

## The Fall 09/Spring 10 FAST DATA: Equity Analysis and Reliability Report

### I. Equity

The statistical tests for these data are tests for ordinal scale responses. When comparing two groups as with gender differences for example, the test is the Mann Whitney U. When there are more than two groups as there are for ethnic group differences, or language differences, the test is the Kruskal Wallis H. In either case, the test statistic is a value of chi square ( $\chi^2$ ). The probability value indicates the likelihood that the particular chi square value could have occurred by chance. Probability values of  $p = .05$  or less are considered statistically significant.

#### A. The Multiple Subjects Site Visitation Project, Fall 09

Testing among MS students by ethnicity yielded no significant differences related to any of the TPEs in this project (Table 1).

**Table 1 : MS Differences in Site Visitation Project Scores by Ethnic Group**

TPE	1	2	4	5	11	13
$\chi^2$ value	7.100	2.886	4.478	3.954	5.462	3.432
Probability	.213	.718	.483	.556	.362	.634

Table 2 provides the results for **gender group differences** and indicates that none of the scoring differences related to gender were statistically significant for MS students.

**Table 2 : MS Differences in Site Visitation Project Scores by Gender Group**

TPE	1	2	4	5	11	13
$z$ value	-1.546	-.616	-1.118	-1.003	-1.110	-1.257
Probability	.122	.538	.264	.316	.267	.209

For language differences among students, the major groups were those who spoke Hmong, Spanish, and a “non-applicable” group. Table 3 indicates that there were significant differences related to language on TPE1, but on none of the other TPEs.

**Table 3 : MS Differences in Site Visitation Project Scores by Language Group**

TPE	1	2	4	5	11	13
$z$ value	6.077	1.088	2.604	1.767	3.943	2.246
Probability	.048	.580	.272	.413	.139	.325

In summary, for Fall 2009 in three different analyses of 6TPEs there was just one significant differences in the Site Visitation Project scores and although unlikely to have occurred by chance, that one difference was quite modest.

### B. The Multiple Subjects Holistic Proficiency Project, Spring 10

Testing for students' **ethnic group** identification yielded **significant differences** in scoring related to ethnicity on two TPEs On TPE 5 (student engagement) those of the “not applicable” ethnic group scored highest and Hispanic students scored lowest. On TPE 12 (professional and ethical obligations) the “not applicable” group scored highest and Hmong students scored lowest.

**Table 4 : MS Differences in the Holistic Proficiency Project Scores by Ethnic Group**

TPE	sci	math	hist	3	5	6	10	12
$\chi^2$ value	7.325	8.686	5.975	7.507	13.501	11.765	7.797	15.929
Probability	.292	.192	.426	.276	.036	.067	.253	.014

Table 2 provides the results for **gender group differences** and indicates that female students performed significantly higher than male students on “pedagogical skills in science.” None of the other scoring differences related to gender were statistically significant.

**Table 5 : MS Differences in Holistic Proficiency Project Scores by Gender Group**

TPE	sci	Math	hist	3	5	6	10	12
$\chi^2$ value	249.0	367.5	258.5	361.5	312.5	300.0	319.5	292.5
Probability	.042	.875	.069	.795	.307	.240	.352	.213

For **language differences** among students, the major groups were those who spoke Hmong, Spanish, and a “non-applicable” group. Table 3 indicates that there were significant differences related to language on only TPE1.

**Table 6: MS Differences in Holistic Proficiency Project Scores by Language Group**

TPE	sci	Math	hist	3	5	6	10	12
$\chi^2$ value	2.669	.691	1.758	1.305	1.573	4.775	1.309	.351
Probability	.440	.875	.624	.728	.665	.189	.727	.950

In summary, for Spring 2010 there were statistically significant scoring differences on just 3 of 24 TPEs on the Holistic Proficiency Project. Two TPE differences were related to ethnicity and one related to gender. There were no language-related differences in scoring.

### C. The Single Subjects Site Visitation Project, Fall 09

Testing for students' differences by ethnic group yielded no significant differences for SS students related to **ethnic group differences** on any of the Site Visitation Project TPEs.

**Table 7 : SS Differences in Site Visitation Project Scores by Ethnic Group**

TPE	2	4	5	11	13
$\chi^2$ value	2.052	2.204	3.364	7.609	3.190
Probability	.842	.820	.644	.179	.671

Table 8 provides the results for **gender group differences** and indicates that none of the SS scoring differences related to gender were statistically significant.

**Table 8 : SS Differences in Site Visitation Project Scores by Gender Group**

TPE	2	4	5	11	13
$\chi^2$ value	2164.0	2165.0	2196.0	2019.0	2063.5
Probability	.822	.826	.942	.353	.479

For **language differences** among students, the major groups were those who spoke Hmong, Spanish, and a “non-applicable” group. Table 9 indicates that there were no significant differences related to language on any of the TPEs related to Site Visitation for SS students.

**Table 9 : SS Differences in Site Visitation Project Scores by Language Group**

TPE	2	4	5	11	13
$\chi^2$ value	2.052	2.204	3.364	7.609	3.190
Probability	.842	.820	.644	.179	.671

In summary, for Fall 2009 in three different analyses of 5TPEs there were no significant differences in the Site Visitation Project scores related to ethnicity, gender, or to primary language status for SS students.

#### D. The Single Subjects Holistic Proficiency Project, Spring 10

Testing for students' by **ethnic group** yielded no significant differences in scoring for SS students related to any of the Holistic Proficiency Project TPEs.

**Table 10: Differences in the Holistic Proficiency Project Scores by Ethnic Group**

TPE	1	3	5	6c	10	12
$\chi^2$ value	1.717	5.980	4.319	2.940	2.891	4.949
Probability	.887	.308	.504	.709	.717	.422

Table 11 provides the results for SS **gender group differences** and indicates that females students performed significantly higher than male students on all TPEs related to the Holistic Proficiency Project *except* “pedagogical skills.” None of the other scoring differences related to gender were statistically significant.

**Table 11: Differences in the Holistic Proficiency Project Scores by Gender Group**

TPE	1	3	5	6c	10	12
$\chi^2$ value	1597.0	1756.0	1658.0	1723.5	1953.0	1633.5
Probability	.001	.015	.004	.010	.159	.004

For **language differences** among SS students, the major groups were Hmong, Spanish, and a “non-applicable” group, presumably English speakers. Table 12 indicates that there were no significant differences related to language on any of these TPEs.

**Table 12: Differences in the Holistic Proficiency Project Scores by Language Group**

TPE	1	3	5	6c	10	12
$\chi^2$ value	7.897	4.139	8.216	6.578	9.211	6.811
Probability	.246	.658	.223	.362	.162	.339

In summary, for Spring 2010 there were no statistically significant scoring differences on the Holistic Proficiency Project related to SS students' ethnicity nor to their language status. There were significant differences in scoring by gender, however, and these appear to be pronounced on TPEs 1, 3, 5, 6c, and 12.

## II. Inter-rater Reliability.

Scores on the different projects ranged from 1 (does not meet expectations), essentially a failing grade, to 4 (exceeds expectations). For each of the four data sets reported here—the Site Visitation (MS & SS) and the Holistic Proficiency Project (MS & SS) a number of students' work were scored a second time to determine inter-rater reliability. The numbers of students involved in each case and the double-scoring results are as follows:

- A. The Site Visitation Project: Multiple Subjects Students, Fall 09.** There were 10 students double-scored in 6 categories resulting in 60 decisions made by both judges. Of that number, 38 of the scores were identical from the first to the second scoring.
- As a percent agreement value, this represents .63.  
Because there were 4 possible outcomes,  $1/4^{\text{th}}$  or 15 of the 60 decisions can be expected to be agreement by chance.
  - None of the disagreements were by more than a single point.
  - Two of the 22 disagreements were about whether the student failed the TPE.
- B. Holistic Proficiency Project: Multiple Subjects Students, Spring, 1020.** There were 13 students double-scored in 8 categories resulting in 104 decisions made by two scorers. In 58 of the cases, both made the same decision.
- As a percent agreement value, this represents .56.  
26 of the 104 decisions can be expected to have occurred by chance.
  - One of the disagreements was by more than a single point.
  - None of the disagreements were about whether the student failed the TPE.
- C. The Site Visitation Project: Single Subjects Students, Fall 09.** There were 14 students double-scored in 5 categories resulting in 70 decisions made by both scorers. Of that number, 42 were identical.
- As a percent agreement value, this represents .60.  
17.5 of the 70 decisions can be expected to be agreement by chance.
  - Two of the disagreements were by more than a single point.
  - None of the disagreements were about whether the student failed the TPE.
- D. Holistic Proficiency Project: Single Subjects Students, Spring, 1020.** There were 20 students double-scored in 6 categories resulting in 120 decisions made by both scorers. In 87 of the cases, both scorers made the same decision.
- As a percent agreement value, this represents .725.  
30 of the 120 decisions can be expected to have occurred by chance.
  - One of the disagreements was by more than a single point.
  - Four of the 33 disagreements were over whether a student failed the project TPE.

In summary, these four exercises demonstrate good consistency in the scoring. In each case, the number of agreements between the first and second judges is substantially higher than could be expected to have occurred by chance. Disagreements of more than a single point were uncommon.

## FAST Data Scoring Trends, Fall 2008-Spring 2010

Because the scores are essentially rankings, data are ordinal scale. The appropriate measure of central tendency is the median, the middle-most score. The scores for these multiple and single subjects students universally have medians of either 2 or 3, which is to be expected with a scoring range of 1-4. What is somewhat surprising is that, except for scores on the comprehensive lesson plan project, whether the data are from Fall 08 or any semester since, the median scores are universally 3s for both multiple and single subjects students. This isn't necessarily problematic unless the raw scores from which the medians are calculated don't vary.

The pattern in scoring over the four semesters represented in these data tends to be quite stable from semester to semester. Scores in the 2-3 range consistently represent 80 - 90% of the scores. This explains why the median score vary little.

As one will note from examining the distributions of raw scores, 4s are quite uncommon. Perhaps this is to be expected as students may still be engaged in the formative stages of the learning process. If a score of 4 is rare, then it seems likely that the lowest scores indicating that students have not yet managed to complete some element of the project successfully will be more common. But in fact they are uncommon. The data upon which this report is based were gathered over four semesters from Fall 08 to Spring 10. For MS students there are 26 tpe categories in the four tasks. With each measured four times over the semesters there were 104 categories. There were *no* scores of 1 on 55 of those 104 occasions (53%). The relative rarity of unacceptable work is the same for SS students.

Such a low level of failure can be explained two ways. Either the teacher preparation students manage to complete all elements of the four projects without the need for remediation at a remarkable level, or the scores are inflated. The report above on inter-rater reliability indicates that the scores have reasonably good reliability, but there are at least questions about validity.

The scoring patterns described above tend to be quite stable. With minor exceptions, the proportions of 1s, 2s, 3s, and 4s tend to vary little over the four semesters for which there are data here,

**Multiple Subjects** *Comprehensive Lesson Plan Project*

tpe	Scoring	Fall 08	Spr 09	Fall 09	Spr 10
Ela1	Median	3.0	3.0	3.0	2.0
	1s (%)	1.0	0	7.4	11.1
	2s (%)	42.7	40.4	34.1	49.2
	3s (%)	47.6	41.5	47.4	39.7
	4s (%)	8.7	18.1	11.1	0
6b	Median	3.0	3.0	2.0	2.0
	1s (%)	2.9	0	4.4	9.5
	2s (%)	29.1	47.9	51.9	54.0
	3s (%)	49.5	43.6	40.7	28.6
	4s (%)	18.4	8.5	3.0	7.9
7	Median	2.0	2.0	2.0	3.0
	1s (%)	0	0	2.2	0
	2s (%)	52.4	53.2	68.9	28.6
	3s (%)	40.8	38.3	25.9	63.5
	4s (%)	6.8	8.5	3.0	7.9
8	Median	3.0	2.0	2.0	2.0
	1s (%)	2.9	0	6.7	3.2
	2s (%)	57.3	60.6	47.4	55.6
	3s (%)	35.9	37.2	42.2	34.9
	4s (%)	3.9	2.1	3.7	6.3
9	Median	3.0	3.0	2.0	3.0
	1s (%)	1.9	0	1.5	4.8*
	2s (%)	42.7	41.5	54.1	27.0*
	3s (%)	47.6	54.3	39.3	65.1*
	4s (%)	7.8	4.3	5.2	1.6*

\*Entry of an invalid number results in a non-100% sum.

Multiple Subjects *Site Visitation Project*

tpe	Scoring	Fall 08	Spr 09	Fall 09	Spr 10
1	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	1.1	2.7
	2s (%)	19.3	17.9	23.4	17.1
	3s (%)	70.2	70.2	69.1	71.2
	4s (%)	10.5	11.9	6.4	9.0
2	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	2.1	0
	2s (%)	27.2	28.6	20.2	21.6
	3s (%)	62.3	61.9	74.5	71.2
	4s (%)	10.5	9.5	3.2	7.2
4	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	2.1	0
	2s (%)	31.6	34.5	36.2	33.3
	3s (%)	58.8	52.4	57.4	57.7
	4s (%)	9.6	13.1	4.3	9.0
5	Median	3.0	3.0	3.0	3.0
	1s (%)	.9	0	2.1	0
	2s (%)	29.8	32.1	37.2	30.6
	3s (%)	58.8	58.3	58.5	63.1
	4s (%)	10.5	9.5	2.1	6.3
11	Median	3.0	3.0	3.0	3.0
	1s (%)	.9	0	1.1	0
	2s (%)	24.6	44.0	38.3	33.3
	3s (%)	65.8	44.0	55.3	61.3
	4s (%)	8.8	11.9	5.3	5.4
13	Median	3.0	3.0	3.0	3.0
	1s (%)	0	1.2	3.2	2.7
	2s (%)	36.0	22.6	24.5	17.1
	3s (%)	50.0	46.4	53.2	68.5
	4s (%)	14.0	29.8	19.1	11.7



Multiple Subjects *Teaching Sample Project*

tpc	Scoring	Fall 08	Spr 09	Fall 09	Spr 10
10,11	Median	3.0	3.0	3.0	3.0
	1s (%)	2.4	1.7	0	1.1
	2s (%)	24.7	12.6	22.3	16.8
	3s (%)	67.1	70.6	68.1	65.3
	4s (%)	5.9	15.1	9.6	16.8
1,9	Median	3.0	3.0	3.0	3.0
	1s (%)	0	1.7	1.1	1.1
	2s (%)	44.7	22.7	20.2	28.4
	3s (%)	52.9	62.2	74.5	56.8
	4s (%)	2.4	13.4	4.3	13.7
2,3	Median	3.0	3.0	3.0	3.0
	1s (%)	0	1.7	1.1	3.2
	2s (%)	51.8	22.7	31.9	27.4
	3s (%)	41.2	61.3	58.5	61.1
	4s (%)	7.1	14.3	8.5	8.4
7,8,9	Median	3.0	3.0	3.0	3.0
	1s (%)	0	.8	0	0
	2s (%)	40.0	21.0	19.1	32.6
	3s (%)	48.2	49.6	67.0	52.6
	4s (%)	11.8	28.6	13.8	14.7
2-4,9	Median	3.0	3.0	3.0	3.0
	1s (%)	1.2	.8	0	0
	2s (%)	45.9	21.0	26.6	26.3
	3s (%)	48.2	62.2	61.7	56.8
	4s (%)	4.7	16.0	11.7	16.8
3,12	Median	3.0	3.0	3.0	3.0
	1s (%)	1.2	1.7	1.1	0
	2s (%)	43.5	18.5	29.8	26.3
	3s (%)	50.6	49.6	55.3	60.0
	4s (%)	4.7	30.3	13.8	13.7
12,13	Median	3.0	3.0	3.0	3.0
	1s (%)	2.4	1.7	1.1	1.1
	2s (%)	36.5	16.0	25.5	27.4
	3s (%)	54.1	60.5	62.8	58.9
	4s (%)	7.1	21.8	10.6	12.6

Multiple Subjects *Holistic Proficiency Project*

tpe	Scoring	Fall 08	Spr 09	Fall 09	Spr 10
1m	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	1.1
	2s (%)	11.1	7.0	10.4	19.4
	3s (%)	65.6	74.8	71.9	68.8
	4s (%)	23.3	18.3	17.7	10.8
1s	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	1.1
	2s (%)	10.0	9.6	14.6	23.7
	3s (%)	68.9	72.2	65.6	61.3
	4s (%)	21.1	18.3	19.8	14.0
1h	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	1.1
	2s (%)	14.4	8.7	10.4	22.6
	3s (%)	67.8	71.3	72.9	64.5
	4s (%)	17.8	20.0	16.7	11.8
3	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	0
	2s (%)	14.4	7.0	18.8	19.4
	3s (%)	66.7	67.0	69.8	67.7
	4s (%)	18.9	26.1	11.5	12.9
5	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	0
	2s (%)	13.3	6.1	12.5	18.3
	3s (%)	50.0	59.1	70.8	66.7
	4s (%)	36.7	34.8	16.7	15.1
6a	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	1.1
	2s (%)	13.3	14.8	9.4	21.5
	3s (%)	57.8	61.7	67.7	63.4
	4s (%)	28.9	23.5	22.9	14.0
10	Median	3.0	3.0	3.0	3.0
	1s (%)	1.1	0	0	0
	2s (%)	15.6	8.7	11.5	18.3
	3s (%)	55.6	67.0	70.8	68.8
	4s (%)	27.8	24.3	17.7	12.9
12	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	1.1
	2s (%)	16.7	9.6	14.6	22.6
	3s (%)	50.0	66.1	69.8	58.1
	4s (%)	33.3	24.3	15.6	18.3

**Single Subjects** *Comprehensive Lesson Plan Project*

tpe	Scoring	Fall 08	Spr 09	Fall 09	Spr 10
6b	Median	3.0	3.0	3.0	2.0
	1s (%)	.9	4.0	3.8	0
	2s (%)	34.3	29.3	40.6	54.4
	3s (%)	49.1	48.5	51.9	40.4
	4s (%)	15.7	18.2	3.8	5.3
7	Median	2.0	3.0	2.0	3.0
	1s (%)	0	2.0	0	0
	2s (%)	59.3	39.4	61.7	19.3
	3s (%)	37.0	43.4	33.1	71.9
	4s (%)	3.7	15.2	5.3	8.8
8	Median	2.0	3.0	3.0	3.0
	1s (%)	7.4	2.0	5.3	1.8
	2s (%)	58.3	25.3	35.3	38.6
	3s (%)	25.0	57.6	51.1	50.9
	4s (%)	9.3	15.2	8.3	8.8
9	Median	3.0	3.0	3.0	3.0
	1s (%)	.9	1.0	2.3	1.8
	2s (%)	42.6	28.3	37.6	26.3
	3s (%)	50.0	56.6	52.6	56.1
	4s (%)	6.5	14.1	7.5	15.8

## Single Subjects Site Visitation Project

tpe	Scoring	Fall 08	Spr 09	Fall 09	Spr 10
2	Median	3.0	3.0	3.0	3.0
	1s (%)	0	1.0	0	0
	2s (%)	25.9	17.7	35.8	48.3
	3s (%)	62.0	57.3	32.1	41.4
	4s (%)	12.0	24.0	32.1	10.3
4	Median	3.0	3.0	3.0	3.0
	1s (%)	.9	1.0	0	0
	2s (%)	20.4	16.7	30.6	44.8
	3s (%)	63.9	52.1	34.3	44.8
	4s (%)	14.8	30.2	35.1	10.3
5	Median	3.0	3.0	3.0	3.0
	1s (%)	.9	1.0	0	0
	2s (%)	29.6	19.8	36.6	34.5
	3s (%)	55.6	55.2	46.3	60.3
	4s (%)	13.9	24.0	17.2	5.2
11	Median	3.0	3.0	3.0	3.0
	1s (%)	.9	1.0	.7	0
	2s (%)	26.9	28.1	32.8	43.1
	3s (%)	60.2	41.7	44.8	44.8
	4s (%)	12.0	29.2	21.6	12.1
13	Median	3.0	3.0	3.0	3.0
	1s (%)	0	1.0	0	0
	2s (%)	14.8	12.5	23.9	32.8
	3s (%)	63.0	47.9	39.6	34.5
	4s (%)	22.2	38.5	36.6	32.8

Single Subjects *Teaching Sample Project*

tpe	Scoring	Fall 08	Spr 09	Fall 09	Spr 10
10,11	Median	3.0	3.0	3.0	3.0
	1s (%)	0	1.1	0	2.2
	2s (%)	8.3	13.3	6.0	14.6
	3s (%)	73.8	70.0	83.1	70.1
	4s (%)	17.9	15.6	10.8	13.1
1,9	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	1.5
	2s (%)	9.5	28.9	20.45	19.7
	3s (%)	77.4	61.1	68.7	63.5
	4s (%)	13.1	10.0	10.8	15.3
2,3	Median	3.0	3.0	3.0	3.0
	1s (%)	0	2.2	0	.7
	2s (%)	17.9	32.2	19.3	24.8
	3s (%)	63.1	52.2	66.3	58.4
	4s (%)	19.0	13.3	14.5	16.1
7-9	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	.7
	2s (%)	19.0	20.0	20.5	17.5
	3s (%)	50.0	56.7	62.7	57.7
	4s (%)	31.0	23.3	16.9	24.1
2-4,9	Median	3.0	3.0	3.0	3.0
	1s (%)	0	2.2	0	0*
	2s (%)	11.9	21.1	18.1	25.5*
	3s (%)	67.9	62.2	59.0	60.6*
	4s (%)	20.2	14.4	22.9	13.1*
3,12	Median	3.0	3.0	3.0	3.0
	1s (%)	0	2.2	0	3.6
	2s (%)	17.9	20.0	22.9	21.69
	3s (%)	60.7	57.8	59.0	60.6
	4s (%)	21.4	20.0	18.1	13.9
12,13	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	2.2
	2s (%)	13.1	23.3	15.7	20.4
	3s (%)	63.1	53.3	65.1	61.3
	4s (%)	23.8	23.3	19.3	16.1

\*These do not sum to 100 because of an invalid score.

Single Subjects *Holistic Proficiency Project*

tpe	Scoring	Fall 08	Spr 09	Fall 09	Spr 10
1	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	2.2
	2s (%)	14.3	20.2	17.2	25.2
	3s (%)	66.7	73.0	65.5	63.0
	4s (%)	19.0	6.7	17.2	9.6
3	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	.7
	2s (%)	25.0	29.2	23.0	29.6
	3s (%)	66.7	59.6	60.9	56.3
	4s (%)	8.3	11.2	16.1	13.3
5	Median	3.0	3.0	3.0	3.0
	1s (%)	1.2	0	0	.7
	2s (%)	20.2	19.1	25.3	29.6
	3s (%)	51.2	67.4	56.3	54.8
	4s (%)	27.4	13.5	18.4	14.8
6c	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	.7
	2s (%)	23.8	20.2	19.5	26.7
	3s (%)	65.5	65.2	57.5	55.6
	4s (%)	10.7	14.6	23.0	17.0
10	Median	3.0	3.0	3.0	3.0
	1s (%)	1.2	1.1	0	2.2
	2s (%)	14.3	24.7	24.1	31.9
	3s (%)	63.1	67.4	63.2	49.6
	4s (%)	21.4	6.7	12.6	16.3
12	Median	3.0	3.0	3.0	3.0
	1s (%)	0	0	0	6.7
	2s (%)	19.0	27.0	19.5	27.4
	3s (%)	56.0	69.7	62.1	47.4
	4s (%)	25.0	3.4	18.4	18.5