

***Report on the Results of the Academic Profile® Exam & Essay  
Spring 2004***

Prepared by  
Melinda Kreth, Gen. Ed. Coordinator and Assoc. Prof. of English

August 1, 2004

Submitted to  
The General Education Council

## **ABSTRACT**

This report discusses the pilot test of The Academic Profile® exam and essay on a convenience sample of CMU students ( $n = 295$ ) in Spring 2004 to assess their General Education knowledge and skill. Exam results revealed that students scored in the mid-range of possible scores and performed about as well as students at 14 comparison institutions, except in the areas of critical thinking, in which CMU students scored lower, and reading comprehension, in which CMU students scored slightly higher. The results of the essays—which were scored by ETS and again by CMU scorers—indicate that CMU students have little difficulty with basic writing skills, such as grammar, spelling, etc., but their writing demonstrates a lack of critical thinking ability and a lack of ability to support claims with appropriate evidence. Several confounding variables may partially account for students' overall poor performance on the exam and essay, i.e., problems with the students, the sampling procedure and testing conditions, the writing environment, the writing prompts, our curriculum, and our teaching methods. Recommendations are offered for future Gen. Ed. assessment.

# **Report on the Results of The Academic Profile® Exam & Essay Spring 2004**

**Prepared by Melinda Kreth, Gen. Ed. Coordinator, August 1, 2004**

## **INTRODUCTION**

In Fall 2003, the General Education Council approved a plan to conduct a pilot study of our General Education program using an instrument provided by the College Board/Educational Testing Service (ETS), i.e., The Academic Profile®, which consists of an exam and an essay. According to ETS,

The Academic Profile is a test of general academic knowledge and skills. It is intended for use by colleges and universities in assessing the outcomes of their general education programs to improve the quality of instruction and learning. The test focuses on the academic skills developed through general education courses rather than on the knowledge acquired about the subjects taught in these courses. It does this by testing college-level reading, and critical thinking in the context of the humanities, social sciences, and natural sciences. Mathematics and writing skills are tested independently of context areas.

This pilot study is CMU's first attempt at assessing students' overall General Education knowledge and skills.

## **METHODS**

We struggled to devise a feasible sampling procedure with effective incentives for student participation, but due to a number of constraints, we ultimately opted for a convenience sample for this pilot study, aware that the results might not be generalizable to the population of 2004 graduating seniors. A letter was sent to instructors of senior seminars, capstones, and upper-level methods courses asking them to either require students to participate as part of the course requirements or to offer extra credit to students who chose to participate.

A total of 295 students took the exam. Of these, there were 235 seniors, 45 juniors, two sophomores, and 13 classified as "other" (probably because these students did not indicate their class standing in the demographics section of the exam).

The exam was administered and the essays were written in CMU's Testing Center and in classrooms at various dates and times over a three-week period in February 2004. An abbreviated version of the exam consisting of 36 questions was used, and students were given 40 minutes to complete the exam. The essays were written by hand, and students were given 50 minutes to write the essays. The essays were written immediately after students completed the exams. A total of 280 essays were obtained and scored by ETS, indicating that 15 students chose not to write the essay. (Note: CMU scored only 279 essays, discarding one that was only four lines long.)

Completed exams and essays were sent to ETS for scoring. However, it was decided by the Gen. Ed. Council, the Office of Academic Affairs, and members of the English Department's

Composition Committee to also assess the essays locally, using criteria drawn from the objectives for CMU's writing competency courses, ENG 101 and ENG 201. Composition faculty also felt that both the ETS results and the local assessment results would help the Composition Program revise the master syllabus for ENG 201.

Our local assessment of the essays was held on July 19 and 21, 1:00 – 5:00 p.m. Nine scorers read and scored the essays: three tenured composition faculty and six graduate students enrolled in the Summer II ENG 618: Teacher Research Methods course (and all of these graduate students were either secondary English teachers or graduate teaching assistants in our Composition Program). Calibration was conducted on a set of eight essays, at which time a high level of inter-rater agreement was established.

The remainder of this report includes a summary and discussion of the results of the exam and essay scores, followed by the conclusions and recommendations.

## RESULTS OF THE EXAM

Overall, students scored in the midrange of possible scores in all areas except critical thinking, in which our students scored lightly below midrange, and reading comprehension, in which students scored slightly above midrange. (Note: The “Writing” scores below do not refer to the essay scores but to scores on exam questions designed to assess students’ declarative knowledge *about* writing. These “writing” scores will be discussed later in this report in conjunction with the essay scores.)

### Summary of Scores (n = 295)

	Possible Range	Mean	95% Confidence Limits	Standard Deviation	25 <sup>th</sup> Percentile	50 <sup>th</sup> Percentile	75 <sup>th</sup> Percentile
<b>Total Score</b>	400 – 500	449.22	447 – 451	18.51	435	447	462
<b>Skills Subscores</b>							
Critical Thinking	100 – 130	112.33	112 – 113	6.72	107	111	116
Reading	100 – 130	119.78	119 – 120	6.28	116	120	125
Writing	100 – 130	115.37	115 – 116	4.34	112	116	118
Math	100 – 130	114.93	114 – 116	5.86	110	115	119
<b>Academic Subscores</b>							
Humanities	100 – 130	115.98	115 – 117	6.39	110	114	122
Social Sciences	100 – 130	114.52	114 – 115	6.46	110	115	121
Natural Sciences	100 – 130	116.00	115 – 117	5.50	116	116	119

The percentiles tell us what proportion of students scored *below* the score indicated in that column. For example, only 25% of students scored at or below 107 on the critical thinking portion of the exam, whereas, 50% scored at or below 111, and 75% scored at or below 116, meaning that only 25% scored above 116, which is just slightly above midrange.

## Proficiency Criteria for Reading/Critical Thinking

### **Level 1**

At level 1, a student can

- recognize factual material explicitly presented in a reading passage
- understand the meaning of particular words or phrases in the context of a reading passage

### **Level 2**

At level 2, a student can

- synthesize material from different sections of a passage
- recognize valid inferences derived from material in the passage
- identify accurate summaries of a passage or of significant sections of the passage
- understand and interpret figurative language
- discern the main idea, purpose, or focus of a passage or a significant portion of the passage

### **Level 3**

At level 3, a student can

- evaluate competing casual explanations
- evaluate hypotheses for consistency with known facts
- determine the relevance of information for evaluating an argument or conclusion
- determine whether an artistic interpretation is supported by evidence contained in a work
- recognize the salient features or themes in a work of art
- evaluate the appropriateness of procedures for investigating a question of causation
- evaluate data for consistency with known facts, hypotheses or methods
- recognize flaws and inconsistencies in an argument

## Proficiency Criteria for Math

### **Level 1**

At level 1, a student can

- solve word problems that would most likely be solved by arithmetic and do not involve conversion of units or proportionality. These problems can be multi-step if the steps are repeated rather than embedded.
- solve problems involving the informal properties of numbers and operations, often involving the Number Line, including positive and negative numbers, whole numbers and fractions (including conversions of common fractions to percent, such as converting "1/4" to 25%).
- solve problems requiring a general understanding of square roots and the squares of numbers.
- solve a simple equation or substitute numbers into an algebraic expression.
- find information from a graph. This task may involve finding a specified piece of information in a graph that also contains other information.

### **Level 2**

At level 2, a student can

- solve arithmetic problems with some complications, such as complex wording, maximizing or minimizing, and embedded ratios. These problems include algebra problems that can be solved by arithmetic (the answer choices are numeric).
- simplify algebraic expressions, perform basic translations, and draw conclusions from algebraic equations and inequalities. These tasks are more complicated than solving a simple equation, though they may be approached arithmetically by substituting numbers.
- interpret a trend represented in a graph, or choose a graph that reflects a trend.
- solve problems involving sets; the problems would have numeric answer choices.

### **Level 3**

At level 3, a student can

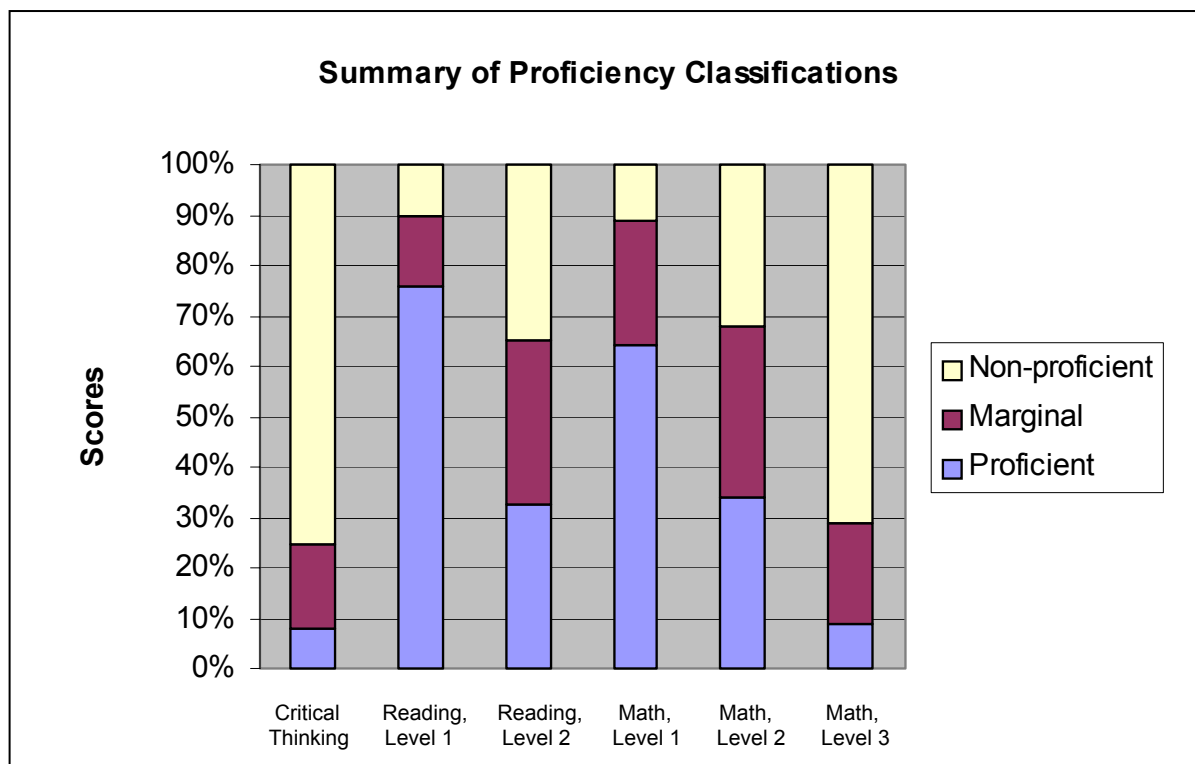
- solve word problems that would be unlikely to be solved by arithmetic; the answer choices are either algebraic expressions or are numbers that do not lend themselves to back-solving.
- solve problems involving difficult arithmetic concepts such as exponents and roots other than squares and square roots and percent of increase or decrease.
- generalize about numbers, e.g., identify the values of (x) for which an expression increases as (x) increases.

- solve problems requiring an understanding of the properties of integers, rational numbers, etc.
- interpret a graph in which the trends are to be expressed algebraically or in which one of the following is involved: exponents and roots other than squares and square roots, percent of increase or decrease.
- solve problems requiring insight or logical reasoning.

### Summary of Proficiency Classifications (n = 295)

These data tells us that our students did well only at the very *lowest* level of proficiency, at a level we might expect of incoming freshmen or first-semester sophomores. But since 80% of the exam participants were graduating seniors and another 15% were second-semester juniors, these data are alarming (though as will be discussed at the end of this report, other factors probably contributed to our students’ overall poor performance on this exam).

Skill Dimension	Proficiency Classification		
	Proficient	Marginal	Non-Proficient
Critical Thinking	8%	17%	76%
Reading, Level 1 (lowest)	76%	14%	10%
Reading, Level 2 (highest)	33%	33%	35%
Math, Level 1 (lowest)	64%	25%	11%
Math, Level 2	34%	34%	32%
Math, Level 3 (highest)	9%	20%	71%



### CMU Scores Compared to those of Other Institutions

Only those institutions testing 30 or more students in a college class were included in the analyses for that college. Before the comparisons are shown, a list of the comparison institutions and the sample size at each institution is provided below:

Andrews University, MI	349
Cleveland State, OH	52
East Tennessee State, TN	2,152
Idaho State, ID	68
Middle Tennessee State, TN	3,965
Mississippi State, MS	498
Tennessee State, TN	2,121
Texas A & M at Commerce, TX	30
Univ. of Georgia at Athens, GA	39
Univ. of Memphis, TN	743
Univ. of Mississippi, MS	58
Univ. of Missouri at Kansas City, MO	1,129
Univ. of Missouri at St. Louis, MO	2,444
<u>Univ. of South Carolina at Columbia, SC</u>	<u>120</u>
<b>Total:</b>	<b>13,768</b>

ETS compared CMU scores to the “weighted number of students” at the other institutions rather than the total number. This weighting prevents the statistics from being dominated by the few institutions that tested a relatively large number of their students (e.g., Middle Tennessee State, Univ. of Missouri at St. Louis, etc.). Weighted  $n = 12,803$ .

### Comparison of Summary Scores

	Possible Range	CMU Means ( $n = 235$ )	Doc Means ( $n = 12,803$ )
<b>Total Score</b>	400 – 500	451.06	449.05
Critical Thinking	100 – 130	112.89	112.10
Reading	100 – 130	120.49	120.3
Writing	100 – 130	115.65	115.60
Math	100 – 130	115.16	114.60
Humanities	100 – 130	116.67	116.10
Social Sciences	100 – 130	115.09	114.80
Natural Sciences	100 – 130	116.31	116.00

There were only minor differences in the mean scores for seniors and the comparison group, so results for seniors only were reported to us by ETS for comparison.

### Comparison of Proficiency Scores

As can be seen in the tables below, CMU scores are nearly identical to the scores for students at the comparison schools, except for Reading, Level 2, where a substantially lower proportion of CMU students than comparison group students scored in the Proficient category (33% vs. 43%). However, the data suggest that CMU students did about as well on the exam as students at the comparison institutions.

Skill Dimension	Proficiency Classification					
	Proficient		Marginal		Non-Proficient	
	CMU	Other	CMU	Other	CMU	Other
Critical Thinking	8%	6%	17%	17%	76%	77%
Reading, Level 1 (lowest)	76%	76%	14%	16%	10%	9%
Reading, Level 2 (highest)	33%	43%	33%	23%	35%	33%
Math, Level 1 (lowest)	64%	61%	25%	26%	11%	13%
Math, Level 2	34%	33%	34%	30%	32%	37%
Math, Level 3 (highest)	9%	8%	20%	19%	71%	72%

CMU (n = 295), Other (n = 12,803)

But given that CMU students performed in the middle of the 100 to 130 range of scores, that seems to mean that our students' performance was no more or less mediocre than that of students at the comparison institutions, and how one interprets *that* depends, in turn, on a number of other factors that are discussed at the end of this report.

The comparisons are also meaningless unless we know 1) the academic quality of the comparison institutions, and 2) the educational environment of each institution, and 3) demographics of the student populations at those institutions

How does one assess the academic quality of the comparison institutions? Certainly, none of the schools on the list is a nationally, let alone internationally, renowned college or university. Only one school, the Univ. of Georgia at Athens, is the flagship school of its state, but only 39 of its students were tested. However, some of the other schools listed are satellite campuses of their state's flagship university. None of the schools on the list has what would be considered high entrance requirements, and many of the schools have a relatively high proportion of students who must complete remedial courses before taking regular college-level courses.

In terms of educational environment, there is little consistency among the schools. Some are urban commuter campuses, while others are located in rural areas. All but one are public institutions.

Finally, there is great variety among the comparison schools in terms of the demographics of their student populations. Some of the schools have much larger proportions of racial/ethnic minorities than does CMU, and several serve largely non-traditional students, unlike CMU, which serves mostly traditional students, many of whom live on campus

## RESULTS OF THE ESSAY AND WRITING PORTION OF THE EXAM

Before examining the results, it will be helpful to review the writing prompts to which students responded, as well as the scoring criteria used locally and by ETS, and the instruction given to CMU scorers.

### ETS Writing Prompts

Note: These prompts were developed by and included in The Academic Profile® testing materials from ETS.

1. Humanities Essay Topic

More and more [we] will discover that we will have to turn to poetry to interpret life for us, to console us, to sustain us. Without poetry, our science will appear incomplete; . . . the best poetry will be found to have a power of forming, sustaining, and delighting us, as nothing else can.

Mathew Arnold, from "The Study of Poetry," 1880

Explain what Arnold is saying in the passage above about the value and the functions of poetry and whether or not you believe that poetry or art in general does in fact serve these functions. Provide relevant examples of works from your course work in the humanities or the arts to support your argument. Examples of humanities and the arts include literature, painting, music, dance, film, and philosophy.

2. Natural Science Topic

In 1905, the French physicist and mathematician Henri Poincaré announced, "The aim of science is not things themselves . . . but the relations between things."

Explain what Poincaré meant and how his statement applies or fails to apply to one of the natural sciences you have studied (for example, chemistry, biology, physics, geology, astronomy). If you agree with Poincaré, provide specific examples of the kinds of relationships this branch of science examines. Or, if you think the statement fails to apply, discuss the kinds of questions this branch of science seeks to answer, or the means it uses to seek the answers.

3. Social Science Topic

The replacement of the power of the individual by the power of a community constitutes the decisive step of civilization. The essence of it lies in the fact that the members of the community restrict themselves in their possibilities of satisfaction, whereas the individual knew no such restrictions. . . . A good part of the struggles of mankind centre round the single task of finding an expedient accommodation—one, that is, that will bring happiness—between the claim of the individual and the cultural claims of the group, and one of the problems that touches the fate of humanity is whether such an accommodation can be reached by means of some particular form of civilization or whether this conflict is irreconcilable. . . . [For} it is impossible to overlook the extent to which civilization is built upon a renunciation of instinct, how much it presupposes precisely the non-satisfaction (by suppression, repression, or some other means?) of powerful instincts.

Sigmund Freud, from *Civilization and Its Discontents*, 1930

Explain what the statement above indicates about the relationship between the individual and civilization. Providing relevant examples from your study of the social sciences, discuss whether you agree or disagree with Freud's view. Examples of the social sciences include political science, psychology, sociology, history, anthropology, economics.

### The CMU Scoring Rubric <sup>1</sup>

- 4 = **An excellent essay.** (passable)
- It offers an effective response to the prompt.
  - Its overall pattern of organization is appropriate; the internal organization of ideas is effective, and transitions are smooth.
  - The argument is well developed, uses appropriate and effective rhetorical strategies, and avoids logical fallacies.
  - The writing follows the conventions of Standard English, and the writing style is clear, concise, and appropriate.
  - The essay has very few, if any, errors in grammar, punctuation, and spelling.
  - The essay is well written, interesting, and easy to read.
- 3 = **A good essay.** (passable)
- It offers a generally effective response to the problem.
  - Both the overall pattern of organization and the internal organization are good, although transitions may not be as smooth as in a "4" essay.
  - The argument is developed adequately, though not as well as a "4" essay; it uses appropriate and effective rhetorical strategies, and avoids logical fallacies.
  - The writing follows the conventions of Standard English, and the writing style is generally clear, concise, and appropriate.
  - It may have some minor mechanical errors or some awkward spots, but basically it is clear, well written, fairly interesting, and easy to read.
- 2 = **A satisfactory essay.** (passable)
- It offers a response that is basically acceptable but that contains problems that mar effectiveness.
  - It uses an acceptable pattern of organization, but it may lack smooth transitions.
  - The writing follows the conventions of Standard English, but there may be minor errors in style, tone, internal organization, or mechanics; logic may not be developed fully.
  - OR the assignment may be a good ("3") paper with a *major* flaw in one of the following: the argument, organization, tone, or writing style.
- 1 = **An unsatisfactory essay.** (unpassable)
- It offers an unacceptable response to the prompt.
  - It may be a satisfactory ("2") assignment with a *major* flaw in one of the following: the argument, organization, tone, or writing style.
  - OR it may show some evidence of attempting to respond to the prompt but has many errors in organization, development, word choice, style, tone, and/or grammar, punctuation, and spelling. None of these alone would necessarily doom the essay; however, together they make the essay unsatisfactory.
  - OR it may be illegible.

---

<sup>1</sup> Based on some of the course objectives for CMU's English 101 and 201 writing competency courses.

### Instructions for CMU Scorers

The following instructions were given to all nine scorers of the essays:

During the two afternoons allotted for this assessment, 279 essays will be read (26 sets of 10, one set of 11, and one set of 8, the last of which will be used as the calibration set). These essays were written by students who participated in the recent pilot administration of The Academic Profile® in February 2004.

In scoring the essays, do the following:

1. Review again the scoring rubric.
2. Read sets sequentially. For example, if you are given set D 1-10 to begin, then you'll want to read set E 1-10 next, then F 1-10, and so on, until you come back around and finish with set C 1-10. (It's like a golf scramble!)
3. Read the essays in each set quickly. On average, try not to spend more than 15 minutes reading and scoring each set of essays. (Actually, some sets may take a bit more time, while others may take less time.)
4. Don't get bogged down trying to figure out the writer's meaning—if it's not clear then give it a low score and move on to the next essay.
5. Do not concern yourself with *why* an essay is written the way it's written—just score it.
6. Consider whether the essay is passable or not. If not, then give it a 4 and move on. If it is passable, then consider where it falls within the 1-3 range.
7. Do not discuss your scoring of the essays with other readers. If you have a question, ask the session administrator (Melinda Kreth).
8. Record your scores for each set on your scoring sheet. Since each essay in each set is coded, (i.e., A-1), record your scores by simply placing check marks (✓) in the desired score columns next to the appropriate essay number.
9. Write your name on the cover label for each set of essays you grade; this will indicate that you have read and scored all the essays in that stack).
10. Don't take a break in the middle of a set—always complete the set you're working on before taking a break.

### Readings per Set

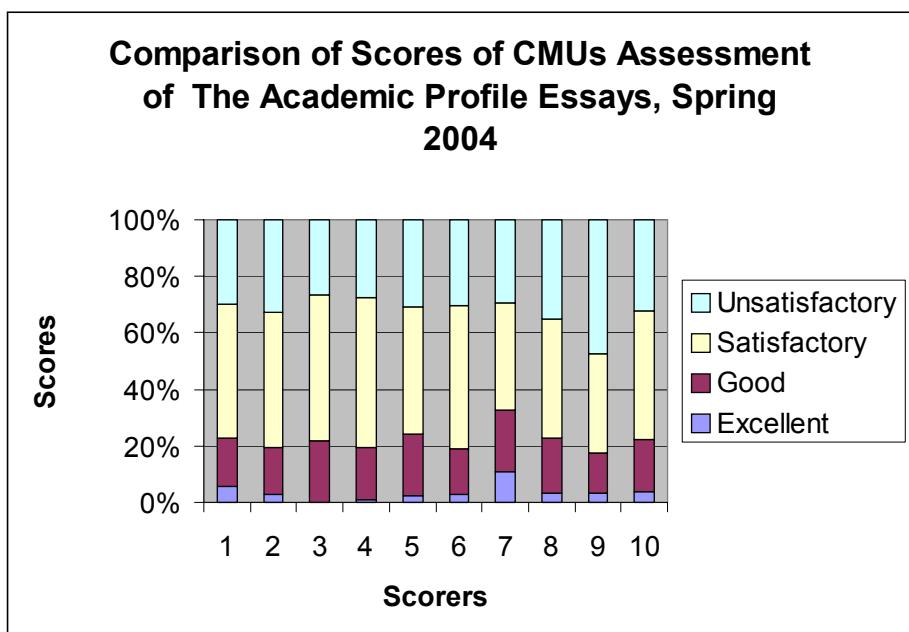
A 1-10	9	H 1-10	9	O 1-10	7	V 1-10	5
B 1-10	9	I 1-10	9	P 1-10	7	W 1-10	7
C 1-10	9	J 1-10	9	Q 1-10	7	X 1-10	5
D 1-10	8	K 1-10	7	R 1-10	6	Y 1-11	5
E 1-10	8	L 1-10	6	S 1-10	5	Z 1-10	6
F 1-10	8	M 1-10	8	T 1-10	5	A 11-20	3
G 1-10	8	N 1-10	8	U 1-10	4	B 11-18	9

### Overall CMU Scores

As can be seen in the following table and bar chart, our local assessment exhibited a high degree of overall inter-rater agreement, and we are confident that our scores are valid with respect to our own scoring criteria.

Scorer	4	3	2	1	Total Scored	Median Score	Mode
①	15 5.6%	45 17.1%	126 47.5%	79 29.8%	265	2.49	2.00
②	6 2.9%	34 16.4%	100 48.1%	68 32.6%	208	2.11	2.00
③	0 0.00%	40 21.6%	96 51.9%	49 26.5%	185	2.05	2.00
④	2 1.1%	33 18.4%	95 53.1%	49 27.4%	179	2.07	2.00
⑤	7 2.5%	61 21.9%	125 44.8%	86 30.8%	279	2.04	2.00
⑥	5 2.7%	30 16.1%	95 51.1%	56 30.1%	186	2.10	2.00
⑦	24 11.0%	47 21.5%	83 38.1%	64 29.4%	218	1.87	2.00
⑧	7 3.2%	42 19.4%	92 42.3%	76 35.1%	217	2.10	2.00
⑨	6 3.2%	26 14.1%	65 35.1%	88 47.6%	185	2.31	1.00
Totals	72 3.8%	358 18.6%	877 45.6%	615 32.0%	1922	2.13	2.00

4 = excellent, 3 = good, 2 = satisfactory, 1 = unsatisfactory (see scoring rubric)

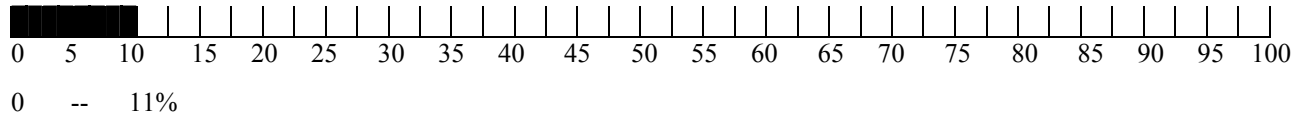


4 = Excellent, 3 = Good, 2 = Satisfactory, 1 = Unsatisfactory

## Range of Scores for CMU Scorers

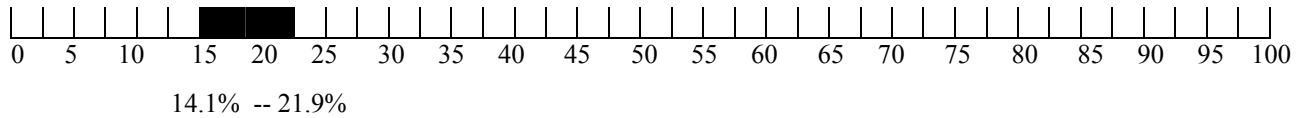
### 4 = Excellent

Of the nine scorers, scorer #3 gave to 0 essays read ( $n = 185$ ) a score of Excellent, while scorer #7 gave 11% of essays read ( $n = 218$ ) a score of Excellent. The remaining scorers fell within this range in terms of the proportion of essays read to which they gave scores of Excellent.



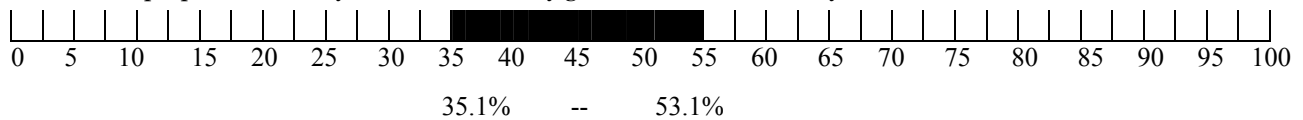
### 3 = Good, 14.1 – 21.9%

Of the nine scorers, scorer #9 gave to 14.1% of essays read ( $n = 185$ ) a score of Good, while scorer #5 gave 21.9% of essays read ( $n = 279$ ) a score of Good. The remaining scorers fell within this range in terms of the proportion of essays read to which they gave scores of Good.



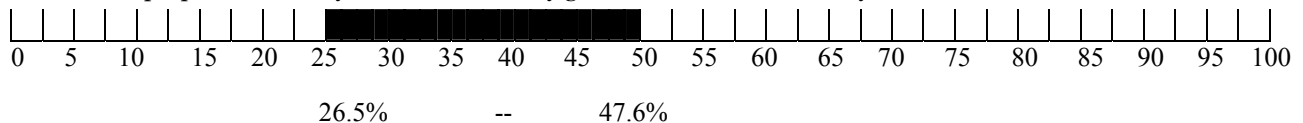
### 2 = Satisfactory, 35.1 – 53.1%

Of the nine scorers, scorer #9 gave to 35.1% of essays read ( $n = 185$ ) a score of Satisfactory, while scorer #4 gave 53.1% of essays read ( $n = 179$ ) a score of Satisfactory. The remaining scorers fell within this range in terms of the proportion of essays read to which they gave scores of Satisfactory.



### 1 = Unsatisfactory, 26.5 – 47.6%

Of the nine scorers, scorer #3 gave to 26.5% essays read ( $n = 185$ ) a score of Unsatisfactory, while scorer #9 gave 47.6% of essays read ( $n = 185$ ) a score of Unsatisfactory. The remaining scorers fell within this range in terms of the proportion of essays read to which they gave scores of Unsatisfactory.



Let's now look at how our local assessment results compare to the ETS assessment of our students' essays.

## Comparison of ETS Scores and CMU Scores of the Essays

ETS scored each essay using the same 4-point scale we used, but ETS scorers used a different set of scoring criteria.

### The ETS Scoring Rubric

4 = **Excellent**. This paper

- demonstrates the ability to use the discourse and analysis appropriate to the academic discipline
- displays a clear understanding of the quotation and the task presented in the topic
- sustains a focused discussion
- uses evidence to support a point, e.g., uses consistently well-developed, well-chosen examples
- demonstrates an awareness of or insight into the complexities implied in the quotation
- avoids sweeping generalizations, clichés, and unsupported assertions
- displays a level of writing skill that supports and enhances the discussion

3 = **Good**. This paper

- demonstrates the ability to use the discourse and analysis appropriate to academic discipline
- displays a clear understanding of the quotation and the task presented in the topic
- sustains a focused discussion
- uses evidence to support a point, e.g., uses a single well-developed example or presents several pertinent, though not thoroughly developed, examples
- displays a level of writing skill that does not interfere with the conveying of information

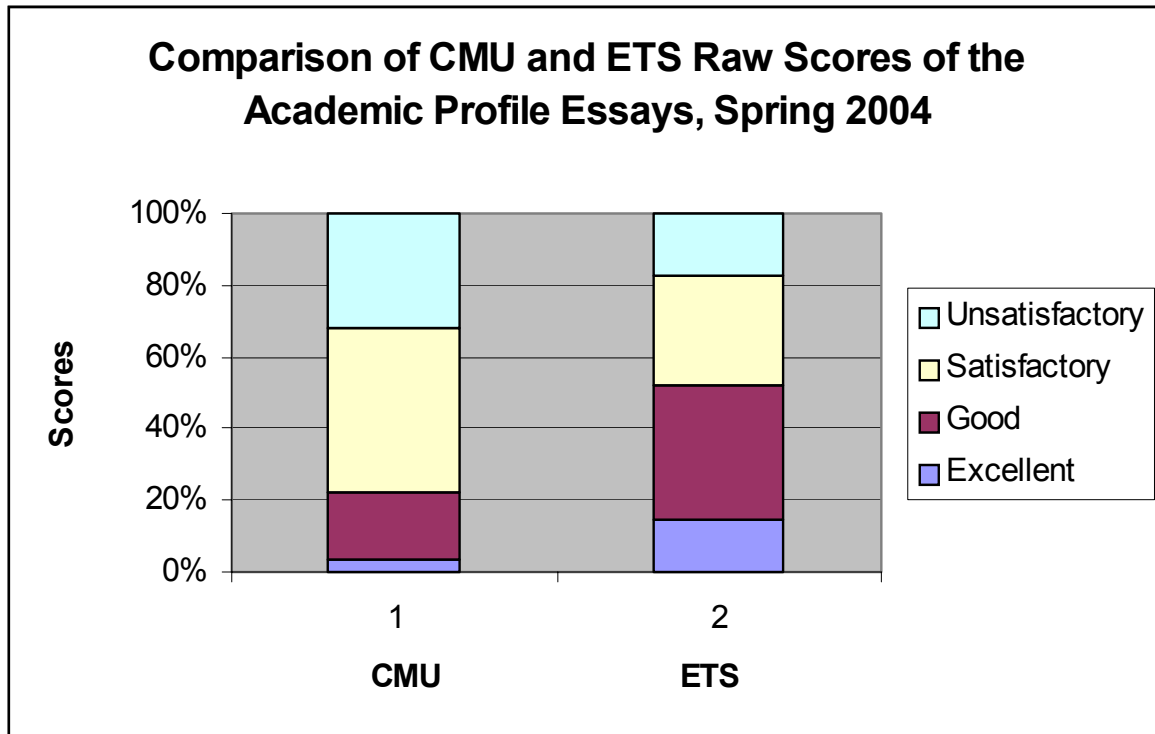
2 = **Satisfactory**. This paper demonstrates an understanding of the quotation but fails to address the task in one or more of the following ways. This paper

- depends on poorly selected or inaccurate examples from coursework
- fails to develop examples adequately
- merely lists (phrases, theories, authors, concepts)
- provides abstractions and generalizations related to the discipline or topic, but fails to develop, explain, or effectively incorporate them into the essay
- addresses only one part of the task
- provides well-developed examples but does not relate them to the topic

1 = **Unsatisfactory**. This paper fails to address the task presented in the topic in one or of the following ways. This paper

- fails to demonstrate understanding of the quotation and/or the task presented by the topic
- is so incoherent that the paper cannot be followed
- depends on feelings, beliefs, or clichés to develop the essay rather than knowledge of relevant coursework
- displays writing deficiencies so severe that the essay does not convey information

As can be seen in the chart and table below, our local assessment actually obtained *lower* scores than the ETS evaluation, with CMU scorers placing a larger proportion of essays into the categories of 2 (Satisfactory) and 1 (Unsatisfactory). ETS scorers placed 52.1% of the essays into the 2 or 1 categories, whereas CMU scorers placed 77.6% of essays into the 2 or 1 categories. Either way, it should be of concern to us that between half and three-fourths of the students who wrote essays performed so poorly. This fact certainly did not escape the notice of the ETS scorers, as was related in an email I received from Bill Wynne, Director, Academic Profile® and Major Field Tests, who wrote, “Our essay scoring staff were particularly surprised by the performance of CMU students on these essays.”<sup>2</sup>



4 = Excellent, 3 = Good, 2 = Satisfactory, 1 = Unsatisfactory

Scorers	4	3	2	1
CMU (n = 295)	3.8%	18.6%	45.6%	32.0%
ETS (n = 12,803)	17.2%	30.7%	37.5%	14.6%

<sup>2</sup> Email received Thursday, July 29, 2004, 5:01 p.m.

As will be discussed later, a number of factors other than lack of ability probably partially account for students' overall poor performance on the Academic Profile® essay, including flawed testing conditions and flawed writing prompts, which might have affected our students differently than students at the comparison institutions.

For ETS, the essays were only one measure of writing proficiency. In addition, the exam portion of The Academic Profile® included several multiple-choice questions *about* writing, and students' responses to these questions formed the basis of the ETS assessment of students' writing "Proficiency Levels."

### **ETS Proficiency Levels for Writing Skills**

#### ***Level 1***

At level 1, a student can

- recognize agreement among basic grammatical elements (e.g., nouns, verbs, pronouns and conjunctions)
- recognize appropriate transition words
- recognize incorrect word choice
- order sentences in a paragraph
- order elements in an outline

#### ***Level 2***

At level 2, a student can

- incorporate new material into a passage
- recognize agreement among basic grammatical elements (e.g., nouns, verbs, pronouns, and conjunctions) when these elements are complicated by intervening words or phrases
- combine simple clauses into single, more complex combinations
- recast existing sentences into new syntactic combinations

#### ***Level 3***

At level 3, a student can

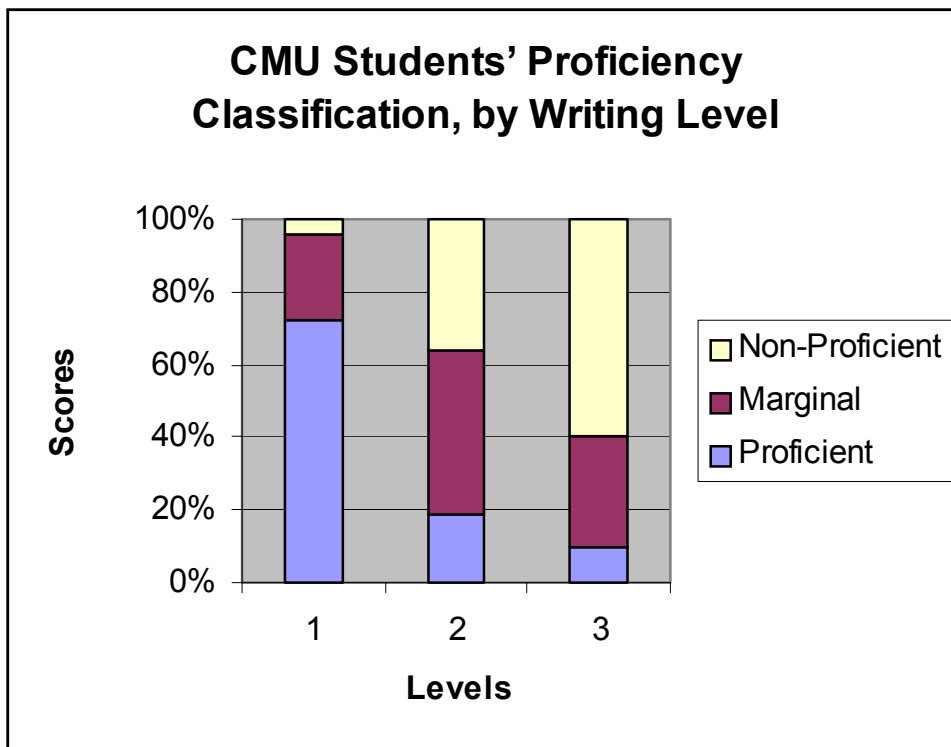
- discriminate between appropriate and inappropriate use of parallelism
- discriminate between appropriate and inappropriate use of idiomatic language
- recognize redundancy
- discriminate between correct and incorrect sentence constructions
- recognize the most effective revision of a sentence

ETS provided a summary of scores for each Proficiency Level as follows:

**CMU Students’ Proficiency Classification, by Writing Level (n = 295)**

The data also tell us that, in terms of their declarative knowledge *about* writing, our students aren’t doing very well, possessing only very basic knowledge.

Writing Skill Level	Proficient	Marginal	Non-Proficient
Level 1 (lowest)	72%	24%	4%
Level 2	19%	45%	36%
Level 3 (highest)	10%	30%	60%



As has been previously mentioned, ETS also compared CMU scores to those of 14 other Doctoral/Research Universities I and II.

### CMU Students' Writing Proficiency Levels Compared to Other Institutions

Skill Dimension	Proficiency Classification					
	Proficient		Marginal		Non-Proficient	
	CMU	Other	CMU	Other	CMU	Other
Writing, Level 1 (lowest)	72%	73%	24%	20%	4%	6%
Writing, Level 2	19%	23%	45%	43%	36%	35%

CMU ( $n = 295$ ), Other ( $n = 12,803$ )

In terms of their declarative knowledge about writing, our students did about as well as students at the comparison institutions, with students at the comparison schools scoring only slightly above CMU students at each level, and the question of whether that's good or bad news depends, in turn, on a number of other factors, which are discussed in the Conclusion section below.

Again, the comparisons between CMU and the comparison schools are meaningless unless we know 1) the academic quality of the comparison institutions, and 2) the educational environment of each institution, and 3) demographics of the student populations at those institutions.

## CONCLUSIONS

With respect to their exam scores, our students scored in the mid-range of possible scores and did as well (or as poorly, depending on one's perspective) as students at the comparison schools.

As for the essay scores, since over a third of the essays scored locally were scored as Unsatisfactory, we might have a problem, but it's not clear whether the problem is widespread or where the cause lies. After all, ETS classified only about 15% of the essays as Unsatisfactory. We must also keep in mind the different scoring criteria used locally and by ETS. The ETS scoring criteria are more focused on content than the CMU criteria.

Why did our students performed as they did on this exam and essay? Many of us involved in planning and administering this assessment project had hoped our students would do better, perhaps much better. There are several possible explanations as to why they didn't do better:

#### 1. *Problems with the students.*

- Perhaps some of the students who took the exam were unintelligent. Although there probably were a few "dummies" among the test participants, it's not likely that the majority were unintelligent, although it certainly depends on how one defines "intelligent."

- Perhaps some of the students who took the exam were ignorant, lacking the content- and procedure-based knowledge to do well on the exam and essay (and if so, then this problem is inextricably connected to possible problems with our curricula and teaching methods).
- Perhaps some of the students who took the exam were angry about being required by their instructors to take the exam. If so, then these students might not have taken the exam seriously. This possibility is directly related to problems with the sampling procedure (see below).
- Perhaps some of the students who took the exam were lazy or didn't care about the exam, and the latter problem might be related to the relatively low stakes for students taking the exam (see below). For example, proctors revealed that a few students didn't even open the test booklet but simply filled in their answer sheet and left.

## 2. *Problems with the Sampling Procedure and Testing Conditions.*

- Perhaps our sampling procedure created negative attitudes among some students. How? Because while some students were offered extra credit by their instructors for taking the exam, other students were *required* by their instructors to take the exam. As noted above, students who were required to take the exam may have felt angry or resentful, leading them not to take the test seriously. Test proctors and members of the Gen. Ed. Council (who attended each testing session to explain to students the purpose of the exam) indicated that some students were actually angry about having to waste their time completing the exam and essay.
- Perhaps instructors didn't do a good job explaining to students what the exam was about and why they were being asked or required to take the exam. Anecdotal evidence suggests that some students didn't know why they were taking the exam (e.g., one student thought he was there to take a test for his capstone course). Even though members of the Gen. Ed. Council provided at the beginning of each testing session a brief overview of the exam and its purpose, it certainly would have been better if all instructors had explained clearly to students the test and its purpose.
- Perhaps the low stakes caused some students to "blow off" the exam. Since scores were not going to be revealed to students, affect their GPA, or appear on their transcripts, many students probably did not take the exam seriously. Anecdotal evidence reported by proctors suggests that this may have been the case. In fact, some students wrote directly on the exam booklet that they thought the exam and essay were "a joke." Also, anecdotal evidence suggests that some students were perplexed about being tested over knowledge and skills that they perceived as unrelated to their major (again, it seems to have escaped their notice that this was a *General Education* exam!) A few students actually wrote statements to this effect directly on their essays.

3. *Problems with the writing environment.* Aside from occasions when students must complete the purely academic task of responding to essay exam questions in class, few college students today write essays by hand but instead use computers to do most of their writing. In fact, many of the English department's composition courses are taught in computer labs, and in these classes, students do all of their in-class and out-of-class writing on computers. Had students been allowed to write The Academic Profile® essays using computers, it is likely that, in some respects (e.g., organization), the essay scores would have been higher.
4. *Problems with the writing prompts.* Those of us who scored the essays locally had difficulty believing that ETS had developed such poor writing prompts.

First, the three prompts are not of equal level of difficulty. While each prompt begins with a quote to which students were asked to respond, the humanities and natural sciences quotes are relatively short, while the social science quote is quite long and more difficult to understand. In addition, the humanities and social science quotes were more abstract than the natural sciences prompt.

Second, the three prompts demand different types of critical thinking skills and elicit different, though not equally appropriate, writing styles and led students to respond to the three prompts in different ways. For example, most students who responded to the humanities prompt wrote first-person narratives with little analysis, citing mostly (and in many cases, solely) personal examples to support their claims, resulting in solipsistic essays of little merit. In contrast, students responding to the natural sciences prompt wrote in either third-person or first-person collective (e.g., "We biologists . . ."), and they cited discipline-specific examples, not personal experiences, to support their claims. Persons who responded to the social sciences prompts tended to use third-person, but as did those who answered the humanities prompt, they tended to cite personal examples or generic examples not tied to any of the social science disciplines cited in the prompt. Overall, responses to the humanities prompt tended to be what scorers referred to as "juvenile" in terms of vocabulary and level of sophistication compared to responses to either of the other two prompts.

Third, while the prompts were of some value in terms of assessing students' writing abilities, the prompts were of limited value for assessing students' *General Education* content-knowledge because many students tended to fall back on what they knew best, i.e., their major courses of study, not their UP courses. In fact, the ETS scoring criteria clearly reflect a disciplinary bias, which seems inappropriate for a *General Education* assessment.

For example, most students who answered the Natural Science prompt were also science majors who identified themselves as such in their essays, and they had no difficulty citing discipline-specific examples to support their claims. Also, most students majoring in fields other than humanities, natural sciences, or the social science (e.g., students majoring in finance) tended to answer the natural science prompt by citing examples from their own discipline rather than a natural

science discipline, and although this was an unacceptable response because it failed to answer the question that was asked, most of these students were able to adequately answer the question in terms of their own disciplines. But this tells us nothing about students' General Education knowledge, only about their knowledge of their own disciplines.

At the other extreme, we can only hope that the students who answered the humanities prompt were *not* humanities majors, because most students who responded to this prompt relied solely on their personal experiences, only a very few citing specific authors/artists and specific works, movements, and/or aesthetic theories as evidence to support their claims. And several students did what scorers referred to as "name-dropping", i.e., students referred in passing to artists such as Bach, Frost, Beethoven, or Hemingway, but did not refer to the authors or their works in any substantive way, and the ETS criteria clearly indicate that all such essays were to receive a score of 1 (Unsatisfactory), no matter how well they were otherwise written. Our local criteria also led us to assign scores to such essays of either 1 or 2, because this kind of "name-dropping" is not rhetorically effective.

So we actually did learn something that is relevant to General Education: based on *these* results in response to *this* prompt, our students seem ignorant of and can't write intelligently about the humanities or fine arts. This is also true, albeit to a lesser degree, with respect to the social sciences.

5. *Problems with Our Curriculum.* Perhaps we're not doing as good a job as we could
  - helping students to develop their critical thinking, reading, writing, and math skills. In particular, critical thinking skills are not explicitly incorporated into our curricula. It's as if we assume that if we require students to read, write, speak, and calculate, then critical thinking skills will naturally develop. Reading, writing, speaking, and calculating are necessary but not sufficient for the development of critical thinking skills.
  - helping students understand and value the relevance of a general, liberal arts education to *all* college graduates.
  - clarifying for ourselves and agreeing on the meaning of that relevance.
6. *Problems with our Teaching Methods.* Perhaps we're not doing as good a job as we could
  - engaging students in active learning.
  - varying our teaching styles to accommodate a range of student learning styles.
  - recognizing that just because one is an expert in one's field doesn't make one an expert *teacher*.

## RECOMMENDATIONS

Although the results of the essay portion of this pilot assessment are disappointing, we must acknowledge that the results might not be generalizable to the population of graduating seniors for 2004, let alone to the overall population of recent CMU graduates, but the results do support the findings of other recent assessments, such as the 2001-2003 NSSE and FSSE, as well as the recent survey of faculty opinions about student writing, which indicated that the most serious problem with our students' writing is not grammar, punctuation, and spelling, but that students' writing demonstrates a lack of critical thinking skills and a lack of ability to support claims logically using context-appropriate evidence.

If we want to obtain reliable, generalizable results from an assessment of learning outcomes (which have yet to be approved for the UP), one option would be to test graduating seniors each year for several years, then perhaps test only every three or four years. Graduation fees could be raised to cover the cost of the annual assessment (and this would be a minimal increase of about \$15.00 to cover the cost of the exams and scoring, and the cost of staffing the testing centers). We could then opt to score only a random sample of 10% of the population. However, for students to take the exam seriously, the stakes must be raised (e.g., by recording scores on students' transcripts, or at least by reporting scores to students themselves). Since the purpose of the assessment would be to obtain data for curriculum evaluation and revision, it is neither necessary nor desirable that students be required to pass the exam in order to graduate.

Finally, the scorers who locally assessed The Academic Profile® essays believe that CMU can do a better job than ETS in developing appropriate writing prompts and that, given the high degree of reliability among CMU scorers and the close agreement between CMU scores and ETS scores, CMU should develop, administer, and score its own writing competency assessments. We also believe that assessment of students' writing be based on a combination of at least two measures: 1) assessment of students' ability to write short essays in response to impromptu, timed-writing situations (preferably written in a computer lab), *and* 2) assessment of their ability to produce—over time and through a process of drafting and revising—more extensive, research-based texts. The latter would be best affected through a portfolio assessment, and some schools require all students to produce an “exit” portfolio as part of their graduation requirements, while at other schools, individual departments or colleges stipulate such a requirement for their majors. Not only do the portfolios provide valuable assessment data, but they also provide students with a valuable asset for their job search.