PPIC | PUBLIC POLICY INSTITUTE OF CALIFORNIA

Instruction Manual for the a–g On Track Model

Version 1.0, April 2013

Julian R. Betts, Andrew C. Zau, and Karen Volz Bachofer

Supported with funding from the Donald Bren Foundation

Contents

Introduction

In California, high school students must complete a set of courses known as the a–g requirements, with grades of C or higher, to be eligible to apply for admission to either of the state's public university systems, the University of California and the California State University. Several large school districts have adopted the a–g courses as requirements to graduate from high school, in a bid to put more of their students on a track to attend a public university after graduation. However, in recognition of the difficulty of these standards, participating school districts typically allow students to satisfy graduation requirements with grades of D or higher. The San Jose Unified School District adopted these standards in 2002, and the same requirement will soon apply to other districts including the San Diego Unified School District (SDUSD), the Los Angeles Unified School District (LAUSD), the San Francisco Unified School District, and the Oakland Unified School District.

The a–g On Track Model is a new tool that California school districts can use to help forecast which students are likely to complete the overall a–g course requirements with grades of D or higher or C or higher. The model also calculates the probability that students will complete individual subject requirements with grades of D or higher in social studies, English Language Arts (ELA), mathematics, science, and world languages. (subjects 'a' through 'e' in the a–g sequence). The model uses various measures of students' academic preparation as of either grade 6 or grade 7, as well as grade-5 scores on the California Standards Test (CST) in science, if available. Early intervention to help these middle and high school students becomes possible if students at risk of not completing the a–g courses can be identified at the end of either grade 6 or grade 7, and if these data are available to the school district. (Using data from grade 6 allows interventions to begin earlier than if data from grade 7 are used, but secondary school districts that do not include elementary schools may not be able to obtain grade-6 data from feeder elementary school districts. Further, in some districts, elementary schools include grade 6, which may make it hard to obtain a grade point average (GPA) for grade 6 as easily as for students in middle or junior high schools.)

The a–g On Track Model is based on models we have estimated using data from SDUSD for students in the graduating class of 2011 who were in grades 6 and 7 during the 2004–2005 and 2005–2006 school years. The model is downloadable without charge. Although confidential student data must be entered by district staff, none of the data need to be sent outside the district or processed online—the spreadsheets are tools with full functionality at the local level.

The model consists of four Excel spreadsheets: two "validation" spreadsheets and two "forecasting" spreadsheets. Table 1 describes their purposes and file names.

TABLE 1
Names of Excel spreadsheets provided as part of the a–g On Track Model, by purpose

Grade on which student characteristics are based	Purpose: Validating the model using data for students who have already graduated	Purpose: Forecasting the probability that students who have not yet reached, or finished, high school will complete the a–g requirements
Grade 6	A-G_On_Track_Model_ Validation_Grade_6.xlsx	A-G_On_Track_Model_Forecasts_Grade_6.xlsx
Grade 7	A-G_On_Track_Model_ Validation_Grade_7.xlsx	A-G_On_Track_Model_ Forecasts_Grade_7.xlsx

The four spreadsheets are quite similar. The validation spreadsheets allow district administrators to test whether the On Track Model is an accurate predictor of a–g completion in their own districts. Administrators enter information on whether students who have already finished high school completed the a–g requirements overall and in the five aforementioned subject areas and then compare the predicted probability of completing the a–g requirements against actual completion rates. The forecasting spreadsheets are identical except that they do not allow input of information on whether the student completed the a–g requirements, because these spreadsheets are intended to predict the future (and therefore unknown) course completion of students who have just completed grade 6 or grade 7.

Validation. Because some of the required student variables, such as GPA, English Learner (EL) status, or participation in special education, may be measured or determined differently from one district to another, we urge districts to test the a–g On Track Model using data from older students in their own districts for whom grade 12 a–g completion results are already known. They can then compare the predictions of the validation spreadsheets against what actually occurred among these students. This validation exercise is crucial, because it provides some direct evidence on how well the model works in the given district.

If this validation exercise shows that the model has good predictive power for students in a given district, then that district may want to use it to forecast whether students in the district will complete the a–g requirements overall, and in the five aforementioned subject areas.

Conversely, if the validation exercise shows that the model does not accurately predict which students completed the a–g requirements in the given district, it probably means that the district will need to develop its own statistical model of a–g completion rates.

Forecasting. To predict completion of the a–g requirements, users will need to enter a small number of student variables, measured either in grade 6 or grade 7 (except for the score on the science CST test, which is administered in grade 5) into the forecasting spreadsheets. These spreadsheets will provide the estimated probability that each student will complete the a–g requirements by the end of grade 12, overall, and in the five aforementioned subjects.

Overview of Data Requirements

For districts wishing to use data from grade 6 to validate or forecast student success on the a–g requirements, the following data must be gathered from the grade-6 records.

Binary—that is (0,1)—indicators:

1. Student is female
2. Student is an English Learner (EL)
3. Student is Reclassified Fluent English Proficient (RFEP)
4. Student is in special education
5. Student was enrolled outside the district in grade 5

Continuous variables for:

1. Annual GPA
2. Percentage of days absent in the school year
3. Grade-6 scaled scores for the student on the CST in mathematics and English Language Arts, and the calendar year (e.g. "2010") in which the grade-6 tests were taken
4. Grade-5 scaled score on the CST in science
5. Calendar year in which the CST science test was administered

Alternatively, if the district wants to use grade-7 data, data should be gathered from grade 7 for all of the above variables, except for the grade-5 scaled score on the CST in science and the calendar year in which the grade-5 CST science test was administered.

The user should then enter these data into the appropriate grade-6 or grade-7 Excel spreadsheet, as described above in Table 1. These data should be entered into the Data Entry worksheet.

We note that there is one column in the Data Entry worksheet that is created automatically. If the science score for the grade-5 CST is missing, then an indicator variable for this will be set to 1 in column L.

For the validation spreadsheets, data on student completion of the a–g course requirements, overall and by subject area for English Language Arts, mathematics, science, social studies, and world languages, are also needed.

We have organized the rest of this section as follows:

1. Preparing Data for Use in the Spreadsheets; Limitations of the Model
2. Inputting Data from Grades 6 and 7 into the Validation Spreadsheets and Interpreting the Results
3. The Grade-6 Forecasting Spreadsheet

1. Preparing Data for Use in the Spreadsheets; Limitations of the Model

Data, from whatever district source, should be in a form that can be exported into an Excel spreadsheet or copied and pasted directly into the provided spreadsheet templates. With one row per student, columns include variables such as ID (any identifying variable the district cares to use), demographics, grades, absences, and test scores. Apart from ID, which is for district use and can contain any sort of alphanumeric student identification number, all variables should be numerical. By exporting data in the same order as listed in the Data Entry worksheet of the Excel spreadsheet, the user can save a substantial amount of time in copying and pasting data.

Limitations

i) Missing data

To produce estimated probabilities that a student will complete the a–g requirements, valid data for each variable must be entered into the spreadsheet. The spreadsheet is set up to check for blank (missing) values, and it will not produce estimated probabilities of passage for observations that are missing any of the variables used to predict outcomes. If a user sees "N/A" for a student's predicted probability of completing the a–g requirements, this means that one or more of the required data fields for that student were blank.

There is one exception to this rule: grade-5 science CST scaled scores. We made this exception because we found that in San Diego, about one-twelfth of students for whom we had complete grade-6 data lacked information on the grade-5 CST science test, probably in part due to student mobility into and out of the district. We did not want to drop such students from our forecasting model. Thus, if the grade-5 science score is missing, an indicator variable for this will be set to 1 automatically. In this case, the spreadsheet will produce predicted probabilities of completing the a–g requirements for that student, as long as other variables, pertaining to grade 6 are all available.

ii) Maximum number of observations

The current version of the spreadsheet allows data for up to 10,000 students to be entered. In the rare case that a district has more students in a cohort than this, individual predictions can still be obtained by dividing the data across several versions of the file. Alternatively, on request, we can provide instructions on how to change the formulas in the various calculation spreadsheets to accommodate more than 10,000 observations.

iii) The current model can accommodate CST scores from 2002 through 2011 only

The On Track Model standardizes the CST scaled scores provided by a district administrator to have mean 0 and variance 1, using the statewide mean and standard deviation from the calendar year in which the CST was administered. The current version will standardize CST test scores in mathematics and English Language Arts for test administrations from 2002 through 2011. The spreadsheet will standardize grade-5 science scores from 2004 through 2011. When 2012 CST norms are released, we will update the spreadsheet accordingly.

2. Inputting Data from Grades 6 or 7 into the Validation Spreadsheet and Interpreting the Results

Data input proceeds similarly in both of the spreadsheet templates. District users should open the file, switch to the Data Entry worksheet, and paste in the requested data. **None of the other worksheets (various results worksheets as well as the Predict and Calculations worksheets) in any of the Excel**

spreadsheets should be altered, as these worksheets process the data provided in Data Entry or present results of that analysis.

Below is a screen shot of the Data Entry worksheet from the grade-6 validation spreadsheet. This worksheet is used to input data for students who have already reached grade 12. A similar Excel file may be used to enter grade-7 information for districts that do not have grade 6 middle schools and thus either lack any grade-6 data or, if elementary school data for grade 6 are available, do not have a counterpart to the GPA that is typically available on middle school report cards.

EXHIBIT 1
Sample Data Entry worksheet with several fictitious observations added

8																Below "1" indicates passed			
9	ID	FEMALE	EL	RFEP	Special Ed	GPA	% Days Absent	Math Scaled Score	ELA Scaled Score	Year of Grade 6 CST Test	Science Scaled Score	Missing Science Scaled Score	Year of Grade 5 CST Science	Out of district in grade 5	Completed a-g with C or better	Completed a-g with D or better	Completed math with C or better	Com Matl or b
10	Sim1	1	0	0	0	2.875	0	323	340	2005	387	0	2004	0	0	0	0	
11	Sim2	1	1	0	0	2.045	2.78	288	353	2005	273	0	2004	0	0	1	0	
12	Sim3	0	1	0	0	2.62	2.22	296	312	2005	282	0	2004	0	0	0	0	
13	Sim4	0	0	0	1	3.665	3.89	276	356	2005	342	0	2004	0	0	0	0	
14	Sim5	0	0	0	0	2.33	13.33	315	356	2005	368	0	2004	0	0	0	0	
15	Sim6	1	0	0	0	1.455	0.56	254	262	2005	263	0	2004	0	0	1	1	
16	Sim7	1	0	0	0	3.335	2.22	370	379	2005		1	2004	1	0	0	0	
17	Sim8	1	0	0	1	2.5	6.11			2005		1	2004	0	0	0	0	
18	Sim9	1	0	1	0		2.78	267	269	2005	278	0	2004	0	0	0	0	
19	Sim10	1	0	0	0	3.5	1.67	391	387	2005	363	0	2004	0	1	1	1	
20	Sim11	0	0	0	0	2.9	8.89	356	337	2005	337	0	2004	0	1	1	1	
21	Sim12	1	0	0	0	3.705	0	404	387	2005	337	0	2004	0	1	1	1	
22	Sim13	0	0	0	0	3.215	2.22	418	454	2005	431	0	2004	0	0	0	0	
23	Sim14	0	0	0	0	2.8	3.33	370	402	2005	377	0	2004	0	0	0	0	
24	Sim15	1	0	1	0	1.9	8.89	303	363	2005	337	0	2004	0	0	1	1	
25	Sim16	1	0	0	0	2.585	1.11	267	265	2005	246	0	2004	0	0	0	0	
26	Sim17	0	0	0	0	3.75	6.67	391	467	2005	387	0	2004	0	0	0	1	
27	Sim18	1	1	0	0	2.415	0.56	284	288	2005	296	0	2004	0	1	1	1	
28	Sim19	0	0	1	0	2.6	1.67	351	330	2005	291	0	2004	0	0	0	0	
29	Sim20	0	0	0	0	2.385	1.67	296	334	2005	342	0	2004	0	0	0	0	
30	Sim21	1	0	0	0	3.9	0	410	402	2005	363	0	2004	0	1	1	1	
31	Sim22	1	0	1	0	2.14	20	254	288	2005	257	0	2004	0	0	0	0	
32	Sim23	1	0	1	0	3.6	0.56	343	334	2005	321	0	2004	0	1	1	1	
33	Sim24	0	0	0	0	3	1.11	292	300	2005	300	0	2004	0	0	0	1	
34	Sim25	0	1	0	0	3.4	0	327	246	2005		1	2004	0	1	1	1	
35	Sim26	1	1	0	0	2.085	5	272	318	2005	300	0	2004	0	0	0	0	

This screenshot shows fictitious data that have been entered for several students. (No data appear in rows 10 and below when the worksheet is first opened.)

Note that FEMALE, EL, RFEP, Special Ed, Enrolled in grade 5, and the completion rates for the a–g courses overall and by subject in columns B through D and M and later are all coded as 1 if the student fits into the stated category and 0 otherwise. For instance, we see in row 10 data for a student who is female, is not an EL student, is not reclassified as Fluent English Proficient (RFEP), is not in special education, and did not complete the overall requirement using either the C or higher or D or higher definitions.

GPA is measured on a 0–4 scale and represents the overall GPA in grade 6.

The percentage of days absent should be calculated by the user as: 100% X (number of days absent/number of days in school year). Thus, if a student missed 9 out of 180 school days, he or she was absent 5 percent of the time and this should be entered as 5, **not** 0.05.

The mathematics and ELA scaled scores refer to each student's scaled scores on the grade-6 CST in the stated subjects. It is important to include in column I the calendar year in which the student was tested. We use this to transform each student's score into a "Z-score," which measures the number of standard deviations above or below the state average the student scored in the given year. The year refers to the *calendar year* in which the test was administered. For instance, in row 11, we see that the student took the grade-6 ELA and mathematics tests in 2005. Similarly, the grade-5 science scaled score should be entered in column K, and in column M, the *calendar year* in which the science CST test was taken should be entered. If the student is progressing between grades at a typical rate, this will normally be one year before the year in which the

grade-6 test scores are observed. In row 34, student "Sim25" has a blank entry for the grade-5 science score, and thus the indicator in column L for this science score being missing is set to 1 automatically by the spreadsheet itself.

After the data have been entered in the Data Entry worksheet, predicted probabilities of passage for each student will appear in the Predict worksheet. There is also a series of worksheets showing results for overall completion of the a–g requirements and completion of individual subject requirements.

The Results Overall worksheet provides several tables that examine results related to overall completion of the a–g course requirements. This worksheet is followed by five other worksheets that are identical to the Results Overall worksheet but which instead show results related to completion of course requirements in individual subject areas. For instance the Results by English (b) worksheet shows results related to whether students completed the English subject requirements. Exhibit 2, below, shows the first five tables of results generated by the a–g On Track Model from the Results Overall table. We have used actual data from the class of 2011, measured in grade 6, for these screen shots.

EXHIBIT 2
Table 1 through 5 from the Results Overall worksheet

Overall completion of a-g

Probability Range	Number of Students by Predicted Completion Rate		Percent of Students by Predicted Completion Rates		Number of Students Completing in Each Predicted Completion Rate Category		Number of Students Failing in Each Predicted Completion Rate Category	
	C or better	D or better	C or better	D or better	C or better	D or better	C or better	D or better
<0.10	1272	282	26.0%	5.8%	57	22	1215	260
0.10-0.19	804	550	16.4%	11.2%	114	66	690	484
0.20-0.29	552	629	11.3%	12.8%	124	139	428	490
0.30-0.39	448	591	9.2%	12.1%	158	205	290	386
0.40-0.49	431	560	8.8%	11.4%	199	255	232	305
0.50-0.59	406	496	8.3%	10.1%	227	293	179	203
0.60-0.69	380	589	7.8%	12.0%	265	418	115	171
0.70-0.79	353	560	7.2%	11.4%	270	430	83	130
0.80-0.89	211	504	4.3%	10.3%	173	419	38	85
0.9 or higher	39	135	0.8%	2.8%	29	106	10	29
Total Students	4896	4896	100.0%	100.0%	1616	2353	3280	2543

Probability Range	Actual Completion Rates of Students in Each Predicted Completion Rate Category	
	C or better	D or better
<0.10	4.5%	7.8%
0.10-0.19	14.2%	12.0%
0.20-0.29	22.5%	22.1%
0.30-0.39	35.3%	34.7%
0.40-0.49	46.2%	45.5%
0.50-0.59	55.9%	59.1%
0.60-0.69	69.7%	71.0%
0.70-0.79	76.5%	76.8%
0.80-0.89	82.0%	83.1%
0.9 or higher	74.4%	78.5%
All Students	33.0%	48.1%

The first two tables show the number and percentage distribution of students in the sample by their predicted probability of passing the a–g requirements overall, based on either a C or higher or D or higher grade requirement. Each row corresponds to a different "predicted probability of passing" group. For instance, the first row of each table shows results for students in the predicted probability range from "< 0.10," which means a probability of passing the a–g requirement of less than 0.10 (that is, 10%).

The last row shows results for students in the group "0.9 or higher," which refers to a predicted probability of passing of 90 percent or higher.

To make the tables easier to read, we listed probability ranges as 0.10–0.19, 0.20–0.29, and so on. In fact, we assign any student with predicted probability equal to or greater than 0.1 and strictly less than 0.2 to the group "0.10–0.19."

The third and fourth tables in the worksheet show the number of students in each predicted probability group who actually passed and who actually failed the a–g requirement overall, using both letter grade requirements.

The last table, appearing below the first four tables in Exhibit 2 above, shows the percentage of students in each probability category who actually completed the a–g requirements by the end of grade 12, overall. (Separate worksheets show the corresponding results by individual subject area.) The bottom row shows the pass rates for the entire student sample. The On Track Model clearly works well for the actual SDUSD data, because the percentage of students completing the a–g requirement rises steeply with the predicted probability of completing the course requirements.

One of our key goals in creating the a–g On Track Model is to help school districts simulate what would happen if they chose a cutpoint—that is, a probability of completing the a–g requirements below which a student would be provided with some form of intervention or assistance. This raises a number of questions: For a given cutpoint, how many students would need support? What percentage of all the students who actually failed to complete the a–g course requirements by the end of grade 12 would have been included in the group targeted for assistance? Conversely, what percentage of students who completed the a–g requirements would have been included in the group targeted for assistance even though they passed without any additional help?

Exhibit 3 shows Tables 6 through 8 from the Results Overall worksheet, calculated using grade-6 data from the SDUSD class of 2011.

EXHIBIT 3
Tables 6 through 8 from the Results Overall worksheet, showing the consequences of using various cutpoints to determine students who should receive academic assistance

44	Cutoff: Choose Students with Predicted Passage Probability Below:	Number of Students to Be Included in an Intervention		% Who Did Not Complete Who Would be Included in Intervention		% Who Completed Who Would be Included in Intervention	
45		C or better	D or better	C or better	D or better	C or better	D or better
46	0.1	1272	282	37.0%	10.2%	3.5%	0.9%
47	0.2	2076	832	58.1%	29.3%	10.6%	3.7%
48	0.3	2628	1461	71.1%	48.5%	18.3%	9.6%
49	0.4	3076	2052	80.0%	63.7%	28.0%	18.4%
50	0.5	3507	2612	87.0%	75.7%	40.3%	29.2%
51	0.6	3913	3108	92.5%	83.7%	54.4%	41.6%
52	0.7	4293	3697	96.0%	90.4%	70.8%	59.4%
53	0.8	4646	4257	98.5%	95.5%	87.5%	77.7%
54	0.9	4857	4761	99.7%	98.9%	98.2%	95.5%
55							

The first table in Exhibit 3 shows the total number of students who would have been identified for assistance. The next table shows the percentage of students who failed to complete the a–g requirements by the end of grade 12 and who would have been included in the group below the cutpoint and would therefore receive assistance through whatever academic intervention the district created. The final table shows the percentage of students who, in fact, completed the a–g courses or one component of the a–g courses who would have been included in the intervention program. Again, these tables show calculations based on SDUSD grade-6 data from the class of 2011.

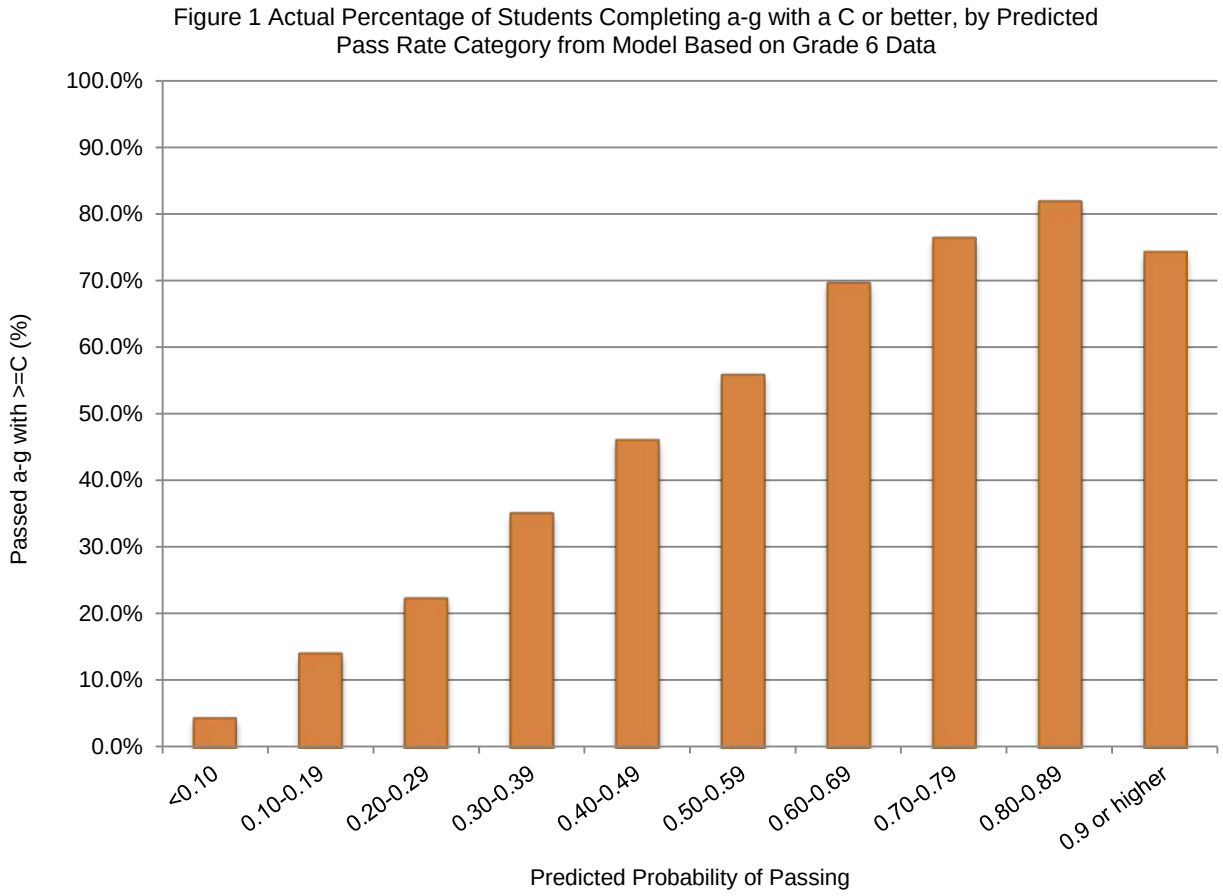
These counts are useful because, by choosing a cutpoint carefully, one could include the majority of students who actually failed to complete the a–g requirements, while excluding most of the students who completed a–g without any additional assistance. For instance, suppose that administrators wanted to design an intervention program to prepare students at risk of failing to complete the overall a–g requirements. The data show that for students with a predicted probability of completing the requirement based on grades of C or higher below 0.5, administrators would have identified 3,507 students. They would have included 87.0 percent of the students who indeed failed to complete the a–g courses and 40.3 percent of students who completed them overall.

To the extent that a district's distribution of a–g course completion does not shift radically from year to year, these retrospective analyses should predict quite well what would happen if such a cutpoint were used to target assistance to students who had just completed grade 6.

Three figures appear below the tabulated results in the Results Overall worksheet. The first figure shows the percentage of students who completed the a–g course requirements with grades of C or higher, plotted against the predicted probability that they would pass. The second figure shows the corresponding graph using only a–g courses completed with grades of D or higher. These graphs are crucial, because if the a–g On Track Model works well for data in the given district, we should see that a very small percentage of students in the bottom groups—those with a "<0.10" or "0.10–0.19" probability of passing—actually complete the requirement by the end of grade 12. Conversely, a large majority of the students in the top groups, such as "0.9 or higher," should have completed the requirements. If so, then the a–g On Track Model may provide good forecasts for students in the given district who have not yet reached grade 12.

Exhibit 4 shows what Figure 1 looks like for the SDUSD sample. The models perform well. (If the model had no predictive power, then the height of the bars would not systematically increase as one moves from left to right in the figures, from the groups with low predicted probability of passing to the groups with high predicted probability.)

EXHIBIT 4
Figure 1 from the Results and Graphs worksheet, showing completion of the a–g courses, by predicted probability of completion with grades of C or higher

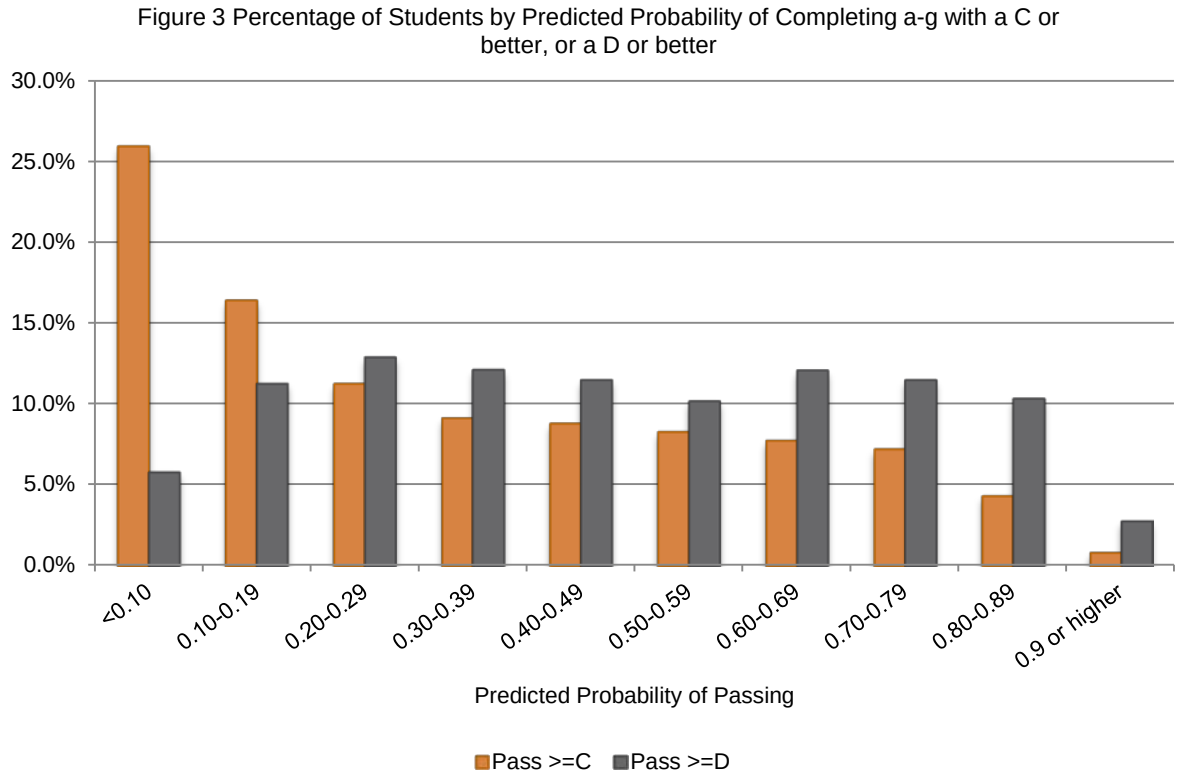


This figure shows sharply increasing a–g completion rates as one moves from the "low-probability" groups at the left of the figures to the "high-probability" groups toward the right of the figure.

The second figure in the Results Overall worksheet shows the corresponding results based on a–g completion that focuses on courses with grades of D or higher.

The last figure in the Results Overall spreadsheet is shown as Exhibit 5, below. This figure provides an example, using SDUSD class of 2011 data from grade 6, of the percentage distribution of students across the ten "predicted probability of passing" groups for overall completion of the a–g courses with C and higher and D and higher grades. Notably, in this example, quite a few students are in the groups with the lowest predicted probability of completing the a–g courses. A second finding that jumps out in this figure is the far smaller percentage of students who are predicted to have trouble with the D and higher requirement compared to the C and higher requirement.

EXHIBIT 5
Figure 3 from the Results and Graphs worksheet, showing the distribution of students by predicted probability of completing the a–g requirements



Each of the tables and figures shown above for overall a–g completion is replicated in separate worksheets that focus instead on a–g completion in specific subject areas. For instance, the Results by English (b) worksheet shows calculations for completion of the English Language Arts course requirements.

In addition to the Results and Graphs worksheet and corresponding worksheets for five specific subjects, another worksheet showing results—named Predict—shows the predicted probability of completing the a–g requirements (overall and by subject area) for each student, using both the C or higher or D or higher definitions.

The Predict worksheet shows, for each student, the predicted probability of completing the a–g courses, overall and by subject area. These results can be used to identify students who are likely to require additional assistance if they are to complete the a–g courses before the end of grade 12.

Exhibit 6 shows the predicted probabilities of completing the a–g courses overall or by subject area for the several fictitious students in the simulated data we entered in the Data Entry worksheet in Exhibit 1.

EXHIBIT 6
Predicted probabilities of completing the a–g requirements from the Predict worksheet, using simulated data

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	predicted overall completion C or better	Actual overall C or better	predicted overall completion D or better	Actual overall D or better	Predicted math >=C	Actual math >=C	Predicted math >=D	Actual math >=D	Predicted English >=C	Actual English >=C	Predicted English >=D	Actual English >=D
2	1	0.00	0	0.00	0	0.00	0	0.01	0	0.00	0	0.04	0
3	2	0.00	0	0.01	0	0.00	0	0.01	0	0.00	0	0.06	0
4	3	0.04	0	0.13	0	0.07	0	0.17	0	0.13	1	0.35	1
5	4	0.06	0	0.16	0	0.09	0	0.19	0	0.24	0	0.44	0
6	5	0.00	0	0.01	0	0.00	0	0.02	1	0.01	0	0.09	0
7	6	0.03	0	0.09	0	0.10	0	0.20	0	0.15	0	0.33	0
8	7	0.08	0	0.23	0	0.10	0	0.27	0	0.26	0	0.47	1
9	8	0.24	0	0.38	0	0.31	0	0.50	0	0.53	1	0.66	1
10	9	0.27	0	0.45	0	0.57	1	0.71	1	0.50	0	0.68	0
11	10	0.15	0	0.29	0	0.40	0	0.57	1	0.26	0	0.56	1
12	11	0.50	1	0.64	1	0.77	1	0.85	1	0.75	0	0.81	1
13	12	0.65	0	0.73	0	0.84	0	0.91	0	0.88	0	0.90	0

Student 8 has predicted probabilities of 0.24 and 0.38 of completing a–g overall using the C or higher and D or higher requirements and, in fact, did not complete under either definition. On the other hand, this student had a predicted probability of completing the English requirements with grades of C or higher or D or higher of 0.53 and 0.66 and did complete the English requirement under both definitions.

3. The Grades 6 and 7 Forecasting Spreadsheets

Using the separate grade-6 and grade-7 forecasting spreadsheets is similar to using the grade-6 and grade-7 validation spreadsheets described above. There is one important difference from the validation spreadsheets—one does not enter data on whether the students actually completed the a–g courses. This is because the forecasting spreadsheet is intended for students who have not yet reached grade 12.

The Results Overall worksheet for the forecasting spreadsheet is much simpler than the one for the validation spreadsheet discussed above. It provides tables on the number and percentage of students by their predicted probability of completing the a–g courses. These tables correspond to the first two tables from the validation spreadsheet shown in Exhibit 2 above. A third table in the Results and Graphs worksheet corresponds to Table 6 in the validation worksheet, shown in Exhibit 3. It shows the number of students who would be included in an intervention if various cutpoints were used to identify students in need of assistance. A fourth table shows the percentage of all students who would be placed in an intervention for various cutpoints. The figure shows the percentage distribution of students by their predicted probability group.

District Feedback: Questions and Suggestions

We are eager to hear from school district administrators about their experiences with the On Track Model and their suggestions for improving it if we update it in the future. We would be particularly interested in seeing the the numerical results in the Results and Graphs worksheet, to see how well the On Track Model works in other districts in California. Of course, we would ask administrators not to send the actual spreadsheets by email, because they may contain identifying student information. However, screenshots of the Results and Graphs section do not pose any confidentiality risks if sent by email.

We make no guarantees that the a–g On Track Model will predict a–g completion rates well in all districts. In particular, if a given district assigns students to special education or to EL or RFEP status, or calculates grades in ways quite different from the practices in SDUSD, the model's accuracy may decline. In such a case, it would make sense for a district to estimate, or have estimated for it, its own tailor-made statistical model.

Suggestions and questions can be sent to Julian Betts at jbetts@ucsd.edu.

Appendix

Using data we gathered from SDUSD, we estimated models that predicted whether graduates in the class of 2011 would complete the a–g requirements based on information for students who were in grade 6 for the first time in 2004–2005 or grade 7 for the first time in 2005–2006. We estimated probit models in which the dependent variable was a binary indicator for whether the student had completed a given set of a–g courses by the end of grade 12, as a function of the student's gender, EL status, RFEP status, enrollment status in grade 5, special education status, the percentage of days during which the student was absent, CST mathematics and ELA test scores in grade 6 or grade 7, GPA in grades 6 or 7, CST science score from grade 5 if available, and an indicator set to 1 if the grade-5 science score was missing. (In the spreadsheet, school district administrators enter the CST scores as scaled scores. The spreadsheet converts these to standardized scores by subtracting the state mean in the given year and dividing by the statewide standard deviation of test scores in the given year. In our models, we set the science score, if missing, to 0, so that the coefficient on the dummy for missing science score proxies for the average science score for those with missing data.)

We estimated 12 probit models for completion of a–g course requirements. These 12 models derive from two grade cutoffs (C or higher and D or higher) interacted with six subject areas (overall completion plus completion in English Language Arts, mathematics, science, social studies, and world languages. Appendix Tables A1 and A2 show the results of these models for grade 6, using the D or higher or C or higher definitions respectively. Most coefficients are significant at the 5 percent level or less, and the signs of coefficients are in the expected directions. Appendix Tables A3 and A4 show corresponding results based on grade-7 data.

Results of these models are similar but not identical to the model shown in the main PPIC report. The best predictor in terms of both statistical significance and the size of the effect is grade-6 or grade-7 GPA: a 1 point increase in GPA is associated with a roughly 15 percentage point increase in the chances of completing a–g using either letter grade requirement. Being an EL student or a participant in special education is negatively associated with a–g completion. CST scores in mathematics and English, and to a much lesser degree science, are positively associated with a–g completion. Interestingly, female students are associated with better outcomes as of grade 6, but this effect weakens somewhat in grade 7, probably because other measured factors, such as grades and test scores, are becoming more strongly correlated with gender.

In the spreadsheet, we calculate the predicted probability of completing a given a–g course requirement using the coefficients generated by the probit models estimated using the SDUSD class of 2011 sample. If the matrix X is the nXk matrix of k explanatory variables (for n students) and β is the corresponding vector of coefficients, the probability of completing the a–g course requirement for student i is set to $\Phi(X_i\hat{\beta})$, where $\hat{\beta}$ refers to the estimated coefficients and $\Phi(X_i\hat{\beta})$ is the cumulative distribution function of a standard normal random variable evaluated at $X_i\hat{\beta}$.

TABLE A1
Probit coefficients for six models of a-g course completion based on grade 6 and grades of D or higher

Variables	Overall	Math	English	Science	Social Studies	World Language
Female	0.0741	0.0142	0.101*	0.0489	0.0321	0.0730
	(0.0407)	(0.0422)	(0.0406)	(0.0437)	(0.0417)	(0.0418)
EL	-0.224**	-0.123*	-0.123*	-0.0856	-0.0604	-0.227**
	(0.0551)	(0.0539)	(0.0522)	(0.0544)	(0.0529)	(0.0533)
RFEP	-0.127*	-0.0839	-0.192**	-0.107	-0.0834	-0.122*
	(0.0537)	(0.0564)	(0.0547)	(0.0600)	(0.0571)	(0.0557)
Special Education	-0.245**	-0.173*	-0.109	-0.199**	-0.164*	-0.347**
	(0.0739)	(0.0703)	(0.0668)	(0.0680)	(0.0668)	(0.0698)
Percent Days Absent	-0.0409**	-0.0341**	-0.0386**	-0.0429**	-0.0407**	-0.0403**
	(0.00568)	(0.00539)	(0.00515)	(0.00523)	(0.00511)	(0.00546)
Grade Point Average	0.488**	0.478**	0.444**	0.368**	0.340**	0.445**
	(0.0329)	(0.0330)	(0.0319)	(0.0337)	(0.0325)	(0.0328)
CST Math Score	0.194**	0.364**	0.106**	0.148**	0.0973**	0.141**
	(0.0344)	(0.0385)	(0.0355)	(0.0395)	(0.0369)	(0.0369)
CST Reading Score	0.117**	0.0823*	0.101**	0.0864*	0.0833*	0.177**
	(0.0382)	(0.0404)	(0.0383)	(0.0416)	(0.0395)	(0.0398)
Grade 5 CST Science Score	0.0544	0.104**	0.0653*	0.0840*	0.0781*	0.119**
	(0.0316)	(0.0335)	(0.0319)	(0.0346)	(0.0329)	(0.0332)
Indicator for No Science Score	-0.286	-0.0938	-0.428**	-0.396**	-0.291	-0.449**
	(0.167)	(0.156)	(0.151)	(0.151)	(0.149)	(0.156)
Not in District in Grade 5	0.0347	-0.124	0.0548	0.0123	-0.0213	0.142
	(0.187)	(0.176)	(0.171)	(0.171)	(0.169)	(0.176)
Constant	-1.304**	-0.846**	-0.671**	-0.0329	-0.115	-0.668**
	(0.101)	(0.0992)	(0.0956)	(0.1000)	(0.0968)	(0.0983)
Observations	5,241	5,241	5,241	5,241	5,241	5,241

Standard errors in parentheses
** p<0.01, * p<0.05

TABLE A2
Probit coefficients for six models of a-g course completion based on grade 6 and grades of C or higher

Variables	Overall	Math	English	Science	Social Studies	World Language
Female	0.143**	0.0567	0.159**	0.0686	0.0323	0.111**
	(0.0440)	(0.0426)	(0.0407)	(0.0417)	(0.0402)	(0.0414)
EL	-0.197**	-0.0212	-0.0972	-0.0918	-0.0659	-0.0885
	(0.0640)	(0.0572)	(0.0548)	(0.0531)	(0.0520)	(0.0541)
RFEP	-0.165**	-0.0126	-0.151**	-0.0695	-0.120*	-0.105
	(0.0578)	(0.0559)	(0.0543)	(0.0561)	(0.0540)	(0.0550)
Special Education	-0.153	-0.132	-0.0258	-0.159*	-0.175**	-0.269**
	(0.0881)	(0.0764)	(0.0719)	(0.0685)	(0.0669)	(0.0718)
Percent Days Absent	-0.0397**	-0.0313**	-0.0343**	-0.0312**	-0.0340**	-0.0336**
	(0.00665)	(0.00583)	(0.00563)	(0.00532)	(0.00518)	(0.00553)
Grade Point Average	0.659**	0.587**	0.589**	0.420**	0.414**	0.524**
	(0.0379)	(0.0344)	(0.0332)	(0.0325)	(0.0317)	(0.0330)
CST Math Score	0.240**	0.456**	0.105**	0.297**	0.129**	0.218**
	(0.0359)	(0.0380)	(0.0347)	(0.0380)	(0.0352)	(0.0364)
CST Reading Score	0.0996*	0.0523	0.197**	0.108**	0.155**	0.153**
	(0.0411)	(0.0404)	(0.0386)	(0.0398)	(0.0381)	(0.0393)
Grade 5 CST Science Score	0.0798*	0.0860*	0.0594	0.111**	0.0569	0.0983**
	(0.0341)	(0.0336)	(0.0319)	(0.0332)	(0.0316)	(0.0327)
Indicator for No Science Score	-0.221	0.00655	-0.294	-0.188	-0.132	-0.408*
	(0.194)	(0.163)	(0.163)	(0.152)	(0.149)	(0.160)
Not in District in Grade 5	0.110	-0.182	0.103	-0.0710	-0.0900	0.156
	(0.214)	(0.184)	(0.183)	(0.172)	(0.169)	(0.180)
Constant	-2.393**	-1.690**	-1.696**	-0.674**	-0.706**	-1.207**
	(0.120)	(0.106)	(0.102)	(0.0978)	(0.0955)	(0.1000)
Observations	5,241	5,241	5,241	5,241	5,241	5,241

Standard errors in parentheses
** p<0.01, * p<0.05

TABLE A3
Probit coefficients for six models of a-g course completion based on grade 7 and grades of D or higher

Variables	Overall	Math	English	Science	Social Studies	World Language
Female	0.0292	-0.0339	0.0743*	0.0198	0.00553	0.0694
	(0.0365)	(0.0375)	(0.0362)	(0.0385)	(0.0370)	(0.0368)
EL	-0.331**	-0.0977	-0.185**	-0.0548	-0.0236	-0.275**
	(0.0545)	(0.0520)	(0.0501)	(0.0519)	(0.0506)	(0.0510)
RFEP	-0.116**	-0.0546	-0.129**	-0.0495	-0.0503	-0.134**
	(0.0440)	(0.0461)	(0.0448)	(0.0486)	(0.0464)	(0.0450)
Special Education	-0.336**	-0.249**	-0.0728	-0.215**	-0.185**	-0.426**
	(0.0648)	(0.0606)	(0.0580)	(0.0587)	(0.0575)	(0.0598)
Percent Days Absent	-0.0417**	-0.0403**	-0.0475**	-0.0447**	-0.0403**	-0.0355**
	(0.00461)	(0.00428)	(0.00413)	(0.00408)	(0.00396)	(0.00420)
Grade Point Average	0.474**	0.456**	0.367**	0.322**	0.326**	0.345**
	(0.0286)	(0.0285)	(0.0274)	(0.0289)	(0.0279)	(0.0278)
CST Math Score	0.156**	0.365**	0.0772*	0.129**	0.0750*	0.155**
	(0.0300)	(0.0325)	(0.0305)	(0.0330)	(0.0312)	(0.0311)
CST Reading Score	0.151**	0.0802*	0.168**	0.0988**	0.0813*	0.188**
	(0.0336)	(0.0346)	(0.0332)	(0.0353)	(0.0338)	(0.0341)
Grade 5 CST Science Score	0.0542*	0.0805**	0.0429	0.0770**	0.0780**	0.102**
	(0.0275)	(0.0289)	(0.0277)	(0.0298)	(0.0284)	(0.0284)
Indicator for No Science Score	-0.266	-0.151	-0.384**	-0.308*	-0.248*	-0.410**
	(0.139)	(0.126)	(0.123)	(0.120)	(0.118)	(0.127)
Not in District in Grade 5	-0.0326	-0.0347	-0.0581	-0.0149	-0.0579	0.134
	(0.148)	(0.135)	(0.132)	(0.130)	(0.128)	(0.136)
Constant	-1.145**	-0.667**	-0.329**	0.159	-0.00456	-0.341**
	(0.0860)	(0.0839)	(0.0812)	(0.0846)	(0.0822)	(0.0825)
Observations	6,712	6,712	6,712	6,712	6,712	6,712

Standard errors in parentheses
** p<0.01, * p<0.05

TABLE A4
Probit coefficients for six models of a-g course completion based on grade 7 and grades of C or higher

Variables	Overall	Math	English	Science	Social Studies	World Language
Female	0.0393	-0.0117	0.118**	0.0449	0.0173	0.0842*
	(0.0397)	(0.0381)	(0.0363)	(0.0369)	(0.0356)	(0.0366)
EL	-0.360**	-0.0399	-0.203**	-0.121*	-0.0924	-0.211**
	(0.0670)	(0.0563)	(0.0535)	(0.0509)	(0.0499)	(0.0523)
RFEP	-0.122*	0.0196	-0.0957*	-0.0537	-0.0500	-0.0823
	(0.0476)	(0.0460)	(0.0443)	(0.0455)	(0.0441)	(0.0445)
Special Education	-0.227**	-0.185**	-0.0309	-0.202**	-0.168**	-0.381**
	(0.0787)	(0.0667)	(0.0618)	(0.0588)	(0.0575)	(0.0621)
Percent Days Absent	-0.0405**	-0.0359**	-0.0374**	-0.0316**	-0.0347**	-0.0271**
	(0.00554)	(0.00475)	(0.00446)	(0.00411)	(0.00406)	(0.00426)
Grade Point Average	0.687**	0.611**	0.538**	0.426**	0.434**	0.464**
	(0.0330)	(0.0297)	(0.0284)	(0.0280)	(0.0273)	(0.0281)
CST Math Score	0.175**	0.401**	0.0819**	0.248**	0.0939**	0.187**
	(0.0321)	(0.0322)	(0.0302)	(0.0315)	(0.0299)	(0.0308)
CST Reading Score	0.146**	0.0877*	0.219**	0.100**	0.133**	0.171**
	(0.0367)	(0.0351)	(0.0336)	(0.0340)	(0.0327)	(0.0339)
Grade 5 CST Science Score	0.0806**	0.0776**	0.0573*	0.115**	0.0629*	0.0935**
	(0.0298)	(0.0291)	(0.0276)	(0.0286)	(0.0273)	(0.0282)
Indicator for No Science Score	-0.291	-0.0830	-0.306*	-0.0792	-0.0947	-0.349**
	(0.172)	(0.140)	(0.135)	(0.122)	(0.120)	(0.131)
Not in District in Grade 5	0.143	-0.0117	0.0727	-0.186	-0.130	0.111
	(0.181)	(0.149)	(0.144)	(0.131)	(0.129)	(0.140)
Constant	-2.337**	-1.667**	-1.430**	-0.608**	-0.693**	-0.965**
	(0.102)	(0.0897)	(0.0856)	(0.0825)	(0.0809)	(0.0838)
Observations	6,712	6,712	6,712	6,712	6,712	6,712

Standard errors in parentheses
** p<0.01, * p<0.05

PPIC | PUBLIC POLICY INSTITUTE OF CALIFORNIA

The Public Policy Institute of California is dedicated to informing and improving public policy in California through independent, objective, nonpartisan research on major economic, social, and political issues. The institute's goal is to raise public awareness and to give elected representatives and other decisionmakers a more informed basis for developing policies and programs.

The institute's research focuses on the underlying forces shaping California's future, cutting across a wide range of public policy concerns, including economic development, education, environment and resources, governance, population, public finance, and social and health policy.

PPIC is a private operating foundation. It does not take or support positions on any ballot measures or on any local, state, or federal legislation, nor does it endorse, support, or oppose any political parties or candidates for public office. PPIC was established in 1994 with an endowment from William R. Hewlett.

Mark Baldassare is President and Chief Executive Officer of PPIC.
Gary K. Hart is Chair of the Board of Directors.

PUBLIC POLICY INSTITUTE OF CALIFORNIA
500 Washington Street, Suite 600
San Francisco, California 94111
phone: 415.291.4400
fax: 415.291.4401
www.ppic.org

PPIC SACRAMENTO CENTER
Senator Office Building
1121 L Street, Suite 801
Sacramento, California 95814
phone: 916.440.1120
fax: 916.440.1121