

# LONGITUDINAL RESEARCH ON WRITING ASSESSMENT

Speakers: Paul Ammon, *University of California, Berkeley*  
Catharine Keech, *San Francisco State University*  
Introducer/Recorder: Juan Aninao, *San Francisco State University*

Paul Ammon and Catharine Keech presented evidence which pointed to difficulties in the longitudinal assessment of writing development. Ammon began by stating that evaluators and researchers in the field, when engaging in longitudinal assessment, assume that students' writing "gets better and better," but he pointed out that writing sometimes "gets worse before it gets better," that "dips" often occur in student performance over a given period of time and instruction. These dips might indeed complicate the task of writing assessment.

Ammon discussed a study in which (over the period of a school year) the writing samples of bilingual children at the third and fifth grade levels were examined by means of both analytic criteria and holistic scoring methods. He found that overall scores generally did not show significant gains over the school year. The sum of the analytic scores was more sensitive to gains, but only among students who began the year at low levels of performance. In attempting to interpret these results, Ammon noted that some of the analytic scores actually went down, a factor which probably accounted for the general impression of little or no gain (as a gain on one dimension was cancelled out by a dip on a different dimension). Such dips could be explained by a number of factors: the student, when given a new version of the "same" task, may redefine it in a more complex way; newly acquired writing strategies may be overgeneralized to situations where earlier strategies would have been more appropriate in attempting something new; the student may experience cognitive overload which brings about deterioration in some aspects of his or her performance.

Ammon articulated some important implications from his research findings: 1) that pre- and post-test comparisons may be problematic, because even though "parallel" topics might be given, the tasks as defined by the writer might be different; 2) that multiple analytic scores are helpful but cannot simply be added up, because gains in one category might be cancelled out by lower scores in other categories, and 3) that, while we lack a theory of writing development sufficient to guide our use of analytic scores for assessment purposes, collection of longitudinal data may be helpful in forming such a theory. Ammon closed by offering a suggestion through which the testing model, with its assumption that writing "gets better and better," might be better aligned with the "teaching model," in which teachers should ideally acknowledge that students' writing often "gets worse before it gets better": In terms of assessment, students should be given credit for the positive changes which have occurred and should not be penalized for the negative changes which may occur at the same time.

Catharine Keech offered evidence of how these "dips" in scores appeared in a longitudinal assessment study of a high school population, and offered some

explanations of why they occurred. During the six years of annual assessment at the high school, improvements in curricula and approaches in writing instruction were being implemented. It would be natural to expect that through instruction and the natural effects of maturity, students' writing would get better, that a step-wise progression would occur from one year to the next for a given age group moving from freshman to senior status. But once again, assessment results showed dips in performance. Keech found that differences in reader standards each year could not fully account for the counter-intuitive sequences which occurred in the longitudinal test scores of the students. A key element was how the students defined the writing tasks: what they attempted in response to the assignment. When the assignment or test question changed, were students able to draw on their knowledge of a variety of approaches? Or did they have access to only one approach which may have worked well for an earlier task but may have been less appropriate to a later assignment? Keech offered the analogy of finding one's way with or without a map. When asked to write narrative, students were already well-familiar with the terrain, had an internal map, and could concentrate on the scenery without getting lost—for most, storytelling functioned as a "familiar well-marked freeway." Others, when they left the freeway, wrote something like oral dyad, not giving the reader a frame of reference or establishing for a reader why they are writing. To be able to "write an essay" successfully, the students need a new map, one through which they could move from a narrative "sequence of events" to a "point by point" form of discursive writing. Without a full understanding of this map and the ways in which it differs from other maps, the students may define the task according to the map they are familiar with, thereby producing narration or oral dyad, and receiving lower scores than on a previous task for which the old map was appropriate. Or, in attempting to "follow the map" closely, students may produce an "essay," but in the process forget that in producing the piece of writing they are supposed to "say something."

Thus, they may produce an essay which follows the format well but which is relatively devoid of insights or substance, and so may receive a lower score than an earlier, less formally-structured piece in which the student freely explored more interesting insights. In addition, the move into formal exposition and argument may require so much attention to map-reading—or figuring out the new discourse level rules—students may lose control over mechanics or fluency at the sentence level. Keech then presented a series of "personal reflection" topics through which students could progress from storytelling, to "tagging on commentary and reflection," to engaging in full analysis of a problem, in concept formation and extended abstract discourse. In concluding, Keech stressed that we need to build into the assessment process ways of recognizing and taking into account when the task has become more difficult—whether as a result of tester's demands or as a product of the student's own ambitions.

During the question and answer period, one audience member brought up the Dartmouth Study and asked how the findings of the two panelists related to that study. It was pointed out that as college students moved from rote learning to abstract understanding, the quality of their writing went down, a finding which confirmed Keech's findings.

(Continued on page 22)

## LONGITUDINAL RESEARCH (continued)

Keech discussed one of her studies in which the error counts stayed the same for pre- and post-test samples for a freshman composition group, but she noted that, in the post-test samples, there was a marked difference in the *kinds* of errors the students made. Another audience member pointed out that to do well on a proficiency test, a timed situation, students must have teachers who will "teach to the test." Keech affirmed this, noting the need for making students "test wise," and added that in the evaluation project she had helped lead, the students could not do well because within the course of their instruction, they had learned the importance of time as a necessary element in the writing process.