ABSTRACT

LIMITED ENGLISH PROFICIENCY, RACE/ETHNICITY AND SOCIO-ECONOMIC STATUS AS INFLUENCES ON SCORES IN LARGE-SCALE ASSESSMENTS

The purpose of this study was to examine the effects of three basic demographic variables on reading test scores for students in the middle elementary grades. Limited English proficiency (LEP), race/ethnicity and socio-economic status (SES) were studied to determine their influence individually and in combination on performance in large-scale assessment programs. Two data sets were analyzed: 1. Grade 4 reading results for the 2003 National Assessment of Educational Progress (NAEP) and 2. Grades 3 and 5 reading results for the 2003 Minnesota Comprehensive Assessment (MCA). The NAEP data contains results for over 179,000 students and the MCA data files consist of results for approximately 60,000 students at each grade level. The NAEP results were analyzed both for the overall national public school sample and for the 50 state-level jurisdictions. The overall NAEP results indicate that all three variables have a significant effect upon reading scores. There are large "gaps" in average test scores among racial/ethnic groups, LEP and Non-LEP students and students eligible and not eligible for the federal free and reduced lunch program. The joint effect of the variables is shown by the fact that the racial/ethnic gaps are dramatically reduced when students are cross-classified by SES and LEP status. The state-level NAEP data reveal a pattern of strong negative correlations between the percent of students in each state eligible for free and reduced lunches and mean state NAEP scores. A similar pattern was observed in the correlation of percent of students classified as LEP for Hispanic and Asian student mean scores. The MCA data showed results parallel to the NAEP data. The gaps between White and Hispanic and White and Asian students were shown to be largely attributable to the incidence of LEP students in the two minority groups. Our analyses clearly demonstrate that reports of "achievement gaps" in assessment results are misleading if SES and LEP status are not taken into consideration.

April 11, 2005

LIMITED ENGLISH PROFICIENCY, RACE/EHNICITY AND SOCIO-ECONOMIC STATUS AS INFLUENCES ON SCORES IN LARGE-SCALE ASSESSMENTS

James S. Terwilliger
and
Paul Magnuson

Minnesota Department of Education

It is widely recognized that race/ethnicity (R/E) and socio-economic status (SES) are variables that have a significant impact upon test scores. Much has been written about the achievement gap between white and minority students. Somewhat less attention has been given to gaps associated with limited English proficiency (LEP) and SES. This paper will demonstrate that LEP has a major impact on test scores and interacts with R/E and SES. This interaction has important implications for the interpretation of test results and the evaluation of individual schools under the regulation of the 2001 No Child Left Behind Act.

INTRODUCTION

No Child Left Behind (NCLB) has raised public awareness of the achievement gap by legislating that schools, districts, and states issue annual summaries of test data broken down by various subcategories. Everyone now has easy access to information about test performance of specific groups of children, notably the results of ethnic groups, students with lower socioeconomic status (SES), Special Education students, and limited English proficient (LEP) students.

Making accurate comparisons between ethnic groups, however, is not always as easy as simply reporting test scores of each ethnic group, since the demographics of ethnic groups differ. Comparing aggregate scores of ethnic groups may indeed lead to the data driven decision making called for by school reformers and federal legislations. Unfortunately, those decisions might be less than satisfactory if the demographic composition of the ethnic groups is not considered.

In this paper, the composition of ethnic groups reported by the National Assessment of Educational Progress (NAEP) and Minnesota' state assessments, the Minnesota Comprehensive Assessments (MCA) is more finely considered, taking into account the percentage of LEP and lower SES children in each subgroup. The data suggest that reporting assessment data by ethnic group without considering LEP and SES is misleading at best, or at worst, serving to reinforce stereotypes rather than dispel them.

In this paper we examine data from two sources:

1. the 2003 National Assessment of Educational Progress (NAEP) grade 4 reading results (See Appendix A for background on NAEP); and
2. the 2003 Minnesota Comprehensive Assessment (MCA) grade 3 and 5 reading results (See Appendix B for background on MCA).

Our original plan was to examine results in both reading and mathematics. However, preliminary exploration of the results revealed that the findings in the two subjects are quite similar. Consequently, presentation of both sets of results is redundant. In addition, the results in each subject are so extensive that treating both would have resulted in an unwieldy paper.

The general structure of this paper is as follows:

Section 2: NAEP results for all national public, N(P), schools.

Section 3: State-level NAEP results for 50 jurisdictions.

Section 4: MCA results for Minnesota grade 3 and 5 students.

Section 5: Overview and conclusions.

NAEP RESULTS FOR ALL PUBLIC SCHOOLS, N(P)

This section presents results for all public school grade 4 students who participated in the 2003 NAEP reading assessment. Over 179,000 students are represented in this data set. Descriptive statistics for the entire sample and for subgroups defined by race/ethnicity, National School Lunch Program (NSLP) status (our proxy for socio-economic status) and English Language Learner status are presented in Table 1.

Insert Table 1 here

It is clear that there are large differences among the subgroups defined by each of the three variables. White students score 30 points above Blacks and 28 points above Hispanics. Students not eligible (NELG) for the NSLP score 28 points above those who are eligible (ELG). Students classified as Limited English Proficiency (LEP) score 33 points below those who are not so classified (NLEP). All of these differences are highly statistically significant ($p < .0001$).

Another way to view these differences is with reference to the variability of the grade 4 reading scores. The standard deviation for the National (P) sample is 37 points. Therefore, the differences cited above range from .75 to .90 standard deviations. These are extremely large by all conventional standards.

Pairwise comparisons among the various subgroups reveal more detail. These are shown in Table 2. The only difference among racial/ethnic groups that is NOT significant is the difference between White and Asian students. The difference between LEP and NLEP students is the largest difference cited and is highly significant statistically. (It is interesting to note that NAEP provides a reporting option for researchers interested in "gaps" among subgroups of interest. Gaps reported by NAEP include racial/ethnic differences and NSLP differences but do NOT include ELL differences.)

Insert Table 2 here

The tests in Table 2 show the influence of each of the three variables when considered in isolation. However, a more important issue is the joint effect of the variables upon test scores. This can be studied by looking at variables in combination.

The three variables are shown in the three possible two-way pairings in Table 3. For each pairing, the mean and percent of students for the cross-tabulations are presented. This permits the examination of the pairwise interaction of the variables. The interaction of race/ethnicity and English language status is presented in the first section of the table. The breakdown of each racial/ethnic subgroup by LEP vs. NLEP is shown. The first obvious outcome is the dramatic differences among the groups with respect to the percent of students classified as LEP. For Whites and Blacks almost no students (one percent) are LEP but for Hispanics and Asians significant percentages are classified LEP (over a third for Hispanics and over a fifth for Asians). These differences give some insight to the origin of the racial/ethnic gaps in Table 1.

Insert Table 3 here

Since the LEP vs. NLEP break is identical among Whites and Blacks, the 30-point gap between the two groups cannot be attributed to English language proficiency. This is confirmed by noting that the White/Black gap for LEP students is 27 points and for NLEP students it is 29 points. These are only slightly different than the overall gap. The situation for Whites vs. Hispanics is somewhat different. The overall gap in Table 1 is 28 points but the differences within the LEP and NLEP groups are 20 and 18 points, respectively. This represents a much more substantial "closing of the gap" than for Whites vs. Blacks. Although the overall gap between White and Asian students is only two points (See Table 1), the breakdown by LEP and NLEP reveals an interesting interaction. The White/Asian gap within the LEP students favors Whites by three points while the gap within NLEP students favors Asians by four points.

The middle section of Table 3 presents the interaction of race/ethnicity and NSLP status. In this case, the major contrast in percent ELG is between White and Asian students as opposed to Blacks and Hispanics. Although substantial numbers in all groups are ELG, the percents for Whites (25) and Asians (39) are dramatically lower than for Blacks (73) and Hispanics (74). Therefore, we should anticipate that NSLP status would have an impact upon both White/Black and White/Hispanic differences. This is borne out by looking at mean differences within ELG and NELG groups. The White /Black difference for ELG students is 19 points and for NELG students it is 21 points. This is a substantial drop from the overall difference of 30 points in Table 1. The White/Hispanic difference for WLG students is 17 points and for NELG students it is 19 points. This is roughly the

same drop as was observed in the case of English proficiency groups. It is interesting to note that the White/Asian differences for ELG and NELG groups follow the same pattern as previously seen for LEP and NLEP groups; the first difference favors Whites and the second favors Asians.

The final section of Table 3 shows the interaction of ELL status and NSLP status. It is clear from the percentages that the incidence of LEP is much higher within the ELG group than for the NELG group, a five to one ratio! The mean gap between ELG and NELG for students who are LEP is 16 points and for students who are NLEP it is 26 points. The first of these is a dramatic change from the overall difference of 28 points although the second is only trivially smaller than the overall difference.

STATE-LEVEL NAEP RESULTS

The previous section used individual data for over 179,000 students who took part in the 2003 grade 4 NAEP reading assessment. NAEP also reports data aggregated at the individual state level. Such data permit an examination of performance patterns that vary from state to state. It also allows a researcher to treat states as individual cases for purposes of correlating outcomes that vary across states, e.g. mean scores and percentage of students who are classified LEP or ELG by race/ethnicity.

Analyses in this section are based upon data from 50 jurisdictions. Hawaii and Alaska are not included in the analysis because they are very unique in their demographic profiles, especially with regard to race/ethnicity. Conversely, two Department of Defense school systems, one for domestic schools (DODD) and one for overseas schools (DESS), are included because they typify the student mix in most states.

The basic data used in this section are quite extensive because a range of outcomes is presented for each of 50 jurisdictions[1]. The data include the values on the different variables for each of the participating states. Electronic versions of the data tables are posted on the website given at the end of this paper.

A total of 21 variables was selected for the state-level analysis. These are described and listed by codes in Chart 1. A brief rationale for the inclusion of each variable is given below.

- The mean for all students in each state (ALL MEAN) is included as a general index of performance. We will correlate this variable with various other variables to show their impact.

- The means for students by racial/ethnic group (W MEAN, B MEAN, H MEAN and A MEAN) are presented to show the differences among groups and to correlate group means with other measures to reveal differential patterns.

- The means by ELL groups (LEP MEAN and NLEP MEAN) are used to demonstrate the influence of English language proficiency.

- LEP percentages corresponding to overall means (P ALL LEP) and racial/ethnic means (P W LEP, P B LEP, P H LEP and P A LEP) are included to examine the relationship of language proficiency to general performance and to performance within each racial/ethnic group.

- Means for NSLP groups (ELG MEAN and NELG MEAN) are employed to demonstrate the influence of socio-economic status on performance.

- The influence of socio-economic status overall (P ALL ELG) and within each racial/ethnic group (P W ELG, P B ELG, P H ELG and P A ELG) is studied through relationships with corresponding means.

- Finally, the interaction of socio-economic status and language proficiency is examined through the percent of students in the two NSLP groups who are classified as LEP (P ELG LEP and P NELG LEP).

[1] Although the 50 jurisdictions are not all states, for convenience we will simply refer to them as states.

Insert Chart 1 here

A summary of descriptive statistics across states is shown in Table 4. This shows the minimum and maximum values, the mean score and the standard deviation for each of the 21 variables. Although there are 50 states in the data set, the sample size for certain variables is less than 50 because the number of students within a subgroup failed to meet the minimum (60) required by NAEP to report data for a state. In particular, note that there are 41 states with Black data, 40 with Hispanic data and only 24 with Asian data.

Insert Table 4 here

It is interesting to compare the results in Table 4 with those in Table 1. First, the means of the state means are slightly different than the corresponding means for the national public school sample. This reflects the fact that each state is weighted equally in Table 4. The states are weighted slightly differently in Table 1 due to differences in sample sizes from state-to-state[2]. The largest discrepancy is the Hispanic mean that is almost five points higher in Table 4 than Table 1. The variation in state means for racial/ethnic groups is smallest for White students and largest for Asian students. This is due in large part to the fact that the sample sizes on which the means are based is much smaller for the three minority groups, especially the Asian groups. As will be noted later, there are also some true "outliers" within the Asian sample. Parallel descriptive summaries for ELL and NSLP subgroups are also in Table 4. As with the racial/ethnic data, the means differ slightly from the corresponding means in Table 1.

[2] Sample sizes range from almost 8,300 in CA to about 1,300 in DESS.

An examination of the minimum and maximum values for means underscores the gaps between groups. Notice the following:

- The minimum W MEAN is higher than the maximum B MEAN.
- The minimum NLEP MEAN is higher than the maximum LEP MEAN.
- The minimum NELG MEAN is higher than the maximum ELG MEAN.

This clearly shows the impact that race/ethnicity, language proficiency and socio-economic status have on grade 4 NAEP reading performance. It is also interesting to note that the maximum value is the same for W MEAN and A MEAN. However, the minimum value for A MEAN is **20 points** lower than the minimum for W MEAN. This results from low scoring "outliers" in the Asian samples. Histograms showing the distributions for W MEAN and A MEAN are presented in Figures 1-2.

Insert Figures 1 and 2 here

A final observation regarding certain of the "percent" variables should be noted. Several variables have standard deviations that are larger than the means, e.g. P W LEP, P B LEP, P ELG LEP and P NELG LEP. In each of these cases the variable has a highly positively skewed distribution with a few extreme high values. As we will see, this positive skew places restrictions on the relationship these variables can have with other variables. Figures 3 and 4 show the histograms for P W LEP and P B LEP, respectively.

Insert Figures 3 and 4 here

The interaction of race/ethnicity and ELL is evident in the percent variables for LEP by racial/ethnic group. The variables are defined to highlight the differential effect LEP status

has on different racial/ethnic subgroups. It is clear that LEP is a major factor for Hispanic and Asian groups but is insignificant for Whites and Blacks. The mean percent of LEP for Hispanics is 30 and for Asians it is over 20. In contrast, for Whites it is less than one and for Blacks it is less than two. When the racial/ethnic groups are aggregated, the overall percent LEP is slightly over five. This value obviously masks important differences among the groups.

The interaction of race/ethnicity and NSLP is reflected in the percent variables for ELG by racial/ethnic subgroups. The differences among the groups present a very different pattern than for LEP. For LEP, the major differences are between White/Black vs. Hispanic//Asian. For NSLP, the major fault line is between White/Asian vs. Black/Hispanic. Notice in Table 4 that P ELG for White and Asian students is 28 and 36, respectively. In contrast, the figure for both Blacks and Hispanics is 70, i.e. there is about a two-fold increase in poverty for Blacks and Hispanics compared with Whites and Asians.

A second way to check the influence of variables is through correlations that link variables to each other. Therefore, we next look at the correlations within each of the sets of variables listed in Table 4. The correlations involving the race/ethnicity variables are presented in Table 5. In this case, the variables of interest are the group means and percent LEP by group.

Insert Table 5 here

There are five means and five corresponding percents. This results in a 5 x 5 correlation matrix. This matrix is shown in Table 5. Note that the five means are listed as column headings and the five percents are listed as row headings in Table 5. It is important to note

that the sample size varies from row to row in the matrix. (N is given along with the p-value for the correlation in each cell.) As previously indicated, the concentration of various racial/ethnic groups varies from state-to-state so the number of students in a state frequently does not reach the minimum required for reporting state-level results, especially for Asian students.

The major diagonal shows the correlations of the five means and the corresponding P LEP variables. The off-diagonals are not of general interest for our present purposes except to note that the P H LEP variable has a consistent strong negative relationship with all the racial/ethnic means. (The reason for this is not immediately clear to the authors.) For the present, we will direct our attention to the correlations on the diagonal.

The correlation of the ALL MEAN with the P ALL LEP is -.49. This indicates a clear tendency for higher overall percents to be accompanied by lower means. However, the overall means and percents represent an aggregation of the means and percents for the four racial/ethnic subgroups. Table 5 reveals that the correlation of subgroup means and percent LEP is close to zero for Whites and for Blacks. But the correlations for Hispanics (-.52) and Asians (-.64) is decidedly negative. The aggregation of the four racial/ethnic groups results in a correlation which acts like a weighted average of the separate group correlations.

Figure 5 is a plot of ALL MEAN against P ALL LEP. As noted previously, the correlation is -.49. It is obvious from the plot that the negative correlation is due primarily to three states with extremely high LEP values: AZ, CA, and NM. If these three states were removed from the plot the correlation would be near zero. These states all have large Hispanic populations with high LEP rates, ranging from 33% in NM to 54% in CA.

Insert Figure 5 here

The negative correlations for Hispanics and Asians are shown in Figures 6 and 7. Figure 6 is a plot of the Hispanic data. In this case there is no obvious outlier. It is interesting to see that states like UT, OR and MN have some of the highest P H LEP rates. Since these states are quite distant from the Mexican-U.S. border, this finding is somewhat surprising. (The high P H LEP for CA was expected.)

Insert Figure 6 here

The plot of Asian data in Figure 7 reveals the influence of three extreme states with high LEP rates and low means: MN, UT and WI. Again, if these three were removed from the data set the correlation would be near zero. These particular states may come as a surprise to many readers. One possible reason is the specific Asian ethnic groups in these states. Both MN and WI have large numbers of Hmong who have recently emigrated from Southeast Asia. Utah, however, has only low percentages of Asian ethnic groups. It would be interesting to compare Asian groups in states like DE, NJ and MD with those in MN, WI and UT both with respect to ethnicity and to recent immigration patterns.

Insert Figure 7 here

Next, we consider the influence of socio-economic status as reflected in NSLP status. Table 6 shows the correlations between ALL MEAN and means for racial/ethnic groups and corresponding percent of students ELG for NSLP. This table is parallel to Table 5. The five means are on the columns and the five percents for ELG status are on the rows of

the 5 x 5 matrix. Again, the correlations of special interest are on the diagonal of the matrix.

Insert Table 6 here

As expected, the correlations are all negative. However, the pattern is quite different than in Table 5. Now, we see an extremely high negative correlation (-.80) linking overall means to overall percent ELG. The magnitude of this correlation is truly remarkable. This is the most powerful relationship between reading performance and any demographic variable we have studied. This can be interpreted in terms of the variance in ALL MEAN associated with P ALL LEP, i.e. **64% of the variability in mean scores across states is "accounted for" by the variation in poverty rates across states.** The plot in Figure 8 is a visual portrayal of this strong negative correlation.

There is a clear pattern in this figure with the New England states (NH, CT, VT, and MA) being the states with the lowest percent ELG and the highest mean reading scores. At the other extreme, states in the South and Southwest (NM, MS, LA, AL, AZ and NV) are those with the highest percent ELG and the lowest reading scores. (CA is the only state among this latter group that is not entirely in the South or Southwest.)

Insert Figure 8 here

Unlike the results for racial/ethnic groups in Table 5, the correlations between group mean scores and percent ELG within each group are very consistent across groups ranging from a value of -.66 for whites to a value of -.52 for blacks and Hispanics. This represents a strong contrast to the ELL results in which W MEAN and B MEAN are virtually unrelated

to the P W LEP and P B LEP. In those instances, the lack of relationship can attributed to the low incidence of LEP within those two groups. (See Figures 3 and 4.)

In the case of NSLP, there is greater variability in P W ELG. The standard deviation for P W LEP is similar to that for P B ELG and P H ELG although it is smaller than P A ELG. However, as might be expected, the mean for P W ELG is much lower than for other groups. (See Table 4.) The plot of W MEAN vs. P W ELG is shown in Figure 9.

Insert Figure 9 here

Inspection of Figure 9 suggests that the relationship is curvilinear rather than linear. Therefore, the true relationship is likely stronger than suggested by the correlation of -.66. Note that the specific states at the lower extremes in Figure 10 are different from those in Figure 9. In particular, the states with the highest P W ELG and lowest W MEAN are WV, KY, OK, NM and AR. (NM is the only one in common with Figure 8.) On the other hand, the two states with the lowest P W ELG and the highest W MEAN are CT and NJ. (CT shares this distinction in both plots.)

An obvious question arises from the fact that the ALL MEAN/P ALL ELG correlation is so much higher than the corresponding correlations within the separate racial/ethnic groups. The explanation lies in the variation of racial mixes across states. Three New England states have over 90% White students: ME (95%), NH (94%) and VT (95%). The ALL MEAN is the **same** as the W MEAN in ME and VT and is only one point lower in NH. Obviously, the P W ELG and P ALL ELG is also virtually identical in these states. Notice that NH is the most extreme state in the upper left corner of Figure 8. VT also is one of the top states. However, in Figure 9, NH and VT have “dropped down” closer to the middle of

the plot. (The same is true of ME although that state is not labeled in the plots due to lack of space.)

At the other extreme, several states have less than 50% White students: NM (32%), MS (45%), LA (44%) and CA (34%). In each of these states the W MEAN and ALL MEAN differ by 15–20 points. At the same time, the percent of students ELG is much greater for minorities than for Whites. Consequently, when one compares Figure 8 with Figure 9, each of these states appears to have "moved" up and to the left in the plot. Consequently, they no longer appear as extremes in Figure 9. This changes the shape of the scatter plot and results in a lower correlation.

The remarkable consistency in the correlation of racial/ethnic means and the percent of students eligible for NSLP suggests that the effect of poverty is quite uniform across racial/ethnic groups. The fact that the correlation for Whites is slightly higher hints at the possibility that poverty may have a larger impact among Whites than among minorities.

Finally, we examine the interaction of NSLP with ELL. First, the incidence of LEP is clearly higher among students who are eligible for NSLP. Note in Table 4 that the mean for P ELG LEP is four times as high as P NELG LEP: 9.47 vs. 1.94. Not surprisingly, the maximum value and standard deviation for P ELG LEP is also four times as large as the corresponding values for P NELG LEP. As noted earlier, the relationship of the mean to the standard deviation on these two variables implies a positive skew. This is confirmed in Figures 10 and 11.

Insert Figures 10 and 11 here

As a final step, we will consider the correlations between NSLP means and the

percent of students classified LEP and NLEP within the two major NSLP groups. The correlation between ELG MEAN and P ELG LEP is -.41 and the correlation between NELG MEAN and P NELG LEP is -.46. These negative relationships are consistent with previous results using race/ethnicity groups. However, in view of the high degree of skew in the two percent variables, it is of interest to know more about these correlations.

Figure 12 shows the scatter plot for ELG students. It is obvious that the small number of extreme cases account for the negative relationship. In particular, CA and three Southwestern states (NM, AZ, and NV) have extremely high P ELG LEP values in combination with low ELG MEAN values. These four "outliers" make the overall relationship negative.

Insert Figures 12 and 13 here

Much the same is true in Figure 13 with the main difference being that the range of P NELG LEP values is much smaller than for P ELG LEP. However, the same states as in Figure 12 tend to account for the negative trend in the relationship.

MINNESOTA 2003 STATE ASSESSMENT RESULTS IN READING

The Minnesota Comprehensive Assessments (MCA) is a test mandated by the state legislature in 1997. The MCAs are given in reading, writing and mathematics to help schools and districts measure progress toward the state's academic standards. Scores on the MCAs in reading and math are used to determine whether a school has made Adequate Yearly Progress (AYP) as required under the No Child Left Behind (NCLB)

legislation of 2001. This paper will address only the 2003 MCA reading results in grades 3 and 5. A more detailed description of the MCA is provided in Appendix B.

Grades 3 and 5 were selected because they are on either side of the grade 4 group that took NAEP. Unlike the NAEP sample, the students at grades 3 and 5 represent all public school students in Minnesota at those two grade levels. A summary of MCA mean scores, the percents of students in various racial/ethnic, NSLP and ELL subgroups and the "gaps" in the scores among groups is presented in Table 7. The overall means and standard deviations are given to provide a context for judging the various subgroup means and the gaps between subgroups.

Insert Table 7 here

The pattern of results for race/ethnicity groups is similar to that for NAEP. The largest gaps occur between Whites and Blacks/Hispanics with a somewhat smaller gap between Whites and Asians. Generally speaking, the performance of Blacks and Hispanics is very similar. It is striking that the gap between Whites and both Blacks and Hispanics is about one standard deviation at both grade levels.

The NSLP differences are not as large as the differences between Whites and Blacks/Hispanics but are larger than the difference between Whites and Asians. The ELL status groups show the largest differences in Table 7. At both grade levels, the gap between NLEP and LEP students is substantially greater than for any of the gaps between racial/ethnic or NSLP groups. The ELL gap exceeds one standard deviation at both grade levels.

Next, we examine the **joint effects** of race/ethnicity, NSLP and ELL. This is done by considering the pattern of means for two variables in combination. The results are shown in Table 9. Plots corresponding to various sections of this table will be presented to provide a visual image of the findings.

Insert Table 8 here

First, look at the race/ethnicity x ELL interaction shown at the top of Table 8. As was true with the NAEP data, we see that there is a sharp division between White/Blacks vs. Hispanics/Asians in the incidence of LEP. Whites at both grade levels are 99% NLEP. Blacks are 88% (grade 3) and 90% (grade 5) NLEP. In contrast, both Hispanics and Asians are over 50% LEP at both grade levels.

The impact of this huge variation in the incidence of LEP across racial/ethnic groups is obvious when we examine the means for LEP and NLEP groups. At both grade levels the NLEP Asian students have **higher** means than the NLEP White students. This is in stark contrast to the racial/ethnic differences in Table 8 where we saw that Whites outperformed Asians by well over 100 points. The differences between Whites and Hispanics are also substantially reduced if we consider only NLEP students. The original gap of 186 points shrinks to 94 points at grade 3 and the gap of 201 points reduces to 99 points at grade 5. However, the White/Black differences are virtually unchanged for NLEP students because the incidence of LEP is almost the same in the two groups. Figures 14 (grade 3) and 15 (grade 5) graphically show the results for the race/ethnicity x ELL status interaction.

Insert Figures 14 and 15 here

Race/ethnicity x NSLP interactions are presented in the middle of Table 8. The pattern in Minnesota is different from the NAEP data in one major respect: the percent of Asian students eligible for NSLP is over 60% in Minnesota compared to less than 40% for NAEP. In both data sets the percent of Blacks and Hispanics eligible is 70-75%. However, Whites in Minnesota have a slightly lower percent eligible (21%) than for NAEP (28%).

Again, we find that the gaps among racial/ethnic groups shrink when we consider only students not eligible for NSLP. The difference between Whites and all other groups is much smaller than those shown in Table 7. The change is particularly dramatic in the case of Whites vs. Asians. The original difference drops from 123 to 42 at grade 3 and from 132 to 21 at grade 5. Figures 16 (grade 3) and 17 (grade 5) show graphic representations of the data on race/ethnicity x NSLP interactions.

Insert Figures 16 and 17 here

Finally, consider the interaction of ELL and NSLP. In Table 7, the NLEP/LEP gap at grade 3 is 209. For students not eligible for NSLP, the gap between NLEP/LEP is 109 and for those eligible the gap is 130 (Table 9). Back on Table 7 again, at grade 5, the overall gap is 243. For students not eligible for NSLP, the gap between NLEP/LEP is 212 and for those eligible it is 151. The changes in this instance are not as dramatic as

previously noted when ELL status and NSLP status were shown in combination with race/ethnicity.

OVERVIEW AND CONCLUSIONS

This paper has demonstrated that presenting test scores as a group average on national or state assessment data, without consideration of variables within the group, may lead to false conclusions about the achievement of a particular group. It is wise to be particularly mindful of how the data might mislead casual consumers – or even not-so-casual consumers – when the data paint ethnic groups in a particular light.

The NAEP data show a significant LEP/NLEP mean difference for reading. The difference varies from one R/E subgroup to the next. The differences are much larger for Hispanic and Asian students than for White and Black students. In other words, the achievement gap between White and Asian and White and Hispanic is overstated when not controlled for LEP. It was also noted that there is a strong negative correlation between average NAEP scores and incidence of LEP students for both Hispanic and Asian subgroups. For these two ethnic groups, states with a large percentage of LEP will, on average, score lower than states with a smaller percentage of LEP. It becomes clear why states with significant numbers of Hispanic and Asian students tend to have mean NAEP scores that rank them low compared with states that have small numbers of students in these two groups. This is especially true for states with large Hispanic populations.

The NAEP data also show that poverty is a strong predictor of test scores, and that poverty affects ethnic groups similarly, with a strong positive correlation between poverty and low scores. Ethnic groups, which, on average, have a greater percent of students in poverty, will, on average, tend to score lower than ethnic groups with a lower percent of students in

poverty. This is a social (in this case economic) difference that is measured by national and state assessments of academic achievement. Academic success is indeed possible with lower SES students – and school reform movements, including the school improvement planning required by NCLB, may indeed help individual students. But in the aggregate, lower SES children are likely to continue under-performing relative to their middle class peers until the social issues contributing to their low performance are addressed, whatever the race of the lower SES children.

At the state level, the NAEP data reveal a number of issues. First, Minnesota, which compares quite favorably with other states when all students are considered, fares poorly when comparisons of the achievement gap in Minnesota versus other states are considered. The general solution is to steer clear of unsophisticated comparison; that message is a goal of this paper. The specific solution is to reconsider how data are reported. For example, reporting NAEP scores by ethnic group, *after* controlling for LEP and NSLP, is a fairer comparison.

Second, within Minnesota it would be quite understandable if the public perception of achievement by ethnic group, based on the state's accountability test, were that Whites fare best and the achievement gap compared to White scores is universally large. Consider this quote regarding Minneapolis schools:

> Despite the improvements for students of color, the results for white students are still 45 percentage points better than Latino students, 41 points better than black students, 40 points better than Asian students and 30 points better than American Indian students (Brandt, October 11, 2004, Minneapolis StarTribune).

The article goes on to ask “Why does the gap persist? Take your pick.” None of the possible answers given include the lack of control for LEP. But as we’ve shown, when one controls for limited English proficiency, non-LEP Asian students score statistically even with Non-LEP White students in Minnesota. Perhaps if we in Minnesota, and elsewhere, routinely controlled for LEP, we would be talking about the achievement gap between Asian students and the other races and ethnicities, instead of White students and the other races and ethnicities. (In fact, throughout the tables and charts in this paper we consistently presented the White score first as the starting point from which to measure the gap. It would be interesting to consider how many readers stopped to consider why the average White score is the default high end of the achievement gap for all races and ethnicities instead of the average Asian score!)

Third, school officials in Minnesota often complain that their NCLB status, which essentially alerts the public to whether or not schools and districts are meeting the state’s targets under NCLB, is too dependent on the demographics of their students and therefore not comparable one school to the next or one district to the next. If a state’s measure of adequate yearly progress under NCLB allowed for the control of LEP and SES, perhaps NCLB would be the stronger for it.

Fourth, the manner in which data are presented does affect how data are used. Data presented in too coarse a fashion do impact decisions made by school and district site teams. For example, schools and districts not making adequate yearly progress for two or more consecutive years are deemed in need of improvement and must write and annually update an improvement plan. The plan must focus on the areas identified by state tests – the

Asian or Hispanic students' math test scores, perhaps. The problem might not be related to race or ethnicity, however. The issue might be that fifty percent of the Asian and Hispanic students don't speak English very well or are suffering from the effects of poverty. The school or district may erroneously focus on ethnic groups when it should instead focus on English acquisition strategies and matters commonly associated with poverty (health, study habits, lack of role models, and so on).

Regardless of the data source, whether national or state, school officials and community members who are consistently presented with rank-ordered data of achievement by ethnic group will most likely form some opinions about those ethnic groups. Those opinions may have the effect of reinforcing ethnic difference s in achievement, since teachers, parents, school board members, principals, even students themselves, believe that the numbers do not lie. The mere fact that there is little if any public opposition to reporting student data by ethnic group without controlling for LEP and SES may suggest that the achievement gap is culturally accepted. School reform efforts need to counter ingrained public opinions with good data. Misleading reporting of data may further establish public opinion, making school reform efforts *more* difficult than they already are.

REFERENCE

Brandt, S. (October 11, 2004). Racial testing chasm persists. Minneapolis StarTribune. Accessed online October 11, 2004 at www.startibune.com/stories/1557/5025891.html.

NOTES

Access this paper and further data associated with this paper at http://education.state.mn.us/html/intro_english_resource.htm.

APPENDIX A: Description of NAEP Reading Assessment

The NAEP Reading Scale

For every subject assessed, NAEP reports how well students of various demographic subgroups performed. (Note that NAEP does not report individual student scores.) For example, results are reported for male students and female students, for students of various racial or ethnic categories, and for students in schools in different regions.

How does NAEP summarize what students in these groups know and can do, in order to be able to compare how the groups performed?

In reading, NAEP creates a scale ranging from 0–500, based on statistical procedures called item response theory (IRT). IRT is a set of statistical procedures useful in summarizing student performance across a collection of test exercises requiring similar knowledge and skills. All NAEP subject–area scales are produced using these procedures.

The reading data are scaled separately by the two contexts for reading (reading for information and reading for literary experience) for grade 4, and the three contexts for reading (reading for information; reading for literary experience; and reading to perform a task) for grade 8, resulting in two or three separate subscales at each grade. The composite scale is a weighted combination of these subscales. IRT information functions are only strictly comparable when the item parameters are estimated together. Because the composite scale is based on three separate estimation runs, there is no direct way to compare the information provided by the questions on the composite scale.

Three different contexts for reading were assessed:	**Students were assessed on four different aspects of reading:**
• **Reading for literary experience:** Readers explore events, characters, themes, settings, plots, actions, and the language of literary works by reading novels, short stories, poems, plays, legends, biographies, myths, and folktales. • **Reading for information:** Readers gain information to understand the world by reading materials such as magazines, newspapers, textbooks, essays, and speeches. • **Reading to perform a task:**[4] Readers apply what they learn from reading materials such as bus or train schedules, directions for repairs or games, classroom procedures, tax forms (grade 12), maps, and so on.	• **Forming a general understanding:**[1] The reader must consider the text as a whole and provide a global understanding of it. • **Developing interpretation:** The reader must extend initial impressions to develop a more complete understanding of what was read. • **Making reader/text connections:**[2] The reader must connect information in the text with knowledge and experience. • **Examining content and structure:**[3] Requires critically evaluating, comparing and contrasting, and understanding the effect of such features as irony, humor, and organization.

APPENDIX B: Minnesota Comprehensive Assessments (MCA)

The MCAs are reading, writing and mathematics tests that help schools and districts measure student progress toward the state's academic standards.

The 1997 legislature mandated a system of statewide testing and accountability for students enrolled in grades 3 and 5. Since 1998, all Minnesota grade 3 and grade 5 students have been tested annually with a single statewide test for the purpose of statewide system accountability. Since 1998, MCAs in grades 7, 10 and 11 have been introduced, with additional tests in grades 4, 6 and 8 planned for the 2005-2006 school year.

Scores on the MCAs in reading and mathematics are used to determine whether schools are making Adequate Yearly Progress (AYP). Pre-determined performance levels are set to decide whether students have attained specified levels of achievement. In 2004 there were five distinct levels of achievement: 1, 2, 3, 4 and 5. Students scoring at Level 2 are deemed "partially proficient" and students scoring at Level 3 are deemed "proficient" which is akin to meeting grade level expectations.

This paper uses mean scores rather than achievement levels as an index of performance. Means use more information than do achievement levels. Also, means are less arbitrary than are achievement levels. The means are for scaled scores on the MCAs. The scaled scores are based upon a transformation of the raw scores using the Rasch Partial Credit Model (RPCM).

The details of the model are beyond the scope of the present paper. However, certain basic features IRT models should be noted:

- A statistical equating procedure is employed to assure comparability of scores from year-to year and across administrations within a year.
- The scaling method results in a wide range of scores from as low as 200 to as high as 2400. Users of the scores are cautioned against over-interpreting differences in scaled scores.
- The MCA scales are established independently by grade and subject. Therefore, it is inappropriate to compare scores across subjects within a grade. The scales are independent .
- The scaled scores on the MCA are *not* vertically aligned so the interpretation of individual score differences across grades is not appropriate.

TABLES AND FIGURES TO ACCOMPANY

LIMITED ENGLISH PROFICIENCY, RACE/ETHNICITY AND SOCIO-ECONOMIC STATUS AS INFLUENCES IN LARGE-SCALE ASSESSMENTS

JAMES TERWILLIGER

AND

PAUL MAGNUSON

MINNESOTA DEPARTMENT OF EDUCATION

APRIL, 2005

TABLE 1

SUMMARY OF NATIONAL (P) 2003 GRADE 4 READING DATA

UNIVARIATE MEANS, PERCENTS AND GAPS

OVERALL MEAN	**216**
OVERALL STD. DEV.	**37.2**

R/E GROUP	MEAN	%	GAP
W	227	59	
B	197	17	30
H	199	18	28
A	225	4	2

NSLP STATUS	MEAN	%	GAP
NELG	229	52	
INA	219	4	10
ELG	201	44	28

ELL STATUS	MEAN	%	GAP
NLEP	219	92	
LEP	186	8	33

CODES:

W =WHITE, B=BLACK, H=HISPANIC, and A=ASIAN

NELG= NOT ELIGIBLE FOR FREE/REDUCED LUNCH
ELG= ELIGIBLE FOR FREE /REDUCED LUNCH
INA= INFORMATION ON ELIGIBILITY NOT AVAILABLE

NLEP= NOT CLASSIFIED AS LIMITED ENGLISH PROFICIENCY
LEP= CLASSIFIED AS LIMITED ENGLISH PROFICIENCY

TABLE 2

SUMMARY OF N(P) PAIRWISE T-TESTS FOR RACIAL/ETHNIC, LANGUAGE LEARNING AND SCHOOL LUNCH SUBGROUPS: 2003 NAEP READING

RACE/ETHNICITY MEANS

	WHITE	BLACK	HISPANIC	ASIAN
WHITE	------			
BLACK	< (-30) P=.0000	-----		
HISPANIC	< (-28) P=.0000	> (2) P=.0046	-----	
ASIAN	= (-3) P=.0765	> (27) P=.0000	> (25) P=.0000	----

ENGLISH LANGUAGE LEARNER (ELL) MEANS

	LEP	NLEP
LEP	-----	
NLEP	> (33) P=.0000	-----

NATIONAL SCHOOL LUNCH PROGRAM (NSLP) MEANS

	ELG	NELG	INA
ELG	-----		
NELG	>(28) P =.0000	----	
INA	> (18) P=.0000	<(-10) P=.0000	-----

TABLE 3

Summary of Means and Percents for Joint Distributions
of N(P) 2003 Grade 4 NAEP Reading Scores

RACE/ETHNICITY X ELL STATUS

	LEP		NLEP	
GROUP	M	%	M	%
WHITE	203	1	227	99
BLACK	176	1	198	99
HISPANIC	183	37	209	63
ASIAN	200	21	231	79

RACE/ETHNICITY X NSLP STATUS

	ELG		NELG		INA	
GROUP	M	%	M	%	M	%
WHITE	212	25	232	70	230	4
BLACK	193	73	211	24	199	4
HISPANIC	195	74	213	22	200	4
ASIAN	209	39	234	57	230	4

ELL X NSLP

	LEP		NLEP	
GROUP	M	%	M	%
ELG	183	15	204	85
NELG	199	3	230	97
INA	187	8	222	92

CHART I

SUMMARY OF VARIABLES FOR STATE-LEVEL ANALYSES

CODE	VARIABLE DESCRIPTION
ALL MEAN	MEAN FOR ALL STUDENTS
W MEAN	MEAN FOR WHITE STUDENTS
B MEAN	MEAN FOR BLACK STUDENTS
H MEAN	MEAN FOR HISPANIC STUDENTS
A MEAN	MEAN FOR ASIAN STUDENTS
LEP MEAN	MEAN FOR STUDENTS CLASSIFIED LEP
NLEP MEAN	MEAN FOR STUDENTS NOT CLASSIFIED LEP
P ALL LEP	PERCENT OF ALL STUDENTS CLASSIFIED LEP
P W LEP	PERCENT OF WHITE STUDENTS CLASSIFIED LEP
P B LEP	PERCENT OF BLACK STUDENTS CLASSIFIED LEP
P H LEP	PERCENT OF HISPANIC STUDENTS CLASSIFIED LEP
P A LEP	PERCENT OF ASIAN STUDENTS CLASSIFIED LEP
ELG MEAN	MEAN FOR STUDENTS ELIGIBLE FOR NSLP
NELG MEAN	MEAN FOR STUDENTS NOT ELIGIBLE FOR NSLP
P ALL ELG	PERCENT OF ALL STUDENTS ELIGIBLE FOR NSLP
P W ELG	PERCENT OF WHITE STUDENT ELIGIBLE FOR NSLP
P B ELG	PERCENT OF BLACK STUDENTS ELIGIBLE FOR NSLP
P H ELG	PERCENT OF HISPANIC STUDENTS ELIGIBLE FOR NSLP
P A ELG	PERCENT OF ASIAN STUDENTS ELIGIBLE FOR NSLP
P ELG LEP	PERCENT OF STUDENTS ELIGIBLE FOR NSLP WHO ARE CLASSIFIED LEP
P NELG LEP	PERCENT OF STUDENTS NOT ELIGIBLE FOR NSLP WHO ARE CLASSIFIED LEP

TABLE 4

SUMMARY OF DESCRIPTIVE STATISTICS FOR STATE-LEVEL VARIABLES: G4R03

CODE	N	MIN	MAX	MEAN	STD. DEV.
ALL MEAN	50	203	228	218.58	6.33
W MEAN	50	217	238	226.24	4.58
B MEAN	41	188	215	199.32	6.89
H MEAN	40	191	220	204.43	7.16
A MEAN	24	197	238	225.63	9.78
LEP MEAN	35	174	204	188.71	8.82
NLEP MEAN	50	205	229	220.22	5.73
P ALL LEP	50	1	30	5.24	5.85
P W LEP	50	0	5	.90	1.02
P B LEP	41	0	11	1.68	2.26
P H LEP	40	8	61	29.83	14.34
P A LEP	24	3	54	21.50	14.11
ELG MEAN	49	191	217	204.33	6.04
NELG MEAN	49	218	238	228.84	4.01
P ALL ELG	49	17	67	40.63	10.62
P W ELG	49	10	53	27.86	9.28
P B ELG	40	44	88	70.40	10.34
P H ELG	39	31	89	69.72	11.39
P A ELG	23	19	71	35.96	15.12
P ELG LEP	49	0	48	9.47	9.51
P NELG LEP	49	0	11	1.94	2.17

TABLE 5

CORRELATIONS AMONG RACE/ETHNICITY MEANS AND PERCENT OF STUDENTS CLASSIFIED AS LEP: G4R03

		ALL MEAN	W MEAN	B MEAN	H MEAN	A MEAN
P ALL LEP	Pearson Correlation	-.487**	-.177	-.004	-.522**	-.262
	Sig. (2-tailed)	.000	.218	.980	.001	.216
	N	50	50	41	40	24
P W LEP	Pearson Correlation	-.378**	-.104	.020	-.367*	-.217
	Sig. (2-tailed)	.007	.471	.903	.020	.309
	N	50	50	41	40	24
P B LEP	Pearson Correlation	-.171	-.044	.116	-.433**	-.531**
	Sig. (2-tailed)	.285	.785	.471	.008	.009
	N	41	41	41	36	23
P H LEP	Pearson Correlation	-.457**	-.549**	-.406*	-.523**	-.554**
	Sig. (2-tailed)	.003	.000	.014	.001	.005
	N	40	40	36	40	24
P A LEP	Pearson Correlation	.023	-.267	-.304	-.179	-.640**
	Sig. (2-tailed)	.915	.208	.159	.402	.001
	N	24	24	23	24	24

TABLE 6

CORRELATIONS AMONG RACE/ETHNICITY MEANS AND PERCENT OF STUDENTS ELIGIBLE FOR NSLP: G4R03

		ALLMEAN	WMEAN	BMEAN	HMEAN	AMEAN
PALLELG	Pearson Correlation	-.804**	-.448**	-.329*	-.231	.239
	Sig. (2-tailed)	.000	.001	.038	.158	.271
	N	49	49	40	39	23
PWELG	Pearson Correlation	-.432**	-.660**	-.177	-.014	-.238
	Sig. (2-tailed)	.002	.000	.276	.931	.274
	N	49	49	40	39	23
PBELG	Pearson Correlation	-.326*	-.289	-.522**	-.060	.167
	Sig. (2-tailed)	.040	.071	.001	.733	.456
	N	40	40	40	35	22
PHELG	Pearson Correlation	-.259	-.140	-.256	-.522**	-.087
	Sig. (2-tailed)	.112	.396	.138	.001	.695
	N	39	39	35	39	23
PAELG	Pearson Correlation	.077	-.031	-.077	-.144	-.534**
	Sig. (2-tailed)	.726	.890	.732	.512	.009
	N	23	23	22	23	23

TABLE 7

Means, Percents and Gaps for Univariate Results: 2003 MCA Reading

GRADE 3
(N = 58,107)

OVERALL MEAN = 1517
OVERALL S. D. = 186

GRADE 5
(N = 61,576)

OVERALL MEAN = 1567
OVERALL S. D. = 202

***RACE/ETHNICITY**

GROUP	MEAN	%	GAP
WHITE	1544	79	
BLACK	1364	8	180
HISPANIC	1358	5	186
ASIAN	1421	6	123

GROUP	MEAN	%	GAP
WHITE	1600	80	
BLACK	1391	8	209
HISPANIC	1399	4	201
ASIAN	1468	5	132

NSLP STATUS

GROUP	MEAN	%	GAP
NELG	1557	69	
ELG	1410	31	147

GROUP	MEAN	%	GAP
NELG	1616	69	
ELG	1448	31	168

ELL STATUS

GROUP	MEAN	%	GAP
NLEP	1528	92	
LEP	1319	8	209

GROUP	MEAN	%	GAP
NLEP	1580	93	
LEP	1337	7	243

*Race/ethnicity percents do no add to 100 because American Indian students are not included in the analysis.

TABLE 8

Summary of Means and Percents for Joint Distributions:
2003 MCA Reading

RACE/ETHNICITY X ELL STATUS

GRADE 3

GRP	LEP		NLEP	
	M	%	M	%
W	1362	1	1546	99
B	1302	12	1372	88
H	1293	59	1450	41
A	1339	61	1550	39

GRADE 5

GRP	LEP		NLEP	
	M	%	M	%
W	1386	1	1601	99
B	1295	10	1401	90
H	1308	53	1502	47
A	1361	55	1602	45

RACE/ETHNICITY X NSLP STATUS

GRP	ELG		NELG	
	M	%	M	%
W	1470	21	1564	79
B	1334	75	1454	25
H	1324	74	1456	26
A	1359	62	1522	38

GRP	ELG		NELG	
	M	%	M	%
W	1511	21	1623	79
B	1359	77	1498	23
H	1359	73	1510	27
A	1397	65	1602	35

ELL STATUS X NSLP STATUS

GRP	ELG		NELG	
	M	%	M	%
NLEP	1438	27	1561	73
LEP	1308	82	1372	18

GRP	ELG		NELG	
	M	%	M	%
NLEP	1475	27	1619	73
LEP	1324	84	1407	16

FIGURE 1: DISTRIBUTION FOR W MEAN:G4R03

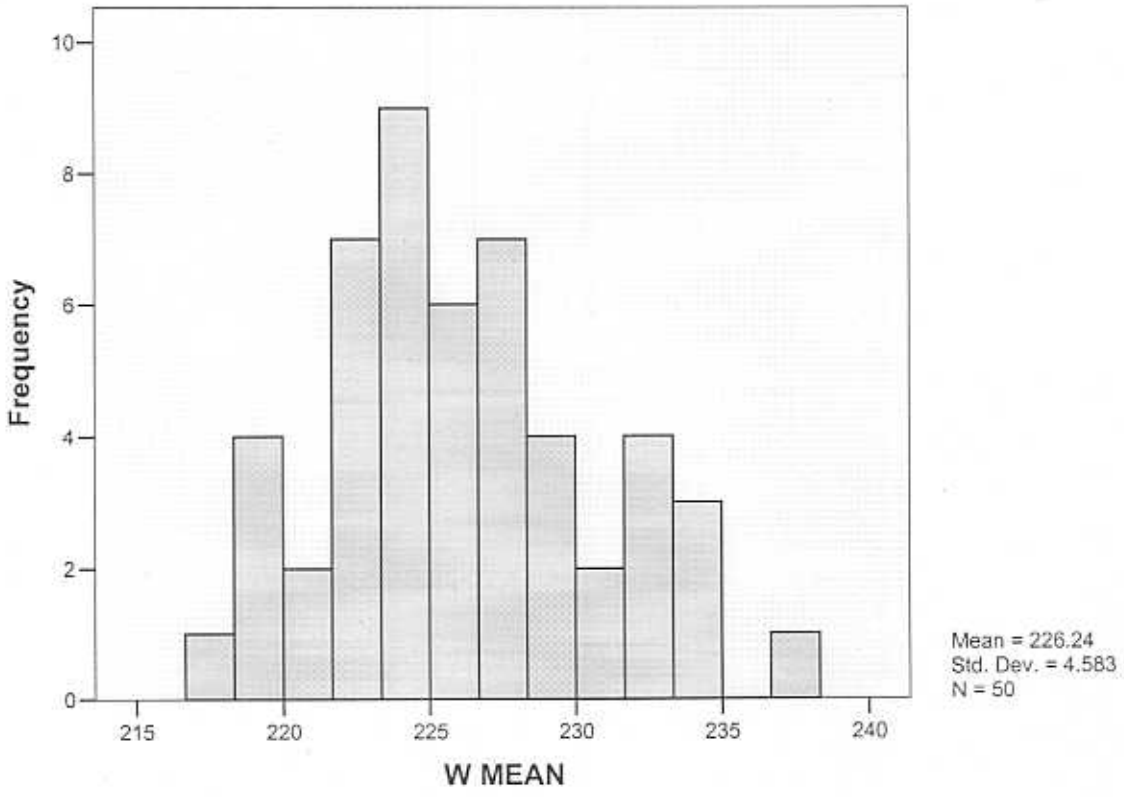

FIGURE 2: DISTRIBUTION FOR A MEAN:G4R03

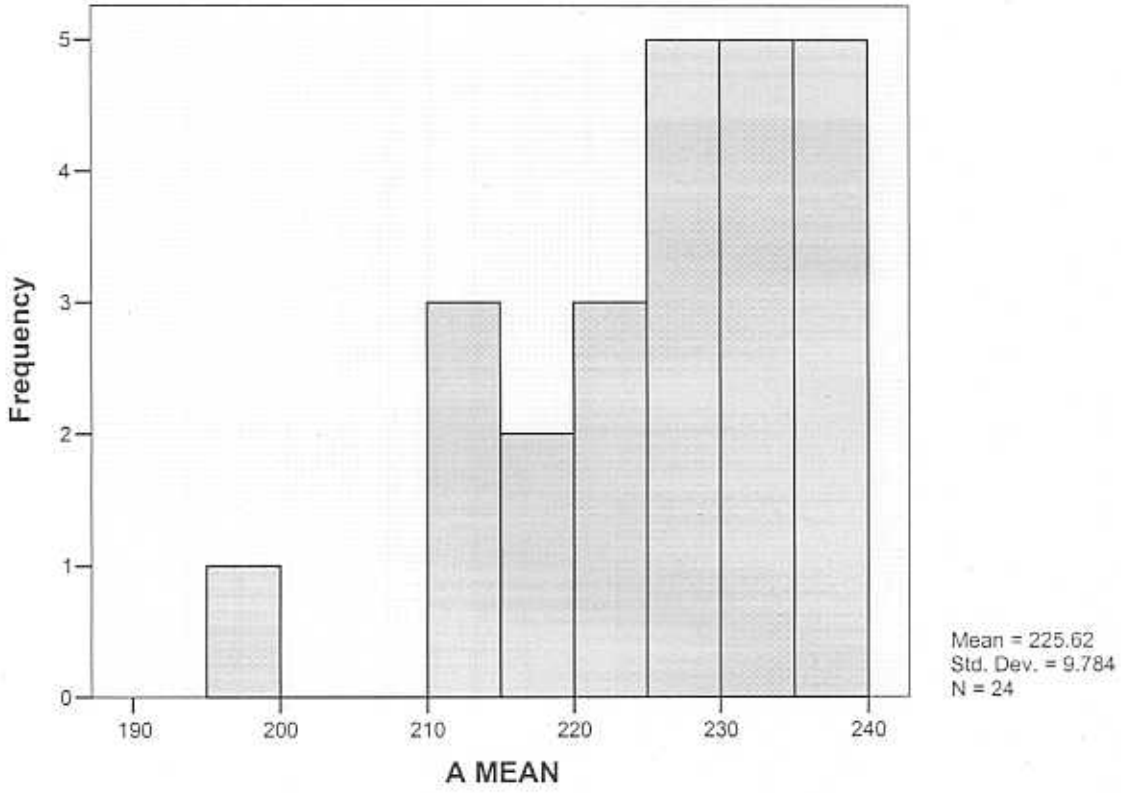

FIGURE 3: DISTRIBUTION FOR P W LEP:G4R03

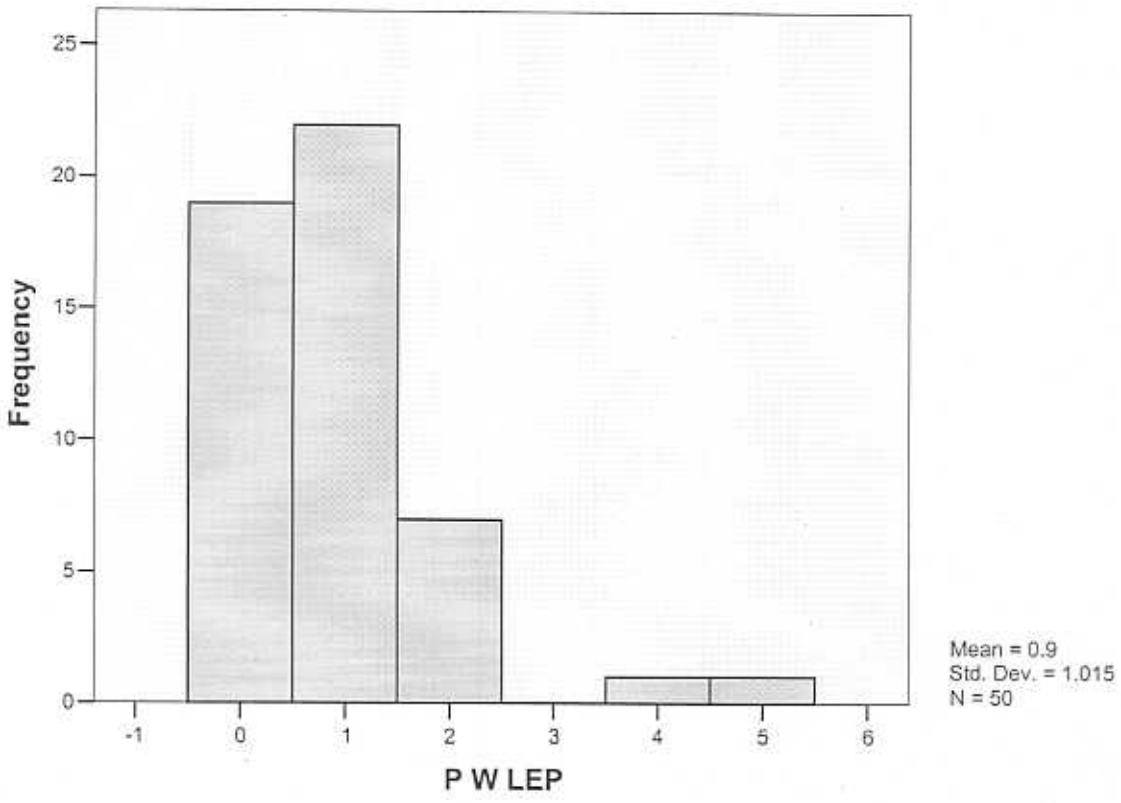

FIGURE 4: DISTRIBUTION OF P B LEP:G4R03

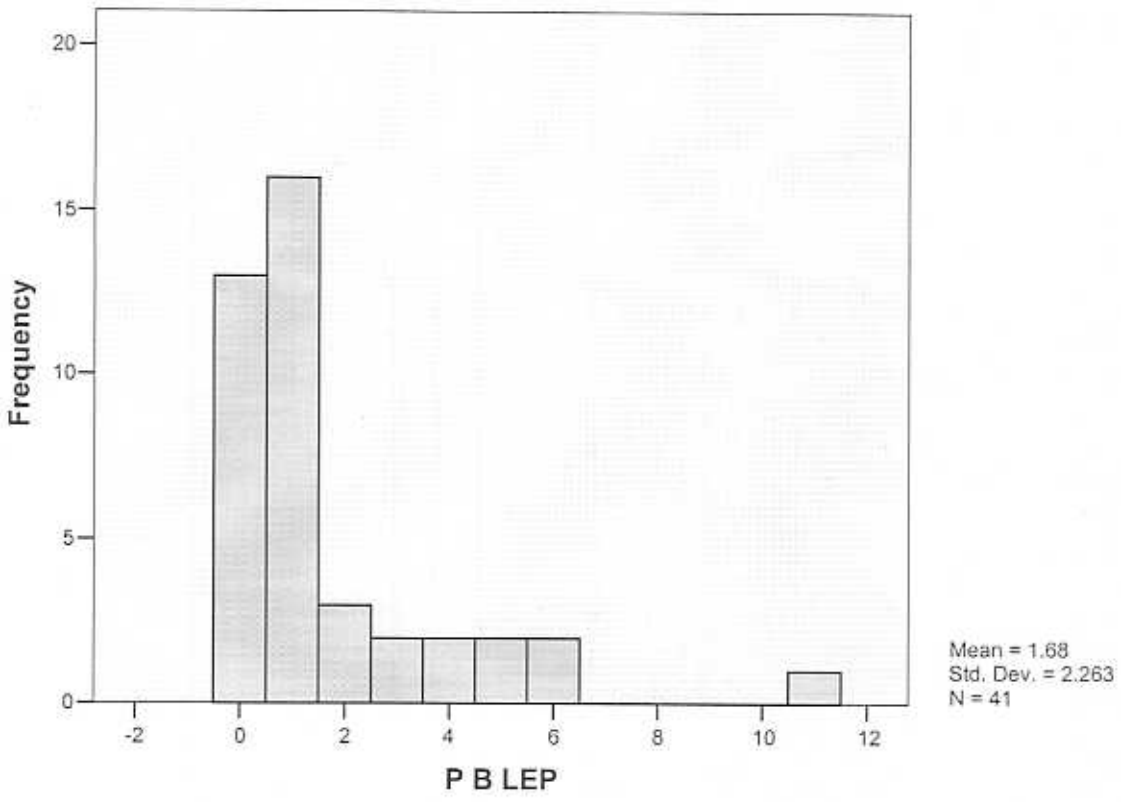

FIGURE 5: PLOT OF ALL MEAN vs. PERCENT ALL LEP:G4R03

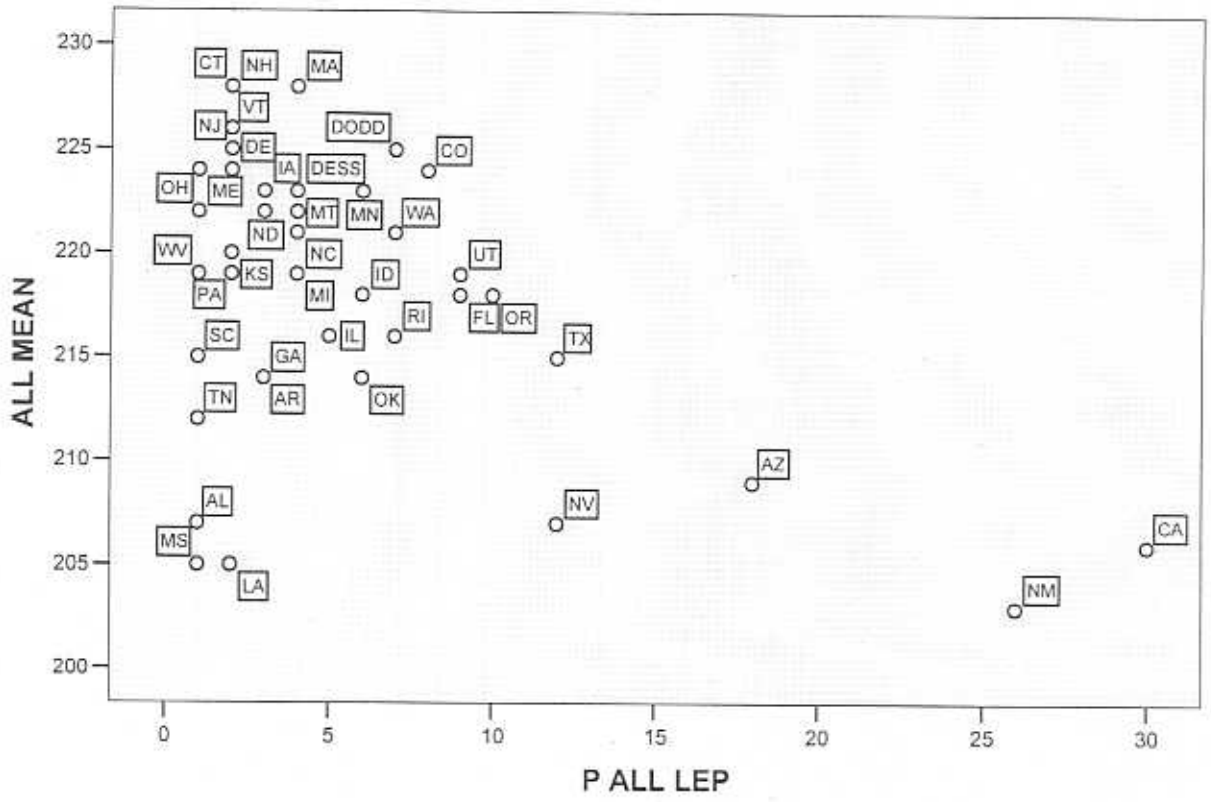

FIGURE 6: PLOT OF H MEAN vs. PERCENT H LEP:G4R03

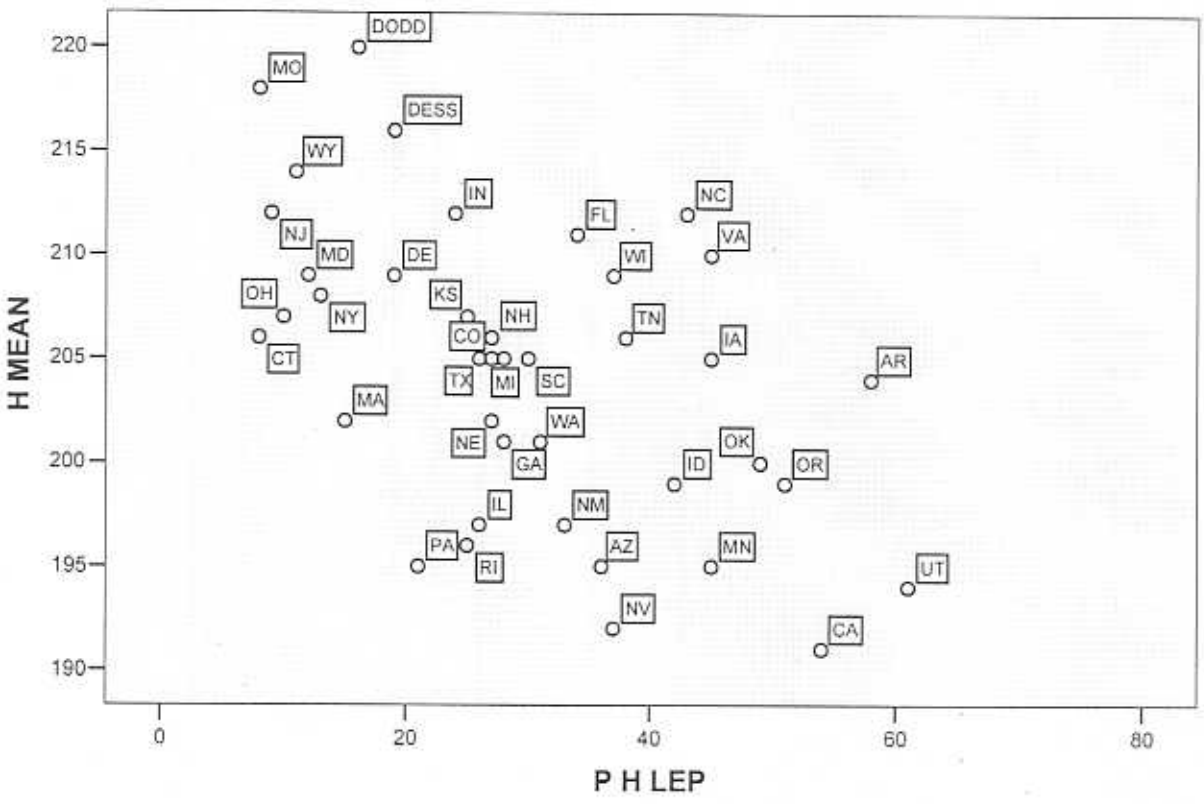

FIGURE 7: PLOT OF A MEAN vs P A LEP:G4R03

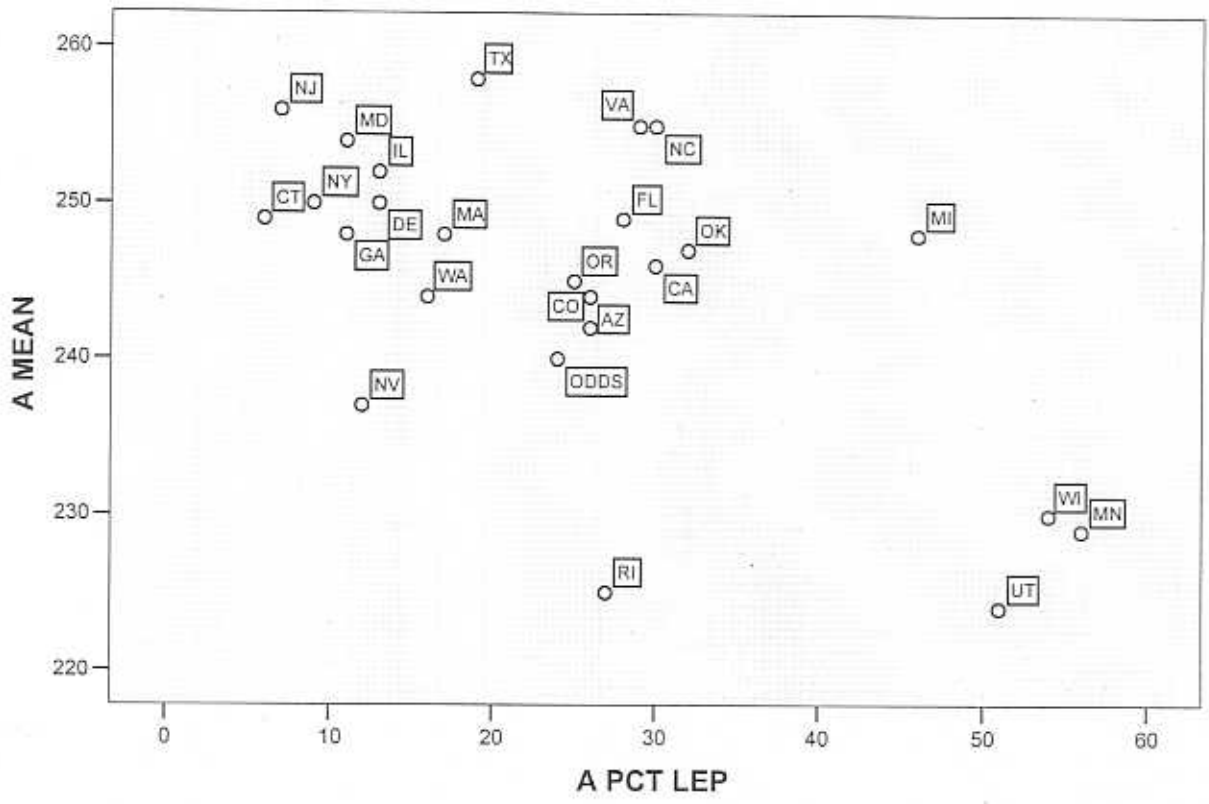

FIGURE 8: PLOT OF ALL MEAN vs. PERCENT ALL ELG: G4R03

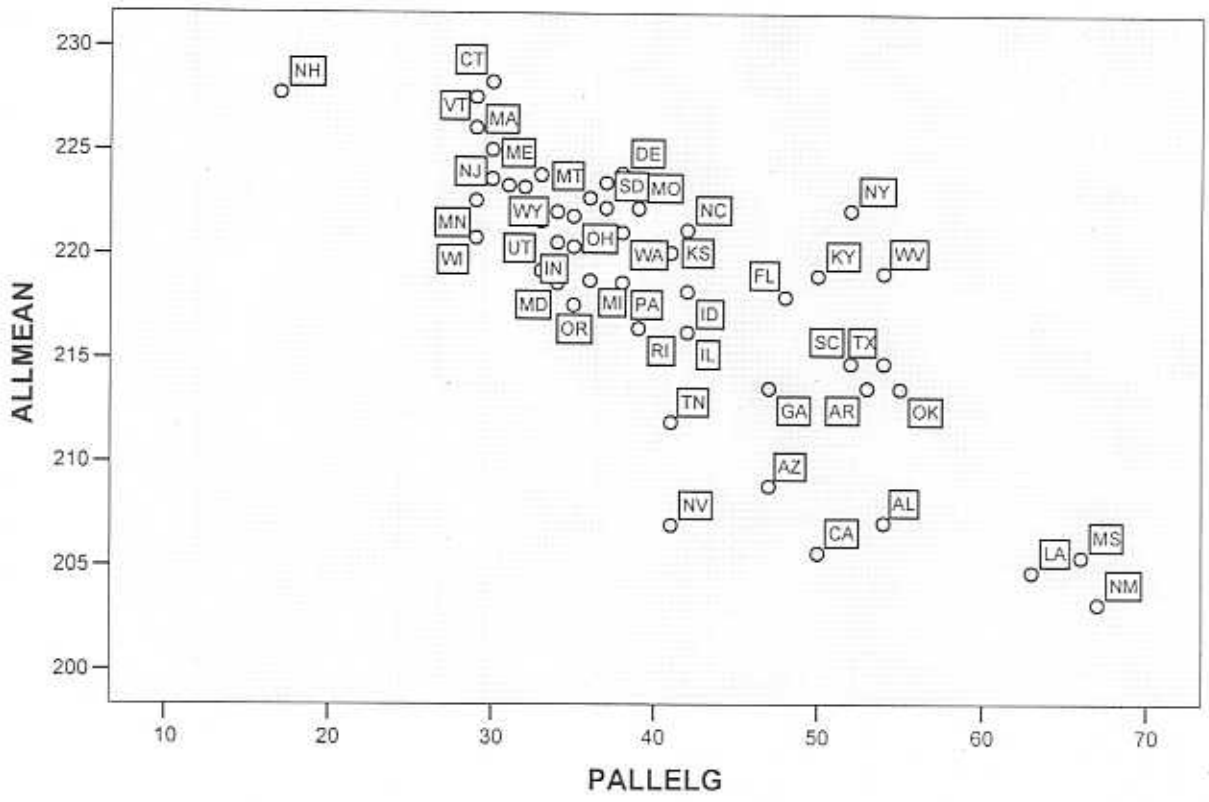

FIGURE 9: PLOT OF W MEAN vs. PERCENT W ELG: G4R03

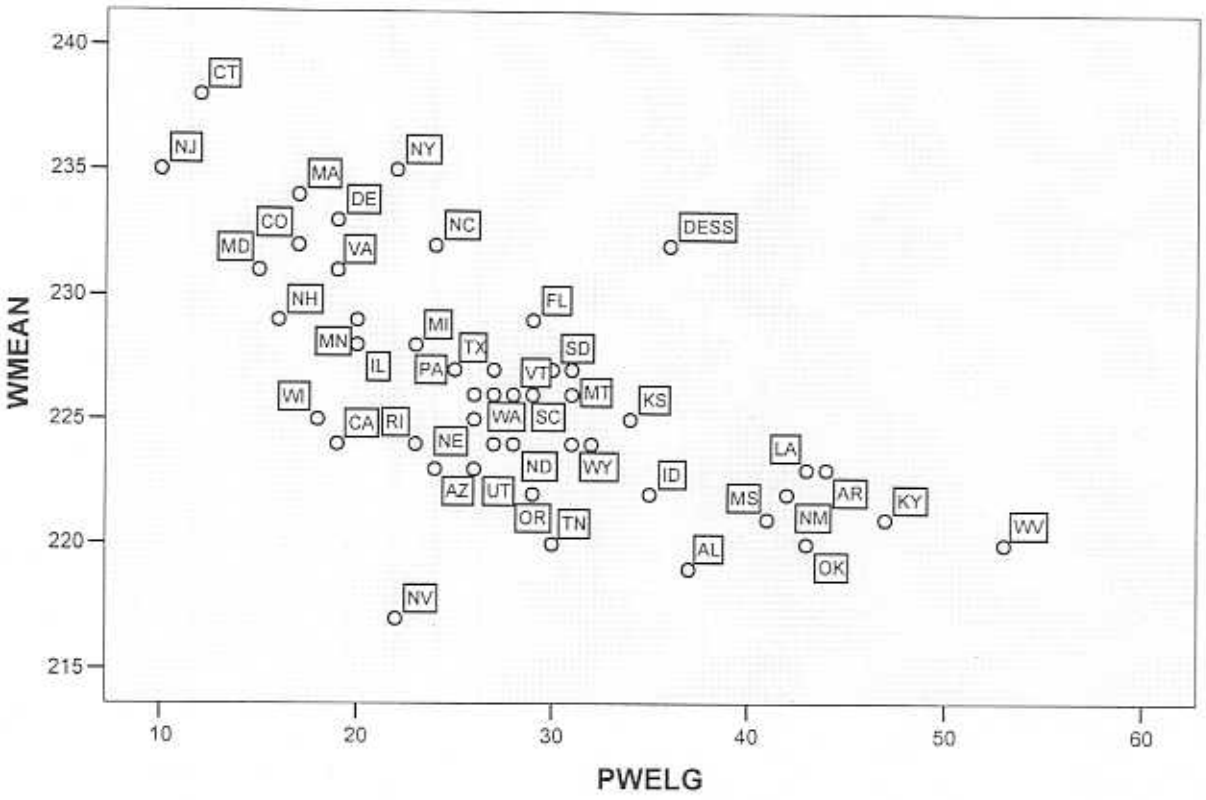

FIGURE 10: DISTRIBUTION FOR P ELG LEP:G4R03

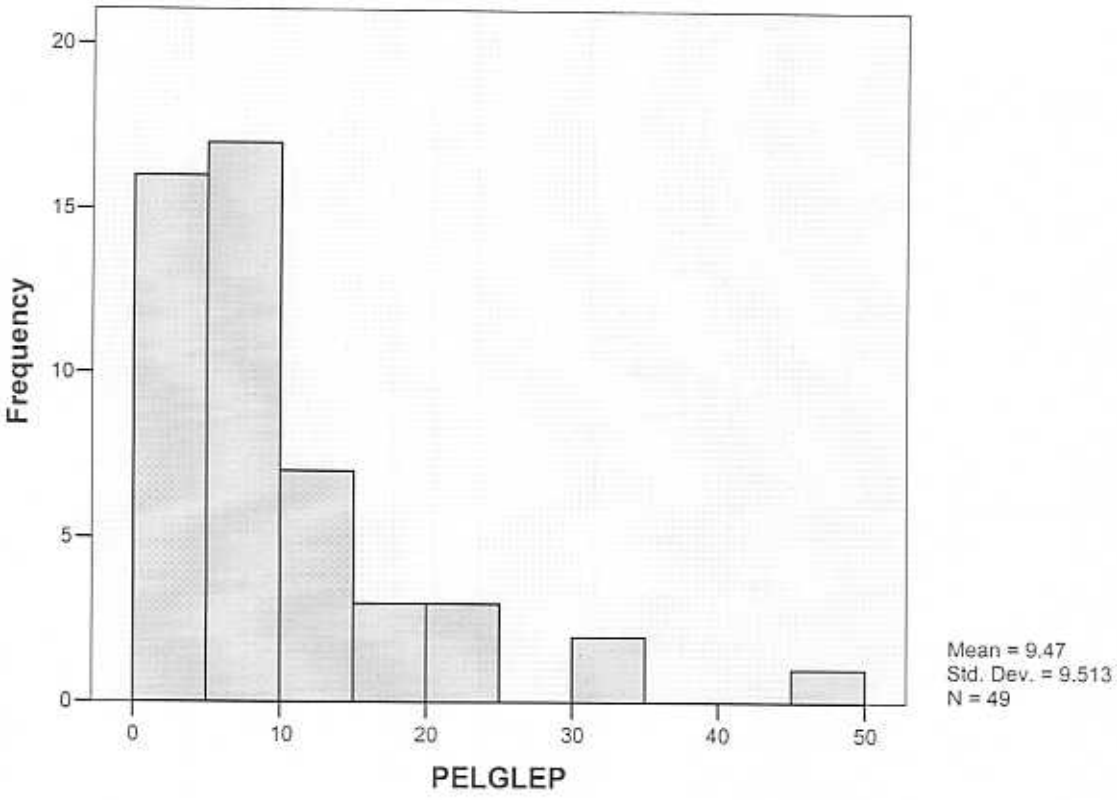

FIGURE 11:DISTRIBUTION FOR P ELG LEP:G4R03

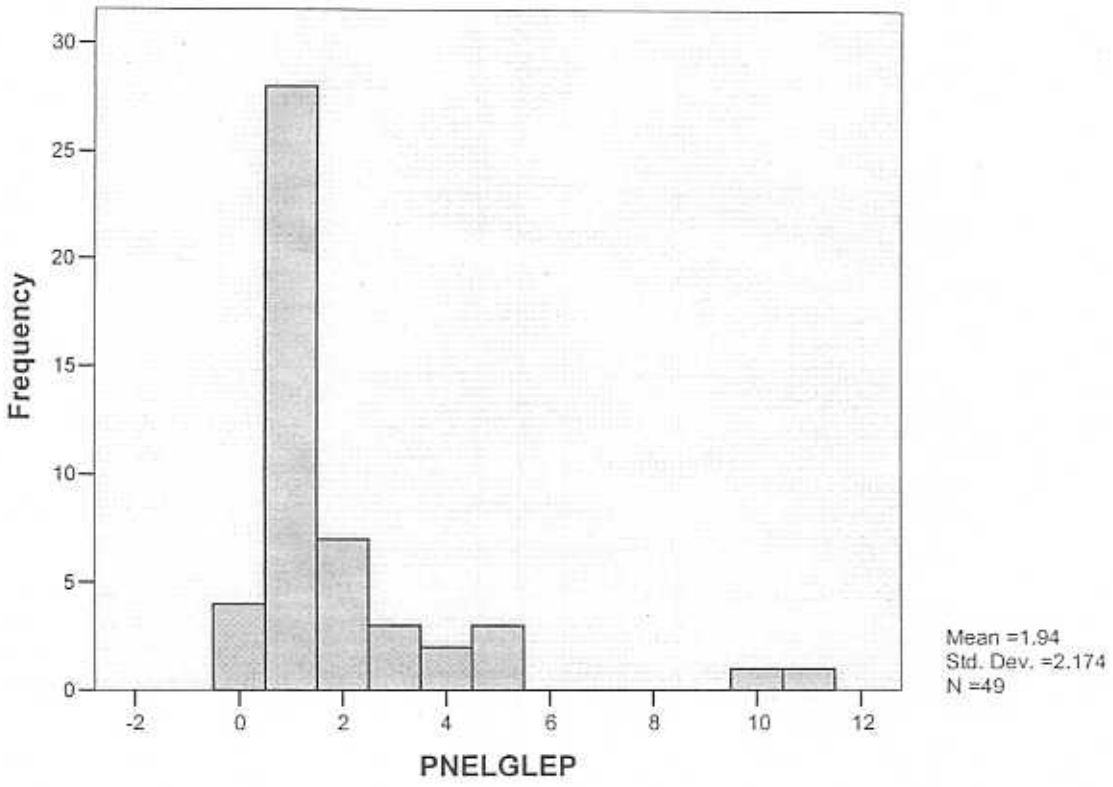

FIGURE 12: PLOT OF ELG MEAN vs. P ELG LEP:G4R03

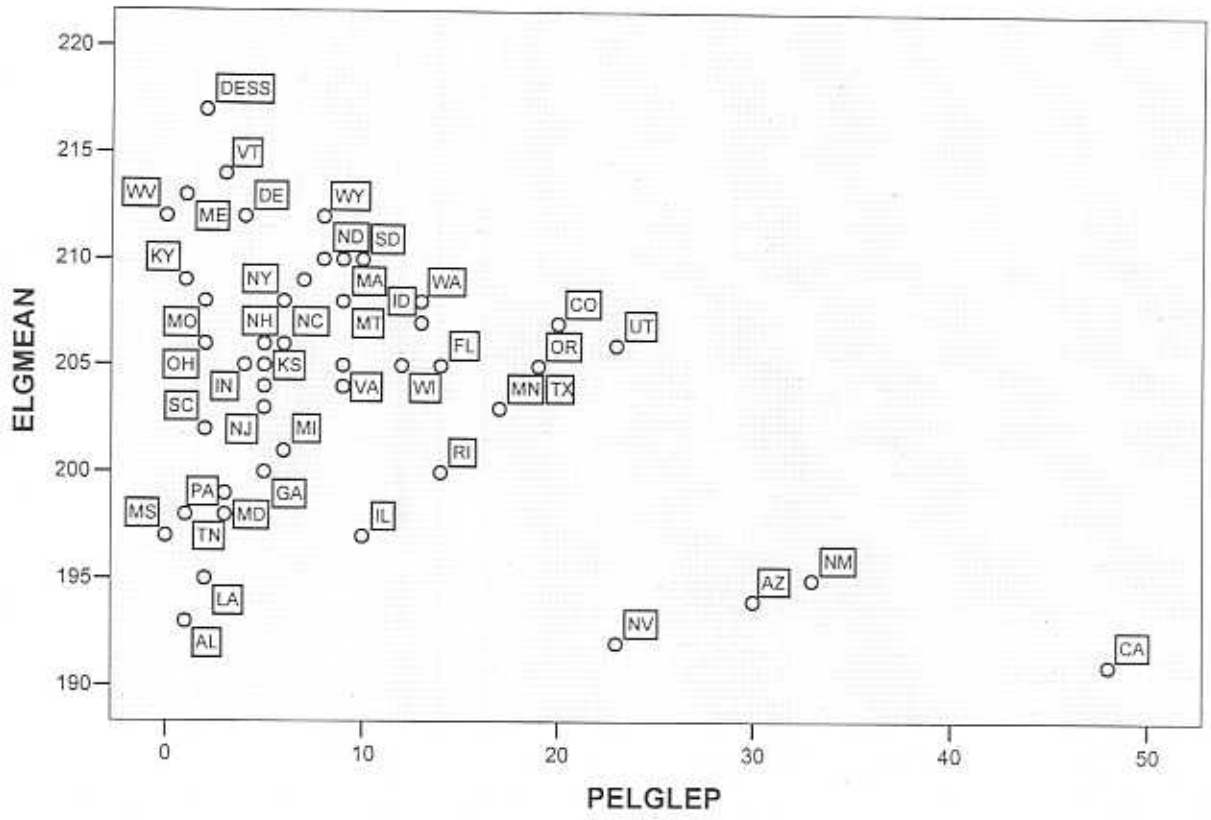

FIGURE 13: PLOT OF NELG MEAN vs. P NELG LEP:G4R03

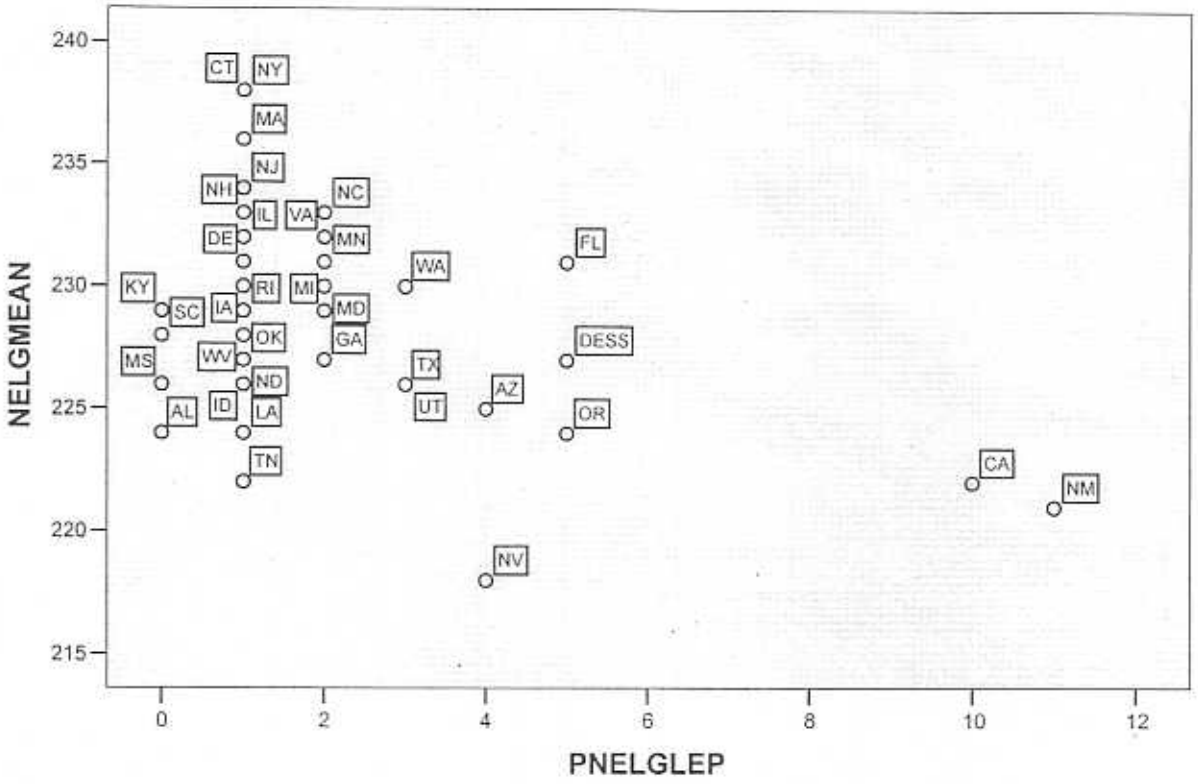

FIGURE 14: 2003 GRADE 3 MCA READING MEANS-RACE/ETHNICITY BY ELL STATUS

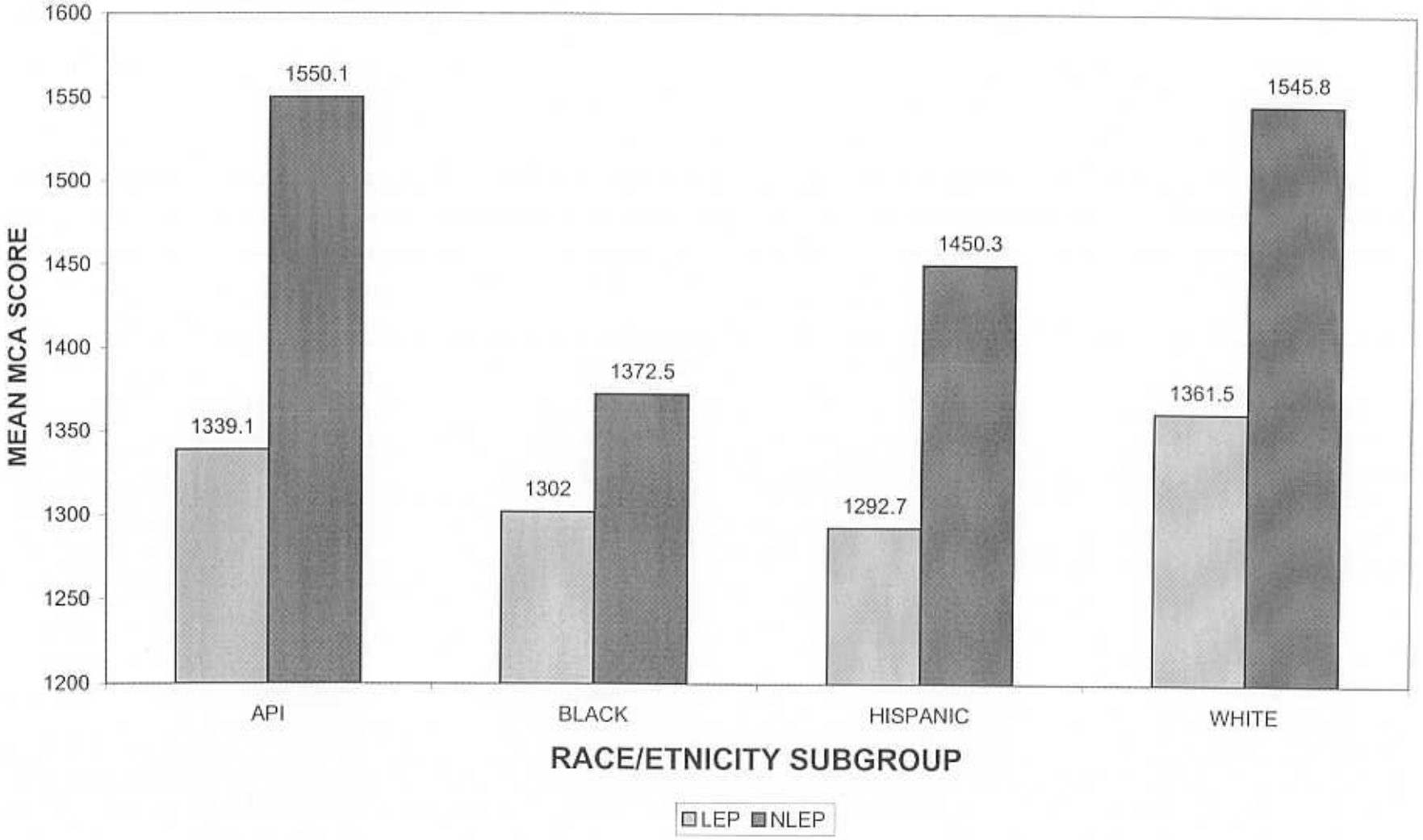

FIGURE 15: 2003 GRADE 5 MCA READING MEANS-RACE/ETHNICITY BY ELL STATUS

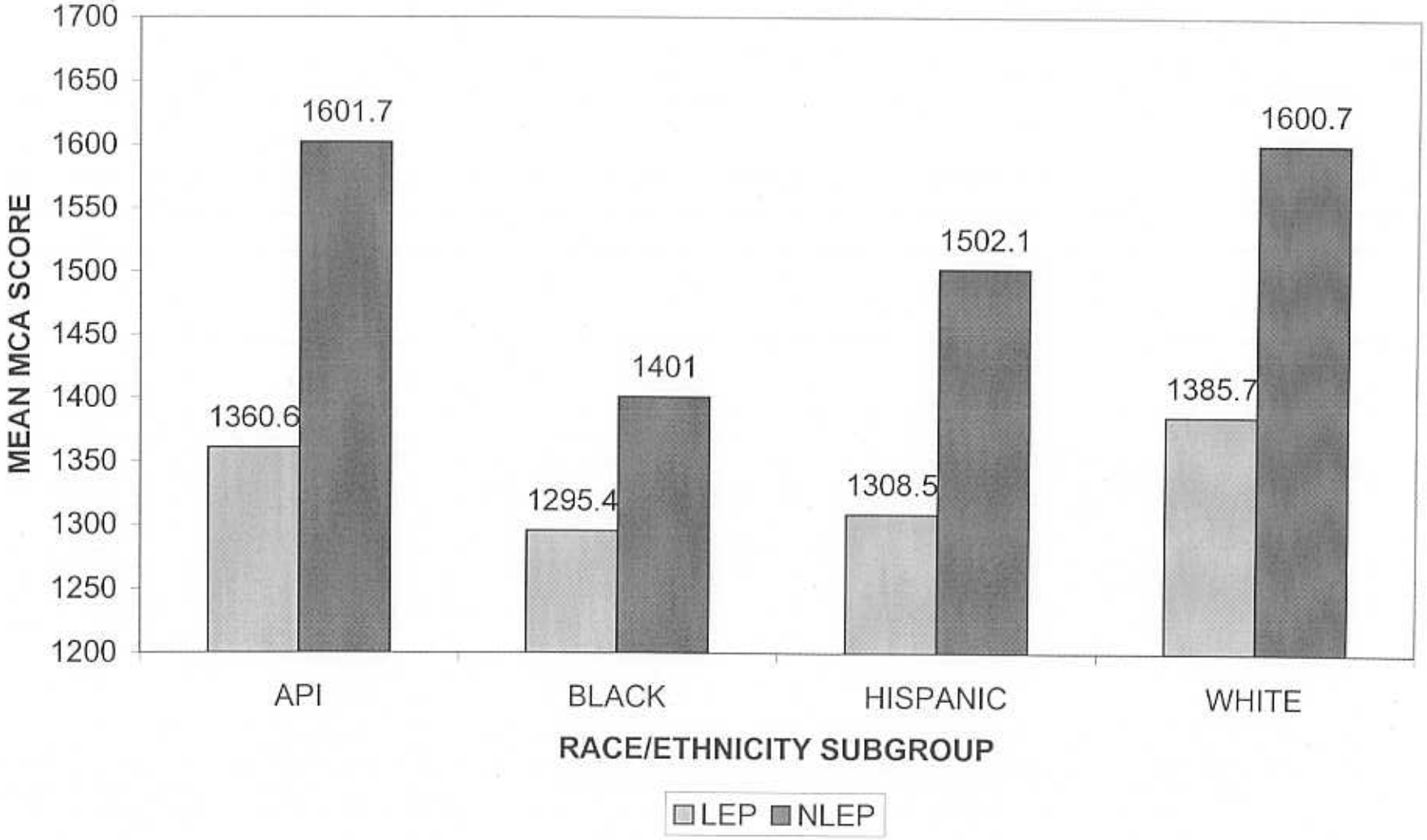

FIGURE 16: 2003 GRADE 3 MCA READING MEANS-RACE/ETNICITY BY NSLP STATUS

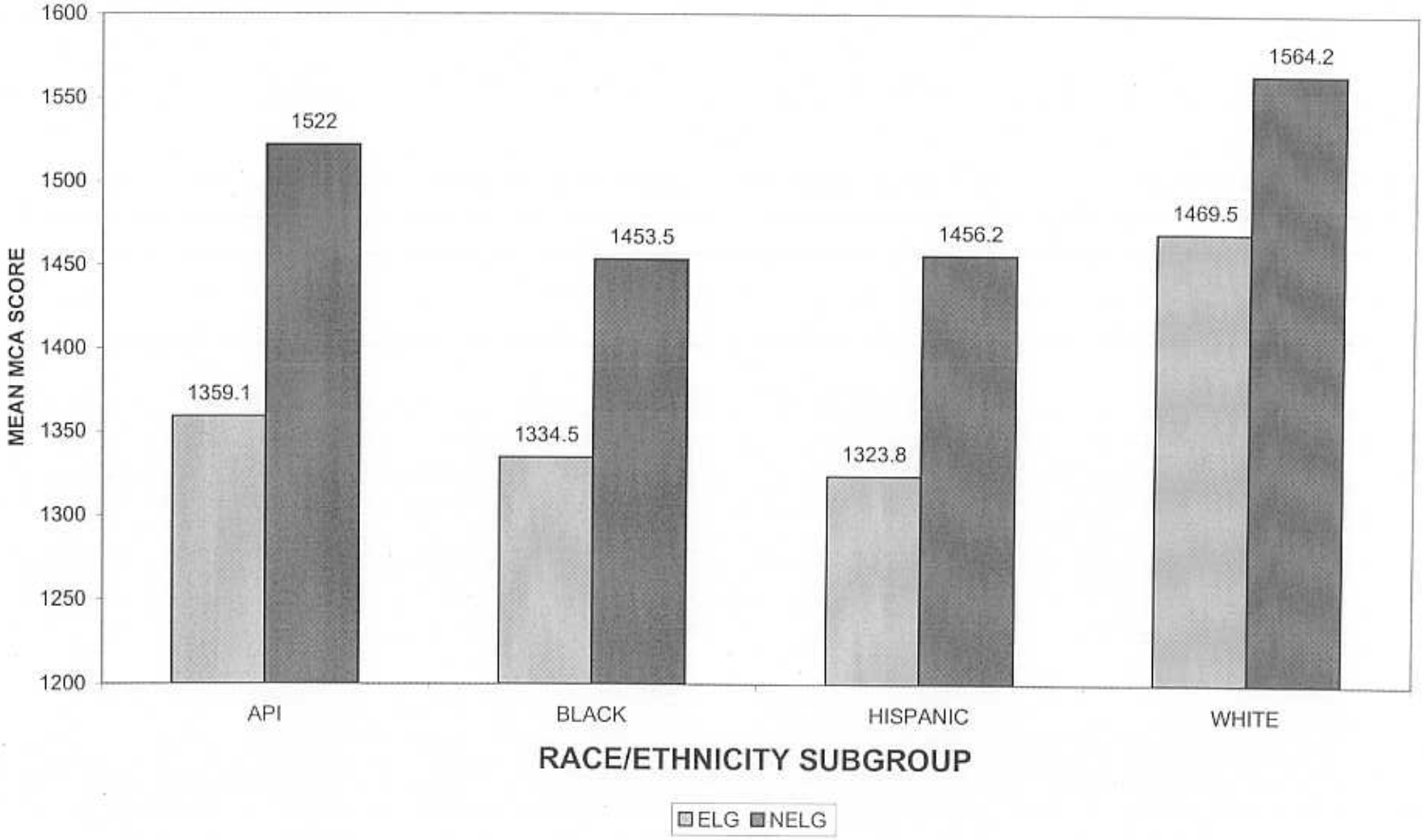

Figure 17: 2003 GRADE 5 MCA READING MEANS-RACE/ETHNICITY BY ELL STATUS