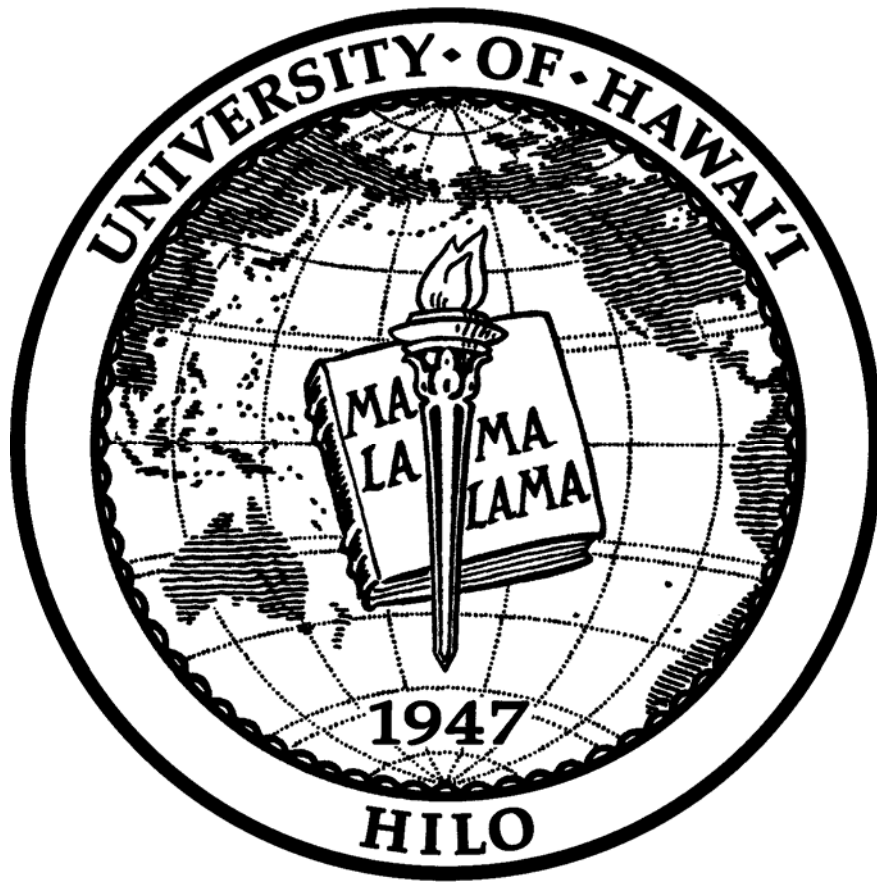


UH Hilo Writing Program Assessment

Proposal # 1



Summer, 2007

UH Hilo Office of Institutional Research

Writing Assessment

Sample Preparation

This instructional assessment project utilized final papers submitted by students enrolled in ENG 100, ENG 100T, or ESL 100 to determine the extent to which course outcomes/proficiencies were met. At semester's end, the writing samples were collected from the respective instructors by a representative of the English department and brought to the Institutional Research Office for randomized sorting prior to the project getting underway.

Writing samples were collected for one hundred forty out of one hundred eighty seven students enrolled in one or another of the identified course/sections at census date (5th week of classes). Samples were randomized and assigned a unique identifier by institutional research office, then returned to the English department representative for use in the departmental "norming sessions" and subsequent "team reading sessions" stipulated in the assessment proposal.

While research design called for all samples to be anonymized and blind-evaluated, the unique identifier assigned to each student paper is meant to support this research effort by enabling us to tie the individual student essay and reader's assessments of those back to instructor, course enrollment, course outcomes, or other institutionally stored data such as class standing, HS GPA, or SAT/ACT scores.

Assessment of Samples

Student papers used in the "norming session" were to be read by as many as 5-7 individuals and evaluated for the proficiencies demonstrated on 6 separate attributes. The intent of the norming session was to assist readers in collectively developing clear objective evaluative standards for scoring proficiency levels on a 5 point scale, 0 to 4. Student papers used in the project "team readings" were each to be read by at least two individuals, with a third reader being called upon when only discrepancies between reader evaluations on a given rubric were encountered.

In all, of the 140 writing samples collected and randomized...4 papers were discarded because of missing pages etc, 6 papers were used in the norming sessions, 110 papers were evaluated by 2 individuals, and 20 papers required a 3rd reader. Analysis presented in the following pages incorporates all of the reader evaluations for all 136 writing samples utilized.

Institutional Research Office received nine separate data files that resulted from the round robin reading sessions. These were compiled into a single base file of 318 rows, and 8 columns (reader ID, written sample ID, and the 6 separate rubric element scores), with each row representing one of 318 rendered readings of the 136 papers (2.33 average readings per). The compiled reader's sample assigned rubric scores were then averaged, the assigned rubric scores were then summed in a new column, the number of valid averaged rubric scores available for each sample was determined and stored in a new column, and an overall mean score for each writing sample was then computed.

It deserves mention here that one class-set of instructor final paper assignments had not apparently been required to demonstrate proficiencies with respect to the documentation rubric. Readers agreed to assign an "NA" to that rubric for those instructor's papers, but to proceed with scoring on the "sources" rubric anyway. This procedure was carried out by and large, but in a few instances the reader scored both "documentation" and "sources" NA. After some consideration of how to remedy or account for these scoring errors, the actual effect upon the analysis was determined inconsequential given the planned methods of analysis and inference.

Basic Research Questions & Data Processing Methods

At the most basic level, this study must attend to two separate considerations. First, we must evaluate the extent to which the readers consistently & objectively evaluate specific attributes of the same written sample (and by extension, the extent to which their average evaluation for each attribute across all samples is consistent with other reader's average attribute-specific evaluations). Fortunately, the use of fine-tuned and clearly expressed rubrics as well as the consensus/3rd party intervention methods undertaken by the department in both the norming and team reading sessions partly alleviate some of these reader validation concerns.

The following provide evidence for ascertaining reader validity:

- Reader's assigned item specific and overall mean scores compared among other reader's + Standard Deviation
- Reader's assigned item specific and overall mean scores compared on the same writing sample + Standard Deviation.

Secondly, the study seeks to ascertain the extent to which the student writing samples reflect the proficiency levels appropriate to college level writing (ENG 100, 100T, ESL 100) course outcomes.

The following criteria are to be used for ascertaining individual and collective proficiency levels:

- Individual student's item specific and overall mean scores evaluated relative to a 2.0 mean score.
- Overall item specific and aggregate mean scores evaluated relative to a 2.0 mean score.

Beyond Basics

A more loosely articulated set of research possibilities unfold as we begin to look across the writing samples for variables that may be of significance in better understanding the basic results. A variety of table and chart formats are accordingly explored for depicting the different ways of viewing results.

Results

Reader Validation

First, to demonstrate that readers had assigned to them a reasonable distribution of writing samples that came from all courses and Instructors, the following displays for each reader the distribution of original student course and instructors beneath the samples read by each.

Table 1

SUBJECT	COURSE_NUMBER	INSTRUCTOR_NAME	Elisabeth	Kristen	Laurie	Luke	Mark	Seri	Susan
ENG	100	Erker, Elisabeth D.	4	4	6	5	4	3	4
ENG	100	Luangphinith, Seri I.	2	2	2	5	6	4	4
ENG	100	Wackerbarth, Susan	5	7	9	10	9	7	7
ENG	100T	Hayashi, Karla S.	8	5	6	2	3	2	1
ENG	100T	Mollegaard, Kirsten	4	2	4	3	4	3	3
ENG	100T	Panek, Mark T.	3	5	2	4	3	1	3
ENG	100T	Sagle, Lauri R.	11	7	11	6	9	15	13
ESL	100	Bailey, Luke	2	8	4	7	7	3	2
ESL	100	Saft, Scott L.	4	5	4	5	3	5	5
Total Readings			43	46	48	48	48	43	42

For each reader involved in the assessment project, the number of papers read and the respective rubric mean scores are presented below.

Rubric 1 - Line of Thought, Reasoning

Rubric 2 - Analysis

Rubric 3 - Essay Form and Structure

Rubric 4 - Grammar, Style, Etc

Rubric 5 - Document Conventions

Rubric 6 - Source Use and Integration

Table 2

Reader	Samples Read	Rubric 1	Rubric 2	Rubric 3	Rubric 4	Rubric 5	Rubric 6	Rubric Mean Total	Overall Rubric Mean
Elisabeth	43	2.47	2.33	2.44	2.33	1.66	1.72	12.93	2.16
Kristen	46	2.33	2.52	2.22	2.22	1.15	1.72	12.15	2.03
Laurie	48	2.50	2.44	2.60	2.42	2.22	2.18	14.35	2.39
Luke	48	2.08	2.23	1.96	2.56	1.60	2.10	12.53	2.09
Mark	48	2.17	2.06	2.15	2.79	2.44	1.96	13.57	2.26
Seri	43	1.74	2.02	1.58	2.00	1.52	1.93	10.80	1.80
Susan	42	2.36	2.24	2.19	2.48	1.83	2.05	13.14	2.19

A measure of variability among the reader rubric mean scores overall is presented below. Overall among the readers generally, the scores they assigned were most variant (widely distributed) on rubric 5, documentation conventions. Variability was least on rubrics 2 and 6, and in the overall rubric mean (the average of all 5 or 6 rubric scores assigned on all papers read).

Table 3

StDevOf Rubric 1	StDevOf Rubric 2	StDevOf Rubric 3	StDevOf Rubric 4	StDevOf Rubric 5	StDevOf Rubric 6	StDevOf Rubric Mean Total	StDevOf Overall Rubric Mean
0.26	0.18	0.33	0.25	0.44	0.18	1.13	0.19

A more rigorous test of reader consistency is in evaluating the average differences between the reader's assigned rubric scores on the same writing sample. To do this, results are first grouped by the sample ID # then the standard deviation between assigned scores on a given sample is computed

for all 136 writing samples. The average of standard deviations among grouped samples is displayed below.

Table 4

Avg StnDev Rubric 1	Avg StnDev Rubric 2	Avg StnDev Rubric 3	Avg StnDev Rubric 4	Avg StnDev Rubric 5	Avg StnDev Rubric 6
0.53	0.55	0.55	0.49	0.65	0.63

The data underlying these averages, the numeric and percent frequency of standard deviations by rubric is presented below.

Table 5

StDev Rubric 1	# Sample	% Samples		StDev Rubric 4	# Sample	% Samples	
0	52	38.2%		0	56	41.2%	
0.41	1	0.7%		0.41	1	0.7%	
0.49	2	1.5%		0.53	1	0.7%	
0.52	1	0.7%		0.55	1	0.7%	
0.53	1	0.7%		0.58	10	7.4%	
0.58	6	4.4%		0.71	45	33.1%	
0.71	49	36.0%		0.79	1	0.7%	
1	7	5.1%	17.60%	0.82	1	0.7%	
1.15	2	1.5%		1	3	2.2%	14.70%
1.34	1	0.7%		1.15	3	2.2%	
1.41	10	7.4%		1.41	13	9.6%	
1.53	4	2.9%		1.53	1	0.7%	
StDev Rubric 2	# Sample	% Samples		StDev Rubric 5	# Sample	% Samples	
0	49	36.0%			34	25.0%	
0.49	1	0.7%		0	39	28.7%	
0.52	2	1.5%		0.38	1	0.7%	
0.53	2	1.5%		0.41	1	0.7%	
0.58	6	4.4%		0.53	1	0.7%	
0.71	50	36.8%		0.58	2	1.5%	
0.89	1	0.7%		0.71	31	22.8%	
1	5	3.7%	18.40%	1	1	0.7%	19.90%
1.15	7	5.1%		1.15	3	2.2%	
1.41	11	8.1%		1.34	1	0.7%	
1.53	1	0.7%		1.41	9	6.6%	
1.73	1	0.7%		1.53	5	3.7%	
StDev Rubric 3	# Sample	% Samples		1.73	3	2.2%	
0	45	33.1%		2.08	1	0.7%	
0.38	1	0.7%		2.12	4	2.9%	
0.41	1	0.7%		StDev Rubric 6	# Sample	% Samples	
0.49	1	0.7%			12	8.8%	
0.52	1	0.7%		0	43	31.6%	
0.58	9	6.6%		0.41	1	0.7%	
0.71	57	41.9%		0.49	3	2.2%	
1	6	4.4%	15.40%	0.55	1	0.7%	
1.1	1	0.7%		0.71	52	38.2%	
1.15	3	2.2%		1	2	1.5%	17.60%
1.41	8	5.9%		1.15	2	1.5%	
1.53	2	1.5%		1.41	11	8.1%	
2.12	1	0.7%		1.53	2	1.5%	
				2.12	6	4.4%	
				2.83	1	0.7%	

Where the standard deviation is 0, readers were in absolute agreement in scoring on that rubric; where the standard deviation is .71, typically two readers have scores on that rubric differing by 1 increment.

Occurrences of standard deviations equal or greater than 1 are highlighted in bold borders in the display below and their % frequency aggregated for ease in comparison.

It can be inferred from this that there was the least variability overall in scoring on rubric 4 (highest % with stnd dev "0", and smallest % with stnd dev >1)

Isolating Variability Among Readers

The average standard deviations among grouped norming session samples versus samples read by 2 or 3 readers is displayed below, These affirm that the least same sample same item scoring variability was found among the largest share of the overall sample. One might have expected the norming session to display the greatest variability in this respect (since the scoring rules were just being negotiated and assimilated), despite the consensus/3rd party intervention methods undertaken by the department in both the norming and team reading sessions. The fact that those standard deviations are less than those found in the samples read by 3 individuals implies that overall variability in reader's scoring is not attributable to uncertain use of the rubrics so much as to peculiar traits of those samples that were kicked out for a third reading.

Overall, there were 71 writing samples with one or more instance of a standard deviation greater than 1.0. Further inquiry could seek to characterize those samples, and determine the cause for reader disagreement being in rubric clarity, or rubric attainment or both.

Table 6

2 readers

110 essays

AvgOfStDevOfLine of Thought - Reasoning	AvgOfStDevOfAnalysis	AvgOfStDevOfEssay Form and Structure	AvgOfStDevO
0.44	0.46	0.49	

Table 7

3 readers

20 essays

AvgOfStDevOfLine of Thought - Reasoning	AvgOfStDevOfAnalysis	AvgOfStDevOfEssay Form and Structure	AvgOfStDevO
0.94	0.99	0.88	

Table 8

5-7 readers Norming Session

6 essays

AvgOfStDevOfLine of Thought - Reasoning	AvgOfStDevOfAnalysis	AvgOfStDevOfEssay Form and Structure	AvgOfStDevO
0.63	0.58	0.48	

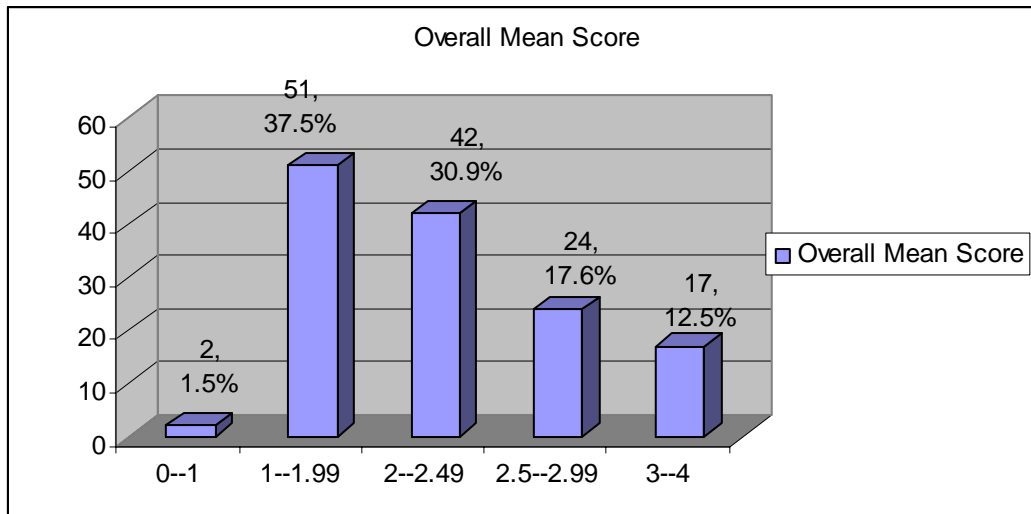
The average of Student Rubric Scores 1 through 6, the average rubric total, and overall rubric average are depicted below. It's evident in Table 9 that the minimum course proficiency of 2.0 is met on 4 of the 6 rubrics, as well as on the overall rubric mean score.

Table 9

Rubric 1 Avg.	Rubric 2 Avg.	Rubric 3 Avg.	Rubric 4 Avg.	Rubric 5 Avg.	Rubric 6 Avg.	Avg. Rubric Total	Avg. Overall
2.23	2.26	2.17	2.38	1.79	1.99	12.36	

The distribution of overall mean scores is displayed below; those below the 2.0 level represent 39% of the total 136 writing samples examined. Scores above 2.0 were 61% of total.

Chart 1

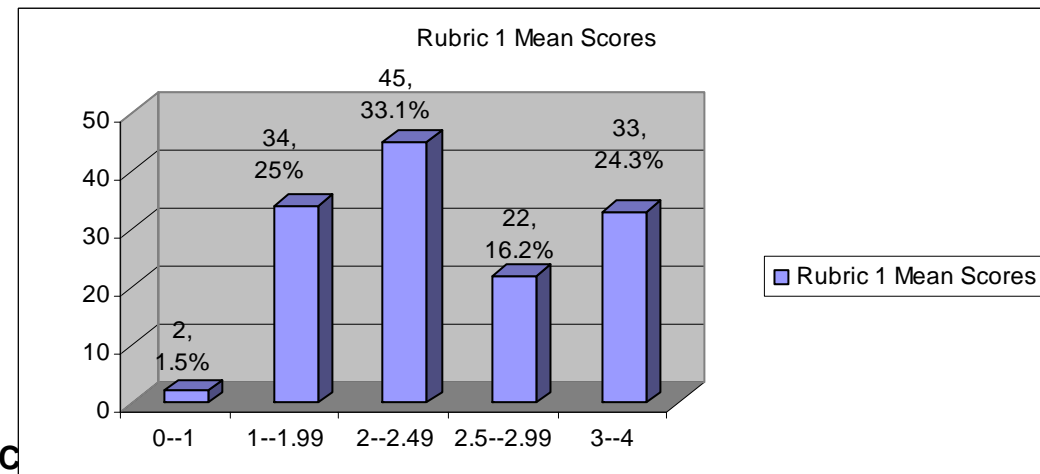


The distribution of overall mean scores for rubrics 1-6 are displayed by range below.

Chart 2 below displays 73.6% of students above the 2.0 mark on rubric 1; charts 3-7 that follow display 72.7% meeting course proficiency on rubric 2, 69.1% on rubric 3, 78.7% on rubric 4, 37.5% on rubric 5, and 61.1% on rubric 6.

Interesting to note on rubric 4, where the lowest variability noted in reader analysis is now seen to correlate with largest share of students scoring in the 3-4 range. Rubric 5 is unusual in having its largest share of students (34%) missing the data altogether, in addition to having an equal number/percent (16.2%) scoring in the top range as the bottom.

Chart 2



C

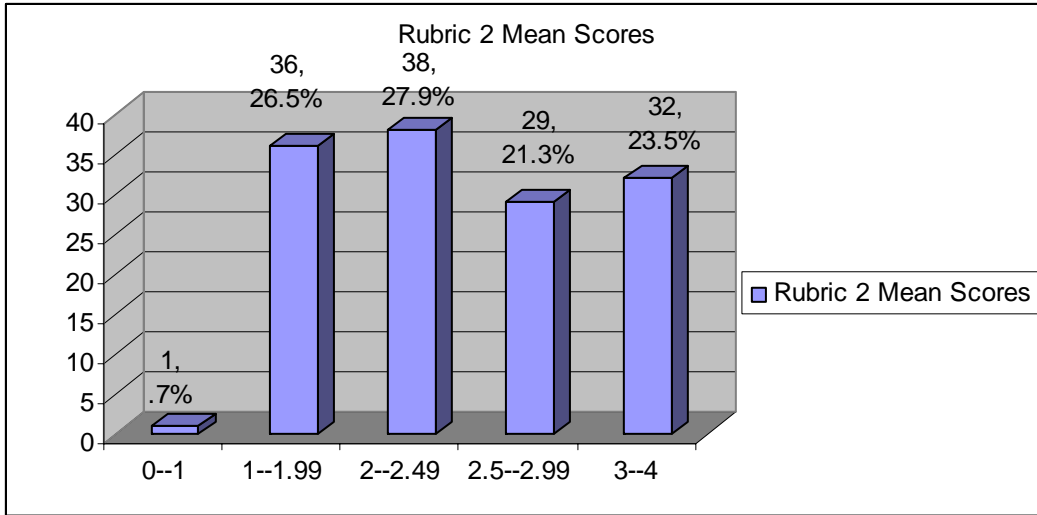


Chart 4

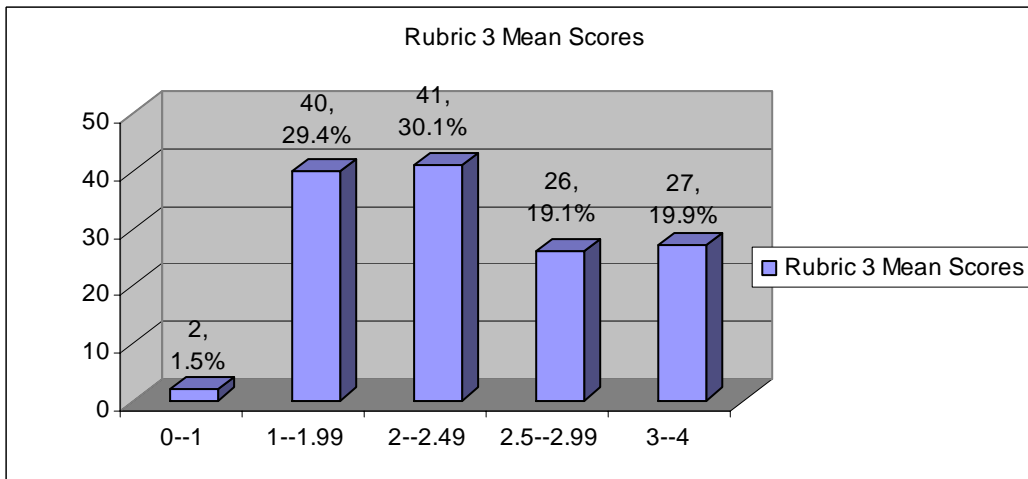
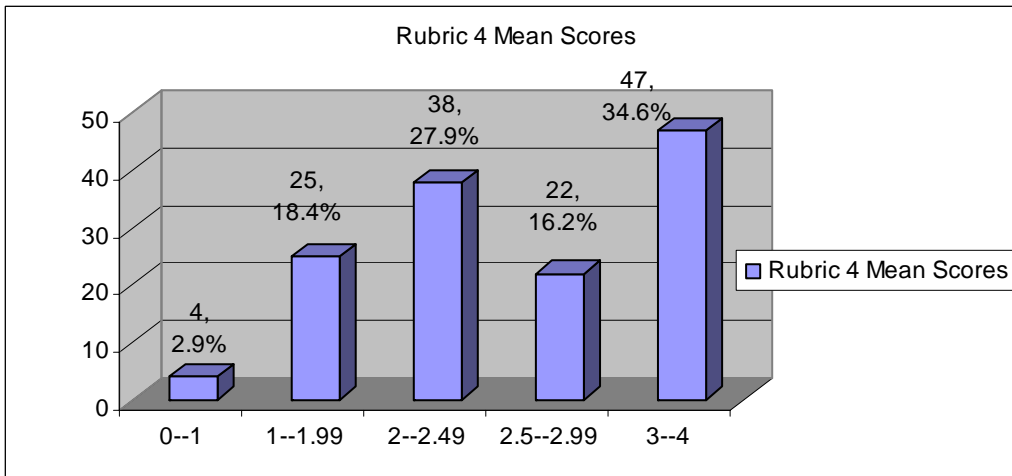


Chart 5



These widely varied rubric score distributions may have curriculum/pedagogical implications.

While missing data are evident in the mean score range distributions below for rubric 5 and 6, it is again clarified that the occasional presence of just 5 scores had been accounted for in calculating the individual's overall mean score presented in chart 1.

Chart 6

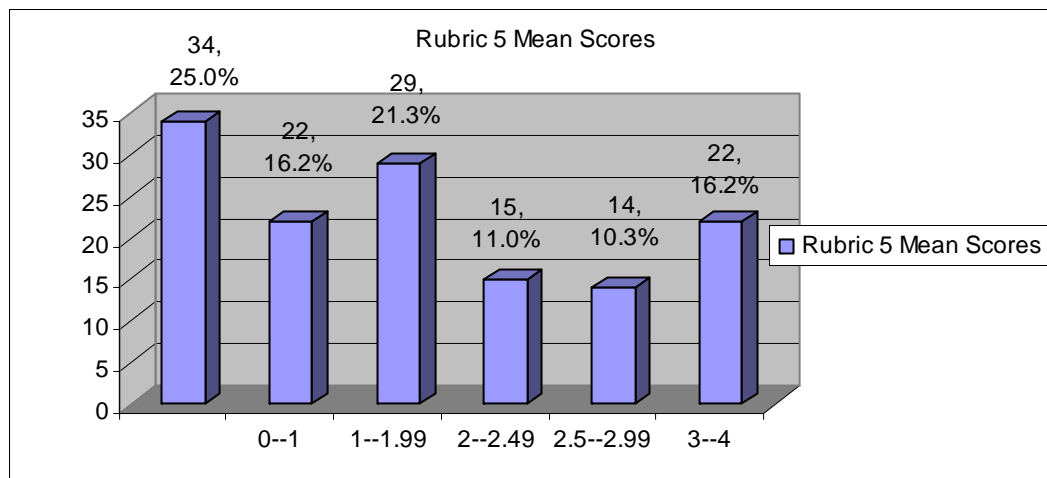
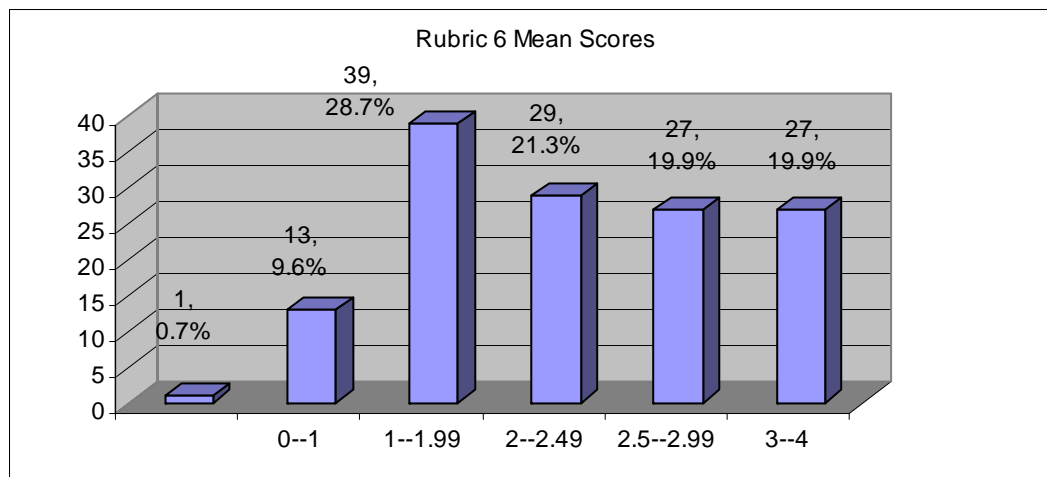


Chart 7



Charts 8 and 9 below explore the relationship between student’s overall rubric score range and student cumulative SAT and HS GPA averages. SAT scores can be seen as having a somewhat less linear correlation with performance on overall rubric score range than is evident for HS GPA. Of course, these are both pre-collegiate indicators of academic performance—and so their reference offers no real insight to inform teaching and learning.

Tables 10 through 17 offer a comparative look at the proportion of 136 students in the total sample versus those 41 students with overall mean score ranges 2.5—2.99 and 3—4 by course enrollment, instructor, student admit type, student high school type, student educational level, student residency, student ethnicity, and student gender. Summarizing those tables, the following are in various degrees proportionately over-represented among top performers: females (by 2%), Caucasians (by 11%), non-residents (by 4%), sophomores (by 8%), foreign HS graduates and US GED holders (by 4% and 2% respectively), transfer students and post-bachelors degree students (by 3% and 2% respectively), students enrolled with Wackerbarth (by 12%), and students enrolled in ENG 100 (by 12%)

Chart 8

SAT Mean Scores by WRT Assessment Overall Range

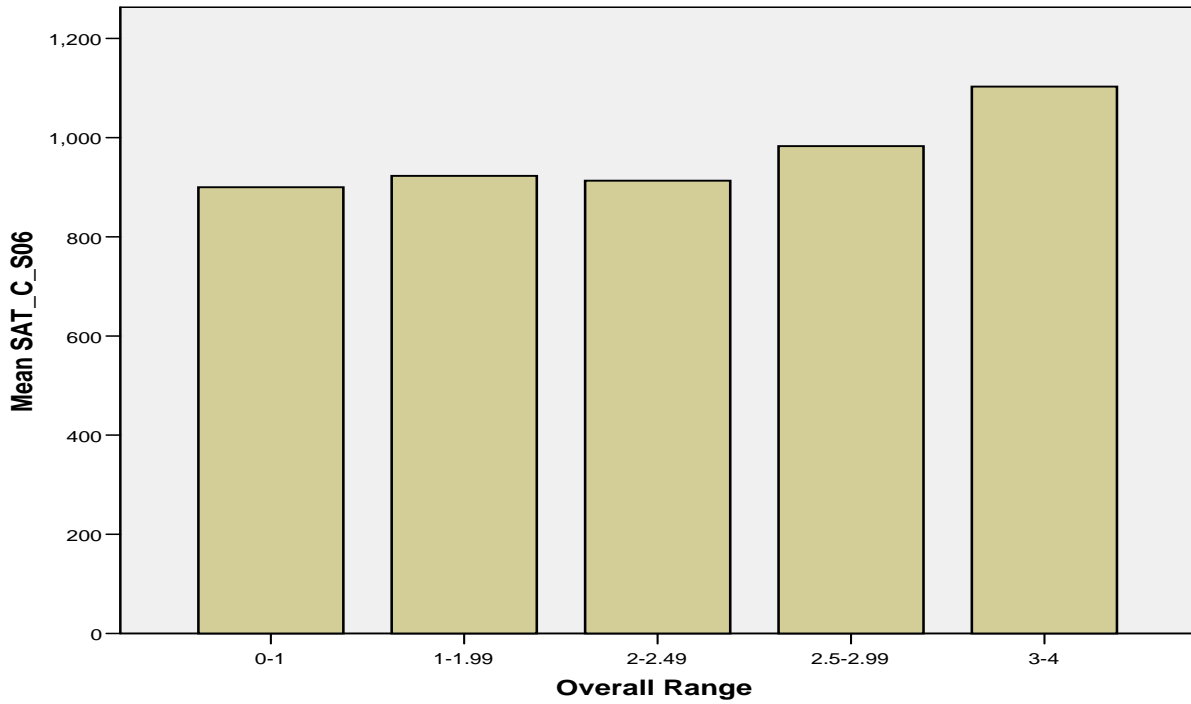


Chart 9 HS GPA Mean Scores by WRT Assessment Overall Range

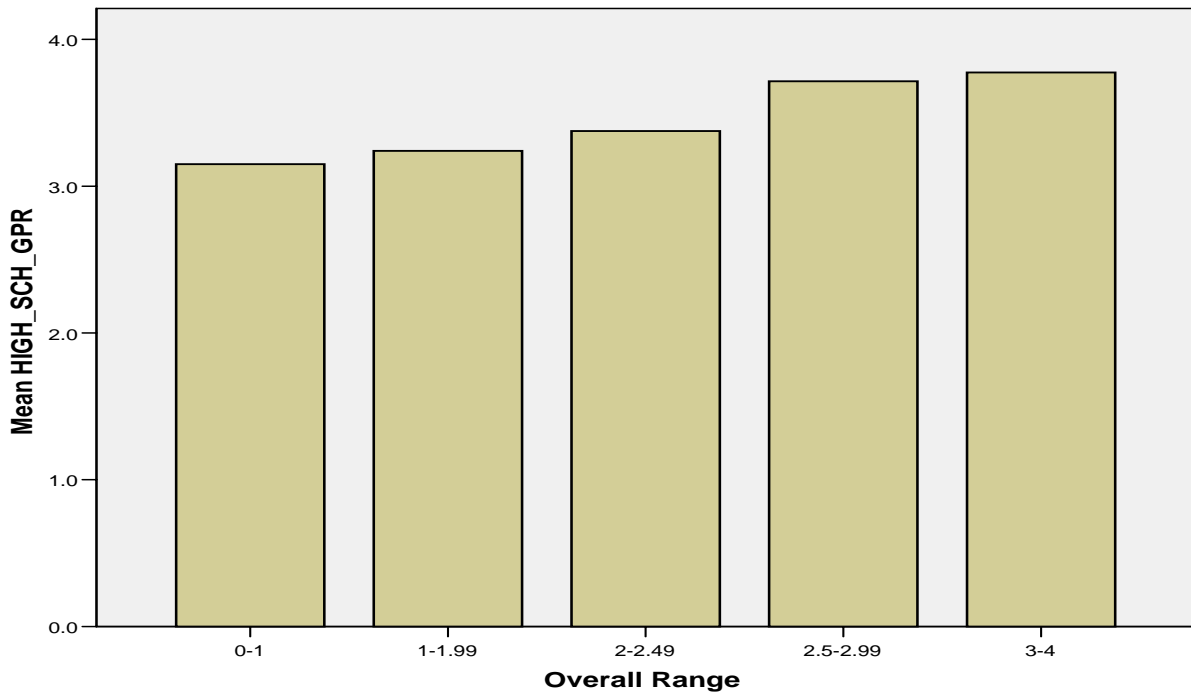


Table 10

Overall and Top Performing Samples by Course Enrolled

(Top performing =overall mean >=2.5)

Subj	Crse	# All	% All	# Top	% Top
ENG	100	46	33.82%	19	46.34%
ENG	100T	65	47.79%	17	41.46%
ESL	100	25	18.38%	5	12.20%

Table 11

Overall and Top Performing Samples by Course Instructor

(Top performing =overall mean >=2.5)

INSTRUCTOR_NAME	# All	% All	# Top	% Top
Bailey, Luke	12	8.82%	2	4.88%
Erker, Elisabeth D.	14	10.29%	3	7.32%
Hayashi, Karla S.	13	9.56%	2	4.88%
Luangphinit, Seri I.	9	6.62%	4	9.76%
Mollegaard, Kirsten	11	8.09%	3	7.32%
Panek, Mark T.	10	7.35%	3	7.32%
Saft, Scott L.	13	9.56%	3	7.32%
Sagle, Lauri R.	31	22.79%	9	21.95%
Wackerbarth, Susan	23	16.91%	12	29.27%

Table 12

Overall and Top Performing Samples by Student Admit Type

(Top performing =overall mean >=2.5)

STYP_ADM	# All	% All	# Top	% Top
Continuing	1	0.74%		
Freshman	86	63.24%	24	58.54%
Post-Bac (M)	7	5.15%	3	7.32%
Returning	3	2.21%	1	2.44%
Transfer	39	28.68%	13	31.71%

Table 13

Overall and Top Performing Samples by High School Type

(Top performing =overall mean >=2.5)

HIGH_SCH_TYPE_IRO	# All	% All	# Top	% Top
	12	8.82%		
CA	5	3.68%		
CFA	2	1.47%	1	2.44%
CO	5	3.68%	2	4.88%
Foreign HS	18	13.24%	7	17.07%
GA	1	0.74%	1	2.44%
GED Diploma	3	2.21%	2	4.88%
Hawaii Private	19	13.97%	4	9.76%
Hawaii Public	59	43.38%	18	43.90%
IA	1	0.74%	1	2.44%
ID	1	0.74%		
IL	1	0.74%	1	2.44%
LA	1	0.74%		
MN	1	0.74%	1	2.44%
NA	1	0.74%	1	2.44%
NJ	1	0.74%	1	2.44%
OR	1	0.74%		
POS	1	0.74%	1	2.44%
WA	1	0.74%		
WI	1	0.74%		
WV	1	0.74%		

Table 14

Overall and Top Performing Samples by Education Level

(Top performing =overall mean >=2.5)

EDLVL_IRO	# All	% All	# Top	% Top
	1	0.74%		
Freshman	82	60.29%	21	51.22%
Sophomore	28	20.59%	12	29.27%
Junior	14	10.29%	3	7.32%
Senior	9	6.62%	3	7.32%
Unclassified	2	1.47%	2	4.88%

Table 15

Overall and Top Performing Samples by Residency

(Top performing =overall mean >=2.5)

RESIDENCY_DESC	# All	% All	# Top	% Top
Converted Resident	3	2.21%	1	2.44%
N/R Hawaiian Exemption	1	0.74%	1	2.44%
N/R Institutional Exemption	3	2.21%	1	2.44%
N/R WUE Exemption	15	11.03%	3	7.32%
Non-Resident	28	20.59%	10	24.39%
Resident	86	63.24%	25	60.98%

Table 16

Overall and Top Performing Samples by Ethnicity

(Top performing =overall mean >=2.5)

ETHNICITY_CATEGORY	# All	% All	# Top	% Top
Native American	1	0.74%		
Asian/Pacific Islander	80	58.82%	21	51.22%
Hispanic	3	2.21%		
Caucasian	35	25.74%	15	36.59%
Other/Mixed	17	12.50%	5	12.20%

Table 17

Overall and Top Performing Samples by Gender

(Top performing =overall mean >=2.5)

GENDER	# All	% All	# Top	% Top
F	86	63.24%	27	65.85%
M	50	36.76%	14	34.15%