

## Evaluation of the Voyager Program for Students with Disabilities

Pete Goldschmidt, Ph.D.

This evaluation examines the first year impact of Voyager. The evaluation focuses on the quantitative impact as indicated by student performance on the California Standards Test<sup>1</sup>. Hence, several important aspects are not covered in this evaluation and ought to be considered. To that extent, two important elements are whether Voyager is being applied to the correct students (i.e. the students for whom the program was designed to serve) and the implementation of Voyager. The latter issue is indirectly addressed as part of this evaluation. Implementation is important to consider as many schools have not implemented the program for much time and therefore it is unlikely to have much of an effect. Also, there is likely to be variation in program implementation quality. Hence, without data on time and implementation quality, these two elements are confounded as potential causes of the variability in impact among schools. The results indicate that students participating in Voyager score about 13 to 16 scale score points below SWD not participating in Voyager, in Mathematics and English Language Arts respectively. However, there is significant variation among schools in the program's impact. These results are fully detailed below.

In order to fully consider the impact of Voyager, both the demographic characteristics of the target population and participants, as well as outcomes are presented below.

---

<sup>1</sup> Standards, or criterion referenced, tests are better suited for evaluation purposes than norm-referenced tests; however, given the large scale nature of the CST, student improvement on specific content may not be readily apparent on CST results. On the other hand, for the Consent Decree and NCLB, the CST is the standard used to hold schools and the district accountable. Excludes CAPA results

## Demographics

In order to determine whether Voyager is serving the students that the program is intended to serve the demographic make-up of participants is presented, and by way of comparison, the LAUSD demographics of SWD are presented as well. It is also important to consider the demographic make-up of program participants if the program is over-represented by a particular subset of students that may or may not be more or less difficult to serve or are not the intended target.

Table 1<sup>2</sup> presents the distribution of low and high incidence disabilities<sup>3</sup>. The complete Voyager dataset included 7,101 students that were classified as having participated in Voyager. The analysis focuses on about 5,300 elementary students who had valid assessment information for both the current 2006-2007 school year as well as the prior year (2005-2006). CST results (presented in table 4) for the 5,300 students are identical to those for these 7,101 students.

Table 1  
SWD Classification by Program Participation

Incidence	Not in		In	
	<u>Voyager</u>	<u>Percent</u>	<u>Voyager</u>	<u>Percent</u>
Low	2,483	23	705	13
High	8,311	77	4,629	87
Total	10,794		5,334	

Notes:

Valid CST scores only.

Elementary school students only.

Table 1 and all subsequent tables and analyses are based on elementary students (grades 2-5)<sup>4</sup> and those that have valid CST information. The demographic

<sup>2</sup> Table 1 and all subsequent tables are based on data in the STAR file. Further, the results are based on students with valid test scores and valid demographic information. In general, while the absolute value of the number of students differs (when using only students with valid test scores) the percentages are nearly identical.

<sup>3</sup> High incidence is defined as SLI and SLD, low incidence is defined as all other disability categories.

<sup>4</sup> 99.6% of Voyager participants are in grades 2-5.

information appears to be representative of all 7,101 Voyager participants and indicates that results are likely representative as well. Table 1 indicates that the Voyager students are substantively more likely to be classified as high incidence than elementary SWD who are not participating in Voyager.

Table 2 presents the grade distribution of Voyager participants used in the analysis.

Table 2  
Grade distribution by Program Participation

Grade	Not in		In	
	<u>Voyager</u>	<u>Percent</u>	<u>Voyager</u>	<u>Percent</u>
2	144	1	64	1
3	3,368	31	1,482	28
4	3,558	33	1,810	34
5	3,724	35	1,978	37
Total	10,794		5,334	

Notes:

Valid CST scores only.

Elementary school students only.

Voyager is approximately evenly split among grades 3, 4, and 5. Only those 2<sup>nd</sup> graders who were also 2<sup>nd</sup> graders in 2005-2006 are include in this analysis as the subsequent analysis requires prior year assessment results. Unconditional results are not affected.

Table 3 presents the general demographic characteristics of students who are participating in Voyager. The results in Table 3 indicate that Voyager is over-represented by Hispanic students and that virtually all of the Voyager participants are classified as low SES (eligible for free or reduced priced lunch).

Table 3  
 General Demographics by Program Participation

Gender	Not in		In	
	<u>Voyager</u>	<u>Percent</u>	<u>Voyager</u>	<u>Percent</u>
Boys	7,298	68	3,585	67
Girls	3,496	32	1,749	33
Ethnicity				
Native America	52	0	15	0
African American	1,492	14	791	15
Asian	245	2	25	0
Filipino	168	2	26	0
Pacific Islander	16	0	2	0
Hispanic	7,206	67	4,400	82
White/Other	1,615	15	75	1
Language Status				
ELL	5,187	48	3,639	68
SES Status				
Low SES	8,901	82	5,322	100
Total	10,794		5,334	

Notes:

Valid CST scores only.

Elementary school students only.

## Results

As noted the evaluation is based on examining the performance differences by students in ELA and Mathematics who participated in Voyager against those SWD who did not participate in Voyager accounting for observable differences among students that could potentially alter mean results. The CST results are presented in scale scores as performance levels tend to hide small differences in performance and eliminate all within level changes. Further, performance levels are not on a continuous scale and cannot be mathematically manipulated or summarized as readily as scale scores. Table 4 displays the unconditional results for 2006-2007 and 2005-2006 CST results by treatment condition (program participation). The results in table 4 indicate that that

students in the program score substantively lower on both ELA and Mathematics. The unconditional difference in performance is about 29 points in ELA and 39 points in Mathematics. Table 4 also highlights why simple unconditional results may provide misleading interpretations of program effects. This is exemplified by the 2005-2006 CST results. The 2005-2006 results are centered around the elementary SWD mean<sup>5</sup> to highlight pre-existing performance differences between voyager participants and non-participants. Hence, the 2005-2006 CST results in Table 4 indicate that Voyager participants scored about 21.5 and 28.6 below the SWD elementary school average, while

Table 4

Unconditional CST results by Program Participation

CST	Not in		In		In - Not
	<u>Voyager</u>		<u>Voyager</u>		
	mean	s.d.	mean	s.d.	
06-'07					
ELA	291.9	48.1	262.9	29.6	-29.1
Math	302.7	72.6	263.3	47.6	-39.4
05-'06 (centered on average elementary SWD performance)					
ELA	10.7	51.6	-21.5	29.2	-32.2
Math	14.1	75.1	-28.6	50.7	-42.8

non-participants scored about 10.7 and 14.1 points above average in ELA and Mathematics, respectively. This indicates that Voyager participants started the 2006-2007 school year substantively behind their non-participant classmates, and account to some extent for current year performance differences.

The analysis needs to consider two important aspects: one, student inputs, as highlighted by table 4, that indicates pre-existing differences; and two, the impact of schools. The first aspect is accounted for by including prior performance, and other available student information (e.g. observable

<sup>5</sup> This centering also eases interpretation of subsequent results.

characteristics presented in tables one through three. The second aspect is accounted for by utilizing a multilevel model that includes two levels; at level one are individual student outcomes and student characteristics and at level two are schools (see appendix A).

Additional aspects that need to be considered to assist in parsing out potential program effects are whether there are joint effects among particular factors. That is, it may be that student scores differ by program participation, grade, and the combination of program participation and grade. In this instance, it would be important to compare the effect of the program among students in the same grade. The grade and program joint effect indicates, in this case, whether the grade a student is in moderates the impact of Voyager. Several joint effects were considered but only the salient ones are presented below. One of the more important joint effects considered is the effect of prior performance and program participation. Given that Voyager students started behind non-Voyager students, it is important to not only account for pre-existing differences, but also examine whether Voyager has a greater impact depending on initial performance levels. It may be, for example, that Voyager works well for students who start significantly behind (or ahead).

Tables 5 and 6 present multilevel model results for ELA and Mathematics, respectively. The results are based on models that are calculated on elementary SWD only<sup>6</sup>. Table 5 presents results from three models. Model 1 summarizes the unconditional results that are used to partition the variation in CST ELA performance into that attributable to students within schools and that attributable to schools. The results from model 1 in table 5 indicate that about 23.4% of the variation in ELA CST scores is attributable to differences among

---

<sup>6</sup> Given that the results are based on models including only SWD, performance gaps between SWD and non-SWD are not estimated. Preliminary models indicate that the gaps do not change based on program participation.

schools, while the majority of the variation in ELA CST results is attributable to differences among students within schools. This indicates that a substantive majority of the variation on ELA CST scores is due to different student inputs and process experienced by students within schools. Still, it is important to account for the natural structure of the data (student nested within schools) as even between-school variation as small as 5% can have significant impact on standard errors.

Table 5  
English Language Arts CST Results

<i>Fixed Effects</i>	Model 1		Model 2		Model 3	
	<u>Estimate</u>	<u>s.e.</u>	<u>Estimate</u>	<u>s.e.</u>	<u>Estimate</u>	<u>s.e.</u>
Grand Mean	285.1	1.07	293.6	1.10	282.7	1.37
Voyager effect			-30.4	1.21 ***	-14.8	2.07 ***
High Incidence SWD					-4.1	0.82 ***
Joint Voyager-High Incidence					-1.7	1.14
Grade effect					3.3	0.46 ***
Joint Grade-Voyager					2.7	0.75 ***
2005-2006 CST effect					0.6	0.01 ***
<i>Random Effects</i>	<u>Variance</u>	<u>Percent</u>	<u>Variance</u>	<u>Percent</u>	<u>Variance</u>	<u>Percent</u>
				<u>Accounted</u>		<u>Accounted</u>
Means						
Between Students	1608	76.6	1535.9	4.5	955.7	40.6
Between Schools	490	23.4	453.6	7.4	90.9	81.4
Voyager						
Between Schools			368.7		82.5	77.6

The student process of interest is Voyager participation. Model 2 includes an indicator for Voyager participation and estimates the effect of Voyager. The results from model 2 indicate that students participating in Voyager are estimated to score approximately 30 points below non-participants (this result is similar to that displayed in table 4, based simply on means). The results from model 2 also indicate that there is significant variability in the voyager effect

between schools (this is indicated by the random effect for Voyager). In fact the standard deviation of the Voyager random effect is about 19 (square root of 368.7). This implies that the range (95% CI) for the Voyager effect is -30 +/- 38. Hence, there is substantial variation in the Voyager effect- which, as noted, is likely due to differences among schools in student inputs, time since implementation and implementation fidelity.

As noted, it is important to account for observed differences in the Voyager and non-Voyager groups. Model 3 accounts for program participation as well as whether the student is classified in a high or low incidence disability category, grade, prior year CST performance, and joint effects for high incidence and grade<sup>7</sup>. The results indicate that accounting for prior year performance, SWD classification, and grade, the difference between students in Voyager and those not in Voyager is -14.8 points. The results imply that, overall, SWD classified as SLD and SLI perform about 4 points lower than SWD with low incidence classifications on the ELA CST, but that Voyager is not differentially effective for the two types of classifications. The results also indicate that students in high grades perform slightly better than students in lower grades in ELA. The results also indicate that Voyager is somewhat more effective in higher grades than in lower grades. The effect is about 2.7 points per grade. The effect of program and grade can be seen in Figure 1.

Preliminary models examined whether Voyager was more effective for students with lower (higher) prior performance (i.e. the joint Voyager prior performance effect). No effect was found.

Including prior performance, grade and SWD classification accounts for approximately 41% of the between student (within school) variability in 2006-

---

<sup>7</sup> Preliminary model also examine student background characteristics, but these do not affect the impact of Voyager and simply add additional variables to the model that unnecessarily complicate interpretation.



2007 ELA CST performance. These predictors account for about 81% of the between school variability and about 77% of the between school variability in the impact of Voyager.

Table 6 presents the results for CST Mathematics and is interpreted in a similar fashion to table 5. Again, model 1 estimates the amount of variability within and between schools. Similar to the ELA results, about 20% of the variation in Mathematics performance is between schools. The unconditional Voyager effect is presented in model 2 and is also quite similar to the difference displayed in table 4. The Mathematics model 3 incorporates the same factors as the ELA model 3. The results are somewhat different from the ELA results. The conditional difference between Voyager and non-Voyager students is about -11 points. Again, this means comparing two students with average prior year performance. Preliminary models examined whether Voyager was differently effect for students with different prior year performance (the joint effect of

Table 6  
Mathematics CST Results

	Model 1		Model 2		Model 3	
	<u>Estimate</u>	<u>s.e.</u>	<u>Estimate</u>	<u>s.e.</u>	<u>Estimate</u>	<u>s.e.</u>
<i>Fixed Effects</i>						
Grand Mean	293.0	1.5	305.45	1.6	309.9	1.94
Voyager effect			-42.01	1.8 ***	-11.3	2.78 ***
High Incidence SWD					-1.9	1.10 *
Joint Voyager-High Incidence					-2.2	1.46
Grade effect					-7.5	0.73 ***
Joint Grade-Voyager					1.5	1.14
2005-2006 CST effect					0.7	0.01 ***
<i>Random Effects</i>	<u>Variance</u>	<u>Percent</u>	<u>Variance</u>	<u>Percent</u>	<u>Variance</u>	<u>Percent</u>
Means				<u>Accounted</u>		<u>Accounted</u>
Between Students	3806.2	80.5	3645.2	4.2	1814.1	52.3
Between Schools	921.1	19.5	900.6	2.2	149.8	83.7
Voyager						
Between Schools			571.8		71.0	87.6

of Voyager and prior year Mathematics CST results), but none were found. Student mathematic performance decreases with grade (this is likely due to the districts overall emphasis on ELA). Unlike ELA, the gap between Voyager and non-Voyager students does not close significantly as grade increases.

The grade and voyager effects for both Mathematics and ELA are summarized in figure 1. Figure 1 demonstrates that as grade increases Voyager ELA students perform better (although still below non-voyager students). A similar, though not statistically significant, pattern exists for Mathematics as well. Again, these results account for prior performance.

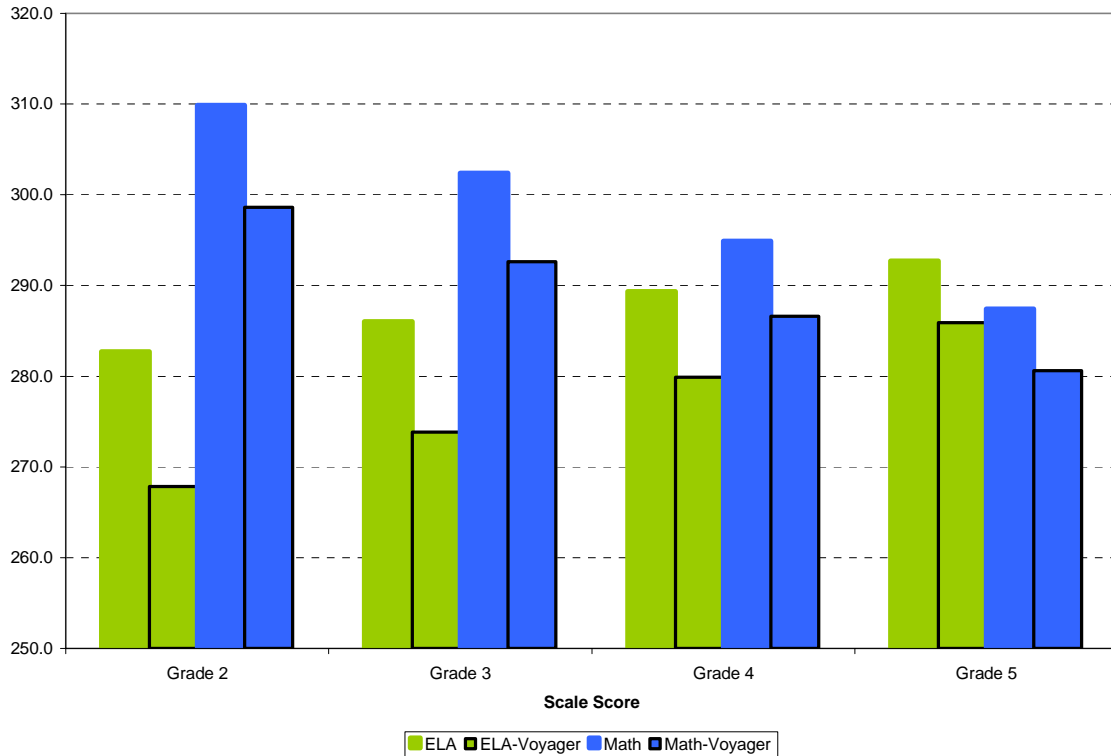


Figure 1: The joint effect of Voyager and grade for ELA and Mathematics

### Discussion

The first year results of Voyager indicate that students who participated in Voyager perform below students who do not participate in Voyager. While it is true that students who participated in Voyager had lower performance in 2005-2006, they did not close the gap by participating in Voyager. Incorporating prior CST results is important when evaluating the impact of Voyager because of the substantive differences between program and non-program students in 2005-2006 CST performance. Further, it is important to consider SWD classification (high vs low incidence) because Voyager serves a greater proportion of high-incidence classified SWD than there are in grades 2-5 in general, thus the comparison must be against like-students. There is some evidence that Voyager

is somewhat more successful for older students (i.e. grade 5 vs grade 3), but not enough to demonstrate positive results.

The evaluation of Voyager would be further strengthened with additional information. It is reasonable to assume that full implementation had not yet occurred and this plays a role in identifying positive Voyager effects. Additional, information would help partition out effects of implementation. Preliminary results indicate that there is significant variation among schools in Voyager effects. These effects are partially accounted for by schools' input characteristics (e.g. average performance of SWD students at a school) but are likely due to implementation fidelity at each school. Follow-up evaluations should include some implementation information, if possible. Minimally, program start date, as well as some indicator of school-wide selection processes (both into Voyager and SWD in general) would help isolate program differences from school to school variation in performance associated with other causes.

Further, formative information might be generated from the quantitative results by examining those schools demonstrating significantly better than average voyager effects. The program, overall, might benefit from observing what takes place at these schools (as opposed to those scoring at the average or below). Of course, this line of focused research should not ignore whether special circumstances related to student inputs (but not modeled) are the overriding reason for a particular school's success with Voyager.

## Appendix A

### Mathematics

Final estimation of fixed effects  
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	309.886028	1.936413	160.031	486	0.000
For VOYAGER2 slope, B1					
INTRCPT2, G10	-11.283333	2.781549	-4.056	486	0.000
For SWD_HI slope, B2					
INTRCPT2, G20	-1.928337	1.099566	-1.754	16121	0.079
For SWDHV slope, B3					
INTRCPT2, G30	-2.200735	1.460295	-1.507	16121	0.132
For GRADE2 slope, B4					
INTRCPT2, G40	-7.493700	0.725323	-10.332	16121	0.000
For GRADE2V slope, B5					
INTRCPT2, G50	1.497236	1.135934	1.318	16121	0.188
For CSTMTHC slope, B6					
INTRCPT2, G60	0.681907	0.008952	76.176	16121	0.000

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	12.23639	149.72918	213	442.40608	0.000
VOYAGER2 slope, U1	8.42907	71.04919	213	272.43230	0.004
level-1, R	42.59165	1814.04907			

## ELA

Final estimation of fixed effects  
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	282.694612	1.366132	206.931	486	0.000
For VOYAGER2 slope, B1					
INTRCPT2, G10	-14.848478	2.068143	-7.180	486	0.000
For SWD_HI slope, B2					
INTRCPT2, G20	-4.054229	0.817061	-4.962	16191	0.000
For SWDHV slope, B3					
INTRCPT2, G30	-1.690335	1.135323	-1.489	16191	0.136
For GRADE2 slope, B4					
INTRCPT2, G40	3.332866	0.458408	7.271	16191	0.000
For GRADE2V slope, B5					
INTRCPT2, G50	2.682427	0.747313	3.589	16191	0.001
For CSTEELAC slope, B6					
INTRCPT2, G60	0.578417	0.012782	45.251	16191	0.000

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	9.53173	90.85396	213	462.97536	0.000
VOYAGER2 slope, U1	9.08367	82.51312	213	341.05537	0.000
level-1, R	30.91544	955.76454			