AMERICAN INSTITUTES FOR RESEARCH®

Year 4 Service Study:
An Alternative Approach

September 13, 2007

Submitted to:

Office of the Independent Monitor
Los Angeles Unified School District

Submitted by:

American Institutes for Research
Jenifer J. Harr
Miguel Socias

I. Introduction

Since 2004-05, the Program Evaluation and Research Branch (PERB) has conducted annual studies of the delivery of special education services in the Los Angeles Unified School District (LAUSD or the "District"), building on AIR's methodology for the first study in 2003-04. The purpose of these studies was to measure the degree of conformity between the actual provision of special education services in LAUSD with what is required by each student's Individualized Education Program (IEP). In summary, the studies attempt to ascertain the extent to which LAUSD students are actually receiving what is required under special education law.

In Years 2 and 3 (2004-05 and 2005-06), AIR conducted validation studies to provide a cross-check of PERB's methodology and findings and assess whether the results were an accurate depiction of service delivery in LAUSD. A major component of this validation process was an independent data entry conducted by AIR staff comparing IEP and log information for a sub-sample of students, using PERB's data entry rules. The results of these checks over the last two years were fairly similar to PERB's. Given the relative stability of the methodology and consistency of AIR's validation results over the last two years, the Office of the Independent Court Monitor (OIM) and AIR agreed to adopt a different approach for Year 4. As in previous years, AIR had ongoing discussions with OIM and PERB to review methodological decisions, with particular emphasis on refining rules for interpreting data from IEPs and logs. In place of the independent data entry of a sample of logs, AIR was charged in Year 4 with developing and testing a computer program to determine compliance with the IEPs using Welligent log information.

We would like to emphasize that this exercise is not a validation of PERB's estimates. Rather, it is a test of the reliability and feasibility of using a computer program to derive these estimates for Welligent IEPs and logs. We further wanted to identify if there were certain types of services for which the program worked particularly well.

This was considered an important alternative to manual interpretations for several reasons. With the growing number of electronic IEPs and logs, it is more cost-efficient to design a computer program to analyze the data, particularly one that can be replicated annually, as opposed to a manual interpretation (which not only involves labor hours in reviewing all the log data by hand, but also resources in preparing, printing, and organizing the hard copy documents). While it was labor-intensive to develop this initial program, it has the advantage of producing estimates immediately. Furthermore, a standardized program ensures that the rules are applied consistently, improves reliability, and enhances transparency (as a program could be reviewed objectively by all involved parties). In the process, we did uncover some degree of inconsistent application of PERB's rules to the data. With a total of 46 rules described in 19 pages of guidelines, the manual process is expected to have some degree of human error. It might be argued that a program would be too rigid to take into account all the subtle cases involved, and would not be able to detect and adjust automatically to new situations. We reason that in a high-stakes accountability process, consistency, transparency, and replication are key

components. With a program, further adjustments can be made and applied retrospectively to all data (as opposed to changes in manual rules, which can only be implemented going forward, unless the logs were to be re-reviewed).

Developing a program on a large scale such as this would have been virtually impossible without the ability to compare the results to PERB's manual interpretation of hundreds of records. Only on a very large scale would we be able to test an array of situations and build a more reliable program.

II. Steps and Challenges

As a first step, we eliminated students for whom PERB used supplemental information (e.g., non-Welligent) to guide its interpretations. Supplemental information could include a revised IEP (consequent to the Welligent extract) or District-provided information on whether the student left the District or exited special education. We also excluded students for whom at least one of their logs obtained by PERB was a paper log (e.g., non-Welligent). The purpose of these exclusions was to create a clean database in which all information was based entirely upon Welligent data. As the purpose of this exercise was not to derive population estimates, but to test a program on a large scale, we were comfortable with dropping students from the original sample. Ultimately, we eliminated 793 students from the original sample of 4,435.

There were several Welligent log databases involved in this process, due to the fact that the extracts were taken from the Welligent at different points in time (Sept/Oct; Oct/Nov; Jan/Feb). These were compiled into a single dataset. There were many student records which had duplicate service information, including the date of service and the provider name. We were advised by PERB that these were valid services (e.g., that a student could receive the same service type more than once a day), and that we should include these in our analyses. As shown in Exhibit 1, RSP makes up the largest proportion of duplicate records.

Exhibit 1. Duplicate records by service type

Service	0	1	2	3	4	Total
			Number of duplicates			
Audiology	1,251	**2**	0	0	0	1,253
VI	3,400	**636**	**126**	**20**	**15**	4,197
School mental health	2,149	**28**	0	0	0	2,177
DHH	1,297	**14**	0	0	0	1,311
Orientation mobility	2,340	**58**	**3**	**4**	0	2,405
APE	26,178	**326**	**3**	0	0	26,507
LAS	10,037	**646**	**24**	0	0	10,707
Inclusion Facilitator	851	**74**	**3**	0	0	928
PT	818	**10**	0	0	0	828
OT	3,569	**98**	0	0	0	3,667
School mental health	1,950	**94**	**3**	0	0	2,047
LRE Counselor	443	**10**	**3**	0	0	456
LAS	427	**4**	0	0	0	431
PKIT	176	**2**	0	0	0	178
RSP	19,677	**8,566**	**2,070**	**444**	**225**	30,982
Total	**74,563**	**10,568**	**2,235**	**468**	**240**	**88,074**

While we followed PERB's direction, we strongly recommend that the time of each service session be included in subsequent log extracts. This will allow us to eliminate redundant records of service, and distinguish those from service sessions that did occur multiple times in a single day.

Given the complexity of the RSP coding rules, and the relatively lower percentage of Welligent logs for this service as well as continuing concern about the reliability of those logs, we did not create a program to estimate compliance for RSP. We pose alternatives at the end of this report for addressing these services. Furthermore, we were unable to include Visual Impairment (VI) and Adaptive PE (APE) services due to the size limitations in Stata Intercooled Edition, the statistical package on which this program is based on.[1] While we did run the program on Pre-School Services, the small number of records (n = 34) resulted in large differences between AIR and PERB figures (e.g., one or two disagreements between the program and PERB's codes would result in significant differences). Consequently, we do not present the results for those services.

In limited cases, we had to exclude a rule or modify it due to a lack of specificity. For instance, PERB made adjustments to the frequency and duration codes "if the track information is wrong from SIS and they sent a partial log." It was not clear to us how to make this determination, and therefore this rule was excluded. In another example, the PERB manual states "only count the holidays/closures if it follows the pattern of

---

[1] This issue would have also occurred with RSP. Stata Special Edition is needed to circumvent this problem.

services." In several cases, the services can have mixed patterns, with sessions occurring on multiple days of the week. We gave credit only if the holiday or school closure fell on the day with the most provided sessions across the eight weeks. This could result in situations in which holidays/closures were not properly credited as they coincided with the second most frequent day. Further specificity is needed to implement this rule for other days, as described in our recommendations.

By far, the most complicated rule to implement in this program related to crediting provider absences if the absence was due to illness, personal necessity, or jury duty. All individual absences for these reasons were credited as service. However, when the absences were consecutive (e.g., the provider was absence twice or more in a row), the rule credited up to two weeks of service, and this could be applied only once. All other consecutive absences were not allowed. For example, if a service was once weekly, the provider could receive credit (with a valid reason) for two consecutive absences; all other consecutive absences would not be counted. For twice weekly services, the provider could receive credit for up to four consecutive absences, and so forth.

To ensure that differences were not due to a program error, we manually checked discrepancies between the program results and PERB's codes for the OT, PT, and LAS logs. Based on our review, we made adjustments to the program when it was clear that a rule had been overlooked or not correctly implemented.

III. Results

Exhibit 2 presents the compliance rates by service, based on the AIR program and PERB codes. It is important to note that the PERB estimates shown below may not align with its Year 4 report, as the results are based on a smaller sample of students. The reader will note the high degree of alignment in the compliance estimates across both approaches. For evidence of service provision, there are virtually no differences, and for frequency and duration, only certain services have noticeable variation (e.g., Language and Speech; Physical Therapy). Overall, the program appears to be a practical and reliable instrument for deriving estimates.

Exhibit 2. Comparison of compliance rates, AIR Program and PERB Manual Interpretations

	Evidence of Log		Frequency		Duration	
	AIR	PERB	AIR	PERB	AIR	PERB
DHH	95.6%	95.6%	83.9%	83.7%	82.3%	82.6%
LAS	82.9%	82.6%	58.1%	60.6%	58.2%	60.3%
LRE	96.0%	96.0%	87.5%	85.8%	79.7%	79.5%
OT	92.2%	92.2%	73.8%	72.4%	67.7%	67.9%
PT	97.8%	97.8%	75.6%	77.2%	85.6%	82.4%
SMH	88.3%	88.1%	71.3%	70.8%	76.5%	74.3%

While the overall compliance rates are fairly similar across AIR and PERB, the rates obscure the detail of the differences between the program and PERB's interpretations. For example, AIR's program may have assigned a "1" (met compliance) and PERB may have recorded "2" (did not meet compliance) for the same student. However, these differences may be canceled out in reverse situations. Below is an example of the duration codes for Physical Therapy (PT) across the AIR program and PERB.

Exhibit 3. Comparison of duration codes for Physical Therapy, AIR Program and PERB Manual Interpretations

		PERB Codes				
		1	**2**	**3**	**5**	Total
AIR Codes	**0**	1	19	0	0	29
	1	70	3	0	0	74
	Drop	0	3	2	65	119
	Total	117	25	2	65	211

AIR codes: 0 = did not meet compliance; 1 = met compliance; Drop = drop student from analysis.
PERB codes: 1 = met compliance; 2 = did not meet compliance; 3 = could not determine duration from log/drop; 5 = drop student from the analysis.

Across the 211 records, there were seven cases in which the program and PERB codes disagreed (note that PERB's code of 3 and 5 were the equivalent to drop).

- AIR 0 and PERB 1: The IEP required 720 minutes across eight weeks (e.g., 90 minutes per week). The program determined that 360 minutes were provided, which included credit for two student absences and one provider illness.
- AIR 1 and PERB 2: In all three of these cases, the program assigned a "1" (compliance met) due to holiday credits per PERB rules.
  - IEP required 240 minutes; 180 minutes provided (including credit for student absence). As all other services occurred on a Friday, the program provided credit for two Friday holidays that were not documented on the log, for a total of 240 minutes.

- o IEP required 120 minutes. The log met duration because only service provided fell on a Friday, and therefore, the log received two Friday holiday credits that were not documented on the log.
- o IEP required 120 minutes. The log met duration because the two provided services occurred on a Thursday, and therefore, the log received a Thursday holiday credit that was not documented on log.
- ▪ AIR drop and PERB 2: All were September 2006 IEPs. The PERB manual stated that if frequency or duration was not met for IEPs completed in September 2006, the default would be to drop from the analysis.

IV. Recommendations

Given the high degree of accuracy observed in the above tables, we consider this initial program to be quite promising for future use. It offers immediate results and is a cost-effective alternative to the labor-intensive manual process. Furthermore, it ensures consistent application of the rules across thousands of records and allows flexibility in refining the rules at any time (or testing the impact of rule changes). Below we present recommendations to consider for Year 5 of the study, based on the program results and our review of the applicability and appropriateness of the existing rules.

- ▪ **Use a program for future estimates.** We recommend that PERB/OIM use a computer program based on the one developed for this task to generate estimates of service provision for Year 5 for the services included in this report. Further deliberation among the involved parties is needed to determine whether the program should analyze information for all special education students in the District or continuing with a sample – and whether the period of analysis would be longer (e.g., entire year) or the existing two-month/8-week timeframe. The advantage of using all students over a longer period is that it provides more precise compliance rates.

  However, such changes will have implications for time-sensitive rules (such as provider absences, holiday credits, etc.), which may result in different estimates than what previous studies have reported. The fact that the Welligent is an evolving database and is continuously updated with newer IEPs poses challenges for a broader time period that will require further consideration. Even though the percentage of students in Welligent has increased substantially over time, the drawback of analyzing Welligent exclusively is that the estimates will not reflect a small proportion of non-Welligent students. These methodological challenges will need to be further considered and resolved.

- ▪ **RSP services.** As noted above, we did not attempt to address RSP logs in this initial program. We pose two alternatives for this service in Year 5: 1) Continue work to incorporate the RSP rules into the program and test its reliability against PERB's 2006-07 results; or 2) Continue to manually check RSP logs by hand for a sample of students over a two-month period. Unfortunately, this second alternative would pose a problem if the period of analysis were increased

substantially. On the one hand, it could be prohibitively expensive to manually check RSP logs for extended time periods. Conversely, restricting this particular service to a brief time window while analyzing the other services over a longer time frame would make these results incompatible. No overall population estimate could be derived in such situation.

- **Provider absences:** We recommend that the rules for crediting provider absences for this outcome be re-considered. The current interpretation means that a provider for a student who requires a daily service could be absent for 10 consecutive service sessions and still receive credit for those sessions, or credit for four consecutive service sessions for a twice weekly service. The provider could also receive credit for more absences beyond that point if the absences are not consecutive. As an alternative, we propose that the total number of credited absences be capped depending on the required frequency of the service. As capacity is developed for using a program in a comprehensive manner, the issue of the provider absences will be less important as the analysis could span a longer timeframe.

- **Rule specificity.** Clearer rules are needed for how to handle appropriately holidays/closures of schools. As noted earlier, PERB's manual states "only count the holidays/closures if it follows the pattern of services." However, "pattern of services" is ambiguous. When the pattern is mixed, there should be clear guidance on when it is appropriate to provide credit. For instance, the rule might stipulate that a Friday holiday credit would apply if services occurred at least five times on a Friday over eight weeks, for a weekly service.

- **Time of service.** Irrespective of whether a program is used to generate estimates in Year 5, we recommend that the time of the service provision be included in the log extract. This information is critical in eliminating true duplicate records of service delivery for services that occur multiple times in a single day.

- **Yearly services.** We also recommend that the Welligent system be modified to allow reporting yearly frequency and duration for services that do not occur on a weekly or monthly basis. Currently, services are designated as yearly if frequency (weekly or monthly) is zero; however, we do not know how accurate this approach is.