DECLARATION OF DR. STEPHEN KLEIN IN THE MATTER OF:
LOS ANGELES UNIFIED SCHOOL DISTRICT
DISMISSAL OF RICARDO GONZALES

Stephen Klein, Ph.D.
August 21, 2012

Introduction

I was retained by the Los Angeles Unified School District (LAUSD) in the above referenced matter to examine whether the scores of Mr. Gonzales' students supported the district's allegation that in 2010 he gave his students unauthorized and inappropriate assistance on the statewide achievement tests.

The next portion of this declaration describes my qualifications to render an opinion in this matter and the data I used in my analyses. I then describe the scores Mr. Gonzales' students received as they progressed from 4th to 6th grade. Next, I discuss the procedures I used and the results I obtained in my analysis of alternative explanations of why Mr. Gonzales' students did so well on the 5th grade tests relative to these same students' performance on the 4th and 6th grade tests they took in 2009 and 2011 (i.e., explanations other than Mr. Gonzales gave his students unauthorized assistance during the exam). Finally, I present my conclusions.

Qualifications

I received a BS from Tufts University and an MS and Ph.D. in Industrial Psychology from Purdue University. From 1964 to 1968 I was a research scientist at the Educational Testing Service in Princeton, NJ. I then became an Associate Professor in Residence at UCLA where I taught courses in statistics, psychometrics, and research design. I also chaired the UCLA Graduate School of Education's Research Methods Division. I joined the RAND Corporation in 1975 where I directed projects in several fields for 31 years. I have worked on numerous cases of alleged cheating and have been qualified as a statistical expert in state and federal courts. I currently consult for professional licensing boards in law, teaching, and dentistry on a variety of matters related to testing. My full CV (which is attached as Appendix C) lists my over 250 publications.

Introduction

California's statewide testing program assesses students in English Language Arts (ELA) and Mathematics (math). I used data from the 2009, 2010, and 2011 administrations of these exams to investigate whether: (a) the scores of Mr. Gonzales' students on the 2010 tests greatly overestimated their scores on the 2011 tests and (b) whether there is any support for alternative explanations for the results; i.e., other than Mr. Gonzales gave his students inappropriate assistance during the 2010 exam that inflated their scores on these tests?

Data Used in the Analyses

I received an EXCEL file via e-mail from LAUSD labeled "899 Cohort 2 data.xlsx." This file contained the ELA and Math scores for the LAUSD students who took those tests as fifth graders in the spring of 2010. This file also contained these students' and their teacher's LAUSD ID numbers, their 2009 scores on the 4th grade forms of the ELA and Math tests, their 2011 scores on the 6th grade forms of these tests and a "flag" to indicate whether or not English was the only language spoken in the student's home. These data enabled tracking the same students over three years of testing (2009 – 2011).

My initial analyses of these data were done in three steps. Step #1 involved identifying all of the roughly 35,000 students who in 2010 had both of the following characteristics: (a) they were in a 5th grade classroom with at least 15 non-special education students and (b) they took California's statewide ELA and math tests as 5th graders in 2010 and again as 6th graders in 2011. I also analyzed these students' scores on the 4th grade tests they took in 2009.

There were 1,620 fifth grade classrooms that satisfied the sampling requirements above. Mr. Gonzales taught the students in one of these classrooms and it is my understanding he gave the statewide tests to all the non-special education students in that classroom. I also understand there was no proctor or observer present when Mr. Gonzales gave the 5th grade tests to his students in 2010.

In Step #2, I used the students' ID numbers to create a computer simulated or "virtual" 4th and 6th grade classroom for every actual 5th grade classroom. The virtual classrooms contained the same students as their actual 5th grade classroom. Thus, the students in a virtual classroom had the same classmates as were in their actual 5th grade classroom in 2010.

Step #3 involved computing the mean ELA and Math scores in each actual and virtual classroom in each year as well as the percentile equivalents of these means. For example, a classroom at the 60th percentile of the 5th grade classrooms in 2010 had a mean score that was as higher than the mean in 60 percent of the district's other 5th grade classrooms.

Comparison of Scores Over Time

The 4th graders who migrated into Mr. Gonzales' 5th grade classroom had mean spring 2009 ELA and Math scores that were somewhat higher than the overall average scores among all of the 1,620 actual classrooms in the analysis. Specifically, the percentile ranks of Mr. Gonzales' 5th grade mean ELA and Math scores were 54 and 63, respectively.

Mr. Gonzales gave the 5th grade tests to his students in the spring of 2010. His classroom's mean scores on those tests were at the 99th percentile; i.e., his classroom was among the top 1% in the district. This is incredible improvement compared to these same students' 4th grade scores a year earlier. However, when these same students took the 6th grade tests given in 2011 (i.e., tests that were administered by their 6th grade teachers), their scores plummeted from the 99th percentile in 2010 to about the 30th percentile in 2011 (see Table 1).

In short, students coming out of Mr. Gonzales' classroom did not repeat their extraordinarily high 2010 scores. In fact, they did not even come close to achieving their 2009 percentile ranks. Instead, the 3-year pattern of scores supports the thesis that (a) the 2010 scores in Mr. Gonzales classroom were grossly inflated and (b) the students in Mr. Gonzales' virtual 6th grade classroom were adversely affected by having Mr. Gonzales as their 5th grade instructor.

Table 1
Mean Scores and Percentile Equivalents of the Mean Scores of Mr. Gonzales' Students When they were in 4th, 5th, and 6th Grade

	2009 4th Grade		2010 5th Grade		2011 6th Grade	
Test	**Mean**	**%ile**	**Mean**	**%ile**	**Mean**	**%ile**
ELA	358	54	414	98	325	31
Math	390	63	519	99	314	24
Total	374	60	467	99	320	29

Alternative Explanations

I examined whether there were plausible alternative explanations for the results in Table 1 (i.e., other than Mr. Gonzales' students received unauthorized assistance in 2010). For instance, were his students' 2010 scores due to chance or those students being assigned to relatively incompetent 6th grade teachers? However, as discussed in Appendix A and B, none of these theories are supported by the data. For example, the likelihood of the observed score patterns being due to chance are less than one in a million.

Conclusion

Mr. Gonzales' 5th graders had scores on their 2010 tests that greatly overestimated their 2011 scores. The difference between these estimates and the students' actual scores were far beyond what could conceivably occur by chance. Something else must have caused the estimates to be so wrong.

The same is true for why the 2010 scores of Mr. Gonzales' 5th graders in 2010 are so much higher than would be expected given those same students' 4th and 6th grade scores. In addition, it is evident from the size and prevalence of the inflated scores in 2010 that they were not due to a few students or an occasional mistake or minor breach in the district's standardized testing procedures. The problem must have been much more widespread to get Mr. Gonzales 5th graders' mean ELA **and** math scores into the 99th percentile and then have them fall into the bottom quarter the following year when those same students took the 6th grade tests.

Mr. Gonzales claimed that his 2011 fifth graders did not do as well on their exams as his 2010 fifth graders did on their exams because the 2010 group was a little more academically able. He based his opinion about differential ability "on tests and personal observations" (July 17, 2002 deposition, page 111, line 8). He did not mention that there was an observer in his classroom during the 2011 but not the 2010 exams.

Finally, there was no evidence that Mr. Gonzales' 5th graders in 2010 were assigned to relatively inept 6th grade teachers in 2011 (see Appendix A). In addition, Mr. Gonzales said he followed all the rules regarding document control. Thus, **the only plausible explanation that remains for the score swings shown in Table 1 is that Mr. Gonzales inappropriately assisted several if not most of his 5th graders during his 2010 administration of the exams. No other explanation fits the data.**

Appendix A
The Chance Hypothesis

I used multiple regression analyses to investigate whether the huge swings in percentile equivalents in Table 2 were likely to be attributable to chance. To do this, I constructed a separate multiple linear regression equation to predict each test score (ELA, math, and total). This equation provides the best fitting (i.e., most accurate) straight line estimate of a "virtual" classroom's mean 2011 sixth grade score based on (a) that classroom's mean 2010 test score and (b) the percentage of students in the classroom who are living in English Only (EO) homes. The EO variable was used to adjust for possible linguistically related background characteristics.

This equation takes the following form: $y' = a + b_1x_1 + b_2x_2$ where:

y' is the predicted 2011 score
a is a computer derived constant for scaling the y' values
x_1 is the classroom's 2010 mean score
x_2 is the percentage of EO students in the classroom
b_1 is the computer derived weight for the 2010 score
b_2 is the computer derived weight for the EO value

The dependent variable that is used to construct the equation is the classroom's mean sixth grade score in 2011. I constructed separate equations to predict the 2011 sixth grade ELA, math, and total scores. The R^2 values associated with these equations were: .88, .71, and .81, respectively. These values are fairly large which suggests that the abilities tested by the 2011 tests are similar to those measured by the 2010 tests.

I used multiple regression analysis to find the best fitting straight line between a virtual classroom's mean 2011 test score and that classroom's combination of mean 2010 scores and the percentage of students who were classified as "EO" (i.e., English Only spoken in the home). There were separate equations for predicting ELA, math, and total scores. I then used these equations to compute the size of disparity between the predicted and actual mean 2011 scores in Mr. Gonzales' classroom.

The difference between a classroom's actual and predicted mean score is called its "residual" score. For example, the actual and predicted mean scores of Mr. Gonzales' students when they took the sixth grade ELA tests in 2011 were 325 and 401, respectively. Mr. Gonzales' students' mean residual ELA score was therefore 325 – 401 = -76.

Of the over 1,500 other fifth grade classrooms studied, none had a residual score that came even close to -76. In short, *Mr. Gonzales' 2010 fifth graders did far worse on the sixth grade ELA exam than what they were expected to do on the basis of their EH percentages and fifth grade scores in 2010.*

The grade 5 math scores of Mr. Gonzales' students overestimated their sixth grade math scores by 111 points. And, their average total sixth grade score (i.e., the average of their mean sixth grade ELA and math scores) was over estimated by 97 points. *Table 2 shows that all of these differences are far beyond what is likely to occur by chance. Thus, the results of this study are consistent with and thereby strongly support LAUSD's allegation that Mr. Gonzales' students received unauthorized and inappropriate assistance.*

Table 2
Statistics used to calculate the probability that the observed difference between the actual and predicted scores could have arisen by chance.

Test	Actual Score	Predicted Score	Residual Score	Standard Error	z-score	Probability p-value
ELA	325	401	76	10.82	7.02	<.00000000001
Math	314	425	111	23.39	4.75	<.000001
Total	320	417	97	16.18	6.00	<.000000001

Residual score = Actual score – Predicted score
z-score = Residual score/Standard error
p-value = probability that the observed difference between the actual and predicted score was likely to arise by chance. For example, a p-value of <.000001 indicates there was less than one chance in 100,000.

Appendix B
Blaming the Sixth Grade Teachers

Mr. Gonzales had 16 fifth grade students who went on to sixth grade and took that grade's tests in the spring of 2011. These students were distributed across 44 sixth grade teachers. Only one of these sixth grade teachers had more than four of Mr. Gonzales' former students. It is difficult to believe that most of these 44 sixth grade teachers were grossly inept, which is what would have had to occur for Mr. Gonzales' students to have done so well on the 2010 tests relative to their 2011 scores. Nevertheless, Mr. Gonzales may try to blame the 6th grade teachers in 2011 for his 2010 students doing so poorly in 2011.

To investigate this thesis, I identified all of the 32,574 sixth graders in the LAUSD database that had all of the following characteristics: (a) they had a 5th grade ELA and math scores on the 2010 statewide exam; (b) they had ELA and Math scores on the 2011 sixth grade statewide exam; (c) they had a fifth grade teacher ID #; and (d) they had the ID numbers for one or more of their sixth grade teachers. I used these students' data to construct the best fitting straight line equation to predict a student's sixth grade ELA, Math, and Total score. The predictors in all three equations were the student's corresponding fifth grade score and English Only (EO) classification.

There were 1,282 sixth graders who had one or more of the 44 teachers that had at least one of Mr. Gonzales' fifth grade students in their classroom. I divided these 1,282 students into two groups. Group #1 consisted of the 16 students who came from Mr. Gonzales' fifth grade classroom. Group #2 consisted of the 1,266 students that were not from his classroom.

I used the (multiple regression) equations described above to obtain a predicted ELA, Math, and Total 2011 score for each of the 1,282 students. Next, I subtracted each student's predicted 2011 score from that student's actual 2011 score. The difference between the actual and predicted scores is called the student's "residual" score. A positive residual score indicates the student did better than expected on the sixth grade tests while a negative residual score indicates the opposite. Mathematically, the mean residual score across all of the 32,574 students used to construct the equations will be very close to zero.

Finally, I computed the mean residual ELA, Math, and Total score in each group. If Mr. Gonzales' fifth graders were assigned to relatively less effective sixth grade teachers, then the students in Group #2 would have a large negative mean residual score, especially in comparison to Group #1s mean residual score.

The mean ELA, math, and total residual scores of the 16 students who came from Mr. Gonzales classroom were: -93, -129, and -108, respectively. All of these values are quite large and statistically significantly different than zero.

The corresponding mean residual scores of the 1,266 students who did not have Mr. Gonzales for their fifth grade teacher (but who had a teacher that did have at least one of his students) were: -14, -2, and -3. All of these values were close to zero, none were statistically significantly different than zero, and all were significantly lower than those for Mr. Gonzales' former students.

Given these results, there is no empirical evidence to support the thesis that the extremely lower than expected scores of Mr. Gonzales' former students on the 2011 statewide tests was due to their having inept sixth grade teachers.

Table 1
Percentile Equivalents of the Mean Scores in Mr. Gonzales' Actual 5th Grade Classroom and his Virtual 4th and 6th Grade Classrooms

Test Score	Percentile Equivalents of the Mean Test Scores in Mr. Gonzales'		
	Virtual 4th Grade Classroom	**Actual** 5th Grade Classroom	Virtual 6th Grade Classroom
ELA	54	98	31
Math	63	99	24
Total	60	99	29

Table 2
Statistics used to calculate the probability that the observed difference between the actual and predicted scores could have arisen by chance.

Test	Actual Score	Predicted Score	Residual Score	Standard Error	z-score	Probability p-value
ELA	325	401	- 76	10.82	7.02	<.0000001
Math	314	425	-111	23.39	4.75	<.0000001
Total	320	417	- 97	16.18	6.00	<.0000001

Residual score = Actual score – Predicted score

z-score = Residual score/Standard error

p-value = probability that the observed difference between the actual and predicted score was likely to arise by chance. For example, a p-value of <.000001 indicates there was less than one chance in 100,000.