



Validating School Improvement Plan (SIP) Goal Measures as Predictors of High School Graduation

Key Findings

1. The differences between the probability of high school completion for those students who met or missed the SIP measures were all statistically and practically significant; in other words, those differences are not occurring due to chance and the magnitude of the effect is large.
2. Regression analysis shows that within MMSD, each of the high school SIP goal measures is an extremely significant predictor of high school completion above and beyond the influence of student demographic characteristics and attendance.

All of the SIP goal measures at the high school level were chosen based on existing research and extensive conversations with experts. However, context is key in educational research, and what is predictive of positive student outcomes in one location is not always predictive of positive outcomes in another. To further validate the use of these SIP goal measures in MMSD, the Research & Program Evaluation Office tested the predictive power of these variables using MMSD data. This report also contributes to the larger literature on predicting high school graduation by testing the predictive power of a set of key variables on the ultimate outcome of high school completion.

We examine the effect of the following variables on the likelihood of high school completion:

1. Two or more course failures during grade 9
2. Cumulative GPA of 3.0 or higher at the end of grade 11
3. Meeting the ACT college readiness benchmark in reading
4. Meeting the ACT college readiness benchmark in math

Data

Population

The students included in these reports come from the Wisconsin Department of Public Instruction's official lists of students in the cohorts expected to graduate in 2011, 2012, and 2013. These cohorts are based on the expectation that students will graduate in four years. Altogether, our dataset included 5,722 students across three cohorts. Not all students were included in each model; for example, we used only students receiving grades from the four comprehensive high schools (East, La Follette, Memorial, and West) to examine the effect of course failures and GPA because only these four schools use these goal measures as part of their SIP due to alternate grading practices for alternative programs.

Outcome – High School Completion

We use high school completion rather than high school graduation as the outcome of interest. High school completion includes both regular diplomas and equivalencies, while graduation includes only regular diplomas. MMSD sets goals around high school completion rather than high school graduation, so we replicate that approach in our analysis.

SIP Goal Measures

Each SIP measure was coded as a binary variable based on whether a student met that condition. For example, a student failing two or more courses in grade 9 would receive a "1" for that variable while a student failing zero or one courses in grade 9 would receive a "0."

Course failures and ACT variables require additional explanation. We defined a course failure as a student receiving an "F" for the semester at West, Memorial, and East, and a student receiving a quarter "F" at La Follette. Although MMSD high schools set goals around proficiency across the EPAS suite, which includes the EXPLORE 9, PLAN 10, and ACT, universal administration of the full suite began only in the spring of 2013. Therefore, for this analysis we use students' highest ACT scores only to calculate whether or not they met college readiness benchmarks.

Control Variables

Demographic variables come from the 2012-13 school year for students who were still enrolled during that school year; for students who dropped out earlier, demographic variables are drawn from their last enrollment in MMSD. We also control for ninth grade attendance in all models. Given the high correlation between attendance and other achievement measures, we felt it was important to be sure that we compared students with similar attendance rates to isolate further the effect of the SIP goal measures. We chose the 94% attendance cutoff because it aligns with district attendance goals.

Methods

To begin, we conducted a series of chi-square tests for significant differences in the probability of high school graduation based on whether students met the SIP goal measures outlined above. In other words, we compared the probability of graduation between students who failed two or more courses in 9th grade and students who failed fewer than two courses, as an example. This allowed us both to test for a statistically significant difference and to estimate the magnitude of the effect of meeting or missing each SIP goal, which we did by calculating effect sizes.

Next, we created four regression models based on the SIP measures of interest:

1. Model 1: Grade 9 course failures as predictive of high school completion
2. Model 2: Grade 11 cumulative GPA as predictive of high school completion
3. Model 3: ACT Reading College Readiness Benchmark as predictive of high school completion
4. Model 4: ACT Math College Readiness Benchmark as predictive of high school completion

Then, we conducted a series of logistic regressions with high school completion as the binary dependent variable and a robust set of demographic characteristics as control variables, with each high school SIP goal measure serving as the independent variable in its own regression. This process allowed us to estimate the predictive power on the likelihood of high school completion of each of the SIP measures above and beyond the effect of demographic characteristics. In short, this approach was designed to confirm that in MMSD, these measures are meaningful predictors of the ultimate goal of high school completion.

Each regression included the following control variables:

Variable	Coding	Sample Description
Gender	Binary, female=1, male=0	49% female
Race	Series of indicator variables for African-American, Hispanic, Asian, and multiracial (white and other races as reference category)	51% white 20% African-American 14% Hispanic 9% Asian 6% multiracial 1% other races
Special education status	Binary, special education=1, not special education=0	18% special education
English Language Learner (ELL) status	Binary, English Language Learner=1, not English Language Learner=0	10% ELL
Income status	Binary, free/reduced lunch=1, none=0	39% low-income
Parent education	Binary, bachelor's degree/advanced degree=1, less than bachelor's degree or missing=0	44% with college-educated parent
Single parent	Binary, one adult in household=1, all other values=0	35% with single parent
Grade 9 attendance	Binary, 94% or above=1, below 94%=0	68% with grade 9 attendance of 94% or above

In addition, each regression included school fixed effects to account for between-school differences.



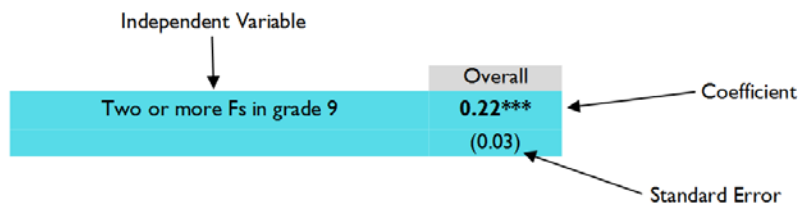
We coded the dependent independent variables as follows:

Variable	Coding	Sample Description
High School Completion	Binary, regular diploma or HSED within four years=1, no diploma or HSED within four years=0	77% completion rate
Model 1: Grade 9 course failures	Binary, two or more failures=1, fewer than two failures=0	26% with two or more failures in Grade 9
Model 2: Grade 11 cumulative GPA	Binary, 3.0 or above=1, below 3.0=0	45% with a 3.0 GPA at end of grade 11
Model 3: ACT Reading college readiness	Binary, met benchmark=1, missed benchmark or did not take ACT=0	52% took ACT, 64% of ACT takers met Reading benchmark
Model 4: ACT Math college readiness	Binary, met benchmark=1, missed benchmark or did not take ACT=0	52% took ACT, 67% of ACT takers met Math benchmark

To test the robustness of our findings across student groups, we conducted these regressions for all students in the three included cohorts, as well as a separate series of regressions for just the students included in the five most common student subgroups chosen as focus groups for school SIP goal-setting: African American, Hispanic, English Language Learner, Special Education, and low-income.

Understanding the Regression Tables

For each regression table in this report, each independent variable shows a coefficient (in this case, an odds ratio), as well as the standard error associated with that coefficient. Standard errors are a measure of the accuracy of an estimate; a smaller standard error implies a more accurate estimate, and in general, standard errors are smaller when sample sizes are larger. Coefficients that are statistically significant are bolded for emphasis and include stars to indicate the level of significance (*=p<0.10, **=p<0.05, ***=p<0.01).



The bottom of each column includes the Pseudo r-squared for that regression model, which is a measure of the percent of variance in the outcome variable (in this case, high school completion) explained by the variables in the model.

Interpreting Odds Ratios

All of the regression results contained within this report use odds ratios as coefficients. An odds ratio represents the change in odds of fulfilling the dependent variable (e.g., completing high school) associated with a change in the independent variable (e.g., meeting the SIP goal measure). In practical terms, imagine this hypothetical example:

- A student who meets the SIP goal measure has an 80% probability of completing high school in four years. Therefore, their odds of graduating are 80%/20%, or 4:1 (simplified to 4).
- A student who does not meet the SIP goal measure has a 55% probability of completing high school in four years. Therefore, their odds of graduating are 55%/45%, or 11:9 (simplified to 1.22).
- To calculate the odds ratio, we would divide 4 by 1.22 to arrive at an odds ratio of **3.27** associated with meeting the SIP goal measure. This finding would indicate that meeting the SIP goal measure is a positive predictor of high school completion.

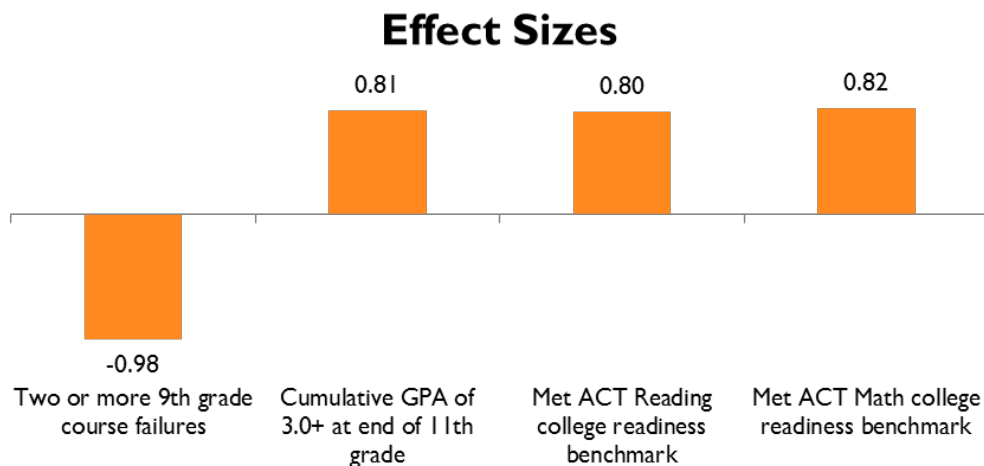
Odds ratios can have the effect of magnifying a positive relationship, as in the example above, where an odds ratio of 3.27 corresponds to a 25 percentage point higher probability of graduating. The odds ratio of 3.27 does not mean that students meeting the SIP goal measure would be 3.27 times more likely to graduate; instead, it means that their odds of graduation are 3.27 times higher, which is a key distinction.



Findings

Before conducting a series of logistic regressions, we conducted chi-square tests to estimate statistically and practically significant differences in the probability of high school graduation based on whether students met or missed the SIP measure. **All differences were statistically significant with $p < .001$; this means that the difference observed between students who met or missed the SIP measure were not occurring due to chance.**

While these tests confirmed the differences were not random, we wanted to further test the magnitude of the difference. The graph below shows the *Cohen's d* effect size associated with each of the SIP measures. Effect sizes help measure the magnitude of the difference between two groups, allowing researchers to look at practically significant differences in addition to statistically significant differences. These effect sizes are calculated using unpaired independent groups.



These effect sizes can be interpreted as the effect of meeting the SIP measure on the probability of high school completion in standard deviations. For example, a student failing two or more courses in 9th grade has a probability of high school graduation that is almost a full standard deviation lower than a student who does not fail two or more courses. A student meeting the ACT Reading college readiness benchmark has a probability of high school graduation that is 0.8 standard deviations higher than a student who does not meet the benchmark. **All of the effect sizes reported above are traditionally considered “large;” this means that the differences observed between students who met or missed the SIP benchmark are practically significant.**

Next, we conducted a series of logistic regressions to test the effect of each SIP goal measure on the likelihood of high school completion. In all four models, the relevant SIP measure was a highly significant predictor of on-time high school completion for students overall, even when controlling for demographic characteristics and grade 9 attendance. In the grade 9 course failures and grade 11 GPA models, the odds ratio associated with the SIP measure was significant at $p < .001$, which means that we can have at least 99.9% confidence that the relationship between the SIP measure and high school completion is not random. For the ACT Reading college readiness benchmark, the odds ratio was significant at $p < .01$, and for the ACT Math college readiness benchmark, the odds ratio was significant at $p < .05$, which means we can have at least 99% and 95% confidence, respectively, that the relationship between each SIP measure and high school completion is not random. **In short, these regressions illustrate that within MMSD, each of the high school SIP goal measures is an extremely significant predictor of high school completion above and beyond the influence of student demographic characteristics and attendance.**

On the subsequent pages, we show the regression results for each of the models and discuss findings in more detail.

**Model #1: Grade 9 Course Failures**

	Overall	African American	Hispanic	ELL	Low income	Special education
Two or more Fs in grade 9	0.22*** (0.03)	0.25*** (0.05)	0.25*** (0.06)	0.30*** (0.08)	0.19*** (0.03)	0.36*** (0.06)
Female	1.78*** (0.18)	2.06*** (0.36)	1.57* (0.37)	1.52* (0.38)	1.92*** (0.26)	1.61*** (0.26)
African American	0.67*** (0.09)	N/A	N/A	0.19 (0.23)	0.79 (0.14)	0.88 (0.18)
Hispanic	1.05 (0.19)	N/A	N/A	0.15* (0.16)	1.38 (0.33)	1.08 (0.36)
Asian	1.53* (0.35)	N/A	N/A	0.23 (0.25)	1.91** (0.56)	0.72 (0.37)
Two or more races	1.25 (0.29)	N/A	N/A	0.21 (0.35)	1.36 (0.43)	1.1 (0.34)
Special education	0.2*** (0.02)	0.35*** (0.06)	0.31*** (0.08)	0.34*** (0.1)	0.28*** (0.04)	N/A N/A
ELL	0.34*** (0.06)	0.75 (0.42)	0.24*** (0.06)	N/A N/A	0.33*** (0.07)	0.69 (0.25)
Low income	0.37*** (0.04)	0.33*** (0.07)	0.5** (0.14)	0.47** (0.14)	N/A N/A	0.4*** (0.07)
Parent is college graduate	1.37** (0.18)	1.04 (0.27)	1.47 (0.55)	1.19 (0.52)	1.33 (0.28)	1.26 (0.25)
Single parent	0.83* (0.09)	0.74* (0.13)	0.66* (0.16)	0.81 (0.23)	0.89 (0.12)	0.97 (0.16)
Grade 9 attendance 94% or above	1.71*** (0.19)	2.08*** (0.41)	1.49 (0.37)	2.4*** (0.64)	2.36*** (0.34)	1.35* (0.24)
Pseudo r-squared	0.3619	0.2349	0.2786	0.1871	0.3007	0.1491

Note: Statistically significant coefficients **bolded** for emphasis. *= $p < 0.10$, **= $p < 0.05$, ***= $p < 0.01$.

Grade 9 course failures are highly predictive of high school completion for students overall and for each of the five subgroups. Overall, those students who had two or more course failures in grade 9 had odds of high school completion 0.22 times as high (or 78% lower) relative to those who failed one or no courses. This result was statistically significant at $p < 0.01$. Course failures are especially predictive for economically disadvantaged students.

**Model #2: Grade 11 Cumulative GPA of 3.0 or Higher**

	Overall	African American	Hispanic	ELL	Low income	Special education
Grade 11 Cumulative GPA 3.0 or higher	3.17*** (0.49)	3.68*** (1.84)	7.00*** (3.45)	3.49*** (1.55)	4.31*** (1.13)	0.51*** (0.12)
Female	1.83*** (0.19)	2.29*** (0.42)	1.57* (0.39)	1.46 (0.4)	2.00*** (0.27)	1.82*** (0.3)
African American	0.6*** (0.09)	N/A N/A	N/A N/A	0.24 (0.29)	0.73* (0.14)	0.7* (0.14)
Hispanic	0.95 (0.18)	N/A N/A	N/A N/A	0.15* (0.16)	1.21 (0.29)	0.9 (0.3)
Asian	1.26 (0.3)	N/A N/A	N/A N/A	0.22 (0.25)	1.55 (0.46)	0.83 (0.44)
Two or more races	1.12 (0.25)	N/A N/A	N/A N/A	0.36 (0.62)	1.36 (0.41)	0.87 (0.26)
Special education	0.18*** (0.02)	0.32*** (0.06)	0.27*** (0.07)	0.32*** (0.09)	0.24*** (0.03)	N/A N/A
ELL	0.37*** (0.07)	0.82 (0.48)	0.27*** (0.06)	N/A N/A	0.38*** (0.08)	0.69 (0.26)
Low income	0.42*** (0.05)	0.41*** (0.09)	0.53** (0.16)	0.53* (0.18)	N/A N/A	0.36*** (0.07)
Parent is college graduate	1.45*** (0.2)	1.28 (0.34)	1.28 (0.51)	0.91 (0.4)	1.42* (0.29)	1.55** (0.31)
Single parent	0.77** (0.08)	0.65** (0.12)	0.63* (0.16)	0.71 (0.21)	0.85 (0.12)	0.89 (0.15)
Grade 9 attendance 94% or above	2.19*** (0.24)	2.78*** (0.54)	1.6* (0.4)	2.69*** (0.72)	3.00*** (0.42)	2.24*** (0.38)
Pseudo r-squared	0.3339	0.2078	0.2651	0.1667	0.2541	0.1253

Note: Statistically significant coefficients **bolded** for emphasis. *= $p < 0.10$, **= $p < 0.05$, ***= $p < 0.01$.

Having a GPA of 3.0 or higher at the end of Grade 11 is highly predictive of high school completion for students overall and for four of the five subgroups. Overall, those students who had a cumulative GPA of 3.0 or higher at the end of grade 11 had odds of high school completion more than 3.17 times higher than those who had a cumulative GPA below 3.0. This result was statistically significant at $p < 0.01$. The notable exception to this trend is that for special education students, having a 3.0 GPA at the end of Grade 11 is associated with a *reduced* likelihood of high school completion. We are working to explore potential explanations for this unexpected finding.



Model #3: ACT Reading College Readiness Benchmark

For these regressions, we restricted the population of included students to those who had taken the ACT.

	Overall	African American	Hispanic	ELL	Low income	Special education
Met ACT Reading Benchmark	2.61*** (0.97)	4.39 (5.31)			2.46 (1.67)	3.46 (4.72)
Female	1.73* (0.54)	4.19** (2.63)			1.27 (0.6)	8.24* (10.24)
African American	0.25*** (0.11)	N/A N/A			0.32 (0.27)	0.86 (1.41)
Hispanic	0.55 (0.29)	N/A N/A			0.33 (0.31)	0.19 (0.32)
Asian	0.53 (0.32)	N/A N/A			0.53 (0.58)	N/A N/A
Two or more races	0.59 (0.39)	N/A N/A	Regression unavailable – all	Regression unavailable – all ELL	0.78 (0.99)	N/A N/A
Special education	0.76 (0.34)	1.22 (0.96)	Hispanic students meeting benchmark	students meeting benchmark	0.56 (0.36)	N/A N/A
ELL	0.76 (0.52)	0.18 (0.25)	completed high school	completed high school	1.29 (1.08)	N/A N/A
Low income	0.96 (0.39)	2.34 (1.52)			N/A N/A	0.33 (0.46)
Parent is college graduate	0.59 (0.22)	0.42 (0.27)			0.5 (0.27)	0.27 (0.36)
Single parent	0.46** (0.15)	0.11*** (0.09)			0.38* (0.2)	0.07** (0.09)
Grade 9 attendance 94% or above	3.7*** (1.2)	6.11*** (4.22)			2.29 (1.16)	1.69 (1.88)
Pseudo r-squared	0.1372	0.2255			0.1111	0.3676

Note: Statistically significant coefficients **bolded** for emphasis. *= $p < 0.10$, **= $p < 0.05$, ***= $p < 0.01$.

Overall, among students taking the ACT, those who met the college readiness benchmark in Reading had odds of high school completion more than 2.6 times higher than those who missed the benchmark. This result was statistically significant at $p < 0.01$. This trend is mirrored for three subgroups, but was not statistically significant.

The lack of significant coefficients associated with the ACT Reading college readiness benchmark for the subgroups outlined does not necessarily mean meeting the benchmark is not predictive for these groups. We restricted these regressions only to students who took the ACT, and because these cohorts graduated from MMSD before universal administration of the ACT, students in these traditionally disadvantaged subgroups took the ACT at relatively low rates. As a result, the regressions for these subgroups are relatively underpowered for ACT-related measures because there are relatively few test-taking students on which to base our analysis. Although the results are not statistically significant, the odds ratios are large and in the expected direction.

In addition, given that taking the ACT was optional for these cohorts, it is likely that meeting the college readiness benchmark after taking the ACT has less predictive power on the probability of graduation than simply choosing to take the ACT in the first place. We ran a regression that included all students in these three cohorts, substituting a binary variable that indicated whether the student had taken the ACT instead of the college readiness variable. We found that choosing to take the ACT was highly predictive of graduation, with ACT takers having odds of high school completion more than 13 times higher than their peers who did not take the test, which supports our theory that test-taking may be the more powerful predictor for these cohorts. For future cohorts with universal ACT participation, we can revisit the predictive power of meeting the college readiness benchmarks.



Model #4: ACT Math College Readiness Benchmark

For these regressions, we restricted the population of included students to those who had taken the ACT.

	Overall	African American	Hispanic	ELL	Low income	Special education
Met ACT Math Benchmark	2.49** (0.93)		2.08 (2.38)		4.29** (3.03)	1.21 (1.49)
Female	1.96** (0.62)		0.54 (0.54)		1.53 (0.75)	8.93* (11.09)
African American	0.27*** (0.12)		N/A N/A		0.39 (0.33)	0.62 (1.07)
Hispanic	0.57 (0.31)		N/A N/A		0.33 (0.31)	0.16 (0.27)
Asian	0.54 (0.33)	Regression unavailable	N/A N/A		0.51 (0.55)	N/A N/A
Two or more races	0.66 (0.43)	– all African-American students meeting benchmark	N/A N/A	Regression unavailable – all ELL students meeting benchmark	0.85 (1.09)	N/A N/A
Special education	0.84 (0.38)		0.29 (0.43)		0.65 (0.42)	N/A N/A
ELL	0.78 (0.53)		0.38 (0.42)		1.54 (1.3)	N/A N/A
Low income	0.86 (0.35)	completed high school	0.11* (0.15)	high school	N/A N/A	0.23 (0.32)
Parent is college graduate	0.58 (0.21)		0.09* (0.12)		0.43 (0.23)	0.22 (0.31)
Single parent	0.45** (0.15)		1.9 (2.76)		0.36* (0.19)	0.06** (0.08)
Grade 9 attendance 94% or above	3.62*** (1.17)		2.77 (2.7)		2.44* (1.26)	1.45 (1.62)
Pseudo r-squared	0.1351		0.2193		0.1278	0.3519

Note: Statistically significant coefficients **bolded** for emphasis. *= $p < 0.10$, **= $p < 0.05$, ***= $p < 0.01$.

Overall, among students taking the ACT, those who met the college readiness benchmark in Math had odds of high school completion nearly 2.5 times higher than those who missed the benchmark. This result was statistically significant at $p < 0.05$. Economically disadvantaged students showed a similar result, with odds nearly 4.3 times higher, while findings for Hispanic and Special Education students were not statistically significant.

The lack of significant coefficients associated with the ACT Math college readiness benchmark for the subgroups outlined does not necessarily mean meeting the benchmark is not predictive for these groups. We restricted these regressions only to students who took the ACT, and because these cohorts graduated from MMSD before universal administration of the ACT, students in these traditionally disadvantaged subgroups took the ACT at relatively low rates. As a result, the regressions for these subgroups are relatively underpowered for ACT-related measures because there are relatively few test-taking students on which to base our analysis. Although the results are not statistically significant, the odds ratios are large and in the expected direction.

In addition, given that taking the ACT was optional for these cohorts, it is likely that meeting the college readiness benchmark after taking the ACT has less predictive power on the probability of graduation than simply choosing to take the ACT in the first place. We ran a regression that included all students in these three cohorts, substituting a binary variable that indicated whether the student had taken the ACT instead of the college readiness variable. We found that choosing to take the ACT was highly predictive of graduation, with ACT takers having odds of high school completion more than 13 times higher than their peers who did not take the test, which supports our theory that test-taking may be the more powerful predictor for these cohorts. For future cohorts with universal ACT participation, we can revisit the predictive power of meeting the college readiness benchmarks.