PROBLEMS AND SOLUTIONS IN USING OPEN-ENDED PRIMARY TRAIT SCORING

Speaker: Michael C. Flanigan, University of Oklahoma
Introducer/Recorder: Chris Anson, University of Minnesota

Michael Flanigan began by outlining his university’s plan for five years of experimental research on the teaching and testing of writing. Much of this research will replicate published research studies, but original research will also be conducted. All of the studies will be controlled experimental studies so that the researchers can be fairly faithful to the original ones and can analyze any differences between the original and new research.

Flanigan discussed one study, already completed, in which his colleague David Mair and he combined the strategies of two studies by George Hilllocks, an experimental study involving teaching extended definition using inquiry and models and a descriptive study dealing with “modes of instruction” (both of which Hilllocks discusses in some detail in this book Research on Written Composition). In the replicated study at Oklahoma, all twenty classes consisted of university freshmen; for nine of the ten teachers it was their second semester of teaching, and approximately 500 students were involved.

Flanigan pointed out that both Hilllocks’ studies because both dealt with significant areas in teaching and writing. Extended definition represents a kind of discourse that permeates almost all thinking and writing. The researchers believe that by replicating such an important study they could get inside the problems of the earlier research, and come to understand it better. The experimental extended definition study also used Hilllock’s open-ended primary trait scoring technique because the researchers wanted to learn to use and understand it better.

After reporting the findings from a small sampling of the data, Flanigan described some problems that he and his colleagues faced as they attempted to use Hilllocks’ open-ended primary trait scoring system and he discussed the modifications they made in it to obtain reliable results. He pointed out that with an open-ended primary trait scoring scale theoretically there is almost no limit to what students can score. Most scoring scales range from 1 to 6 (as in the holistic score for the ECT), 2 to 8 (as in CLEP), 1 to 5 (as in CORE scoring) and so forth. In open-ended primary trait scoring, the limit for a talented student is probably dictated by time and the variation and limitations imposed by the writing called for. In the papers scored in this study, the top score was 28.

The traits for which students could receive scores were: (1) properly putting an item in a class; (2) creating criteria for the class; (3) giving examples; and (4) providing contrasting examples to clarify and limit each criterion. Points were not given for differentiae as in Hilllock’s original study; instead, class and differentiae were combined (on the advice of Hilllocks when the study was set up). Hilllocks’ scorers had had problems reaching agreement on this point. Students could receive 2 points for the class, 2 for each criterion, 2 for each example, and 2 for each contrastive example. Obviously the more criteria, examples and contrastive examples students could come up with, the higher their score. In initial training, scorers had problems staying close together in the higher ranges, so Flanigan modified his tolerance of acceptability by allowing scores in the range 1 to 10 to differ by as point, 11 to 20 to differ by 2 points, and 21 up to differ by 3 points. Scores within that range were averaged; scores that did not meet acceptable standards were read by a third reader. If the third reading fell within range of either of the other two readers, then those scores were averaged. If there still was no agreement, a fourth and fifth reader scored the paper, and the paper and the range of scores were given to the researchers and a score was determined. For example, one paper was scored 6 and 8; a third reader gave it 10; the fourth reader gave it 9, and the fifth reader gave it 7. Its final average was an 8. Only seven papers required the fourth and fifth reader. Often, readers had problems keeping clearly in mind the kinds of criteria the writers were developing. To simplify the process, any one clear criterion could be accompanied by a number of examples and contrastive examples. If no criterion was given, only one example could be counted. If an undeveloped example or string of general examples was given, a score of 1 was given.

Flanigan concluded that open-ended primary-trait scoring offers real promise, for it allows for a kind of differentiation that closed, limited systems do not. However, researchers who use the system will probably have to modify it to get consistent, reliable scores. They will also have to plan their research so that the traits they are describing and scoring are clear, well-defined, and fully conceptualized by their scores. The session ended with the speaker giving participants six papers that had been scored by three readers and leading participants through a guided scoring session.

THE IMPLICATIONS OF THE RHETORICAL DEMANDS OF COLLEGE WRITING FOR PLACEMENT

Speaker: Kathryn Fitzgerald, University of Utah
Introducer/Recorder: Linda Jorn, University of Minnesota

Kathryn Fitzgerald gave participants attending this session a chance to analyze student writing in terms of rhetorical evaluation criteria developed at the University of Utah. These criteria are intended to do the following: (1) describe the rhetorical situation college students face when asked to write an essay that will be assessed and used to place them in a freshman writing course; (2) assert that the evaluation of the rhetorical situation provides valid criteria for placement of students into various levels of freshmen writing courses; and (3) shape the discussion...
of the student writing by providing a holistic view of writing.

Before handing out samples of student writing, Fitzgerald discussed the theoretical background for developing the rhetorical criteria. She also reviewed some of the common problems of holistic scoring, emphasizing the fact that holistic scoring does not consider the different purposes of writing. (For example, persuasive vs. self expressive writing). The rhetorical evaluation criteria developed at the University of Utah were designed to alleviate some of the problems encountered when using holistic scoring. The criteria help readers consider the purpose of students' writing and identify the internal and external purposes of the writing situation.

Fitzgerald pointed out that students' internal and external purposes complicate the writing situation for them. At the University of Utah, faculty feel that the purpose for students' writing needs to come from the students (i.e., internal), but in academia the purpose often comes from the instructor and is motivated by grades (external). The student has to think up his or her purpose for writing and must shape this purpose to serve the academic external purpose. Therefore, the student's purpose is always dual. These internal and external purposes are in essence the rhetorical situation and they need to be taken into account when faculty evaluate writing, particularly when this evaluation is used to place freshmen into English courses. Students' ability to handle this complex rhetorical situation informs instructors of the students' readiness for college writing.

Next Fitzgerald described how the rhetorical expectations of University of Utah professors were determined and used to develop the rhetorical evaluation criteria. These criteria consist of the following categories:

Category 1: The writer's relationship to college readers and writers. **Expectations:** The most proficient writers recognize that any single piece of college writing is part of an ongoing written discussion about a topic and that they are expected to make a contribution to the discussion. They recognize that an authority (i.e., professor, test giver) identifies issues for discussion.

Category 2: The writer's relationship with his or her subject matter. **Expectations:** College writers control their subject matter, pressing it into service to support their internal and external purposes.

Category 3: The writer's relationship to the conventions of the genre. **Expectations:** College writers employ syntactical units appropriate to their thought, precise vocabulary, and the mechanics and spelling of standard written American English.

University of Utah students are given placement essay directions that explain the external rhetorical situation; and they have 45 minutes to plan, write, and revise their essays.

After reviewing the theoretical background and the criteria, participants used these criteria to evaluate and discuss some student writing. Fitzgerald pointed out that readers are told to pay attention to content and reasonableness, that there are no hard and fast rules, and that judgment is a balancing act of various criteria and expectations of each institution. Readers at the University of Utah look at the quantity of students' writing as relative to every piece of writing. In summary, Fitzgerald stated that these rhetorical evaluation criteria force readers to evaluate writing for its purpose, help readers define good college writing, and address the need to teach students about the effect that the rhetorical situation has on their writing.

**USING VIDEO IN TRAINING NEW READERS OF ASSESSMENT ESSAYS**

**Speaker:** George Cooper, University of Michigan

**Introducer:** Terence Collins, University of Minnesota

Large scale testing programs face a recurring problem of reader consistency and reliability. In this presentation, George Cooper demonstrated how the English Composition Board at the University of Michigan uses a video format for training new readers, and for disseminating information about the ECB's procedures to various campus constituencies. While Cooper presented alone, his remarks were prepared with Liz Hamps-Lyon.

In its placement readings, members of Michigan's ECB teams are guided by statements of criteria clustered under three headings: "structure of the whole essay," "smaller rhetorical and linguistic units," and "conventions of standard English surface features." Students write essays in response to prompts that define a situation and provide several choices of opening sentences. Two important characteristics of the 6000 student essays, then, are that topic choice is limited and orientation toward the topic is guided through provision of choices for essay openings. Further, the essays are rated for placement: recommendations fall into one of the following categories: exempt (7%), Introductory Composition (82%), and tutorial (11%). These recommendations reflect scores of 1, 2-3, and 4. While criteria for quality are outlined to readers, no specific calibration of trait content for the four point range is provided.
of the student writing by providing a holistic view of writing.

Before handing out samples of student writing, Fitzgerald discussed the theoretical background for developing the rhetorical criteria. She also reviewed some of the common problems of holistic scoring, emphasizing the fact that holistic scoring does not consider the different purposes of writing. (For example, persuasive vs. self expressive writing). The rhetorical evaluation criteria developed at the University of Utah were designed to alleviate some of the problems encountered when using holistic scoring. The criteria help readers consider the purpose of students' writing and identify the internal and external purposes of the writing situation.

Fitzgerald pointed out that students' internal and external purposes complicate the writing situation for them. At the University of Utah, faculty feel that the purpose for students' writing needs to come from the students (i.e., internal), but in academia the purpose often comes from the instructor and is motivated by grades (external). The student has to think up his or her purpose for writing and must shape this purpose to serve the academic external purpose. Therefore, the student's purpose is always dual. These internal and external purposes are in essence the rhetorical situation and they need to be taken into account when faculty evaluate writing, particularly when this evaluation is used to place freshmen into English courses. Students' ability to handle this complex rhetorical situation informs instructors of the students' readiness for college writing.

Next Fitzgerald described how the rhetorical expectations of University of Utah professors were determined and used to develop the rhetorical evaluation criteria. These criteria consist of the following categories:

Category 1: The writer's relationship to college readers and writers. Expectations: The most proficient writers recognize that any single piece of college writing is part of an ongoing written discussion about a topic and that they are expected to make a contribution to the discussion. They recognize that an authority (i.e., professor, test giver) identifies issues for discussion.

Category 2: The writer's relationship with his or her subject matter. Expectations: College writers control their subject matter, pressing it into service to support their internal and external purposes.

Category 3: The writer's relationship to the conventions of the genre. Expectations: College writers employ syntactical units appropriate to their thought, precise vocabulary, and the mechanics and spelling of standard written American English.

University of Utah students are given placement essay directions that explain the external rhetorical situation; and they have 45 minutes to plan, write, and revise their essays.

After reviewing the theoretical background and the criteria, participants used these criteria to evaluate and discuss some student writing. Fitzgerald pointed out that readers are told to pay attention to content and reasonability, that there are no hard and fast rules, and that judgment is a balancing act of various criteria and expectations of each institution. Readers at the University of Utah look at the quantity of student writing as relative to every piece of writing. In summary, Fitzgerald stated that these rhetorical evaluation criteria force readers to evaluate writing for its purpose, help readers define good college writing, and address the need to teach students about the effect that the rhetorical situation has on their writing.

**USING VIDEO IN TRAINING NEW READERS OF ASSESSMENT ESSAYS**

**Speaker:** George Cooper, University of Michigan

**Introducer:** Terence Collins, University of Minnesota

Large scale testing programs face a recurring problem of reader consistency and reliability. In this presentation, George Cooper demonstrated how the English Composition Board at the University of Michigan uses a video for self-monitoring within the reader cadre, for training new readers, and for disseminating information about the ECB's procedures to various campus constituencies. While Cooper presented alone, his remarks were prepared with Liz Hamps-Lyon.

In its placement readings, members of Michigan's ECB teams are guided by statements of criteria clustered under three headings: "structure of the whole essay," "smaller rhetorical and linguistic units," and "conventions of standard English surface features." Students write essays in response to prompts that define a situation and provide several choices of opening sentences. Two important characteristics of the 6000 student essays, then, are that topic choice is limited and orientation toward the topic is guided through provision of choices for essay openings. Further, the essays are rated for placement: recommendations fall into one of the following categories: exempt (7%), Introductory Composition (82%), and tutorial (11%). These recommendations reflect scores of 1, 2-3, and 4. While criteria for quality are outlined to readers, no specific calibration of trait content for the four point range is provided.