through the process the panel uses to develop the rubric.

Phyllis Liston began by describing what the exam coordinators learned in the process: (1) implementing, coordinating and gaining community-wide acceptance for exit exams is "a lot harder than it looks"; (2) communication at all levels is essential; (3) low-level mistakes can cause high-level difficulties; money when needed is found; and (4) holistic grading works well. In addition, the exam needs full administrative and faculty support. As the director of the writing competency exam, Liston found the administrative duties to be a full-time responsibility requiring personnel assistance.

Liston explained Ball State's "3/3/3" exam process. Students sign up for the exam three weeks before the exam date and are given an instruction sheet detailing the exam process, the exam question, how to prepare for the exam, and where to go to receive help preparing for the exam. On the exam date, students are given three hours to write approximately three pages in response to the exam question. Students must pass the test to graduate. After two attempts, they are required to enroll in-and repeat until they pass—an upper division writing course. The second opportunity to take the exam constitutes an automatic appeal. Exit from the course is by portfolio prepared by the students with the help of their instructors. Portfolios are evaluated by two or more readers other than the classroom teacher/coach. No student takes the exam more than twice.

John Mathew explained the training process for holistic graders by taking participants through a mini grading workshop. We read and ranked three sample essays high, middle, and low. Then we read, ranked, and integrated into the previous essays three more, and did the same for two additional essays. We then discussed our ranking of one of the essays in terms of its strengths and weaknesses. Finally we were presented with the six-point rubric developed by the panel for the particular exam and were asked to rate the essay.

In an actual reading, graders read ten papers at a time, assess, record and score, and pass the papers on to a second reader. Papers with scores that do not match are given to a third reader. All pass decisions are made by the University Provost under the advisement of the panel and other administrators after all exams for the quarter have been scored. The panel acknowledges a high reader calibration and suggests a main reason for it is that readers do not know the cut-off point for failing, and therefore are more objective and not sympathetically influenced to pass a borderline paper.

Linda Pelzer described the rubric design process. The panel develops a new rubric for each exam by reading and sorting all essays written for the exam into high, middle, and low categories. After sorting, they discuss the categories and write about them, and then draft a six-point rubric—one that is quite detailed and descriptive and that includes specific examples from student papers to illustrate the rubric's categories. A six-point rubric is used because it eliminates a middle score and because a four-point rubric would not be specific enough to encompass the aspects of the writing they wish to assess. The panel takes care and time in designing the rubric to make it clear and specific in order for readers to reach consensus and to withstand criticism from students, parents and faculty. Rubrics are kept on file at the University library. One indicator of the success of the rubric is that students who fail the exam and wish to contest it usually reach agreement after examining the rubric and evaluating their own writing against it.

Although the writing competency examination project is bigger than the panel first anticipated, they agree that it is worth the work.

PROFICIENCY TESTING: ISSUES AND MODELS

Speakers: George Gadda, University of California, Los Angeles
Mary Fowles, Educational Testing Service, New Jersey

Introducer/Recorder: Adele Hansen, University of Minnesota

George Gadda opened the discussion with a statement concerning general issues in developing a proficiency testing program. Proficiency testing, like achievement testing, measures success in a particular domain. There are several motivations for proficiency testing: to certify individual achievement exclusive of grades, to validate a program's effectiveness, or to screen before certification of passing to the next level of instruction. The choice of purpose governs the rest of the assessment program. Proficiency tests may be used to exempt students from further work; to prove value added in a course program; to permit passage, graduation or certification; or to identify those who need further instruction.

Gadda noted that test-makers should define the domain of the test by describing the kind of written ability being assessed and that we should make a public statement concerning the criteria used for judgment. Tests used for advancement should be a well-defined part of the curriculum, with samples and grading criteria clearly described. Ideally, scorers should be those people who are testing and using the results. In addition, we need to determine what will happen to those who don't pass. Gadda noted that proficiency tests should not be a "roadblock." He concluded by stating that we should strive for high reliability and validity in our testing because proficiency tests need to withstand legal challenges.

Mary Fowles remarked that we need an increased understanding of what is to be tested and that the "community" must share the same standards. She referred to a project in Rhode Island, where a state administrator
decided to work on literacy beginning in the third grade. ETS was asked to construct a test that encouraged good writing. They worked with local administrators and teachers from every school district in the state to formulate a writing test which was administered to all 3rd graders. The test featured a pre-writing section and then an essay test. It also included an editing phase, where students were given specific questions about content.

Fowles described how scorers were trained: Every district in the state was represented in training sessions, and benchmark papers were identified and then used to train local raters. After the results were tabulated, the teachers returned to the classroom and showed examples of good papers to the students and discussed the scoring criteria. Next, the state decided to develop a portfolio of such "assignments" to validate the scores on the "test" and to enhance teaching.

In the discussion that followed, questions were raised concerning the "read and respond" type of test. Gadda agreed that such a test does assess reading as well as writing, but that there is a connection and such tests are useful to determine the students basic ability to do university level work. He added that such tests seem most fair, because all students begin with the same information and the students can then better understand the testing situation. He cautioned that such tests should always be pre-tested to discover if the reading is "accessible and interesting" and if the assignment elicits more than one response, because this can affect raters' evaluations.

CREATING, DEVELOPING, AND EVALUATING A COLLEGE-WIDE WRITING ASSESSMENT PROGRAM

Speaker: Suzy Groden, University of Massachusetts, Boston
Introducer/Recorder: Geoffrey Stre, University of Minnesota

In this session, Suzy Groden reported on the University of Massachusetts' writing assessment program, on-going since 1978. She described how it was developed, changed, and validated. The exam is a "rising junior exam," required of students after 68 credits (or within first semester for transfer students). Called a test of writing proficiency, the exam really tests reading, writing, and critical thinking because students have to respond to questions on texts (or "reading sets") with which they are provided one month prior to the exam. Students are either judged proficient or must remediate their writing skills.

The idea behind the test is to teach students what the faculty want them to know in various core curriculum courses, courses designed to include elements of critical analysis and reading/writing associated with that discipline. Each reading set is 20 pages and focuses on a controversial topic associated with a specific discipline. Students chose one of three sets, from natural sciences and mathematics, the social sciences, or the humanities, and they read the set for a month. After the first exam was given in 1978, a sample for students became available and a student manual was developed.

Groden stated that one problem in the exam is the lack of a penalty for those who fail. The exam is graded by readers who are trained in one morning and then read exams all afternoon. A student needs two readers to agree in order to pass, and three readers to agree in order to fail. But there are actually no practical penalties now associated with failing the exam: students can still take upper-division courses if they fail, and there is now an alternate way to demonstrate proficiency—a portfolio.

During the course of subsequent years, changes occurred in the context of the exam. After Groden and the university's ESL Director became involved, an interest in writing and the acquisition of language found its way into the readings. Policies surrounding the implementation of the test were gradually loosened. The use of the portfolio alternative was extended, particularly to ESL students. Also, the range of writing samples included in the portfolio was expanded to include more than just the traditional analytical paper: lab reports, for example, would be accepted. Students were allowed three hours to write the exam, rather than just two. And in one of her more striking findings, Groden found students wrote more easily when they switched from the standard-size blue books to the larger, 8- and 1/2-inch size (that being the standard in which they most frequently composed). The exam committee also spent more time thinking about readings and questions; the exams became more complicated, involving ideas about the nature of knowledge. What ultimately evolved were two possible questions, one for the non-intellectual and one for the more challenging intellect. Finally, they also offered an evening session for taking the exam.

There were also many changes over the years which Groden termed losses. Faculty involvement waned, with more and more responsibility for grading falling to the exam committee. The school changed, taking in fewer freshmen and more transfer students, with the exam becoming a kind of graduation test. Funding dried up, causing the university to retreat from its core curriculum and limit the number of its core courses, and, hence, severing the relationship between the curriculum and writing proficiency.

One area in which the Massachusetts exam developers were successful was in establishing grading criteria. Sending exam samples to national experts, Groden received strong validation and agreement on their criteria. The experts, however, were critical of the number of questions, feeling they made different cognitive demands and were unfair. The committee is continuing to revise the exam.