UTAH VALLEY STATE COLLEGE

DEPARTMENT OF ENGLISH AND LITERATURE

Report 01-03

COMPOSITION PROGRAM ASSESSMENT REPORT

SPRING 2004

Introduction 1
Program Design 1
Database Design and Data Collection 1
Spring 2004 Procedures 2
Statistical Analysis and Report 3
Operational Results 3
Findings—Descriptive Statistics 4
Data Analysis 8
Correlation of Ratings 8
Conclusions 9
Recommendations 10
Primary Trait Matrix 2
Student Paper Cover Sheet 4
Assessor Instructions 5

Composition Program Assessment, Spring 2004

Introduction

Program Design

In August, 2003 the English and Literature Department formed a committee to design, implement and execute an assessment of the three major courses in the composition program: English 1010, 2010, and 2020. By early fall, the committee had decided to base the assessment on a holistic measure of traits: topic control, support, voice, word choice, organization, sentence fluency, conventions, and scholarship. The matrix that describes the traits and values is at Appendix 1 *(see Trait Matrix, Appendix, page A-2).*

Initially, the committee favored a collection of student papers similar to a portfolio; however, the logistics of incorporating a portfolio collection was somewhat daunting and the idea of two or more student papers from each course was abandoned.

The committee determined that the primary outcome for the assessment program in 2004 would be statistical analysis leading to recommendations or suggestions for individual instructors. Analysis of data would allow instructors to see how their evaluation of student papers varied from the department norm, and therefore, offer suggestions of improving evaluation. The data would also identify differences from department norms on individual traits for each student. A trend below or above department norms might suggest an area for more instruction or more focus.

To achieve this goal, the committee decided on a two-step assessment procedure. First, each classroom instructor would assign a paper from guidelines set by the committee. At semester's end, instructors would collect and evaluate 100 percent of assigned student papers, referencing the evaluation matrix mentioned above. The second evaluation would be conducted by someone other than the instructor of record and would also reference the matrix.

Deadlines for instructors to collect and conduct the initial evaluation were set as well as the date for the assessment evaluation.

Database Design and Data Collection

The committee felt that an electronic data collection would be the most efficient and to that end, we designed a SQL relational database. We also decided that cross-platform considerations would suggest a web-based data collection protocol. The database operates on Microsoft SQL Server 2003, the web services are authored in C#, and the administrator's application is windows-based and was authored in C++.

The database uses three primary tables: students, courses, and users (faculty). The student table stores student identification, course and section number, instructor code, evaluator code, and the raw scores for the traits from both the instructor's evaluation and the assessors evaluation. The courses table includes the assigned instructor, course and section number, and term offered. The users table includes first and last name, identification number, faculty status, assessor assignment.

The web service provides the interface between the database and the teachers and assessors. Users may login as teachers and conduct the initial evaluation of their classes. The system queries the database and extracts class rosters for the specific user. When a user logs in as an assessor, the system queries the student table in the for assigned papers. The method of assigning papers to an assessor is discussed later.

The administrator's application provides three main functions: user maintenance, sample sizing and paper assignment, and status. The administrator uses the first screen to add or delete faculty and to set their type (contract, or adjunct or concurrent). The second screen provides an input section that the administrator can use to select the number of papers to assess for each course type and to adjust the total number of papers each assessor will eventually evaluate. The third screen monitors the status of the ongoing assessment effort. Display grids show the administrator how many papers each instructor of evaluator has completed as well as an overall percentage of the total records finished.

Spring 2004 Procedures

In March, the system administrator sent a request to Information Services for a SIS data pull. The request asked for two electronic files. The first file provided comma delimited course number, section number, term, and instructor name. The second file presented student first and last name, identification number, course, section number, and term. After writing the appropriate SQL import statement the files were appended to the courses and student tables in the database.

The administrator groomed the data for three days. The primary concern after importing the data was to properly index the two main tables so the instructor of record name and identification numbers were accurately available in the student table.

After cross referencing the user table, courses table, and student table to ensure accuracy, we opened the web service to test several fictitious teachers and assessors. From the first request to the SIS programmers until we were ready to begin initial evaluation was approximately one week.

In preliminary testing with the Assessment Committee, we learned that conducting an assessment on a paper without the benefit of the assignment prompt was exceptionally difficult. As a result, we asked all of the instructors to prepare a written prompt. The prompt was added to the database in the student table for later use with cover sheets.

Preparing student paper cover sheets was our next task, *(see Paper Cover Sheets, Appendix, page A-4)*. Using MS Word and a SQL query on the student table, we were able to generate a mail merge that printed the student id, course number and section, and the prompt.

The mail merge produced 3752 cover sheets, sorted by course and section. We also prepared individualized instructions for entering data as well as a copy of the evaluation matrix. Sorting, printing, and preparing envelopes for this semester's program consumed about 9 hours over two days.

The last phase of the data collection effort began on April 27th as we asked the administrator's application to assign assessors to evaluated papers. First, the application

randomly picks a number between 1 and the total number of students in all three courses (3752 in this case). Then it selects the first faculty member alphabetically. The the student record selected has been evaluated and the instructor id is not the same as the selected faculty member, the paper is assigned to that faculty member. This process is repeated until the first faculty member has been assigned the correct number of papers to evaluate. When satisfied, the application selects the next faculty member and the process is repeated again. The process ensures an assessor cannot evaluate any of his own students' papers. The entire sorting process takes the system approximately 3 minutes.

After sorting, we run another mail merge. This merge prepares instructions and cover sheets for each assessor, ***(see Assessor Instruction, Appendix, page A-5)***.

This year the department secretary collected evaluated papers from the instructors. As the papers came in, she put our pre-printed divider sheets that identified each section in the folders with the papers. When we were ready to prepare the assessors' packages, we used the identification numbers found on the paper cover sheets to fill the list of papers to be assessed on the assessor's instruction sheets. The sorting process took about 6 hours.

After four days of assessing, the server was attacked by a computer hacker and we had to suspend operations while we rebuilt the server. We had no additional assessment completions after the database was restored.

Statistical Analysis and Report

The SQL database makes direct queries from Microsoft Excel a simple and powerful matter. Furthermore, it permits us to use the statistical tools found in Excel to facilitate the analysis. Complete discussion of the statistical tools and the resulting inferences is in a later section of this report.

Operational Results

Initial Evaluation

Though our expectations for participation might have been excessively high, we might have predicted about a 70 percent completion rate from about 90 percent of our teachers on the initial evaluation of student papers. Table 1 depicts our actual participation

Table 1: Initial Evaluation Participation

Course	Students Enrolled	Papers Evaluated	Percentage
ENGL 1010	1578	925	58.6
ENGL 2010	1660	1013	61.0
ENGL 2020	514	269	52.3

Table 2 shows the percentage of instructors that completed at least 70% of their assigned papers. You might note that 40 of 58 adjuncts completed the majority of

their assigned papers, while 20 of 24 full-time instructors completed at least 70% of

Table 2: Number Completing Initial Evaluation by Type Instructor

	Papers Completed versus assigned		
	More than 70%	Less than 50%	0%
Adjunct	40	1	17
Contract	20	3	1

assigned papers. Teaching assignments placed a rather wide range of demands on our instructors. For instance, one adjunct completed 119 of 131 papers; another finished 0 of 3. The data suggests that any participation most often resulted in near completion.

Assessment

The first indication that we might not enjoy enthusiastic participation surfaced when we asked for prompts. Instructors submitting prompts electronically were returned cover sheets requiring only that they pass them to the appropriate student. This action saved countless hours for instructors willing to take a few minutes and email a prompt. *Nonetheless, only 37 of the 190 sections (19%) offered this semester had prompts stored in the database.* Only four additional sections had prompts attached to the cover sheets by instructors. Unfortunately, only *186 of the 865 papers assessed (21.5%) had prompts.* Consequently, our assessors' task was made substantially more difficult. *For a sample prompt, see Paper Cover Sheets, page A-4.*

The administrator determined that we required 319 papers from 1010; about 325 from 2010 and about 240 from 2020. He also assigned 85 instructors assessor duties. To avoid difficulties with assigning papers to assessors we rounded the sample size to 340, 340, and 255 respectively. Thus, each instructor would assess 11 papers. Table 3 details our participation success with the assessment phase.

Table 3: Assessment Completion by Type Course

	1010	2010	2020
Total Assigned	340	340	255
Papers Unavailable	111 of all types lost		
Total Completed	236	238	159

Eleven instructors did not complete any assessment (121 papers); papers lost their cover sheet or were somehow not available in 111 cases. We assessed 865 of 935 or 92%. Statistically, we have a good sample and should be able to draw reasonable inferences from the data.

Findings—Descriptive Statistics

Descriptive statistics provide explanatory insight into your dataset. Primarily, we use descriptive statistics to tell us about the data: it's characteristics, size, distribution of

data, how many datapoints lie where and so on. Later, we'll use inferential statistics to make value statements concerning the relationships between data and causal factors. The next four tables present a variety of descriptive datapoints

Aggregate Scores

During design phase, the committee chose to weight the traits:

Table 4: Trait Weights

Trait	Weight
Control of topic	3
Support	3
Conventions	2
Organization	2
Scholarship	2
Word choice	1
Sentence fluency	1
Voice	1

The weighting factors and decision to divide the measurement into 6 categories resulted in a possible aggregate score of 100. The following table presents the weighted aggregate scores for all papers evaluated and assessed.

TEVAL = Weighted score assigned by instructor

AEVAL = Weighted score assigned by instructor

Table 5: Weighted Aggregate Scores

	TEVAL			AEVAL		
Course	Average	Median	Std Dev	Average	Median	Std Dev
1010	63.66	63	15.52	58.5	57	15.23
2010	74.70	75	15.87	62.26	63	14.83
2020	71.34	72	15.89	68.34	70	14.71

Correlation scores between instructor of record and the assessors were

0.299 for Engl 1010 with 235 sampled,

0.248 for Engl 2010 with 237 sampled, and

0.068 for Engl 2020 with 158 sampled.

The relationship between instructor rating and assessor rating is proved in 1010 and 2010 and rejected in 2020. This significant finding will be discussed later *(see "Correlation of Ratings" on page 8)*

Trait Scores

We'll begin the presentation of trait data with English 1010. Tables 6 through 8 present the data for all three courses:

Table 6: English 1010

	Teachers			Assessors		
	Average	Median	StdDev	Average	Median	StdDev
Topic	3.924	4.000	1.020	3.640	4.000	1.112
Support	3.919	4.000	1.110	3.568	4.000	1.137
Organization	3.835	4.000	1.108	3.653	4.000	1.075
Conventions	3.792	4.000	1.089	3.542	4.000	1.131
Voice	3.979	4.000	1.078	3.496	4.000	1.361
Word Choice	3.886	4.000	1.010	3.547	4.000	1.041
Sentence Fluency	3.826	4.000	1.083	3.597	4.000	1.124
Scholarship	3.614	4.000	1.275	3.242	3.000	1.240

Table 7: English 2010

	Teachers			Assessors		
	Average	Median	StdDev	Average	Median	StdDev
Topic	4.752	5.000	0.969	3.874	4.000	1.125
Support	4.592	5.000	1.058	3.895	4.000	1.152
Organization	4.479	5.000	1.074	3.811	4.000	1.024
Conventions	4.303	4.000	1.162	3.622	4.000	1.079
Voice	4.433	5.000	1.233	3.702	4.000	1.312
Word Choice	4.399	4.000	1.138	3.803	4.000	1.006
Sentence Fluency	4.311	4.000	1.127	3.790	4.000	1.105
Scholarship	4.450	5.000	1.268	3.483	4.000	1.238

Table 8: English 2020

	Teachers			Assessors		
	Average	Median	StdDev	Average	Median	StdDev
Topic	4.415	5.000	1.069	4.101	4.000	1.080
Support	4.472	5.000	1.135	4.264	4.000	1.034
Organization	4.365	5.000	1.122	4.258	4.000	1.092
Conventions	4.038	4.000	1.232	4.019	4.000	1.094
Voice	4.258	4.000	1.057	3.818	4.000	1.405
Word Choice	4.138	4.000	1.133	4.126	4.000	0.891
Sentence Fluency	4.208	4.000	1.126	4.119	4.000	0.944
Scholarship	4.252	4.000	1.169	4.082	4.000	1.253

Comparative Data: Adjunct versus Contract faculty

Adjunct faculty taught 133 sections with 2568 students. They were able to evaluate 1549 papers, and 634 papers were subsequently assessed. Contract faculty taught 47 sections with 940 students. They were able to evaluate 660 papers, and 164 papers were subsequently assessed. The tables 9 & 10 break down our assessment efforts according to the type of instructor.

Table 9: Adjunct Faculty Results

	Evaluation Results			Assessment Results		
	Average	Median	Std Dev	Average	Median	Std Dev
Engl 1010	64.62	62	14.25	57.19	56	17.67
Engl 2010	75.05	76	16.69	60.89	65	20.32
Engl 2020	72.09	72	15.72	63.59	66	21.5

Table 10: Contract Faculty Results

	Evaluation Results			Assessment Results		
	Average	Median	Std Dev	Average	Median	Std Dev
Engl 1010	62.64	63	16.87	56.54	57	16.71
Engl 2010	75.74	75	13.52	58.45	57	14.45
Engl 2020	67.63	70	15.11	73.21	72	9.32

DATA ANALYSIS

This report relies on the statistical tools found in Microsoft Excel. Standard Deviations, Means, Medians, Correlations, t-tests, and other values were derived from importing database information into Excel, then using the statistical tools Excel to compute the values.

Correlation of Ratings

Instructor Versus Assessor

The data supports the claim that assessors evaluated all papers somewhat more sternly than did the instructors.

Adjunct versus Contract Faculty

Adjunct: Assessors agreed with adjunct instructors concerning their 1010 papers. Though assessors were consistently more critical (as noted above), the differences between instructor and assessor were not statistically significant. The assessors' larger standard deviation for 1010 suggests a larger range of scores, but since the median and average have little difference, we can hypothesize the difference is due to the more critical view of assessors in general.

2010 results were more telling. Assessors found 2010 papers significantly poorer than instructors did. The standard deviation was more than 20 and the median was 11 points lower than the instructors'. Although both groups found 2010 better than 1010, the difference between the groups suggests much less agreement between initial and assessed evaluation. Since 2010 had the fewest number of prompts, the variety in assessor expectation may be the primary causal factor.

Contract: Contract faculty also graded their papers higher than the assessors. 1010 and 2020 scores closely paralleled those of the adjuncts (though a bit lower overall). However, assessor scores for 2010 were some 18 points lower than the contract instructors, compared to 10 points lower for the adjunct/assessor difference. Clearly, contract faculty felt their 2010 students were somewhat better than the final assessment results confirm.

The data does not support any significant differences between contract and adjunct faculty.

English 1010

After arriving at standard deviations, means, and medians for the data returned for 1010, the t-test was performed to prove or disprove covariance among the various topics. Apparently, there is no statistical evidence that any value is dependent on any other value. Even though the difference between instructor and assessor rating for papers is statistically significant, there is no correlation to suggest why the difference exists.

English 1010 aggregate scores revealed a median of 63 for teachers and 57 for assessors. The standard deviation for the two groups suggests agreement in relative scoring. And even though assessors were somewhat more critical, their evaluation closely

parallels the instructor ranking. Individual value scores were significantly lower than corresponding values in 2010 and 2020.

Assessors and teachers alike ranked 1010 papers lower than the more advanced courses. This instrument was not designed to determine why that was so; nonetheless, 1010 papers appear to be less polished and successful than the more advanced papers. The variety in instruction and the absence of prompts may point to causes why 1010 papers received lower assessment results.

English 2010

English 2010 aggregate scores revealed medians of 75 for instructors and 63 for assessors. The standard deviation for instructors was 15.9, while assessor standard deviation was 14.8. Interestingly, 2010 papers appear to have received the highest evaluations, yet the spread of scores and the significantly lower assessor score suggests less agreement between teachers and assessors. Indeed, the assessor scores, ranging on an average of 12 points less than the instructors' evaluation, suggests assessors were less satisfied with 2010 than expected.

Instructors felt that 2010 students did well with Support, Organization, and Scholarship. Assessors disagreed, ranking Support, Topic, and Conventions high and Scholarship the lowest.

Instructors were somewhat pleased with students' performance as measured by the traits. As a rule, instructors rated individual traits 0.5 to 0.6 higher than the assessors and averaged about 4.4. Clearly, the assessors were less impressed and their marks suggest student performance in 2010 to be marginally acceptable.

As with 1010, the absence of prompts may have contributed to the statistical differences in 2010 papers.

English 2020

English 2010 aggregate scores revealed medians of 72 for instructors and 70 for assessors. Differences between instructors and assessors throughout the 2020 dataset are statistically insignificant. The close spread of data suggests that instructors and assessors agreed. The data also suggests less variety in paper types or forms. Interestingly more 2020 papers had prompts than any other. We might infer that our assessors' task was somewhat easier with the prompt, or that the prompt aligned the assessor with the instructor.

2020 instructors believed their students did well with Topic, Support, and Voice. Assessors ranked each of these traits substantially lower. However, there is virtually no difference between instructor and assessor for the remaining traits.

CONCLUSIONS

1. Overall participation by our instructors was lackluster. We evaluated only 58% of 1010, 61% of 2010 and 52% of 2020.
2. Instructors did not provide prompts in enough cases to assist assessors in ranking papers.

3. The range of difference for 2020 papers between instructor and assessor was the most narrow, and 2020 papers appeared to be held in the highest regard; 2010 was next and 1010 received the lowest scores.
4. The range of scores for 1010 suggests widely divergent teaching methods and paper assignments.
5. Assessors scored papers lower than instructors.
6. There is no quantitative difference between adjunct and contract faculty regarding paper scores.
7. Student performance between 2010 and 2020 suggests more uniformity of instruction among presumably equally prepared 2020 students.
8. The prompts submitted vary radically from section to section and from course to course. We might infer lack of standardization that exceeds that expected of healthy academic freedom.
9. Aggregate scores show our students are writing "C" papers. Assessors think our 2010 students are failing, with 1010 student doing marginally better.

RECOMMENDATIONS

1. The department should entertain discussions among all who teach 1010 that will generate course goals that will yield a more uniform and improved outcome. Prompts and assessed papers were so wildly divergent, that we can assume that our instruction in 1010 is somewhat disorganized and certainly less focused than we desire.
2. The department should conduct practice scoring sessions with an eye toward normalizing 2010 paper expectations and results.
3. Assessing without a prompt is an unfortunate complication. The department writing coordinators should make every effort to solicit prompts and have discussion among all who teach composition regarding written prompts' criticality, regarding both writing instruction and departmental assessment efforts.
4. Mechanically, the assessment effort was reasonably successful. The electronic aspect performed well; preparing instructors for their initial evaluation duties appeared to succeed; preparing for assessment following instructor evaluation was time consuming and difficult. Therefore, the infrastructure supports another assessment cycle in Spring, 2005.
5. Participation, by any standards, was disappointing. Department members who took the time to engage the system appear to have done an admirable job, both with completeness and with critical attention. However, entirely too many instructors were unable or unwilling to participate. The department should survey all concerned to determine if there were any causal factors that could be improved.

APPENDIX

Primary Trait Matrix

Student Paper Cover Sheet

Assessor Instructions

Primary Trait Matrix

	1-2 Not Proficient	3-4 Moderately Proficient	5-6 Proficient
IDEAS Control of topic (Weight 3)	- The writer's position is unclear, and the central issue of the paper is unfocused. - Writer refers to the concepts under discussion in a vague and imprecise way. - The student's argument is single-minded; he/she tends toward premature closure. - The concepts expressed are superficial, irrelevant to the topic, trivial or unelaborated. - Contradictions among generalizations are neither addressed nor resolved.	- The writer's position is fairly clear, but more work is needed in focusing the position. - The writer includes several perspectives, but has trouble articulating differences and/or controlling the relationships among them. - The train of thought expressed in generalizations is, for the most part, logically consistent.	- Writer understands and precisely identifies the terms and concepts at issue in the discussion. - Writer articulates her/his own position regarding the issue (thesis). - The writer presents a variety of views and controls the tension among them. - The student precisely articulates the point or generalizations he/she is making.
Support (Weight 3)	-There is little attempt to develop points with supporting evidence.	- The evidence appears to be pertinent, but the writer does not explain its relevance. - The evidence does not reach the most concrete levels of specificity. - The paper relies on anecdotal evidence alone.	- The student's evidence is pertinent to the point she/he wishes to make. - The evidence reaches a very concrete level of specificity. - Outside sources beyond anecdote are used as evidence.
ORGANIZATION (Weight 2)	- Ideas are randomly put together with little apparent structure evident. - Ideas are presented in a haphazard manner. - The paper appears illogical in its construction and ordering or ideas.	- A deliberate and sound organizational plan is suggested, but may be slightly confusing or ineffective. - The paper leads to a logical conclusion. - The paper contains a logical progression and ordering of ideas.	- A deliberate and sound organizational plan is evident. - The paper contains an inviting beginning and a logical and effective conclusion. - The logical progression and ordering of ideas allows for the most important points to stand out. - Ideas are randomly put together with little apparent structure evident. - Ideas are presented in a haphazard manner. - The paper appears illogical in its construction and ordering or ideas.
CONVENTIONS (Weight 2)	- Errors in spelling, punctuation, capitalization, and/or grammar and usage interfere with meaning.	- Some distracting errors in spelling, punctuation, capitalization, and/or grammar and usage.	- Few to no errors in spelling, punctuation, capitalization, or grammar and usage.
WORD CHOICE (Weight 1)	- Many words are imprecise, and word usage is confusing to the extent that locating meaning is difficult. - Most nouns and verbs are vague, and there are few descriptive words. - Student uses biased and pejorative language.	- Functional and appropriate words connect the audience to the subject matter. - Sentences contain moderately specific nouns, verbs and descriptive words. - Student avoids biased and pejorative language.	- Accurate and precise words are well suited for connecting the audience to the subject matter. - Sentences contain vivid sensory images with specific nouns, action verbs, and well-chosen descriptive works. - Student avoids biased and pejorative language and instead uses languages that is respectful of the audience.

VOICE (Weight 1)	- The level of formality, tone or diction fail to show an understanding or knowledge of a college reader's expectations. - The register is colloquial and conversational, more like an e-mail to a friend than a college paper. - The voice of the persona might seem argumentative, emotional, preachy, closed-minded, or prone to the use of ridicule or sarcasm. - Diction is vague or imprecise.	- Level of formality and diction is unevenly appropriate to the rhetorical situation. - Voice and tone are generally appropriate, but some unevenness appears. - Diction is adequate, but not especially sophisticated.	- The student chooses a level of formality (register), tone and diction appropriate to academic discourse. - The voice and tone are appropriate for a reasonable and thoughtful audience. Humor or irony is appropriate. - Diction is precise; sophisticated and scholarly words are used.
SENTENCE FLUENCY (Weight 1)	- There is little apparent order to the arrangement of ideas, with unconnected jumps from topic to topic. - Sentences are monotonous; more than half have similar lengths, beginnings, or patterns. - Numerous fragments, comma splices, run-ons and incoherent sentences distract the reader.	- Transitions and connections between ideas may be unsophisticated, but there is evidence of student's attempt to articulate relationships. - Sentences are somewhat varied in length and structure; grammatical /structural errors are rare.	- Ideas are organized and transitions articulated so that the reader can follow a logical progression of thought. - A rhythm is created across sentences; the flow is pleasing to the ear. - A rich variety of sentence lengths, beginnings and patterns enriches meaning. Repetition, if used, is applied for good effect.
SCHOLARSHIP (Weight 1)	- Outside sources are not used or not credited. - Credibility of sources is questionable. - Sources are not integrated into the student's own text.	- Sources are acknowledged informally. - Sources are credible, but quality is sporadic. - Sources are sporadically well integrated into the student's own text.	- The student cites outside sources appropriately, including in-text references and list of works cited. - Sources are scholarly and credible. - Sources are integrated into the student's own text.

The primary traits are adapted from the *Provo School District Assessment Handbook of Instructions* for high school students (2002 edition) and the Utah Statewide Assessment CriteriaBWriting 1010 and 2010" (Revised 2003). The UVSC English Assessment Committee believes that using the six traits from the Provo School District accommodates students arriving at UVSC who are familiar with the six traits language as a result of their high school experience. Using language from the Utah Statewide Assessment acknowledges the standards used for assessing writing proficiency in institutions of higher education in Utah.

The six categories of traits (ideas, organization, conventions, voice, word choice, and sentence fluency) come directly from the Provo School District categorization of traits. The division of proficiency levels into Amore proficient,@ Amoderately proficient,@ and Aless proficient@ is adopted from the Utah Statewide Assessment. The statements used to describe proficiency levels come primarily from the Utah Statewide Assessment, but additional statements from the Provo School District have been included where needed to complete descriptions of the traits. In addition, the UVSC Assessment Committee added statements in order to assess 2010 and 2020 writing assignments that require more in-depth research than is addressed by the Provo School District traits or the Utah Statewide traits.

STUDENT PAPER COVER SHEET

Course: English 1010-001

Student: 555-12-1212

Prompt:

Write an essay on a controversial issue. Learn more about the issue and take a position on it. Present the issue to readers and develop an argument for the purpose of confirming, challenging, or changing your reader views on the issue. Ensure that you base the paper on at least 6 research sources and that you refute at least one oppositional point of view.

Composition Program Assessment—Spring 2004

Assessor Instructions

The Assessment Director established samples for the composition courses: ENGL 1010 = 340; ENGL 2010 = 340; and ENGL2020 = 255—this from some 3752 total papers. The result is that each participating assessor will read and evaluate 11 papers. The Assessment Management Program randomly assigned your user code to four 1010 papers, four 2010 papers and three 2020 papers, excluding those papers you were the instructor of record.

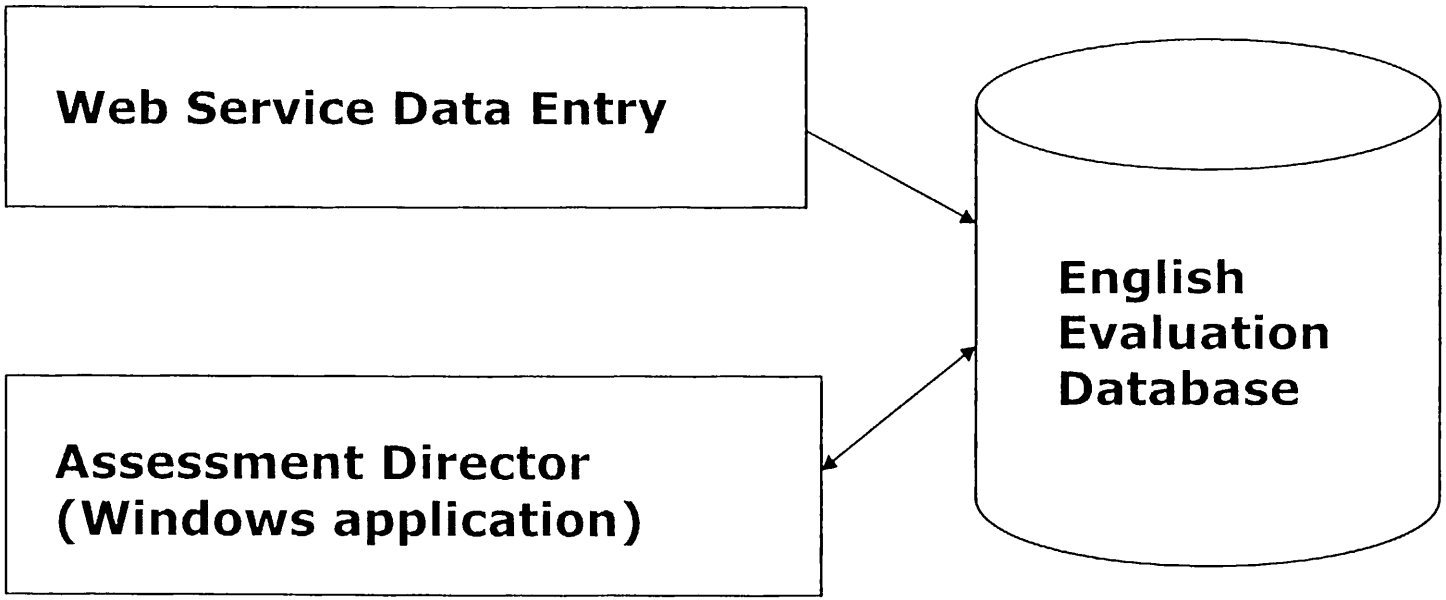
Course	Student	Score these categories 1-6 with 6 being highest score							
		Ideas	Support	Organization	Conventions	Voice	Word choice	Sent. Fluency	Scholarship
ENGL 1010	122-45-7898								
ENGL 1010	987-65-4321								
ENGL 1010	101-20-3033								
ENGL 1010	654-98-1234								
ENGL 2010	789-45-6123								
ENGL 2010	521-55-4567								
ENGL 2010	528-12-5948								
ENGL 2020	457-87-6543								
ENGL 2020	432-56-1245								
ENGL 2020	872-49-6578								

To perform your assessment duties:

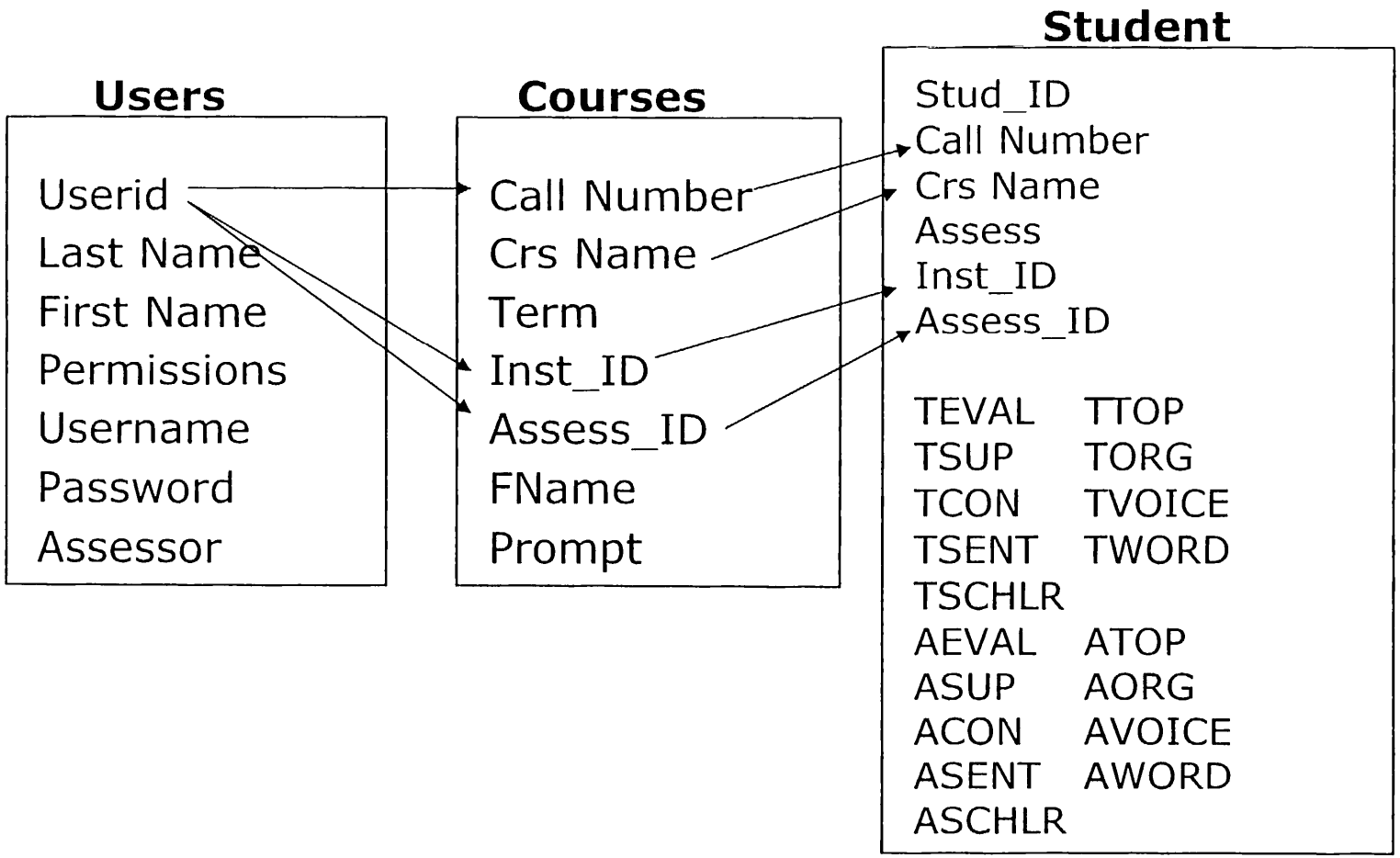
1. Ensure you have each of the papers listed above.
2. Read and assess the papers with the grading grid or input scores directly on the website.
3. Enter ***161.28.110.117/EnglishEvaluation*** on your brower's URL line
4. ***Be sure to select "Assessor" as your role***
5. Your Userid = ***«USERID»***, and your password is ***«PWORD»***
6. After completing the assessment, please dispose of the papers in an appropriate manner.

UVSC English Department Assessment Report

System Design



Database Design: Tables



Assessment Activities

- Request SIS course and student data
- Import SIS data into database
- Groom data (confirm accuracy)
- Input department users
- Prepare student paper cover sheets and instruction packets
- Conduct teacher evaluation of papers

Assessment Activities (cont)

- Collect evaluated student papers
- Assign users to assessor duties
- Perform record sort for assessors
- Sort student papers into assessor packets
- Conduct assessment evaluation

Assessment Activities (cont)

- Download data to MS Excel
- Utilize Excel statistical package to perform calculations
- Analyze results
- Prepare report
- Present findings

Operational Statistics

Course	Enrolled	Evaluated	Percentage
1010	1578	925	58.6
2010	1660	1013	61.0
2020	514	269	52.3
Totals	3752	2207	58.8

Operational Results: Prompts

Course	Papers	Prompts	Evaluated	Assessed
1010	925	536	406	132
2010	1013	155	91	21
2020	269	156	96	43
Totals	2207	847	593	196

Operational Results: Prompts

	Evaluated		Assessed	
Course	No/prompt	w/prompt	No/prompt	w/prompt
1010	389	536	104	132
2010	858	155	217	21
2020	113	156	116	43

Operational Results: Prompts

Course	Assessed with prompts
1010	56%
2010	9%
2020	27%

Data Results: 1010

Teacher evaluation			Assessor evaluation		
Average	**Median**	**Std Dev**	**Average**	**Median**	**Std Dev**
63.66	**63**	**15.52**	**58.5**	**57**	**15.29**

Data Results: 2010

Teacher evaluation			Assessor evaluation		
Average	**Median**	**Std Dev**	**Average**	**Median**	**Std Dev**
74.70	**75**	**15.87**	**62.20**	**63**	**14.83**

Data Results: 2020

Teacher evaluation			Assessor evaluation		
Average	**Median**	**Std Dev**	**Average**	**Median**	**Std Dev**
71.34	**72**	**15.85**	**68.34**	**70**	**14.76**

Data Results: 1010 Traits

	Teachers			Assessors		
	Average	**Median**	**Std Dev**	**Average**	**Median**	**Std Dev**
Topic	3.924	4	1.020	3.640	4	1.112
Support	3.919	4	1.110	3.585	4	1.137
Organization	3.835	4	1.108	3.653	4	1.075
Conventions	3.792	4	1.089	3.542	4	1.131
Voice	3.979	4	1.078	3.496	4	1.361
Word Choice	3.886	4	1.010	3.547	4	1.041
Sent. Fluency	3.826	4	1.083	3.597	4	1.124
Scholarship	3.614	4	1.275	3.242	3	1.240

Data Results: 2010 Traits

	Teachers			Assessors		
	Average	**Median**	**Std Dev**	**Average**	**Median**	**Std Dev**
Topic	4.752	5	0.969	3.874	4	1.125
Support	4.592	5	1.052	3.895	4	1.152
Organization	4.479	5	1.074	3.811	4	1.024
Conventions	4.303	4	1.165	3.622	4	1.079
Voice	4.433	5	1.233	3.702	4	1.312
Word Choice	4.399	4	1.138	3.803	4	1.006
Sent. Fluency	4.311	4	1.127	3.790	4	1.105
Scholarship	4.450	5	1.268	3.483	4	1.235

Data Results: 2020 Traits

	Teachers			Assessors		
	Average	**Median**	**Std Dev**	**Average**	**Median**	**Std Dev**
Topic	4.415	5	1.060	4.101	4	1.086
Support	4.475	5	1.135	4.264	4	1.034
Organization	4.365	5	1.122	4.258	4	1.095
Conventions	4.038	4	1.232	4.010	4	1.094
Voice	4.258	4	1.057	3.818	4	1.405
Word Choice	4.138	4	1.133	4.126	4	0.896
Sent. Fluency	4.208	4	1.126	4.119	4	0.944
Scholarship	4.252	4	1.160	4.082	4	1.253

Findings and Conclusions

1. Overall participation by our instructors was lackluster. We evaluated only 58% of 1010, 61% of 2010 and 52% of 2020.

2. Instructors did not provide prompts in enough cases to assist assessors in ranking papers.

3. The range of difference for 2020 papers between instructor and assessor was the most narrow, and 2020 papers appeared to be held in the highest regard; 2010 was next and 1010 received the lowest scores.

4. The range of scores for 1010 suggests widely divergent teaching methods and paper assignments.

Findings and Conclusions (cont)

5. Assessors scored papers lower than instructors.

6. There is no quantitative difference between adjunct and contract faculty regarding paper scores.

7. Student performance between 2010 and 2020 suggests more uniformity of instruction among presumably equally prepared 2020 students.

8. The prompts submitted vary radically from section to section and from course to course. We might infer lack of standardization that exceeds that expected of healthy academic freedom.

9. Aggregate scores show our students are writing "C" papers. 2010's aggregate scores suggest marginal performance near the high "D" range.

Recommendations

1. The department should entertain discussions among all who teach 1010 that will generate course goals that will yield a more uniform and improved outcome.

2. The department should conduct practice scoring sessions with an eye toward normalizing 2010 paper expectations and results.

3. The department should conduct additional discussions concerning the traits so all instructors and teachers share an understanding of what kinds of writing deserve what kinds of scores.

Recommendations (cont)

4. The department writing coordinators should make every effort to solicit prompts and have discussion among all who teach composition regarding written prompts' criticality, regarding both writing instruction and departmental assessment efforts.

5. The computer applications performed well; preparing instructors for their initial evaluation duties appeared to succeed; preparing for assessment following instructor evaluation was time consuming and difficult. Therefore, the infrastructure supports another assessment cycle in Spring, 2005.

6. The department should survey all assessment participants to determine if there were any factors that resulted in the lack of participation

Recommendations (cont)

1. Retain Joe Strange as your consultant.

   Provide transporation and two-day visit (Houston to Orem) to prepare program and data for next cycle (Feb, 2005) and reliability/validity tests

2. Retain Damon Bahr as a statistical consultant

   Design reliability and validity test (Fall 2004)

   Conduct advanced statistical analysis on the next cycle's data (concurrent with next cycle)