Center for Education Policy Research
HARVARD UNIVERSITY

STRATEGIC DATA PROJECT

SDP FELLOWSHIP
CAPSTONE REPORT 2016

Developing an Early Warning Indicator System in a High-Poverty Urban Context

Tara Chiatovich, Passaic Public Schools
Elizabeth Rivera Rodas, Passaic Public Schools

Executive Summary

Passaic Public Schools aims to ensure that at least 80% of its students graduate from high school on time. In order to develop targeted interventions that support this goal, they wanted to know: What are the characteristics associated with an increased risk of failing to graduate on time?

To predict on- time graduation, Strategic Data Project (SDP) Fellows built a logit model using Limited English Proficiency (LEP), Special Education status (IEP), test scores, free and reduced-price lunch status, attendance, GPA, gender, demographic information and retention data for Grades 4–11. While many early warning indicator systems focus on high school data, SDP Fellows built their model by starting with fourth-grade data, then added each subsequent year's data to determine whether early or later grades' predictors were more powerful. Statistically significant predictors of failing to graduate on time include gender (male), Grade 5-8 test scores, attendance, GPA, having an IEP (especially before high school) and LEP status (only in high school). Predictors from Grades 5 to 8 are more powerful predictors of on-time graduation than high school predictors alone. Now, Passaic can create interventions that target these predictors to increase the on-time graduate rate for all of their students.

Strategic Data Project Fellowship Capstone Reports

Strategic Data Project (SDP) Fellows author capstone reports to reflect the work that they led in their education agencies during the two-year program. The reports demonstrate both the impact fellows make and the role of the SDP network in supporting their growth as data strategists. Additionally, they provide recommendations to their host agency and may serve as guides to other agencies, future fellows and researchers seeking to do similar work. *The views or opinions expressed in this report are those of the authors and do not necessarily reflect the views or position of the Center for Education Policy Research at Harvard University.*

The Passaic Board of Education, a high-poverty urban school district in New Jersey, has in recent years witnessed an increase in its on-time graduation rate—that is, the percentage of high school students who graduate within four years of starting ninth grade. As shown in Table 1 below, it went from 59.9% for the 2011 graduating cohort to 79.1% for the class of 2015. Despite the improvement of nearly 20 percentage points, this most recent graduation rate was a disappointment because it just missed the target of 80%. As such, the Passaic Board of Education wishes to develop interventions to further support students' on-time graduation. To best tailor these interventions to the students they are intended to serve, the district needed to use its own data to uncover characteristics associated with an increased risk of failing to graduate on time.

Table 1

On-time Graduation Rate of Passaic Students, 2011–2015 Cohorts

Cohort	Rate	N
2011	59.9%	740
2012	61.6%	737
2013	73.2%	713
2014	77.1%	654
2015	79.1%	708

Literature Review

A large body of research has identified factors associated with dropping out of school. Some are demographic and, as such, are not factors that a school or district can change through interventions. Race/ethnicity is one such factor. According to the U.S. Department of Education's National Center for Education Statistics (Kena et al., 2015), the 2013 dropout rate among 16- to 24-year-olds in the United States was 5% for Whites, 7% for Blacks, and 12% for Hispanics. NCES, through the work of Chapman, Laird, Ifill, and KewalRamani (2011), has also identified socioeconomic status as linked to dropping out. In 2009, the dropout rate of students from low-income families was 7.4%, which is more than five times the dropout rate of 1.4% for students from high-income families. Knowing which characteristics are tied to an increased risk of dropping out can be useful for identifying students who may be in need of extra support.

Of particular interest to school districts are mutable student characteristics associated with an increased risk of dropping out. Districts may be able to design interventions to change these factors, thereby boosting the likelihood that students will graduate. In Montgomery County Public Schools in Maryland, for example, students who were suspended in or out of school at least once had a dramatically higher likelihood of dropping out. Students suspended in ninth grade were twice as likely as their peers to drop out, and those suspended in third grade were up to nine times as likely to drop out (West, 2013). This same study found additional associations between attendance and dropping out—students missing three or more days of school were up to twice as likely as other students to drop out—and between GPA and dropping out—students whose GPAs were below 3.0 were up to five times as likely to drop out. Similarly, in Providence, Rhode Island, schools, researchers found that each course failure in 10th grade was tied to a drop of 14% in the probability of graduating, and students whose attendance rate was 10 percentage points higher than their peers were 68% to 83% more likely to graduate (Becker, Hall, Levinger, Sims, & Whittington, 2014).

But to have any chance of successfully intervening to help students at risk of dropping out, districts need to be able to *accurately identify* those students. Accuracy can fail in two respects: Students who are able to graduate without additional support may be misidentified as at risk of dropping out, and students who will drop out may not be identified as at-risk. Whereas incorrectly identifying students as at-risk can cause districts to spend money without getting a return on their investment (Dynarski & Gleason, 2002), failing to identify students who actually will go on to drop out limits the ability of districts to offer them needed support and thus limits the likelihood that interventions will improve graduation rates (Bowers, Sprott, & Taff, 2013).

Despite these negative consequences of misidentifying potential dropouts, the accuracy of dropout flags—characteristics used to identify which students are at risk of dropping out—is generally low (e.g., Bowers et al., 2013; Dynarski & Gleason, 2002). Early Warning Indicator Systems—a data-based tool designed to identify students who are at risk of not meeting key academic milestones—typically do a poor job of determining which students would benefit from interventions to help them graduate from high school. Clearly, if a system using a state or district's own data falls short in this regard, relying solely on factors linked generally to dropping out of high school found in the research literature to identify at-risk students would be insufficient. To have any real prospect of boosting graduation rates, states and school districts need to analyze their own data to discover which factors among their *own* students are tied to dropping out of high school.

The current study uses several years of data from former Passaic, New Jersey, students. The factors it identifies as linked to the probability of on-time graduation are based on the population for which they are intended. They should therefore provide better predictive power when applied to other students in the same district, enabling the district to target future interventions to a greater percentage of students who actually are at risk of dropping out. In this manner, students who actually need support to graduate are more likely to get it with less waste of district funds on interventions provided to students who would still graduate without them.

The District

Passaic is a mid-sized urban school district in New Jersey serving a high proportion of minority and low-income students. Its students typically number around 14,000 in an academic year. In addition to kindergarten through Grade 12, Passaic offers two years of free pre-kindergarten classes. It has a number of elementary schools but only one middle school and one high school, with a school for gifted and talented students serving Grades 2–8. Passaic is one of 35 districts in the state belonging to District Factor Group A, the lowest of eight classifications based on socioeconomic factors such as median income, rate of unemployment, and rate of poverty of residents in the district (State of New Jersey Department of Education, 2014).

Table 2 presents descriptive statistics on Passaic students as reported to the New Jersey State Department of Education for the 2015–2016 school year. The district has a high proportion of English language learners (23.49%), and the vast majority of students—over 99%—are non-White. More than one-tenth of students are in the special education program. Passaic is among New Jersey's lower-performing districts on state standardized tests.

Table 2

Demographics of Passaic Students in 2015–2016 School Year (N=14,605)

Group	Percentage
Race/Ethnicity	
Hispanic	92.28%
Asian	1.87%
Black	4.87%
White	0.78%
Other	0.20%
Special Education	12.34%
Limited English Proficiency	23.49%

Available Data and Analysis

The 3,478 students for the study had anticipated graduation dates from 2011 through 2015. All had reached at least the ninth grade, had not transferred out of the district to attend school in another district during high school, and had not transferred into the district after ninth grade. The state's data contained an exit code indicating whether students graduated, transferred out of the district, or dropped out. We used to the exit code to create an indicator variable for on-time graduation (within four years of starting high school), which served as the outcome for logistic regression models.

The data compiled by the state also included information on each student's race/ethnicity, gender, completion of an individual education plan (IEP) for any given year; date of entrance to the limited English proficiency (LEP) program (if applicable); date of exit of the LEP program (if applicable); eligibility for free or reduced-price lunch (FRPL) for any given year; standardized test scores in English language arts (ELA), math, and science for fourth and eighth grade; and repetition of a grade in the prior year for any given year. We excluded from our analyses 399 students who had IEPs at any point and who graduated from high school despite not having passed the required high school exit exam for New Jersey, the HSPA. Since some special education students had different requirements for graduating high school than their general education peers, on-time graduation would mean something different for them than for the population of students as a whole, which would affect results.

We supplemented data supplied by the state with data from the district's data storage and retrieval system, PowerSchool, to determine the number of days students were present, the number of days they were enrolled in the district, course enrollment (e.g., the number of math, ELA, and overall credits completed), and grades earned in their courses. These data were available for each given year for a select set of school years. Additionally, we recorded which students had more than one row of data for a given grade (e.g., had two separate sets of values for fourth grade) in the state's data to supplement information on grade repetition provided by the state.

As its name suggests, one important purpose of an Early Warning Indicator System is to identify as early as possible which students are at risk of dropping out of school. In the case of the current study, however, the available data varied by graduation cohort and grade level and therefore limited how early we could predict the risk of not graduating on time. As a general rule, later grade levels and later cohorts had the most data available, whereas early cohorts—and, for some variables, early grades for all cohorts—had the vast majority of data missing. As an additional complication, some information provided by the state did not distinguish between missing data and students who did not belong to the group (e.g., IEP, repeated grade); these variables were excluded from analyses if the number of students identified as belonging to the group was small (fewer than 30) for all cohorts.

In light of these limitations, we wanted to build multiple sets of models to inform the district about specific student characteristics and performance indicators that occur at different points in a child's schooling. For the bulk of the students in our sample, fourth grade is the earliest point that the state issues standardized tests for math, English language arts, and science, and eighth grade marks the last grade before students enter high school. Thus, we decided to focus on these grades when building our models, which include only predictors that take place before students enter high school.

For the models in the high school grades, we focused on Grades 9, 10, and 11, since those grades, unlike Grade 12, give the district some time to take action before the goal of on-time graduation has been reached or has failed to be reached. Fewer than 10 students in the sample had data from before Grade 4, so the Grade 4 variables are specific to this grade only. Some variables for Grade 8, however, accumulated information for Grades 5 through 8. Specifically, any indicator variables for Grade 8 represented whether a student belonged to the group for any year from Grades 5–8. (See Endnotes 1 and 2 for details on variable values.)

Using the data described above, we created a dataset with one row per student with variables broken out by grade. To illustrate, a student who had no missing data for the LEP variables had one LEP variable for each of the following: (a) Grade 4; (b) Grades 5–8; (c) Year 1 of high school; (d) Year 2 of high school; and (e) Year 3 of high school. This format avoided the nesting that occurs with person–period datasets, where each student has one row of data for each year in the study.

To make the best use of all available data, we ran two separate sets of analyses: one for all cohorts and one for the 2015 cohort alone. For both sets of analyses, we used all available (i.e., non-missing) cases. This approach allowed us to take advantage of the greater statistical power available for some predictors while still allowing us to examine the predictors relating to performance and behavior in high school that were only available for the 2015 cohort.

Time-Invariant Variables[1]

The variables that were consistent through all grade levels for students were: (a) on-time graduation, defined as the student graduating within four years of starting high school; (b) gender (male); (c) race/ethnicity—Black, Asian, Hispanic, White, and Other; and (d) graduation cohort, from 2011 to 2015, defined as the expected year of graduation, which was four school years after the student's ninth grade year. Table 3 below gives descriptive statistics on each of these variables.

Table 3

Percentages for Time-Invariant Variables (*N* = 3,478)

Variable	Percentage
On-time Graduation	70.36%
Male	48.04%
Race/Ethnicity	
Black	7.91%
Asian	3.48%
Hispanic	87.55%
White	1.04%
Other	0.03%
Cohort	
2011	20.67%
2012	20.50%
2013	18.49%
2014	18.48%
2015	20.07%

* = Count of students belonging to group was too small to include variable in analyses.

Grade-Specific Variables[2]

The majority of variables used in analyses were specific to a particular grade level or range of grade levels. For Grades 4 and 5 through 8, they included: (a) LEP for the given grade(s); (b) IEP for the given grade(s); (c) ELA, math, and science state standardized test scores for Grade 4 or Grade 8, converted to z-scores by subtracting the mean and dividing by the standard deviation for ease of interpretation; and (d) repeated grade in Grade 4 or Grades 5–8. In addition, FRPL eligibility was available for Grades 5–8. Note that variables pertaining to group belonging (e.g., LEP, IEP, repeated grade) for Grade 4 indicate whether students belonged to the group in Grade 4 only, whereas for Grades 5–8 they indicate whether students belonged to the group at any point in Grades 5–8.

For Grades 9, 10, and 11, the available variables were: (a) LEP for the given grade; (b) IEP for the given grade; (c) FRPL for the given grade; (d) percentage of days present for the given grade, computed by dividing the number of days the student was present by the number of days the student was enrolled in the district and then multiplying by 100; (e) overall GPA for the given grade; (f) GPA for ELA classes for the given grade; (g) GPA for math classes for the given grade; and (h) on-track indicator, defined as having the expected number of course credits overall and in math and ELA given the student's year in high school.

Table 4 below provides information on the variables that were available by cohort and by grade level, along with a description of each one. It is important to note that predictors involving a student's behavior and performance in high school—percentage present, GPA overall and by subject, and on-track status—were available for all years of high school for only the 2015 cohort. Not shown is that most variables had a sizeable portion of missing values.

Table 4

Descriptive Statistics and Availability of Data by Grade and Cohort

Variable	Availability of Data by Cohort 2011	2012	2013	2014	2015	Mean (SD)/ % across All Cohorts
Grade 4						
LEP				X	X	10.91%
IEP					X	0.25%*
ELA Standardized Test Scores			X	X	X	198.00 (25.27)
Math Standardized Test Scores						202.61 (31.85)
Science Standardized Test Scores				X	X	199.14 (28.77)
Repeated Grade 4					X	00.09%*
Grades 5–8						
LEP	X	X	X	X	X	23.28%
IEP		X	X	X	X	13.81%
ELA Standardized Test Scores		X	X	X	X	202.65 (20.49)
Math Standardized Test Scores		X	X	X	X	199.76 (36.70)
Science Standardized Test Scores		X	X	X	X	208.50 (25.06)
FRPL		X	X	X	X	96.11%
Repeated Grade 5–8	X	X	X	X	X	22.10%
Grade 9						
LEP	X	X	X	X	X	14.89%
IEP	X	X	X	X	X	15.03%
FRPL	X	X	X	X	X	88.77%
Percentage Present					X	94.64% (5.37)
GPA					X	2.87 (0.85)
ELA GPA					X	2.94 (1.55)
Math GPA					X	2.57 (1.47)
On-Track Indicator					X	72.28%
Grade 10						
LEP	X	X	X	X	X	15.77%
IEP	X	X	X	X	X	13.79%
FRPL	X	X	X	X	X	86.44%
Percentage Present				X	X	93.98 (5.63)
GPA				X	X	3.47 (0.83)
ELA GPA				X	X	3.39 (1.30)
Math GPA				X	X	3.26 (1.54)
On-Track Indicator				X	X	72.73%
Grade 11						
LEP	X	X	X	X	X	16.42%
IEP	X	X	X	X	X	13.30%
FRPL	X	X	X	X	X	84.39%
Percentage Present			X	X	X	93.22 (6.51)
GPA			X	X	X	2.51 (0.80)
ELA GPA			X	X	X	2.25 (1.07)
Math GPA			X	X	X	1.83 (1.20)
On-Track Indicator			X	X	X	76.05%

*** = Count of students belonging to group too small to include variable in analyses.**

Analysis Plan

We carried out the steps for building the models separately for the data from (a) each grade (five grade levels); (b) all cohorts; and (c) the 2015 cohort only (two cohort groups). This resulted in a total of 10 separate sets of models. The specific steps were as follows:

1. Build a model with male, demographics, and grade-level specific information on FRPL, grade retention, IEP, and LEP.
2. Add in predictors measuring academic performance when available.
3. Remove the predictors that were not statistically significant from the model for parsimony, resulting in a final model that only included statistically significant time-invariable predictors and predictors specific to the grade level(s).

Following the creation of one final model for each grade level—for all cohorts and for the 2015 cohort separately—we carried out the following additional set of steps to build a cumulative model representing all grades:

1. Add the predictors from the final Grade 4 model to the final model for Grades 5–8.
2. Remove non-significant predictors for parsimony.
3. Add the statistically significant predictors from the cumulative model covering Grade 4 and Grades 5–8 to the final model for Grade 9.
4. Once again, remove non-significant predictors for parsimony.
5. Repeat the process outlined above, adding statistically significant predictors from cumulative models for prior grades to the final models for Grade 10 and then again for Grade 11.

In this manner, we developed a model—for both the data overall and just the 2015 cohort—that was completely cumulative with the possibility of including predictors from all of the grades covered in our models, provided that they remained statistically significant upon controlling for the remaining predictors in the model. The purpose of having a final model with only predictors specific to the grade or grade range for the model as well as an additional model with predictors from prior grades was to give the district as much actionable information as possible.

Results: What Predicts On-Time Graduation

Since the results reported here are from logistic regression models, we give findings in terms of odds ratios, which can be difficult to interpret. An odds ratio greater than 1.0 signals that the variable is associated with a greater likelihood of on-time graduation, and an odds ratio less than 1.0 signals that the variable is linked to a lesser likelihood of on-time graduation. To help with interpretation, we include graphs of predicted probabilities for each model. Graphs of models that had male or race/ethnicity as statistically significant predictors hold male constant at the mean and show results for Hispanic students, who represented the vast majority of students. For graphs showing the predicted probability according to whether students passed or failed state standardized tests, 150 was used as a failing test score and 225 was used as a passing test score.

Results Using Available Data from All Cohorts

Table 5 presents the results from the final model for each grade level using data from all cohorts. Below, we interpret the results separately for each model. The table points to some patterns across the grades, holding constant predictors present in the given model. First, both fourth and eighth grade math test scores were linked to on-time graduation, in that students with higher test scores in both grades were generally more likely to graduate on time. Students' LEP status predicted on-time graduation, but only in high school. In other words, students with limited English proficiency in high school were, on the whole, less likely to graduate on time, but limited English proficiency prior to high school was unrelated to on-time graduation after adjusting for other predictors in the models. Across all grades, males had lower predicted probability of on-time graduation than females. Having an IEP in any of the grades studied was linked to a lower estimated likelihood of on-time graduation. The one caveat to this pattern was in Grade 4, where the count of students with IEPs was too low to include this variable in the model.

With the exception of Grade 4, race/ethnicity related to the probability of on-time graduation, with Asian students being more likely overall to graduate on time than Black and Hispanic students. (See the separate interpretation of each model below for additional information on race/ethnicity.) The Grade 4 model had considerably fewer students in it than the other models, limiting its ability to detect links with on-time graduation and providing a possible explanation for why some trends across other grades did not hold for the fourth grade. A variable that was not significant in any model was FRPL status, which was unrelated in all grades to the likelihood of on-time graduation.

Table 5

Results from Final Models for All Cohorts Predicting On-Time Graduation

	All Cohorts				
Predictors	**Grade 4**	**Grades 5–8**	**Grade 9**	**Grade 10**	**Grade 11**
Intercept	5.37*** (0.62)	2.68*** (0.55)	2.36*** (0.38)	3.02*** (0.52)	3.81*** (0.68)
Male	0.42*** (0.06)	0.53*** (0.06)	0.67*** (0.06)	0.65*** (0.06)	0.66*** (0.06)
Asian	--	27.38** (28.51)	52.76*** (53.78)	9.45*** (4.64)	5.81*** (2.38)
Hispanic	--	1.56* (0.32)	1.30 (0.21)	1.27 (0.22)	1.39 (0.25)
White	--	1.14 (0.70)	1.86 (1.13)	1.40 (0.77)	1.88 (1.24)
FRPL in specified grade(s)	NA	--	--	--	--
IEP in specified grade(s)	NA	0.46*** (0.10)	0.22*** (0.04)	0.19*** (0.04)	0.19*** (0.05)
LEP in specified grade(s)	--	--	0.25*** (0.05)	0.48*** (0.05)	0.43*** (0.05)
Repeated year in specified grade(s)	NA	0.25*** (0.05)	NA	NA	NA
ELA test z-score for specified grade	--	1.52*** (0.13)	NA	NA	NA
Math test z-score for specified grade	1.32** (0.14)	1.72*** (0.15)	NA	NA	NA
Science test z-score for specified grade	1.29* (0.13)	--	NA	NA	NA
N for model	1,084	1,868	2,583	2,606	2,591

Note. ELA, math, and science tests were administered in eighth grade for the Grade 5–8 model. For the Grade 5–8 model, the LEP, IEP, and repeated year variables represent students who qualified in any grade from fifth through eighth grade. See Endnotes 1 and 2 for details on variable values. Standard error appears in parentheses.
NA = variable not available for the specified grade.
-- = variable dropped from model due to non-significance.
* p < .05. **p < .01. ***p < .001.

Model for Grade 4, all cohorts. For Grade 4, being male was associated with a lower likelihood of graduating on time, and higher fourth grade math and science test scores were associated with a higher likelihood of graduating on time. Specifically, the fitted odds that a male student would graduate on time were .42 times the fitted odds that a female student would graduate on time with other predictors in the model held constant. For each positive difference of one standard deviation on Grade 4 math and science standardized test scores, the fitted odds of on-time graduation were 1.32 and 1.29 times those of students with lower scores, respectively, controlling for the other predictors in the model. Race/ethnicity, LEP, and Grade 4 ELA standardized test scores were not statistically significant, and, as noted in Table 4 above, having an IEP, FRPL status, and repeating Grade 4 did not have sufficient counts to be evaluated in the model. Figure 1 provides a visual for the model.

Model for Grades 5–8, all cohorts. For the eighth grade model, being male, having an IEP in Grades 5–8, and repeating a year in Grades 5–8 were all linked to a lower likelihood of on-time graduation. Higher scores on the Grade 8 ELA and math standardized tests were associated with a higher likelihood of on-time graduation, as were Asian and Hispanic race/ethnicity compared to Black race/ethnicity. The fitted odds of graduating on time for male students were .52 times those for female students, and for students with an IEP in Grades 5–8 the fitted odds were .46 times the fitted odds of students without an IEP in any of those grades. For each positive difference of one standard deviation in the Grade 8 ELA and math standardized tests, the fitted odds were 1.51 and 1.72 times those of students with lower test scores, respectively.

For the race/ethnicity variable, we ran a Wald test on the null hypothesis that the coefficients for all race/ethnicity groups were jointly 0, which rejected ($p < 0.01$). We ran additional Wald tests to see which groups differed significantly from others in the probability of graduating on time. The fitted odds of Asian and Hispanic students graduating on time were 27.38 and 1.56 times the fitted odds of Black students graduating on time, respectively; Asian students had a higher probability of graduating on time than either Hispanic or White students. Other differences in the probability of on-time graduation were non-significant between the race/ethnicity groups. These results held constant the other predictors in the model. LEP and FRPL status in Grades 5–8 and Grade 8 ELA standardized test scores were not statistically significant. Figure 2 graphically represents predicted probabilities for the model.

Model for Grade 9, all cohorts. The results for the ninth, 10th, and 11th grade models were all similar to each other: being male, having an IEP, and designation as LEP were all linked to a lower probability of on-time graduation, and Asian students had a higher probability of on-time graduation than Black students. The Grade 9 model revealed that the fitted odds of on-time graduation were .67 times as high for male students as for female students, .22 times as high for students with an IEP as for students without an IEP, and .59 times as high for LEP students as for non-LEP students in Grade 9.

A Wald test rejected the null hypothesis that the coefficients for the race/ethnicity groups were jointly 0 ($p < .001$). The fitted odds of Asian students graduating on time were 52.76 those of Black students, and additional Wald tests revealed that Asian students also had significantly higher fitted odds of on-time graduation than either Hispanic or White students. Other differences between race/ethnicity groups in the probability of on-time graduation were not statistically significant, nor was FRPL status in Grade 9. These results partial out the other predictors in the model. Figure 3 provides a visual for the model.

Grade 10 model, all cohorts. For the Grade 10 model, the fitted odds of on-time graduation were .65 times as high for male students as for female students, .19 times as high for students with an IEP in Grade 10 as for students without an IEP in Grade 10, and .48 times as high for LEP students as for non-LEP students in Grade 10. A Wald test of the null hypothesis that all coefficients for race/ethnicity were jointly 0 rejected ($p < 0.001$); the fitted odds of Asian students graduating on time were 9.45 times the fitted odds of Black students graduating on time, with additional Wald tests revealing that Asian students also had significantly higher probability than White and Hispanic students of graduating on time. Other differences between race/ethnicity groups in the probability of on-time graduation were not statistically significant, and neither was FRPL status in Grade 10. These results control for the other predictors in the model. Figure 4 displays predicted probabilities for the model.

Grade 11 model, all cohorts. Finally, for the Grade 11 model, the fitted odds of on-time graduation were .66 times as high for male students as for female students, .19 as high for students with an IEP as for students without and IEP, and .43 times as high for LEP students as for non-LEP students in Grade 11. As with the other high school models, the Wald test that the coefficients for race/ethnicity were jointly 0 rejected ($p < .001$); the fitted odds of Asian students graduating on time were 5.81 times the fitted odds of Black students graduating on time. Additional Wald tests demonstrated that the fitted odds of Asian students graduating on time were significantly higher than those of Hispanic students, but other differences in the likelihood of on-time graduation between race/ethnicity groups were not statistically significant. FRPL status in Grade 11 was not statistically significant either. These findings hold constant the other predictors in the model. Figure 5 graphically displays predicted probabilities.

Cumulative model for all grades, all cohorts. As described in the analytic plan above, we also built a cumulative model by adding the statistically significant predictors from the model for the prior grade level to the final model for each given grade level. This process for the models using data from all available cohorts resulted in only the predictors from the model for Grades 5–8 being statistically significant. In other words, when the statistically significant predictors from the model for Grade 4 were added to the model for Grades 5–8, the Grade 4 predictors were no longer statistically significant. Similarly, when statistically significant predictors from the model for Grades 5–8 were added to each of the final models for the high school grades, the predictors specific to the high school grades were no longer statistically significant. Thus, the model for Grades 5–8 also serves as the final cumulative model when using data for all available cohorts.

Results Using Data from the 2015 Cohort Only

Table 6 provides the results from the final model for each grade level using data from the 2015 cohort only. We interpret the results separately for each model below. Looking at the results across grades for the 2015 cohort points to some trends, with findings holding constant the predictors that are in the given model.

Scores on both the Grade 4 and the Grade 8 math test were associated with on-time graduation; students with higher scores on either test had a greater likelihood overall of graduating on time. Having an IEP was linked to a lower likelihood of on-time graduation across all grades, with the exception of Grade 4, where the number of students with an IEP was too low to include that variable in the model. For all three grades for which GPA was available, having a higher overall GPA was linked to a higher likelihood of on-time graduation. GPA in ELA and math, however, was not related to on-time graduation after adjusting for other predictors in the model. Finally, for Grades 9 and 10, attendance rate was related to on-time gradation, with a higher percentage of days present predicting a greater likelihood of graduating on time. Neither race/ethnicity nor FRPL status was linked to on-time graduation in any grade for the 2015 cohort.

Table 6

Results from Final Models for 2015 Cohort Predicting On-Time Graduation

	2015 Cohort				
Predictors	**Grade 4**	**Grades 5–8**	**Grade 9**	**Grade 10**	**Grade 11**
Intercept	6.95*** (1.43)	10.09*** (2.19)	0.00* (0.00)	0.00** (0.00)	0.22** (0.12)
Male	0.28** (0.07)	0.26*** (0.06)	--	--	--
Asian	--	--	--	--	--
Hispanic	--	--	--	--	--
White	--	--	--	--	--
FRPL in specified grade(s)	NA	--	--	--	--
IEP in specified grade(s)	NA	0.41* (0.16)	0.09*** (0.05)	0.04*** (0.02)	0.02*** (0.01)
LEP in specified grade(s)	--	--	--	--	--
Repeated year in specified grade(s)	NA	0.27*** (0.09)	NA	NA	NA
ELA test z-score for specified grade	--	--	NA	NA	NA
Math test z-score for specified grade	1.77*** (0.25)	1.98*** (0.36)	NA	NA	NA
Science test z-score for specified grade	--	1.55* (0.29)	NA	NA	NA
GPA for specified grade	NA	NA	3.41*** (0.65)	3.56*** (0.73)	4.85** (1.60)
ELA GPA for specified grade	NA	NA	--	--	--
Math GPA for specified grade	NA	NA	--	--	--
Percentage present for specified grade	NA	NA	1.10* (0.05)	1.08* (0.03)	--
On-track at end of specified grade	NA	NA	__	__	7.62*** (4.16)
N for model	399	516	456	450	444

Note. ELA, math, and science tests were administered in eighth grade for the Grade 5–8 model. For the Grade 5–8 model, the LEP, IEP, and repeated year variables represent students who qualified in any grade from fifth through eighth grade. See Endnotes 1 and 2 for details on variable values. Standard error appears in parentheses.
NA = variable not available for the specified grade.
-- = variable dropped from model due to non-significance.
* p < .05. **p < .01. ***p < .001.

Model for Grade 4, 2015 cohort. The model for Grade 4 for the 2015 cohort revealed that male students were less likely than female students to graduate on time, and that higher scores on the Grade 4 math standardized tests were associated with a higher probability of on-time graduation. Specifically, the fitted odds of male students graduating on time were .28 times those of female students. Each positive difference of one standard deviation was linked to 1.77 times the fitted odds of on-time graduation compared to students with lower scores. Race/ethnicity, male, LEP, and Grade 4 ELA and science standardized test scores were not statistically significant, and, as noted in Table 4 above, having an IEP, FRPL status, and repeating a year in Grade 4 did not have sufficient counts to be included in the model. Figure 6 gives a visual display of these findings.

Model for Grades 5–8, 2015 cohort. The model for Grades 5–8 using only data from the 2015 cohort revealed that male students, students with an IEP in Grades 5–8, and students who repeated a year in Grades 5–8 had significantly lower probability of on-time graduation than their peers, while higher Grade 8 math and science standardized test scores were both linked to significantly higher probability of on-time graduation. The fitted odds of on-time graduation were .26 times as high for male students as for female students, .41 times as high for students with an IEP in Grades 5–8 as for students without an IEP, and .27 times as high for students who repeated a year in Grades 5–8 as for those who did not. For each positive difference of one standard deviation on Grade 8 math and science standardized test scores, the fitted odds of on-time graduation were respectively 1.98 and 1.55 times those of students with lower scores. These results partial out the other predictors in the model. Race/ethnicity, LEP in Grades 5–8, Grade 8

ELA standardized test scores, and FRPL status in Grades 5–8 were not statistically significant for the 2015 cohort. Figure 7 visually interprets the model.

Model for Grade 9, 2015 cohort. The Grade 9 model for the 2015 cohort demonstrates that having an IEP in ninth grade was linked to a lower probability of on-time graduation and that a higher percentage of days present and GPA in ninth grade were linked to a higher probability of on-time graduation. Students with an IEP in ninth grade had fitted odds of on-time graduation that were .09 times those of students without an IEP. For each positive difference of one percentage point in the percentage of days present, the fitted odds of graduating on time were 1.10 times those of students with a lower percentage of days present. For each positive difference of one point in students' ninth grade GPAs, the fitted odds of graduating on time were 3.41 times those of students with lower GPAs. These findings hold constant the other predictors in the model. On-track status, race/ethnicity, LEP, FRPL status, and ELA and math GPA in Grade 9 were not statistically significant for the 2015 cohort. Figure 8 provides a visual for the model.

Model for Grade 10, 2015 cohort. Similar to the Grade 9 model for the 2015 cohort, the Grade 10 model for this cohort showed that having an IEP in 10th grade corresponded to a lower likelihood of on-time graduation and that a higher percentage of days present and higher GPA were associated with a greater likelihood of on-time graduation. Students with an IEP in 10th grade had fitted odds that were .04 those of students without an IEP. Each positive difference of one percentage point in percentage of days present was associated with 1.08 times the fitted odds of graduating on time compared to students with a lower percentage of days present in 10th grade. Each one point positive difference in GPA in 10th grade corresponded to 3.56 times the fitted odds of on-time graduation compared to students with a lower GPA. These results are net of the other predictors in the model. On-track status, race/ethnicity, LEP, FRPL status, and ELA and math GPA in Grade 10 were not statistically significant for the 2015 cohort. Figure 9 displays predicted probabilities from the model.

Model for Grade 11, 2015 cohort. For the 2015 cohort, the Grade 11 model showed a lower likelihood for 11th graders with an IEP to graduate on time than those without one. Being on track at the end of 11th grade and having a higher GPA in 11th grade related to a higher likelihood of on-time graduation. Students with an IEP in 11th grade had fitted odds of graduating on time that were .02 times those of students without an IEP, and on-track 11th graders had fitted odds of on-time graduation that were 7.62 times those of their peers. Each positive difference of one point in GPA for Grade 11 was linked to 4.85 times the fitted odds of on-time graduation, compared to students with lower GPAs. These findings hold constant the other predictors in the model. For the 2015 cohort, percentage of days present, race/ethnicity, LEP, FRPL status, and ELA and math GPA in Grade 11 were not statistically significant. Figure 10 gives a visual for the model.

Cumulative model, 2015 cohort. The cumulative model using statistically significant predictors from all grades for the 2015 cohort revealed a lower probability of on-time graduation for students with an IEP in 10th grade. It also revealed a higher probability of graduating on time for students with higher GPAs in Grades 9, 10, and 11 and for students with higher science test scores in Grade 8. Students with an IEP in 10th grade had fitted odds of on-time graduation that were .02 times those of students without an IEP in 10th grade. Each positive difference of one point in Grade 9, 10, and 11 GPAs was associated with 1.94 times, 1.85 times, and 6.18 times the fitted odds of graduating on time, respectively, compared to students with lower GPAs in those grades. Each difference of one standard deviation in Grade 8 science test scores was associated with 2.07 times the fitted odds of on-time graduation, compared to students with lower Grade 8 science test scores. For the 2015 cohort, none of the statistically significant predictors from the model for Grade 4 (male, Grade 4 math test scores) were statistically significant when added to the cumulative model. Other predictors that were statistically significant from the models for Grades 5–8, Grade 9, Grade 10, and Grade 11 lost statistical significance when added to the cumulative model for the 2015 cohort. Figure 11 provides a graphic interpretation of the model.

Table 7

Results from Cumulative Model, 2015 Cohort, Predicting On-Time Graduation

Predictors	**Cumulative for 2015 Cohort**
Intercept	0.01*** (0.01)
IEP, Grade 10	0.02*** (0.02)
GPA, Grade 11	6.19*** (2.20)
GPA, Grade 10	1.85* (0.48)
GPA, Grade 9	1.94* (0.52)
Grade 8 science test z-score	2.07* (0.61)
N for model	410

Note. Standard error is shown in parentheses.
* $p < .05$. ***$p < 0.001$.

Lessons Learned

In Passaic, factors present before students start high school matter when gauging their probability of graduating on time. In this study, the predictors of on-time graduation from Grades 5–8 trumped predictors from the high school grades. That is, predictors from Grades 9, 10, and 11 no longer had a discernable link to on-time graduation after adjusting for the Grade 5–8 predictors. One likely contributing factor is that the analyses for Grades 5–8 combined information on IEP, LEP, and grade repetition over several grades instead of examining just one grade at a time, which likely made them more powerful predictors. Additionally, a limited number of predictors—LEP, IEP, and FRPL—were available for the high school grades, which in turned limited analyses for these grades. Nevertheless, it is noteworthy that having an IEP prior to reaching high school, which is linked to a lower likelihood of graduating on time, does a better job of predicting on-time graduation than does having an IEP in any single year in high school.

Similarly, analyses that made use of predictors involving high school behavior and performance for the 2015 cohort showed that data that precede high school enrollment can contribute to predictions of on-time graduation over and above data from the high school grades. Specifically, higher Grade 8 science test scores were linked to a greater probability of on-time graduation, even after adjusting for GPA in Grades 9, 10, and 11. An additional lesson from these analyses is that GPA in high school does a better job of predicting on-time graduation than attendance or LEP in high school; after adjusting for GPA in Grades 9, 10, and 11, neither attendance nor LEP status in the high school grades related significantly to on-time graduation. That GPA in high school would matter is not surprising since students must pass their classes to graduate. However, many studies have found a link between attendance and high school graduation (e.g., Beckeret al., 2014; Doss, 1986; Ekstrom, Goertz, Pollack, & Rock, 1986; West, 2013), so it is helpful for the district to know that GPA does a better job than attendance of identifying which students are at risk of not graduating on time.

An important takeaway from this project is the need for data collection at the district level. Many variables collected and stored by the state—such as state standardized test scores and whether students have had IEPs or have participated in the LEP program—were predictive of on-time graduation. Nevertheless, three out of four of the variables in Grades 9, 10, and 11 that remained statistically significant in analyses that made use of data from all grade levels for the 2015 cohort were from the district's own data system, not from the state. Moreover, the variables that were actionable at the high school level—including GPA, attendance, and whether students were on-track—were all from the district's own data system. These variables were limited to only the 2015 cohort, however, because the district did not collect data on them for enough years to include other cohorts as well. A number of factors have been linked to high school graduation in other research, including disciplinary action such as suspension (e.g., Becker et al., 2014; Doss, 1986; Ekstrom et al., 1986; West, 2013). These additional factors are not being collected and stored by the district in a way that would make them reliable and available for analyses. This limits the district's ability to accurately identify which students are at risk of not graduating on time and to intervene to help them.

Next Steps and Challenges

A number of next steps would help the district continue the work of identifying students in need of support to graduate on time. First, the analyses can be rerun at the end of the 2015–2016 school year with the same variables used here. This would add an additional cohort to the models run on all cohorts and include analyses that make use of predictors involving high school behavior and performance currently available only for the 2015 cohort. Second, the district can add to its data collection information on disciplinary action such as suspension at all grade levels, thus making it available to include in future analyses. Finally, the most important and obvious next step is for the district to make full use of the work presented here by designing and implementing interventions to support students who are at risk of not graduating on time. This will clearly require substantial effort and resources and may take more than one school year to put into place, but it will be worthwhile. Moreover, how Passaic carries out this step can be informative to other school districts seeking to act on the information gleaned from data.

As an important note of caution, these analyses show correlation and not causation. We do not yet know, for example, whether a successful intervention to boost math skills in fourth and eighth grade would improve students' chances of graduating on time. Another area that merits attention is the distinction between student factors that lend themselves to intervention, such as test scores and GPA, and those that do not, such as gender and race/ethnicity. Even though the school district cannot intervene to change a factor like gender, we have included such variables in our analyses for two main reasons: to provide as much information as possible to the Passaic Board of Education and to ensure that the findings on student factors that lend themselves to interventions, such as the link between test scores and on-time graduation, are adjusted for these background characteristics. The focus of Passaic's efforts to boost on-time graduation should rest on what they can change, and they should use data to monitor efforts to gauge whether their interventions have the intended effect.

The biggest challenge to continuing the work presented here is the quality and continued availability of data. Some of the earliest predictors of whether students graduate on time come from state standardized tests—the NJAsk—that are no longer administered. The state now uses the PARCC test for ELA and math. This challenge is not an immediate one since several additional cohorts of students will have taken NJAsk ELA and math tests in Grades 4 and 8. Nevertheless, analyses that examine common ground between PARCC and NJAsk would be helpful.

An additional data challenge comes from the way that the state stores some of its data, which can render impossible the task of distinguishing between missing data and students who do not belong to a particular group. A prime example is the LEP variable. We created this variable by flagging any student who had a start date for the LEP program before the school year of interest but whose end date for the LEP program came after the school year. Students who were missing a start date were presumed not to be in the LEP program, but records of tests of English proficiency sometimes contradicted our conclusions. Thus, these variables contained a lot of noise that limited the accuracy of our analyses. This data challenge presents another reason for the district to add to the data it collects and stores in its own data system.

References

Becker, J., Hall, L., Levinger, B., Sims, A., & Whittington, A. (2014). *Student success and college readiness: Translating predictive analytics into action.* Cambridge, MA: Strategic Data Project.

Bowers, A., Sprott, R., & Taff, S. (2013). Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity and specificity. *The High School Journal, 96*(2), 77–100.

Chapman, C., Laird, J., Ifill, N., & KewalRamani, A. (2011). Trends in high school dropout and completion rates in the United States: 1972–2009 (NCES 2012-006). Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch

Doss, D. A. (1986). *Ninth grade course enrollment and dropping out*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. Retrieved from http://files.eric.ed.gov/fulltext/ED273678.pdf

Dynarski, M., & Gleason, P. (2002). How can we help? What we have learned from recent federal dropout prevention evaluations. *Journal for Education of Students Placed at Risk, 7*, 43–69.

Ekstrom, R., Goertz, M., Pollack, J., & Rock, D. A. (1986). Who drops out and why? Findings from a national study. *Teachers College Record, 87*, 356–373.

Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., … Dunlop Velez, E. (2015). *The condition of education 2015* (NCES 2015-144). Washington, DC: National Center for Education Statistics.

State of New Jersey Department of Education (2014). *District factor groups*. Retrieved from http://www.state.nj.us/education/schools/achievement/dfg.htm

West, T. C. (2013). *Just the right mix: Identifying potential dropouts in Montgomery County Public Schools using an early warning indicators approach*. Rockville, MD: Montgomery County Public Schools. Retrieved from http://new.every1graduates.org/wp-content/uploads/2014/01/Just-the-Right-Mix_MCPS_West2013.pdf

Notes

[1] The values for time-invariant indicator variables used in the models are as follows:

- **On-time graduation:** 0 = did not graduate on time, 1 = graduated on time
- **Male:** 0 = female, 1 = male
- **Black, Asian, Hispanic, White:** 0 = student does not belong to race/ethnicity group, 1 = student belongs to race/ethnicity group

[2] The values for grade-specific indicator variables used in the models are as follows:

- **LEP for Grades 4, 9, 10, and 11:** 0 = not LEP in specified grade, 1 = LEP in specified grade
- **LEP for Grades 5–8:** 0 = not LEP ever in Grades 5–8, 1 = LEP at some point in Grades 5–8
- **IEP for Grades 4, 9, 10, and 11:** 0 = no IEP in specified grade, 1 = IEP in specified grade
- **IEP for Grades 5–8:** 0 = no IEP ever in Grades 5–8, 1 = IEP at some point in Grades 5–8
- **Repeated grade for Grade 4:** 0 = did not repeat Grade 4, 1 = repeated Grade 4
- **Repeated grade for Grades 5–8:** 0 = did not repeat any grade during Grades 5–8, 1 = repeated one or more grade during Grades 5–8
- **FRPL for Grades 5–8:** 0 = did not qualify for FRPL any year during Grades 5–8, 1 = qualified for FRPL for at least one year during Grades 5–8
- **On-track indicator:** 0=off-track at the end of the specified grade, 1=on-track at the end of the specified grade

Figure 1: On-Time High School Graduation
by Performance on Grade 4 Math and Science Tests

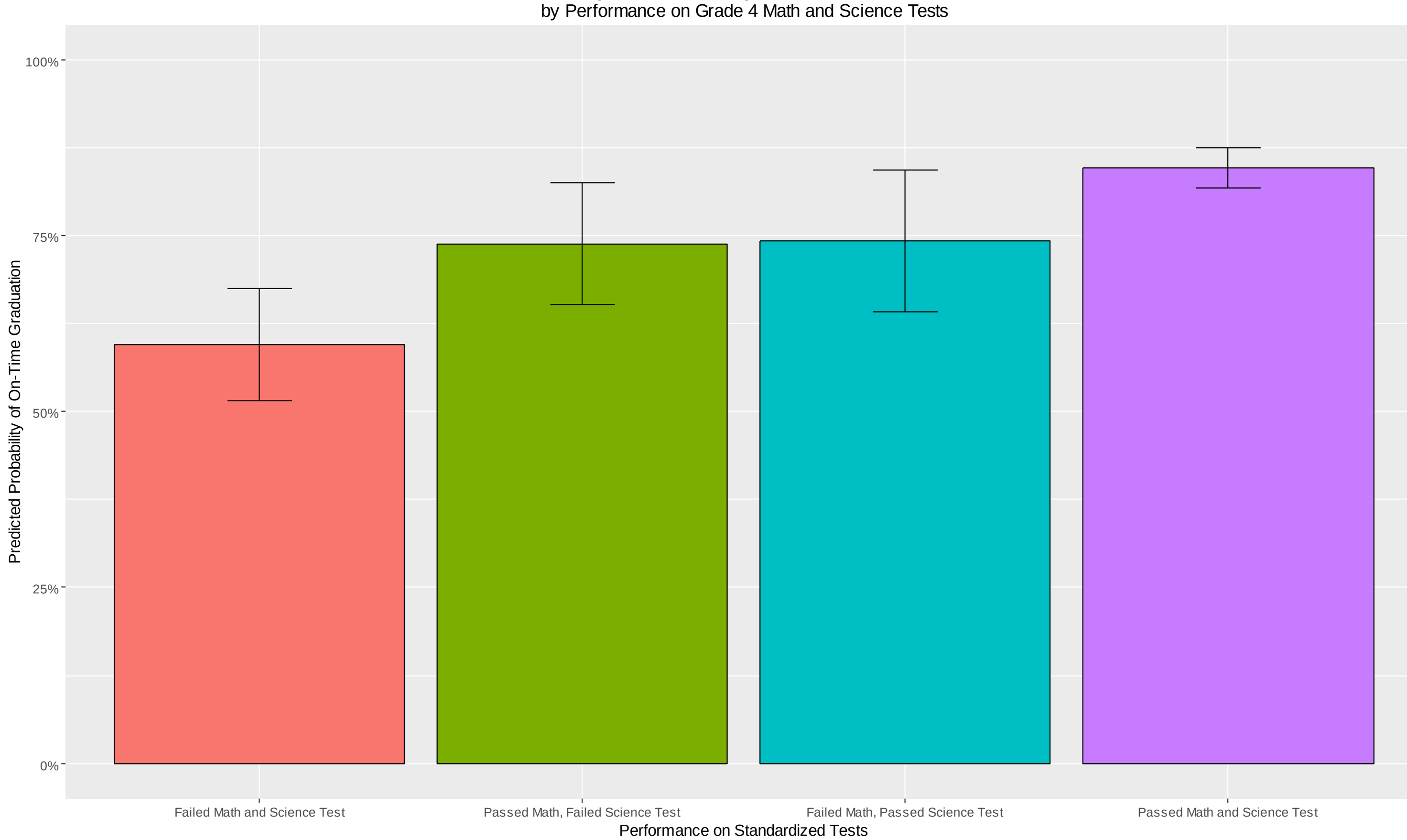


Figure 2: On-Time High School Graduation
by Special Education Status, Grade Repetition,
and Performance on Standardized Tests in Grades 5-8

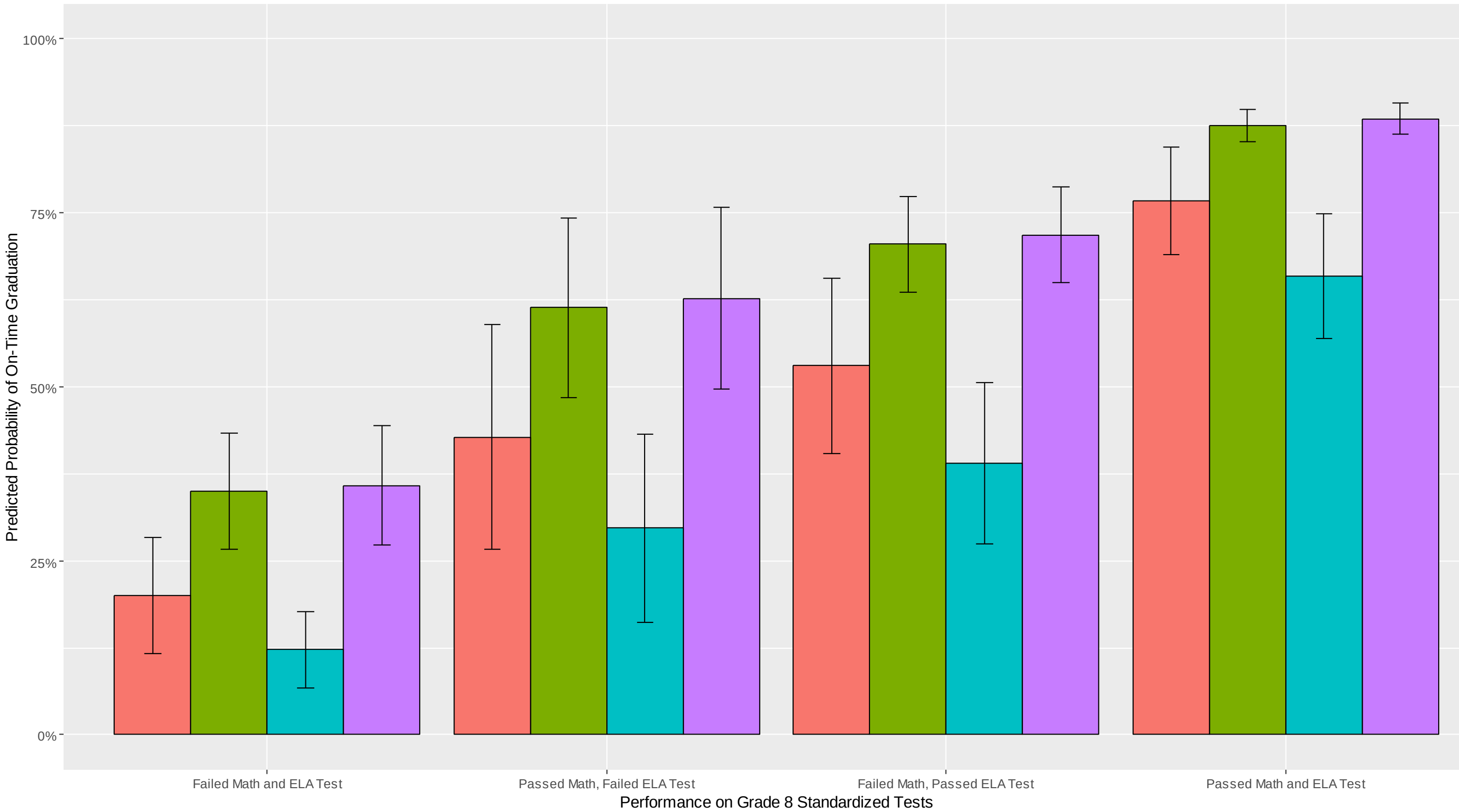
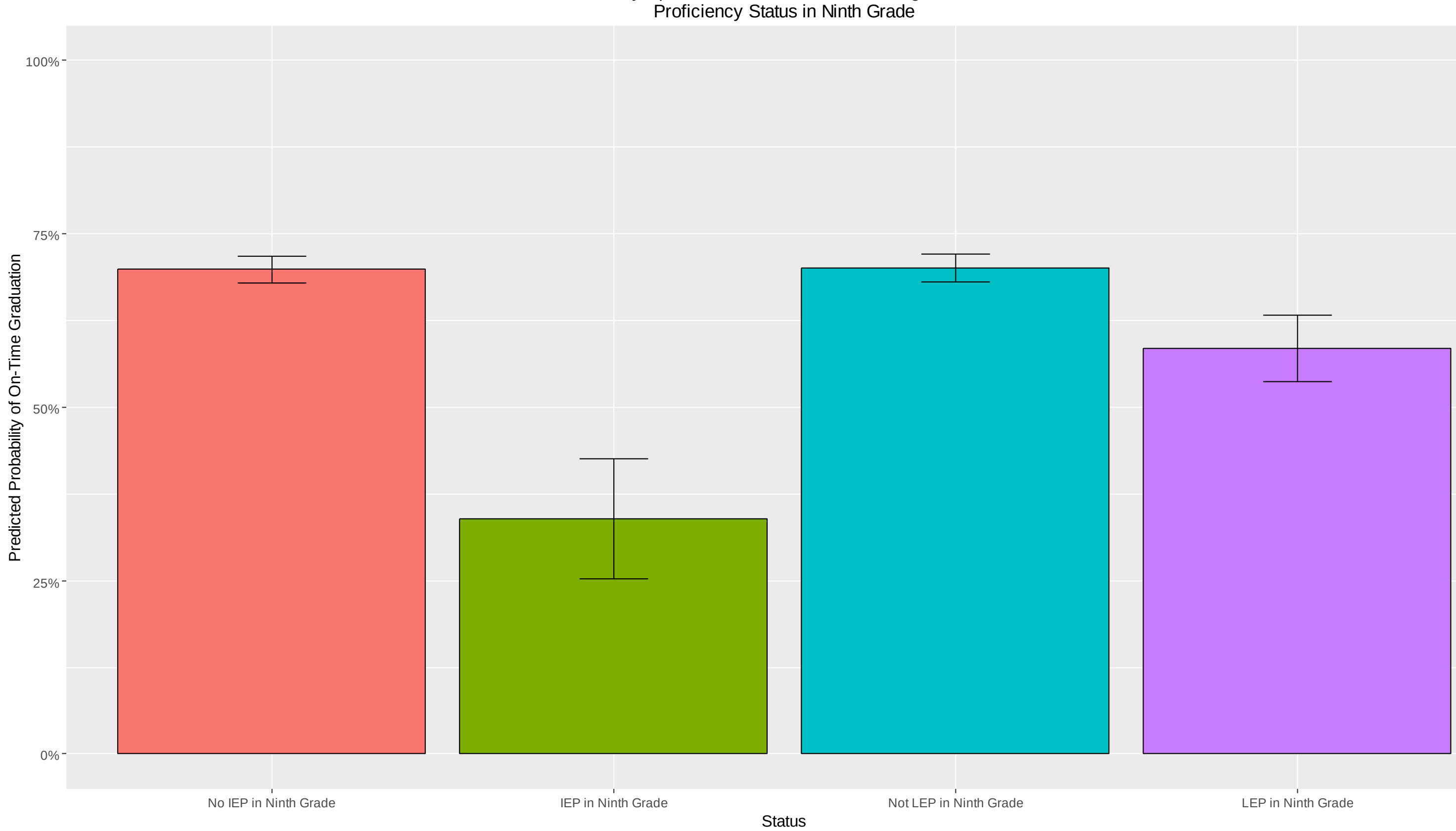


Figure 3: On-Time High School Graduation
by Special Education and Limited English
Proficiency Status in Ninth Grade



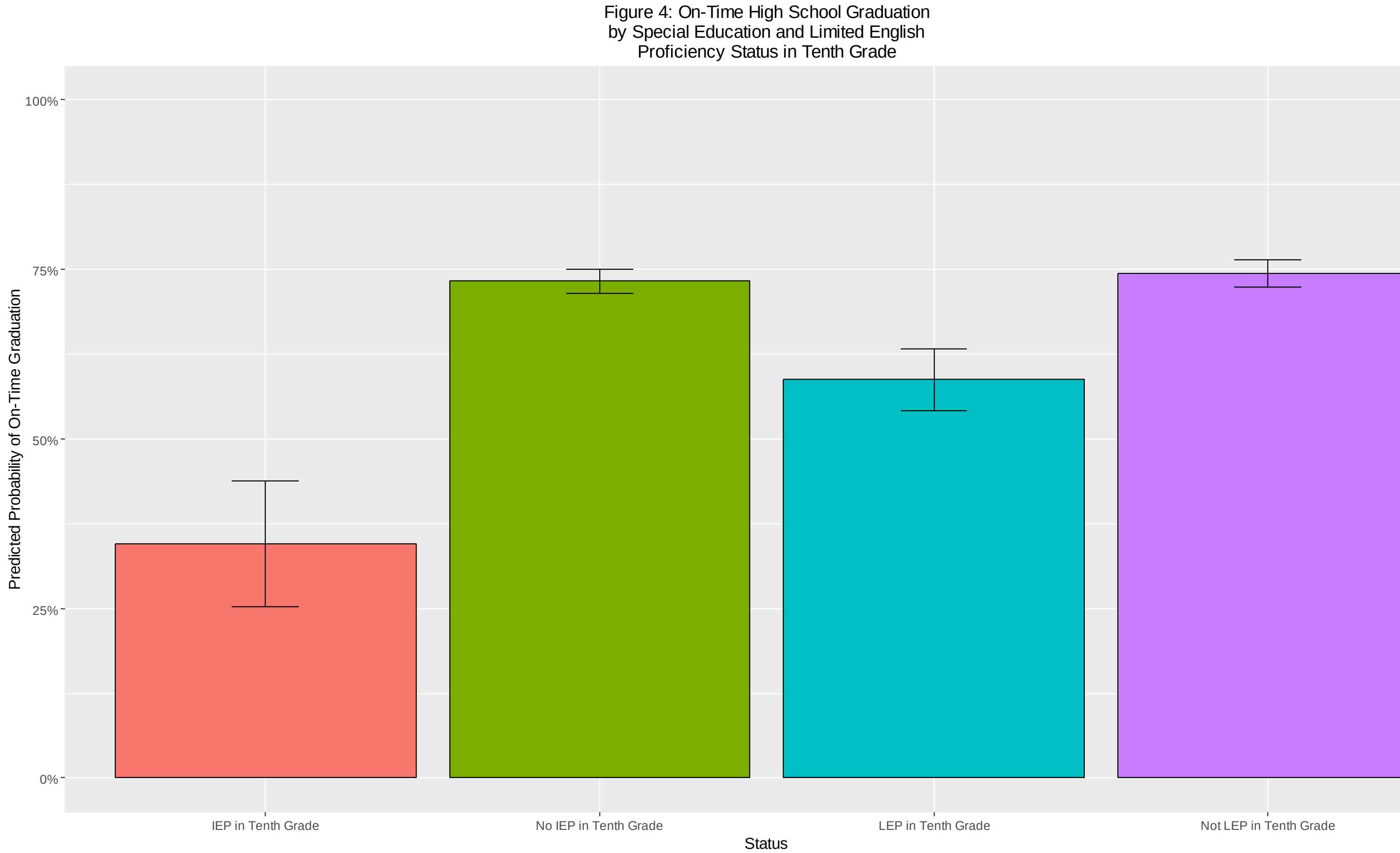


Figure 4: On-Time High School Graduation
by Special Education and Limited English
Proficiency Status in Tenth Grade
Predicted Probability of On-Time Graduation
100%
75%
50%
25%
0%
IEP in Tenth Grade
No IEP in Tenth Grade
LEP in Tenth Grade
Not LEP in Tenth Grade
Status

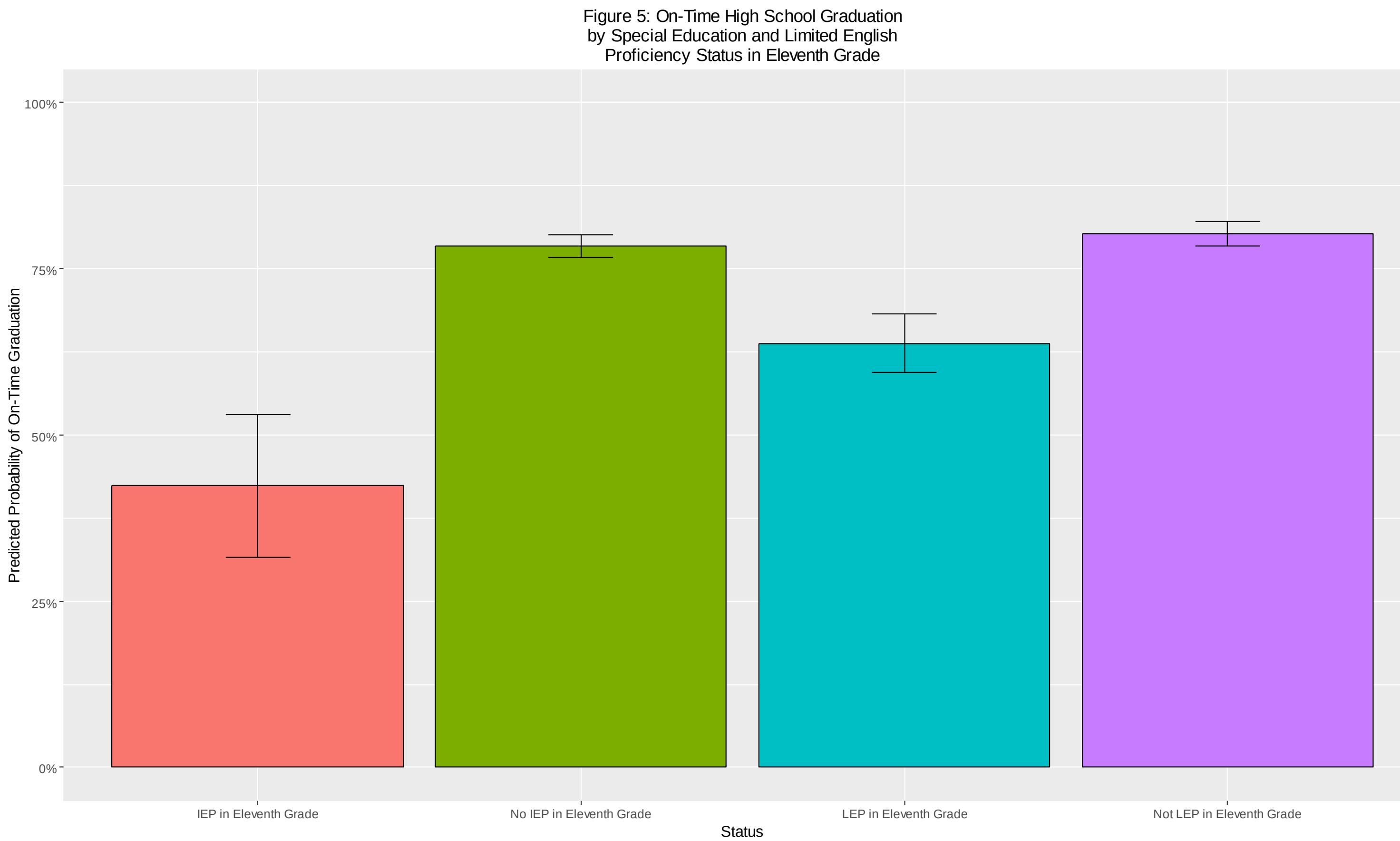


Figure 5: On-Time High School Graduation
by Special Education and Limited English
Proficiency Status in Eleventh Grade
100%
75%
50%
25%
0%
Predicted Probability of On-Time Graduation
IEP in Eleventh Grade
No IEP in Eleventh Grade
LEP in Eleventh Grade
Not LEP in Eleventh Grade
Status

Figure 6: On-Time High School Graduation by Performance
on Grade 4 Math Test for 2015 Cohort

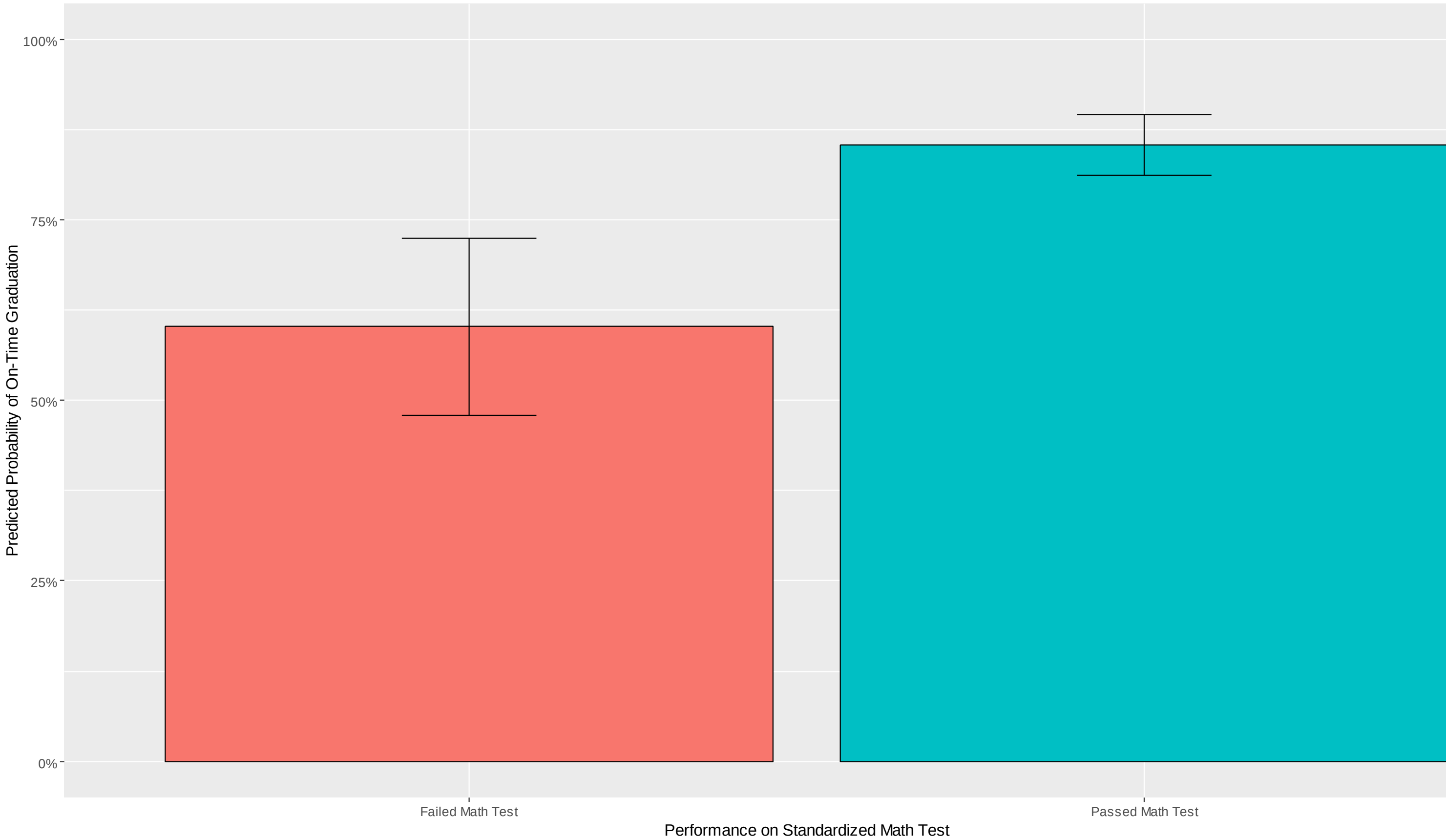


Figure 7: On-Time High School Graduation by Special Education Status, Grade Repetition, and Performance on Standardized Tests in Grades 5-8 for 2015 Cohort

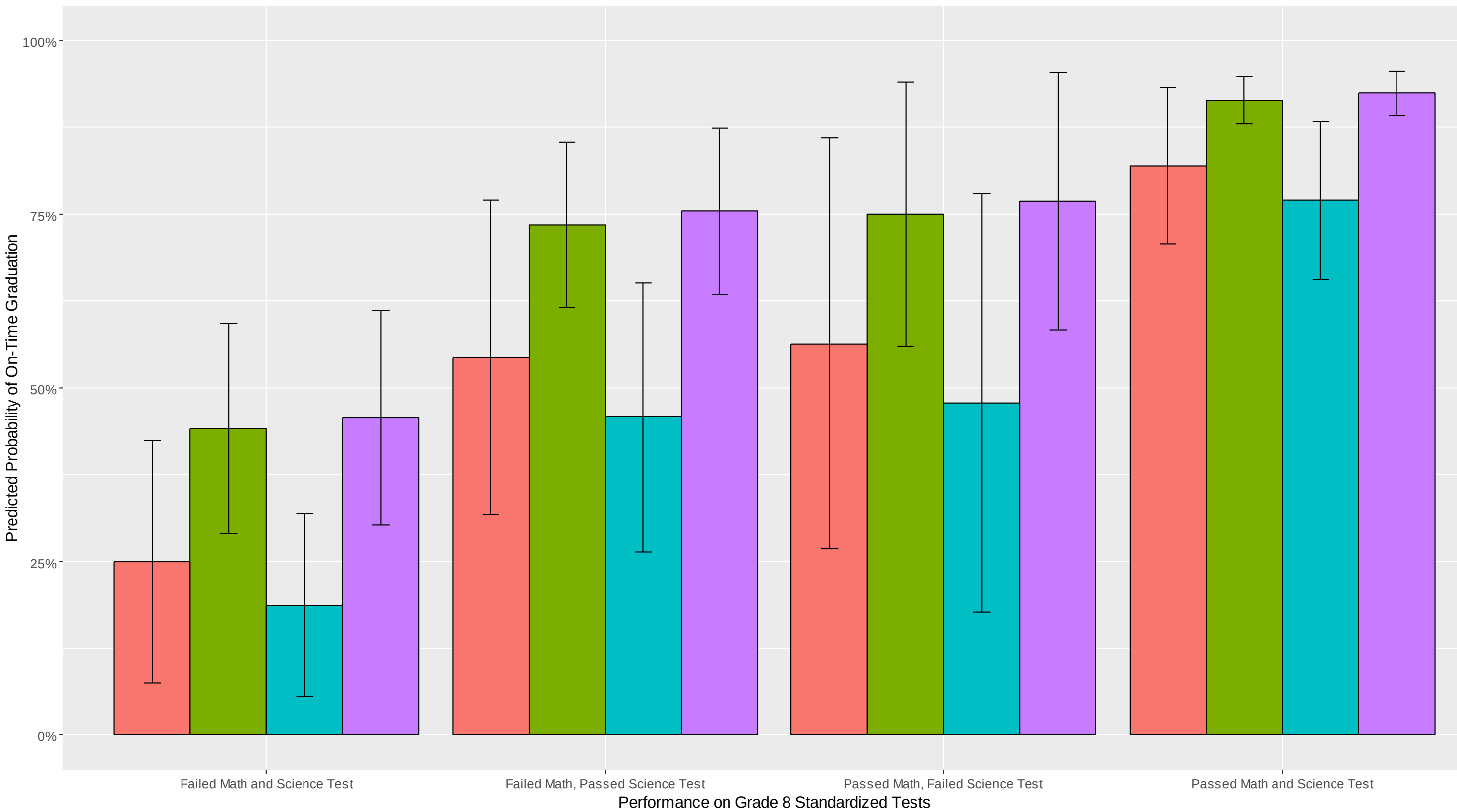


Figure 8: On-Time High School Graduation
by Special Education Status, Percent Attendance
and GPA in Ninth Grade for 2015 Cohort

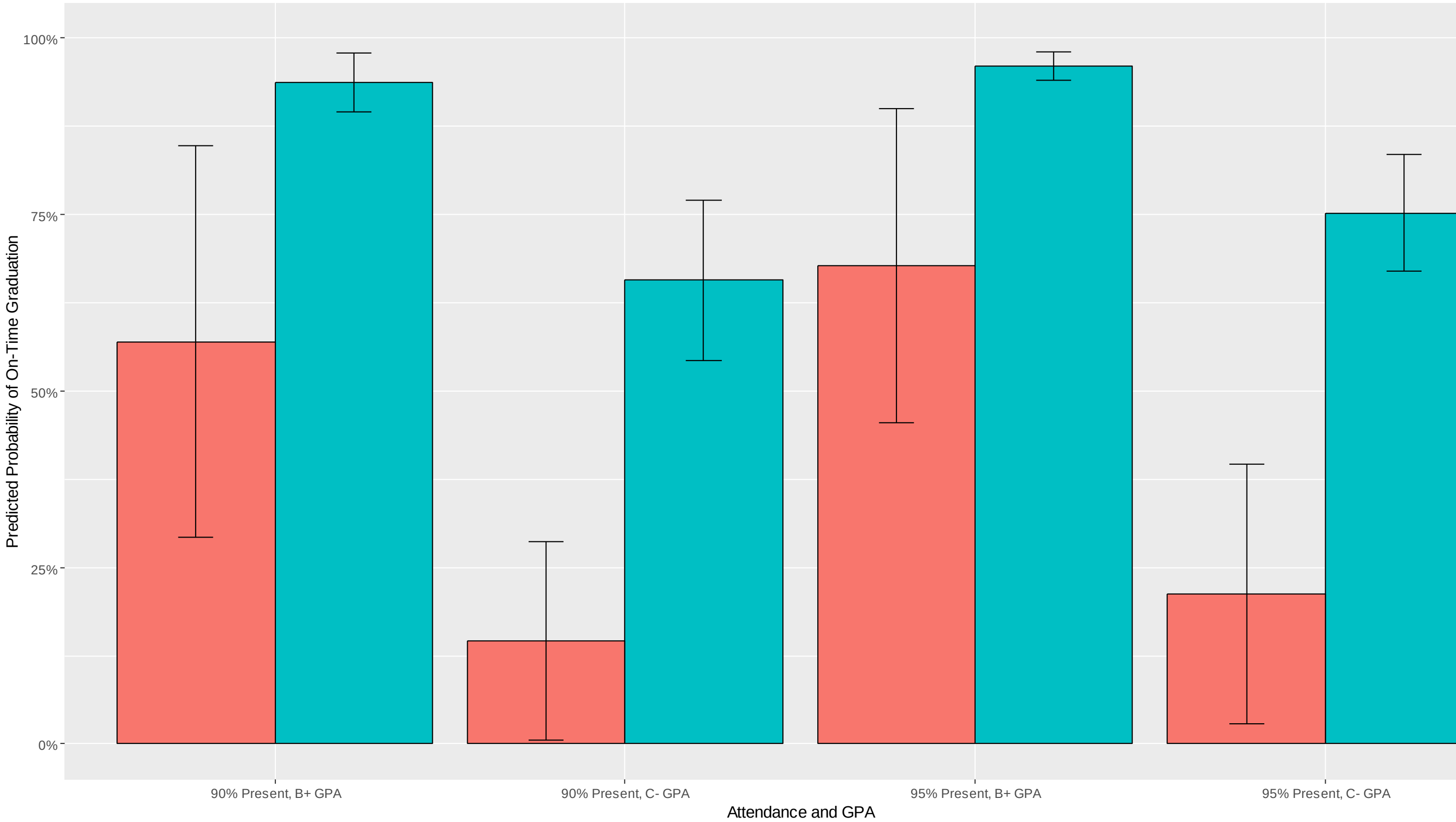
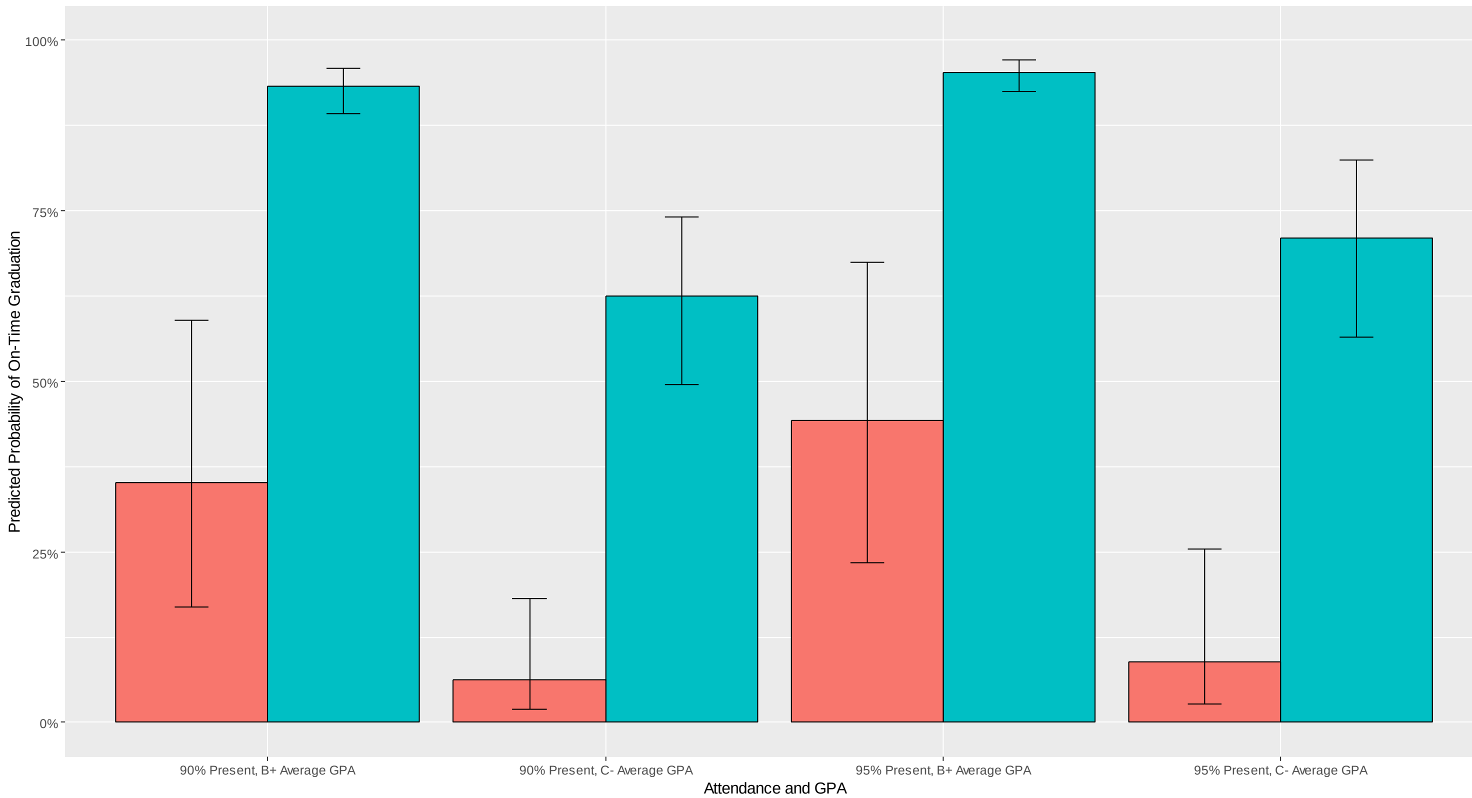


Figure 9: On-Time High School Graduation
by Special Education Status, Percent Attendance
and GPA in Tenth Grade for 2015 Cohort



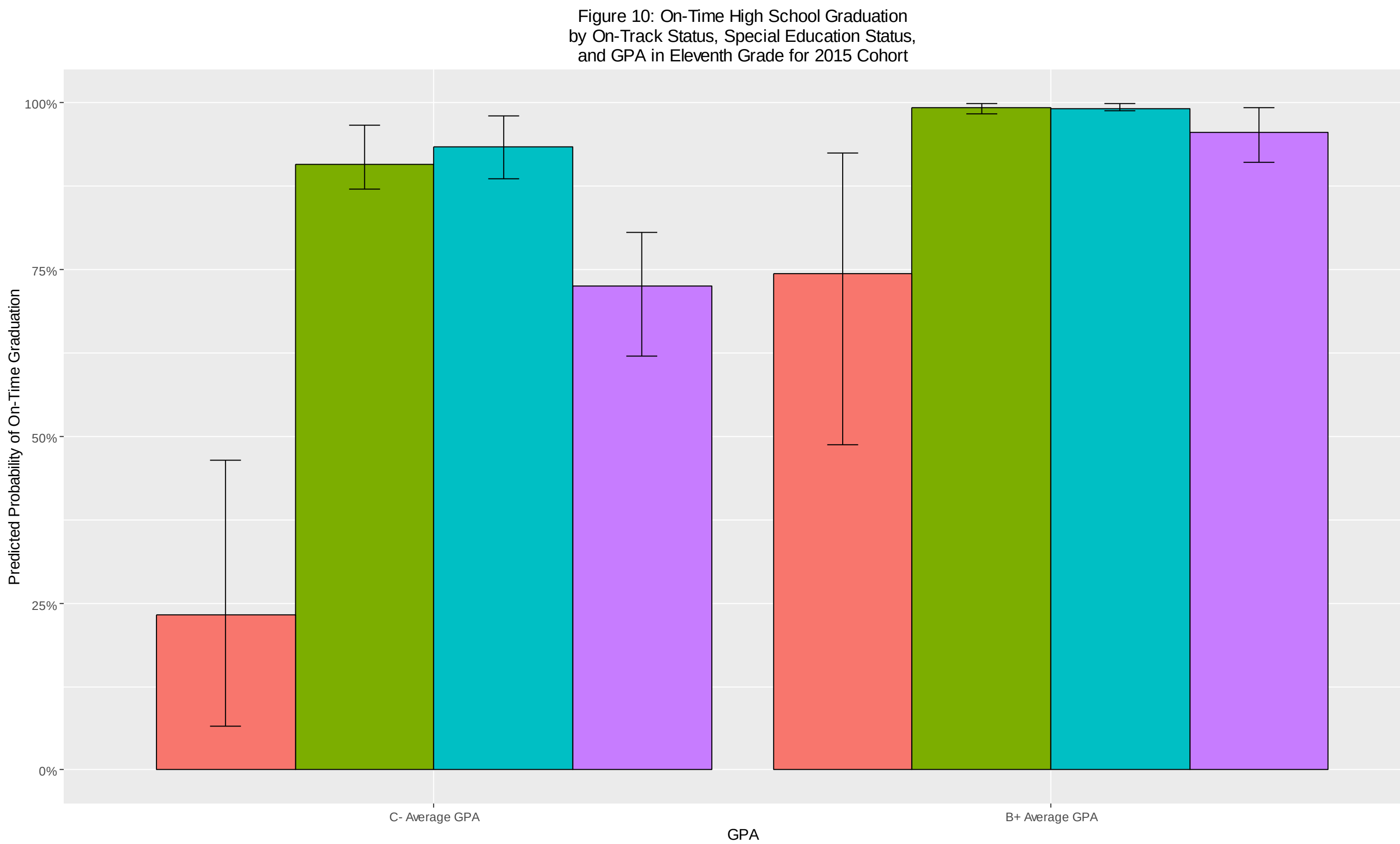


Figure 10: On-Time High School Graduation
by On-Track Status, Special Education Status,
and GPA in Eleventh Grade for 2015 Cohort
Predicted Probability of On-Time Graduation
100%
75%
50%
25%
0%
C- Average GPA
B+ Average GPA
GPA

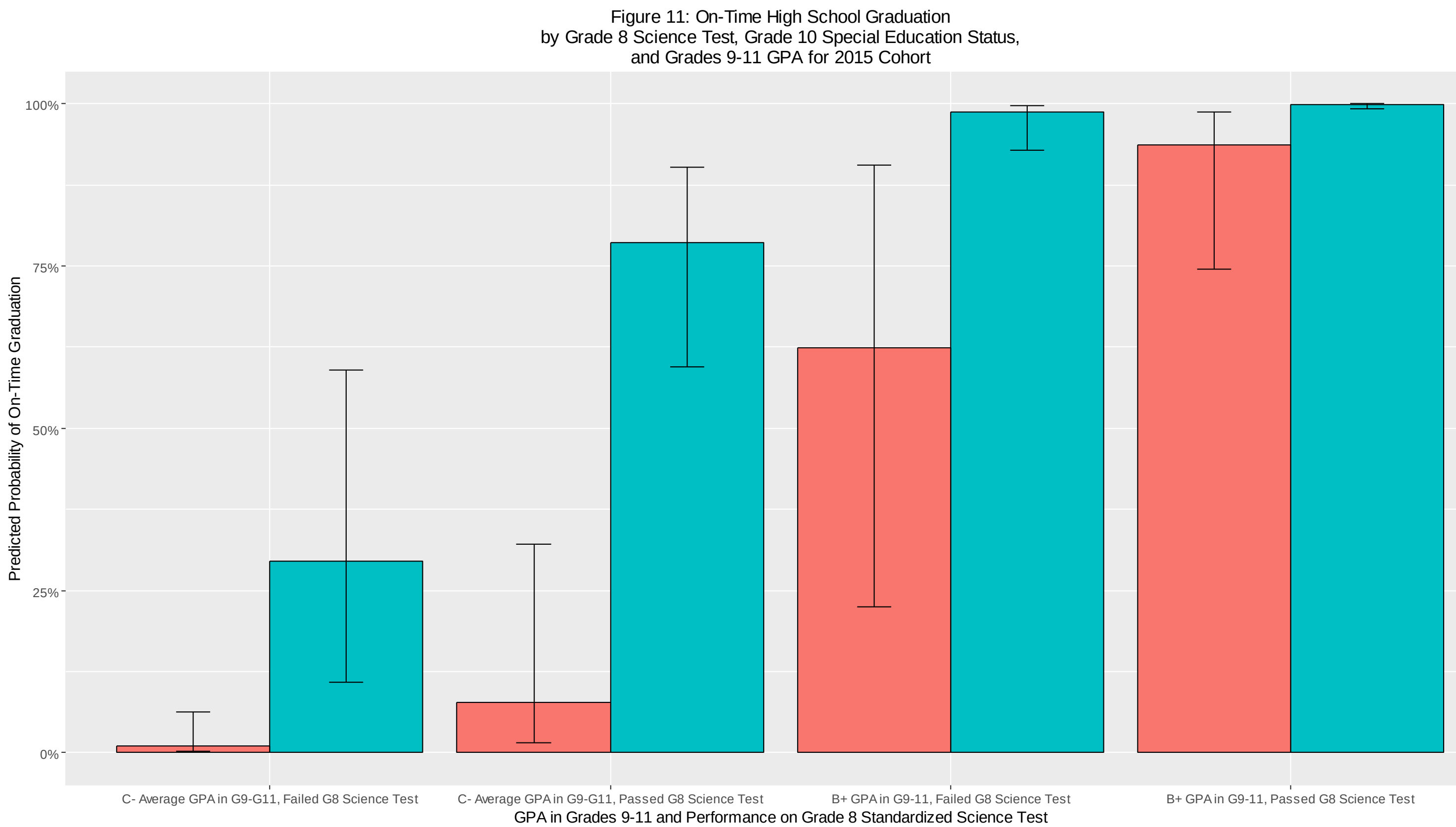


Figure 11: On-Time High School Graduation
by Grade 8 Science Test, Grade 10 Special Education Status,
and Grades 9-11 GPA for 2015 Cohort
Predicted Probability of On-Time Graduation
100%
75%
50%
25%
0%
C- Average GPA in G9-G11, Failed G8 Science Test
C- Average GPA in G9-G11, Passed G8 Science Test
B+ GPA in G9-11, Failed G8 Science Test
B+ GPA in G9-11, Passed G8 Science Test
GPA in Grades 9-11 and Performance on Grade 8 Standardized Science Test