		7
	completely dominates	ACT science	20
	generally dominates	race/ethnicity	17
	conditionally dominates		3

[1] A comparison of multiply imputed dominance results and regression with multiply imputed data for the individual high school subject GPAs and ACT section scores is not provided because the logic between any differences in these analyses was explained in the previous comparison. Were this comparison to be included, we would most likely make slightly different conclusions due to the differences in methodology noted in the section titled Comparison of Multiply Imputed Dominance Results and Regression With Multiply Imputed Data.

Predictor1	Relationship	Predictor2	No. of imputations
ACT math	generally dominates	gender	20
	conditionally dominates	family income	13
	generally dominates		7
ACT reading	is conditionally dominated by	ACT science	16
	is generally dominated by		4
	generally dominates	race/ethnicity	20
	generally dominates	gender	20
	generally dominates	family income	20
ACT science	generally dominates	race/ethnicity	20
	generally dominates	gender	20
	generally dominates	family income	20

Table 12 contains the strongest level of dominance achieved between high school subject GPAs and each predictor and also the number of imputations the relationship held. In no case did the subject GPAs completely dominate any of the other predictors for all imputed datasets, but in some cases, complete dominance was observed for some datasets, and weaker forms of dominance were attained for others. Examining GPA in specific high school subjects shows that across all 20 imputations, English GPA at least generally dominated math, natural science, and social studies GPAs, as well as the demographic factors race/ethnicity, gender, and family income; English GPA also tended to dominate the four ACT section scores. Table 12 also shows that math GPA was often dominated by ACT section scores while at the same time tending to dominate the GPAs for the other subjects and the demographic variables. Natural science GPA tended to be dominated by the ACT section scores and some GPAs for the other subjects but showed some dominance over the demographic variables. Finally, social studies GPA showed varied dominance, tending to be dominated by ACT section scores but also dominating the demographic variables.

Table 12. Strongest Level of Dominance Achieved by High School Subject GPAs

Predictor1	Relationship	Predictor2	No. of imputations
English GPA	generally dominates	ACT English	15
	is generally dominated by		3
	conditionally dominates		1
	is conditionally dominated by		1
	conditionally dominates	ACT math	10
	generally dominates		9
	is generally dominated by		1
	conditionally dominates	ACT reading	18
	generally dominates		2
	conditionally dominates	ACT science	17
	generally dominates		3
	completely dominates	race/ethnicity	13
	conditionally dominates		7
	completely dominates	gender	17
	generally dominates		2
	conditionally dominates		1
	completely dominates	math GPA	12
	conditionally dominates		7
	generally dominates		1
	completely dominates	natural science GPA	16
	conditionally dominates		3
	generally dominates		1
	completely dominates	social studies GPA	17
	conditionally dominates		3
	completely dominates	family income	18
	conditionally dominates		2
Math GPA	is conditionally dominated by	ACT English	10
	is generally dominated by		10
	is generally dominated by	ACT math	9
	is conditionally dominated by		7
	generally dominates		4
	generally dominates	ACT reading	11
	is generally dominated by		5
	conditionally dominates		4
	generally dominates	ACT science	11
	conditionally dominates		4
	is generally dominated by		4
	is conditionally dominated by		1
	generally dominates	race/ethnicity	10
	conditionally dominates		5
	completely dominates		3
	is generally dominated by		2
	generally dominates	gender	15
	conditionally dominates		5

Predictor1	Relationship	Predictor2	No. of imputations
Math GPA	conditionally dominates	natural science GPA	8
	completely dominates		4
	generally dominates		4
	is completely dominated by		2
	is conditionally dominated by		1
	is generally dominated by		1
	conditionally dominates	social studies GPA	10
	completely dominates		5
	generally dominates		4
	is generally dominated by		1
	completely dominates	family income	10
	generally dominates		7
	conditionally dominates		3
Natural science GPA	is conditionally dominated by	ACT English	18
	is generally dominated by		2
	is conditionally dominated by	ACT math	15
	is generally dominated by		4
	generally dominates		1
	is generally dominated by	ACT reading	9
	generally dominates		7
	conditionally dominates		3
	is conditionally dominated by		1
	is generally dominated by	ACT science	10
	generally dominates		7
	is conditionally dominated by		2
	conditionally dominates		1
	generally dominates	race/ethnicity	16
	conditionally dominates		2
	is generally dominated by		2
	generally dominates	gender	18
	conditionally dominates		1
	is generally dominated by		1
	completely dominates	social studies GPA	6
	generally dominates		5
	conditionally dominates		3
	is completely dominated by		3
	is generally dominated by		2
	is conditionally dominated by		1
	generally dominates	family income	15
	conditionally dominates		4
	completely dominates		1
Social studies GPA	is conditionally dominated by	ACT English	19
	is generally dominated by		1
	is conditionally dominated by	ACT math	13
	is generally dominated by		7
	is generally dominated by	ACT reading	15

Predictor1	Relationship	Predictor2	No. of imputations
Social Studies GPA	generally dominates	ACT reading	4
	is conditionally dominated by		1
	is generally dominated by	ACT science	16
	generally dominates		3
	is conditionally dominated by		1
	generally dominates	race/ethnicity	13
	is conditionally dominated by		3
	conditionally dominates		2
	is generally dominated by		2
	generally dominates	gender	18
	is generally dominated by		2
	generally dominates	family income	10
	conditionally dominates		8
	is generally dominated by		2

Table 13 contains the strongest level of dominance achieved between demographic predictors and the number of imputations for which the relationship held. Examining the demographic variables shows that race/ethnicity generally dominated gender consistently but was generally dominated by the ACT section scores (Table 11) and high school subject GPAs (Table 12). Race/ethnicity also shows conditional dominance over family income but was at times dominated by it as well. Gender was generally dominated by ACT section scores, subject GPAs, and race/ethnicity. Finally, family income was generally dominated by most of the other predictors including race/ethnicity and gender.

Table 13. Strongest Level of Dominance Achieved by Demographics

Predictor1	Relationship	Predictor2	No. of imputations
Race/ethnicity	generally dominates	gender	20
Family income	is conditionally dominated by	race/ethnicity	16
	is generally dominated by		4
	is generally dominated by	gender	18
	generally dominates		2

In summary, ACT English and math scores tend to be the most dominant ACT section predictors overall, often dominating the ACT reading and science scores, high school subject GPAs, and demographic variables. English GPA is the most dominant subject GPA, often dominating the other subject GPAs, ACT section scores, and demographic variables. Race/ethnicity tends to dominate gender but is dominated by the ACT section scores and subject GPAs. Family income is generally the weakest predictor, often dominated by all the other variables. This pattern of results suggests that cognitive variables such as ACT section scores and subject GPAs are stronger predictors than these demographic variables.

Discussion

Students' transition from high school to college is characterized by shifts in educational expectations and performance standards. Predicting first-year grade point average (FYGPA) is crucial for college admissions and appropriately allocating college student support services. High school GPA (HSGPA) and scores from standardized tests, such as the ACT, are commonly used to predict students' performance in their first year of college. Much research has shown that both HSGPA and standardized test scores consistently are strong predictors of FYGPA, with some studies indicating that using both measures together offers better predictions than using either measure alone. Importantly, HSGPA tends to be a particularly strong predictor because it reflects the academic behaviors that contribute to students' academic success in college.

Recent studies have further examined the predictive power of HSGPA and standardized test scores, finding that HSGPA tends to be a somewhat stronger predictor than ACT scores of FYGPA across high schools and student populations. Some research has shown that the predictive validity of these measures varies according to demographics such as race/ethnicity and that HSGPA demonstrates a stronger relationship than ACT scores. The relationship between these predictors and long-term outcomes such as graduation rates also highlights the importance of accurately predicting FYGPA because higher FYGPA has been linked to timely degree completion.

In this study, I present an approach to evaluating the relative importance of these predictors by performing both a multiple imputation and dominance analysis, thereby comparing the contribution of all predictors across a combination of all possible subset models. Missing data are common in educational contexts, so multiple imputation presents an opportunity to use the available data to fill in gaps in extant datasets. This allowed me to use the entirety of a dataset, which can help foster more robust research conclusions. Combining this approach with dominance analysis provides a richer dataset in which to evaluate predictor importance.

Also in this study, I explored the relative importance of traditional precollege academic achievement measures for predicting FYGPA, including overall HSGPA, high school subject GPA, ACT Composite and section scores, and student demographics. Examining these predictors through dominance analysis (as opposed to traditional regression methodology) presented an opportunity to offer new insights into the relative importance of overall HSGPAs as well as subject-specific GPAs and ACT scores. Through the methodology employed in this study, I aim to provide a more robust and nuanced understanding of the importance of factors that significantly predict college FYGPA.

The dominance analysis revealed complex relationships between GPAs, ACT scores, and student demographics in predicting FYGPA. Both HSGPA and ACT Composite score dominated the demographic factors race/ethnicity, gender, and family income, but the dominance between HSGPA and ACT Composite score was less clear. In fact, the relationship between HSGPA and ACT Composite score was far more nuanced because across imputations neither consistently dominated the other. This finding runs contrary to what previous studies have found when employing a single regression model methodology. The methodology used in this study presents a new perspective on the importance of HSGPA and ACT scores.

In contrast, three demographic variables showed consistent patterns of being less important predictors than HSGPA and ACT scores. Examining the importance of these demographic variables showed that race/ethnicity tended to dominate both gender and family income and that family income was the least important predictor of the three.

An examination of ACT section scores also revealed specific patterns in their predictive importance. Of the four ACT section scores, ACT English score consistently emerged as the strongest ACT section score predictor of FYGPA. ACT math score, though strong, was a less important predictor than ACT English score, but ACT math score did show dominance over ACT science score and the demographic variables. While ACT reading score was a less important predictor of FYGPA than the other ACT section scores, it still dominated the demographic variables, indicating that even the relatively weaker ACT section scores are better at predicting FYGPA than student demographics are. Overall, ACT English and math scores were the most dominant ACT sections for predicting FYGPA.

When looking at high school subject GPAs, I found that English GPA dominated all the other subject GPAs (including math, natural science, and social studies), the demographic factors, and, in some cases, the ACT section scores in math, reading, and science. The fact that ACT English score was the strongest ACT section predictor and that English GPA was the strongest subject GPA predictor reinforces the strength of English-related performance as a predictor of FYGPA. While math GPA was dominated by the ACT section scores, math GPA showed dominance over the other subject GPAs and demographic variables. The weakest (i.e., the least dominant) predictors of FYGPA were family income and gender. Family income and gender were frequently dominated by the other achievement variables and by race/ethnicity. Collectively, these results highlight that academic readiness measures—such as ACT scores, HSGPAs, and high school subject GPAs—are stronger predictors of academic success in the first year of college than demographic factors.

This study highlights the important role that high school academic performance—particularly HSGPA and ACT scores—plays in predicting college FYGPA. The dominance of these achievement measures over demographic variables stresses the practical importance for colleges to focus on academic achievement metrics while making admissions decisions and providing targeted support to students. For students, the results of this study emphasize the importance of a strong core academic preparation in high school, particularly in English and math, which consistently emerged as the most dominant predictors of FYGPA. For colleges, these findings provide insights that can guide data-driven decisions when refining admissions strategies and developing student support programs. Recognizing the nuanced relationships among GPAs, ACT scores, and FYGPA can present colleges with an opportunity to better identify students who would benefit from additional academic resources, particularly those students whose ACT scores and HSGPAs are inconsistent with each other.

The results of this study show the practical importance of using a combination of HSGPA and standardized test scores in the college admissions process, and ACT also advocates for a holistic approach that considers the broader context of a student's academic journey. While developing their admissions practices and support services, colleges should consider the findings of this study because they can help postsecondary institutions harness the predictive power of high school academic achievement while simultaneously meeting their enrollment goals.

Limitations

Two limitations are worth mentioning. The first is that for this analysis, I used data from one state, a single cohort, and from public institutions only. This limits the generalizability of the findings to populations beyond these parameters. Although this study presents an important first look at predictors of FYGPA using dominance analysis, the study should be replicated with other data sources.

The second limitation is that this study relied on self-reported HSGPA. Although research supports the use of self-reported HSGPA for analytical purposes, student bias may have affected these self-reports, and this bias could potentially underestimate the predictive value of HSGPA. Nonetheless, replicating this study with other datasets that have complete data with official HSGPA is worth exploring.

References

Allensworth, E. M., & Clark, K. (2020). High school GPAs and ACT scores as predictors of college completion: Examining assumptions about consistency across high schools. *Educational Researcher*, *49*(3), 198–211. https://doi.org/10.3102/0013189X20902110

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Beard, J., & Marini, J. (2018). *Validity of the SAT for predicting first-year grades: 2013 SAT validity sample*. College Board. https://files.eric.ed.gov/fulltext/ED582459.pdf

Camara, W., Kimmel, E., Scheuneman, J., & Sawtell, E. A. (2003). *Whose grades are inflated?* College Board. https://files.eric.ed.gov/fulltext/ED563203.pdf

Curabay, M. (2016). *Meta-analysis of the predictive validity of Scholastic Aptitude Test (SAT) and American College Testing (ACT) scores for college GPA* [Master's thesis, University of Denver]. Electronic Theses and Dissertations, 1225. https://digitalcommons.du.edu/cgi/viewcontent.cgi?article=2225&context=etd

Demeter, E., Dorodchi, M., Al-Hossami, E., Benedict, A., Slattery Walker, L., & Smail, J. (2022). Predicting first-time-in-college students' degree completion outcomes. *Higher Education*, *84*(3), 589–609. http://dx.doi.org/10.1007/s10734-021-00790-9

Gast, T. Y., & Burt, I. (2013). Exploring the predictive value of the ACT among first-year college students on academic probation. *Journal of College Orientation, Transition, and Retention*, *21*(1), 46–54. http://dx.doi.org/10.24926/jcotr.v21i1.2859

Gershenfeld, S., Ward Hood, D., & Zhan, M. (2016). The role of first-semester GPA in predicting graduation rates of underrepresented students. *Journal of College Student Retention: Research, Theory & Practice*, *17*(4), 469–488. https://doi.org/10.1177/1521025115579251

Kobrin, J. L., Camara, W. J., & Milewski, G. B. (2002). *Students with discrepant high school GPA and SAT I scores.* College Board. https://files.eric.ed.gov/fulltext/ED562878.pdf

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, *75*(1), 63–82. https://doi.org/10.3102/00346543075001063

Larson, J. R., Jr., & Scontrino, M. P. (1976). The consistency of high school grade point average and of the verbal and mathematical portions of the Scholastic Aptitude Test of the College Entrance Examination Board, as predictors of college performance: An eight year study. *Educational and Psychological Measurement*, *36*(2), 439–443. https://doi.org/10.1177/001316447603600224

Loria, R. N., & Sanchez, E. I. (in press). *Validity of ACT Composite score and high school GPA for predicting probability of timely degree completion: Examining first-year college GPA as a mediator* (ACT Research Report R2419). ACT.

Luchman, J. (2024). domir: Tools to support relative importance analysis. R package version 1.2.0. https://CRAN.R-project.org/package=domir

Marini, J. P., Westrick, P. A., Young, L., Ng, H., Shmueli, D., & Shaw, E. J. (2019). *Differential validity and prediction of the SAT: Examining first-year grades and retention to the second year.* College Board. https://files.eric.ed.gov/fulltext/ED597325.pdf

Noble, J. P., & Sawyer, R. L. (2004). Is high school GPA better than admission test scores for predicting academic success in college? *College and University*, *79*(4), 17–22. https://www.researchgate.net/publication/234702285_Is_High_School_GPA_Better_than_Admission_Test_Scores_for_Predicting_Academic_Success_in_College

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.

Sanchez, E. I., & Buddin, R. (2016). *How accurate are self-reported high school courses, course grades, and grade point average?* ACT. https://www.act.org/content/dam/act/unsecured/documents/5269-research-report-how-accurate-are-self-reported-hs-courses.pdf

Sanchez, E. I., & Mattern, K. (2018). When high school grade point average and test scores disagree: Implications for test-optional policies. In J. Buckley, L. Letukas, & B. Wildavsky (Eds.), *Measuring success: Testing, grades, and the future of college admissions*, pp. 118–144. Johns Hopkins University Press.

Sanchez, E. I., & Moore, R. (2022). *Grade inflation continues to grow in the past decade*. ACT. https://www.act.org/content/dam/act/unsecured/documents/2022/R2134-Grade-Inflation-Continues-to-Grow-in-the-Past-Decade-Final-Accessible.pdf

Sanchez, E. I. (2024). *Changes in predictive validity of high school grade point average and ACT® Composite score after the COVID-19 pandemic*. ACT. https://www.act.org/content/dam/act/secured/documents/R2328-Changes-in-Predictive-Validity-of-HSGPA-and-ACT-Composite-Score-After-COVID-19-2024-09.pdf

Sawyer, R. (2013). Beyond correlations: Usefulness of high school GPA and test scores in making college admissions decisions. *Applied Measurement in Education*, *26*(2), 89–112. https://www.tandfonline.com/doi/pdf/10.1080/08957347.2013.765433

Shaw, E. J., & Mattern, K. D. (2009). *Examining the accuracy of self-reported high school grade point average*. College Board. https://files.eric.ed.gov/fulltext/ED562616.pdf

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Warren, J. M., & Goins, C. L. (2019). Exploring the relationships between high school course enrollment, achievement, and first-semester college GPA. *Journal of Educational Research and Practice*, *9*(1), 386–399. https://files.eric.ed.gov/fulltext/EJ1278226.pdf

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M. R., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT scores, high school grades, and SES. *Educational Assessment*, *20*(1), 23–45. https://www.tandfonline.com/doi/pdf/10.1080/10627197.2015.997614

Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and SAT scores: The role of student ethnicity and first language. *Educational American Research Journal*, *42*(3), 439–464. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=35d30b858d07140efe538a6383b9f674609e33b7

Appendix: Linear Regression With Multiple Imputation Results With Standardized Coefficients

Predictor		Estimate	Standard error	*p* value
(Intercept)		2.84	0.05	< 0.001
HSGPA		0.35	0.08	< 0.001
ACTC		0.31	0.02	< 0.001
Family income	$36K–$60K	0.10	0.04	0.013
	$60K–$100K	0.14	0.04	< 0.001
	> $100K	0.19	0.04	< 0.001
	Missing	0.11	0.15	0.464
Race/ ethnicity	Black	−0.17	0.05	0.002
	American Indian/Native Hawaiian/two or more races	−0.15	0.08	0.066
	Asian	0.27	0.08	0.002
	Hispanic	0.06	0.06	0.328
	Prefer not to respond/missing data	0.03	0.14	0.821
Gender	Male	−0.25	0.03	0.000
	Other/missing	−0.13	0.14	0.361

ABOUT ACT

ACT is transforming college and career readiness pathways so that everyone can discover and fulfill their potential. Grounded in more than 65 years of research, ACT's learning resources, assessments, research, and work-ready credentials are trusted by students, job seekers, educators, schools, government agencies, and employers in the U.S. and around the world to help people achieve their education and career goals at every stage of life. Visit us at www.act.org.