The Fall 08/Spring 09 FAST DATA: Equity Analysis and Reliability Report

I. Equity

A. Comprehensive Lesson Plan: Multiple Subjects Students, Fall 08

The appropriate statistical test used to test for significant differences in ordinal scale scores among multiple groups is Kruskal Wallis. Testing for students' differences by ethnic group yielded one significant χ^2 value for the TPEs in this project, TPE 6B. Table 1 indicates the χ^2 values for each TPE, and the probability that such a value could have occurred by chance. The score differences related to ethnic groups are significantly different on TPE 6B. The others are not significantly different.

Table 1 : MS Differences in Comprehensive Lesson Plan Scores by Ethnic Group

TPE	1ELA	6B	7	8	9
χ^2 value	5.273	13.947	4.935	1.206	10.691
Probability	.383	.016	.424	.944	.058

The appropriate test for two groups, either gender differences, or differences between native and non-native speakers of English for such data is Mann Whitney. Table 2 provides the results for Gender group differences and indicates that none of the differences related to gender were statistically significant on the CLP.

Table 2 : MS Differences in Comprehensive Lesson Plan Scores by Gender Group

TPE	1ELA	6B	7	8	9
z value	-.109	-1.787	-.918	-1.488	-.829
Probability	.913	.074	.359	.137	.407

For native speakers of English v. non-native speakers, TPE 6B and 9 were statistically significant (Table 3) with the native-speakers receiving higher scores in both cases.

Table 3 : MS Differences in Comprehensive Lesson Plan Scores by Native, v. Non-Native Speakers of English

TPE	1ELA	6B	7	8	9
z value	-.670	-2.675	-1.126	-.037	-2.689
Probability	.503	.007	.260	.971	.007

This result isn't surpising since it seems safe to assume that those who are more comfortable with the language in which the CLP must be produced might score higher.

In summary, for Fall 2008 there were occasional significant differences in the Comprehensive Lesson Plan scores related to multiple subjects students' ethnic group membership (6B), and their native English status (6B and 9), but not their gender.

B. Comprehensive Lesson Plan: Single Subjects Students, Fall 08

The test for significant differences in ordinal scale scores among multiple groups is Kruskal Wallis. Ethnic group differences in scores received on the CLP were not significant for single subjects students.

Table 4: SS Differences in Comprehensive Lesson Plan Scores by Ethnic Group

TPE		6B	7	8	9
χ^2 value		2.673	12.154	4.489	5.076
Probability		.849	.059	.611	.534

Analysis of differences between the gender groups and native speakers v. non-native speakers of English results are in Tables 5 and 6.

Table 5 : SS Differences in Comprehensive Lesson Plan Scores by Gender Group

TPE		6B	7	8	9
z value		-3.385	-3.259	-3.403	-3.675
Probability		.001	.001	.001	.001

Among single subjects candidates differences between the scores of males and females were significant—statistically females students consistently received higher scores than male students on each of the TPEs involved in the project.

We asked whether the differences reflected pre-existing differences in measures of verbal ability. Since all credential candidates must pass the California Basic Educational Skills Test (CBEST), we used CBEST reading, writing, and math scores from that test to determine whether there were corresponding differences in scores on those measures for these same students. Differences in CBEST reading, and writing scores were not statistically significant for these students (for reading, $t = -.926$, $p = .357$, for writing, $t = -1.537$, $p = .127$). Differences in math scores (which may have little to do with the skills measures by the Comprehensive Lesson Plan Project) *were* significantly different, favoring males ($t = 2.311$, $p = .023$). Inasmuch as the CBEST scores are a measure of basic reading, writing, and math ability, there appear to have been no statistically significant initial differences between the genders. However, there are gender differences in TPE scores, a fact that warrants further exploration. The issue is why do female students consistently receive higher scores than male students.

The differences among single subjects students between the Comp. Lesson Plan Scores of native speakers of English and non-native speakers were not statistically significant (Table 6).

Table 6 : SS Differences in Comprehensive Lesson Plan Scores by Native, v. Non-Native Speakers of English

TPE		6B	7	8	9
z value		-.935	-.850	-.176	-1.224
Probability		.350	.395	.860	.221

In summary, for Fall 2008 there were no significant differences in the Comprehensive Lesson Plan scores related to the students' ethnic group membership, or their native English status.

There were significant differences according to gender, however. Female students did significantly better on this project than males. This difference appears not to reflect initial differences in reading or writing ability for which CBEST scores indicate non significant differences between the males and the females in this group. These results indicate the need to examine the way this project is managed to ensure that there is no gender bias in the way this assignment is administered or scored.

C. Teaching Sample: Multiple Subjects Students, Spring 09

Kruskal-Wallis analysis indicates that the Teaching Sample scores for multiple subjects students were not significantly different by ethnic group (Table 7).

Table 7 : MS Differences in Teaching Sample Scores by Ethnic Group

TPE	8,10,11	1,9	2,3	7,8,9	2,3,4,9	3,12	12,13
χ^2 value	4.971	7.876	5.623	7.380	2.945	6.068	2.446
Probability	.419	.163	.345	.194	.708	.300	.785

Table 8 provides the results for Gender group differences in the Teaching Sample scores for multiple subjects students in Spring 09.

Table 8 : MS Differences in Teaching Sampe Scores by Gender Group

TPE	8,10,11	1,9	2,3	7,8,9	2,3,4,9	3,12	12,13
z value	-1.224	-2.075	-.172	-1.343	-.305	-.711	-.191
Probability	.221	.038	.863	.179	.761	.477	.848

Among multiple subjects candidates differences between the scores of males and females were not statistically significant, except on TPEs 1,9 where females' scores were significantly higher than males'.

Table 9 provides the Native Speaker results for the Spring 09 Teaching Sample data for multiple subjects students. Results indicate no significant differences.

Table 9 : MS Differences in Teaching Sample Scores by Native, v. Non-Native Speakers of English

TPE	8,10,11	1,9	2,3	7,8,9	2,3,4,9	3,12	12,13
z value	-1.612	-.229	-.328	-.594	-.210	-.970	-1.450
Probability	.107	.819	.743	.553	.834	.332	.147

In summary, for Fall 2008 there were no significant differences in Teaching Sample scores related to students' ethnicity, nor to their native English status. There were significant differences in males' versus females' scores on TPEs 1 and 9.

D. Teaching Sample: Single Subjects Students, Spring 09

Kruskal-Wallis analysis indicates that the Teaching Sample scores for single subjects students of different ethnic groups were not significantly different (Table 7).

Table 7 : SS Differences in Teaching Sample Project Scores by Ethnic Group

TPE	8,10,11	1,9	2,3	7,8,9	2,3,4,9	3,12	12,13
χ^2 value	4.201	4.609	3.235	3.558	2.916	3.271	4.526
Probability	.521	.465	.664	.615	.713	.658	.476

Table 8 provides the results for Gender group differences. Table 9 provides the Native Speaker results for the Spring 09 Teaching Sample data for single subjects students:

Table 8 : SS Differences in Teaching Sample Project Scores by Gender Group

TPE	8,10,11	1,9	2,3	7,8,9	2,3,4,9	3,12	12,13
z value	-.364	-.727	-1.120	-.112	-.116	-.339	-.566
Probability	.419	.163	.345	.194	.708	.300	.785

Among single subjects candidates differences between the scores of males and females were not statistically significant. Neither were the differences between the Comp. Lesson Plan Scores of native speakers of English and non-native speakers statistically significant (Table 3).

Table 9 : SS Differences in Teaching Sample Project Scores by Native, v. Non-Native Speakers of English

TPE	8,10,11	1,9	2,3	7,8,9	2,3,4,9	3,12	12,13
z value	-1.452	-1.571	-.539	-1.467	-.290	-.397	-.731
Probability	.146	.116	.590	.142	.772	.691	.465

In summary, for Fall 2008 there were no significant differences in the Comprehensive Lesson Plan scores related to single subjects students' ethnic group membership, their gender, or their native English status.

II. Inter-rater Reliability.

In each case, work by multiple and single subjects students on the Comprehensive Lesson Plan Project and the Teaching Sample Project was scored a second time in order to determine inter-rater reliability. Scores ranged from 1 (does not meet expectations), essentially a failing grade, to 4 (exceeds expectations). The numbers of students involved in each case and the double-scoring results are as follows:

A. The Comprehensive Lesson Plan: Multiple Subjects Students, Fall 08. There were 18 students double-scored in 5 categories resulting in 90 decisions made by both judges. Of that number, 61 were identical.

- As a percent agreement value, this represents .68. Because there were 4 possible outcomes, 25%, or 15.25 of the 61 agreements can be expected to have occurred by chance.
- Of the 29 disagreements, 1 was by more than a single point.
- Nine of the 90 decisions (10%) resulted in differences over whether the student successfully completed the project

B. Comprehensive Lesson Plan: Single Subjects Students, Fall 08. There were 16 students double-scored in 4 categories resulting in 64 decisions made by both judges. Of that number, 44 were identical.

- As a percent agreement value, this represents .69. Because there were 4 possible outcomes, 25%, or 11 of the 44 agreements can be expected to have occurred by chance.
- Of the 20 disagreements, just 3 were more than a single point.
- Just 3 of the 64 decisions (4.69%) resulted in differences over whether the student successfully completed the project

C. The Teaching Sample Project: Multiple Subjects Students, Spring 09. There were 19 students double-scored in 7 categories. Of the resulting 133 decisions made by the two raters for the multiple subjects Teaching Sample project, Spring 09, 85 were identical.

- As a percent of agreement, this represents an inter-rater reliability value of .64, substantially better than the chance which is .25.
- Of the 48 scoring disagreements, 2 (1.5%) were by more than a single point.
- None of the disagreements were regarding whether the student had successfully completed the assignment.

D. The Teaching Sample Project: Single Subjects Students, Spring 09. There were 8 students double-scored in 7 categories. Of the resulting 56 decisions both judges made, in common for Multiple Subjects students on the Comprehensive Lesson Plan Project, Fall 08, 47 were identical.

- As a percent agreement value, the 47 agreements represents an inter-rater reliability value of .84. Because there were 4 possible outcomes, one would expect 14 agreements to occur by chance (.25).

- Of the 9 disagreements, just 1 (1.5%) was by more than a single point and that one was also a disagreement about whether the student successfully completed the project.