

## Memo

From: Mark Lencho  
Member, *ad hoc* Committee on Writing Assessment for General Education  
To: Steve Friedman  
Chair, Ad Hoc General Education Writing Assessment Team  
Re: Results from the Spring 2002 Writing Assessment  
Date: May 20, 2002

- **Background**

The assessment of student writing from the fall 2001 semester represents the fourth round of writing assessment aimed at evaluating the students' mastery of the set of objectives for the General Education program at University of Wisconsin-Whitewater. Previous assessments have helped in the development and refinement of the rubric used for the current assessment, according to which we evaluate student writing in terms of the three primary traits "thinking," "voice," and "literacy."

What began in 1999 as the assessment of student writing in the General Education capstone course "World of Ideas" (WOI) was broadened in the spring of 2001 to include the assessment of short, impromptu writing from World of Ideas and Freshman English 101. For the current assessment, the comparative dimension was developed a step further. For the first time, the assessment team of Assistant Professor Emily Hipchen, Associate Professor Michael Longrie, and myself evaluated term papers—sustained writing in the range of four to seven pages and incorporating research—from both 101 and 390, mixed together and all addressing similar topics.

- **Collection Procedures**

Unlike previous rounds of assessment where papers were solicited from all sections of World of Ideas, the current assessment is based on papers submitted by Dr. Karen Buckley, who arranged for her students in 101 and 390 to work on similar research projects concerning the impact of the World Trade Center attack on the arts or education. Dr. Buckley submitted a combined 57 papers, representing a comprehensive set of student work from a section of 101 and a section of 390. These papers were then mixed together before their distribution to the assessment team, information identifying writer and section having first been indexed and then held back. From the stack of 57 papers, 28 were randomly chosen for the assessment.

- **The Scoring Procedure**

Sets of papers were distributed to each reader on the assessment team, along with scoring rubrics and a set of instructions reviewing the procedure for scoring the papers according to the rubric. Benchmark papers from previous assessments were also included for each scoring level, along with commentary in which raters had recorded their scoring rationale. Readers then read and scored each term paper, supplementing scores with justifying commentary. Each of the three primary traits was scored on a six-point scale, where all scores were whole number ratings without additional annotation. The overall score for each paper reflected the average of the paper's scores for each

primary trait. After reading and scoring the set of papers independently, the three scorers of the assessment team met for two one-hour-and-one-half long sessions to discuss the papers and to compare ratings. Occasionally, but rarely, the discussions led to some refinement in the raters' scoring.

- **Agenda Established by the Scoring**

Each round of scoring leads us back to the ongoing agenda for writing assessment: a) Are papers trending the same, better or worse from semester to semester, insofar as can be ascertained by inspecting assessment team ratings (as evidenced by a comparison of average overall scores spanning the three assessments)? b) Are there differences in the overall evaluation of “thinking,” “voice,” and “literacy” as we compare the current assessment with the preceding one? c) Are scorers consistent as a group (as evidenced by strong inter-rater agreement coefficients)? Or do scorers diverge from each other in quirky or unpredictable directions? d) Do the primary traits as they have been defined represent discrete and independent analytical components of writing (as evidenced by significant standard deviations for each scorer on each primary trait)? Or is it the case, for example, that a high score on “thinking” on a given paper will automatically entail a high score in “voice” and “literacy” on the same paper for some or all scorers? e) Does the six point scoring scale adequately tease out qualitative differences among the papers (as evidenced by respectable standard deviations within the set of scores for each primary trait and within the set of overall scores)? Or is there clumping around a certain segment of the scoring scale, so that a part of the discriminating function of the scale is not used and the papers are not optimally differentiated?

- **Summary of the Data**

The assessment team scored 28 papers. Out of this total, 21 papers were written in Freshman English (680-101), while seven papers came from World of Ideas (GENED-390).

Overall average score for all papers and raters combined: **2.94**

Overall scores by raters:

Mike Longrie: **3.14**

Mark Lencho: **2.88**

Emily Hipchen: **2.79**

Overall scoring of “thinking”: **2.82**

Overall “thinking” scores by raters:

Mark Lencho: **2.89**

Mike Longrie, Emily Hipchen: **2.79**

Overall scoring of “voice”: **3.06**

Overall “voice” scores by raters:

Mike Longrie: **3.36**

Emily Hipchen: **2.96**

Mark Lencho: **2.86**

Overall scoring of “literacy”: **2.90**

Overall “literacy” scores by raters:

Mike Longrie: **3.28**

Mark Lencho: **2.89**

Emily Hipchen: **2.61**

Average score by writers in 101: **2.86**

Average score by writers in 390: **3.17**

Improvement index from 101-390: **.31**

101 averages by raters:

Mike Longrie: **3.10**

Emily Hipchen: **2.76**

Mark Lencho: **2.71**

390 averages by raters:

Mark Lencho: **3.38**

Mike Longrie: **3.29**

Emily Hipchen: **2.86**

Improvement index per rater:

Mark Lencho: **+.67**

Mike Longrie: **+.19**

Emily Hipchen: **+.10**

Improvement index per rater per primary trait:

Mark Lencho

Thinking: **+.53**

Voice: **+.76**

Literacy: **+.72**

Mike Longrie

Thinking: **+.40**

Voice: **+.10**

Literacy: **.00**

Emily Hipchen

Thinking: **+.10**

Voice: **-.14**

Literacy: **+.53**

- **Noteworthy**

Papers for this assessment were drawn at random from a mixed set of 101 and 390 term papers. To insure genuine random sampling, no effort was made to regulate the percentage contribution from the two sources. As it turned out, the sampling is rather unbalanced, with the great majority of papers coming from the freshman class. According to our scoring, freshman can expect to improve from the high range of “seriously flawed” to the low range of “limited” by the time they are juniors or seniors. Most students are not writing adequately, in terms of our assessment.

Mike Longrie saw the most improvement in “thinking,” Emily Hipchen in “literacy,” while I rated the papers as improving most in “voice.” In contrast, Mike saw no improvement in “literacy” and Emily scored the 390 papers as showing a slight decrease control of “voice” as compared to the freshman papers. Emily was not sure that her scores in voice were an artifact of her greater familiarity with freshmen writing, or if they were a reflection of genuine diminution in this quality during the course of student progress through their undergraduate programs. Overall, Mike and Emily rated the papers as strongest in “voice,” while I rated them weakest in voice. We showed the greatest convergence on “thinking,” where our overall scores were within .1 of each other, and where Mike’s improvement index and my own were only different by .13. Compared to the last round in which we scored a comparably sized batch of WOI papers, these fall 2001 WOI papers averaged approximately .3 to .4 lower in every category.

- **The Next Step**

Supplementing our findings by providing a statistical analysis of inter-rater and primary trait performance promises to provide further insight into the ongoing assessment agenda outlined above.