THE POWER OF READING, OR THE CONSEQUENCES OF TESTING LANGUAGE SKILLS

Speaker: Stephen Krashen, University of Southern California
Introductor/Recorder: Lawrence A. Luck, Cuyhoga Community College, Ohio

Stephen Krashen devoted his presentation and follow-up discussion to research done on the effects of "reading exposure," which he defines as the reading children and adults do for fun and personal interest. The results of the research done so far indicate that people who read for pleasure do significantly better in all areas of language use: reading comprehension, vocabulary development, grammatical correctness, writing style, and spelling. These findings point to the importance of having the opportunity to read for pleasure without the additional baggage of assigned readings, comprehension questions, book reports, vocabulary drills—all the trappings of standard academic readers. Krashen maintains that it is this opportunity which is lacking in most educational systems.

Krashen emphasized the correlation between reading and writing. He noted that it is not the number of books read which is a predictor of academic writing success; rather, it is the level of sophistication of that reading which predicts how successful a reader and writer one will become. But even a reader who never goes beyond the level of comic books will achieve minimal success in writing, and greater success than one who has never done pleasurable reading. Krashen cited several research studies which support the link between reading and writing. Surprisingly, not a great deal of research has been done in this area. The first research Krashen discussed is on sustained silent reading (SSR) programs, in which school children pick their own reading material and read for varying amounts of time each day, though they are not held accountable for their reading. Most of the studies Krashen referred to show SSR to be superior to traditional classroom reading skills programs. In one particularly noteworthy experiment, fourth through sixth grade students received SSR as their total English program. By the end of the second year, SSR students were rated superior in all aspects of English tested to students taught in conventional ways. Krashen concluded that, at worst, SSR produces the same results as a standard language arts program. At best, if the system operates a year longer, it works better than conventional programs.

Citing another series of studies that show the superiority of self-selected reading over assigned readings in a traditional reader, Krashen pointed out that students are more enthusiastic and better disciplined in self-selected reading programs. In studies where there was no comparison group, gains in reading comprehension and other skills were compared to norms on standardized tests. Here again, the results in the majority of cases were positive for self-selected reading. Even in the area of spelling, Krashen reported that, although some schools spend a great deal of time teaching spelling, the time spent produces no appreciable gains in ability. In fact, according to one study, by the seventh and eighth grades, those students who had received no systematic spelling instruction scored the same on spelling tests as students who had been exposed to several different spelling programs in the preceding school years. Krashen's conclusion is that the time given over to spelling and the teaching of other reading skills in many school districts could more profitably be devoted to self-selected programs of SSR. It would also remove a lot of the burden now placed on teachers.

Krashen is convinced that reading for pleasure and interest is the most important educational component and the most reliable predictor of future academic success. Reading is the one skill which has the most predictable and positive effect on the development of all language skills. Since Krashen believes that writing styles and the conventions of writing are gained through pleasurable reading, he concludes that tests of writing are actually tests of reading.

PLACEMENT TESTING: ISSUES AND MODELS

Speaker: William Smith, University of Pittsburgh, Pennsylvania
Introductor/Recorder: Sue K. Ham, University of Akron, Ohio

As he introduced his topic, William Smith reminded his audience that when they use placement tests, they must first assume that "places" exist; that is, there must be various levels of instruction in which to enroll students to optimize their progress. There are also many ways to derive such placement, the most common being SAT scores and/or direct assessment with an essay. Even though placement essays may be graded holistically, such assessment is not necessarily the same as purely holistic rating. Since the placement essays may represent an abnormal distribution of the population, one cannot expect to always see bell-curved data when examining test results.

Smith suggests seven steps in a basic schematic for a placement program: (1) Determine your program's theory of composition and purpose(s) and teaching method(s). (2) In relation to that theory, analyze the student population's writing abilities. (3) Determine what the composition groups will be—how many groups there will be and what each group will be taught. (4) Design an assessment method which is consistent with the theory. Depending on your theory and your purpose, that method could be a grammar test, SAT scores, or an essay. (5) Design the actual assessment tool. (6) Field-test the assessment tool with students. (7) Revise the test based on the results of the field tests. Other important items that any placement program must consider include the time to allow for testing, the date of the test, the money available for testing and scoring, and how high or low to make the pass/fail cut-off. Since such considerations and the needs of individual institutions and individual students vary so greatly and change over the years, one must not be content to do anything just once; it takes years to develop an appropriate program, and one must be content with, even enjoy, the sometimes slow process of discovering what it is one needs to know.

Smith pointed out that it is usually necessary to try out several different types of writing assignments before deciding what kind of topic best fits the particular goals
of the placement program. At Pittsburgh, Smith tried three different assignments: (1) Students were given an open topic; they were asked to write about a time when they were creative. (2) Students read a passage by D. H. Lawrence that described a time when he was creative. They were then asked to write about how Lawrence was creative and asked to respond in terms of their own creativity. (3) Students were given three passages by separate authors (all about being creative) and asked to read and respond to them. On the first assignment, the high-scoring students were rated far superior to the middle and low writers. On the second assignment, the high, middle, and low-scoring students could still be distinguished, but their scores were much closer. The third assignment, which required the reading of three passages, easily produced the distinct scorers, while both the average and superior writers scored higher and closer to each other. Such results prompted Smith to point out that a school needs to try out various topics and discover which ones best suit the student population, and that the department actually wants to consider. Smith suggested that teachers adapt topics that have proved more successful in the lower room. Such topics should be fairly representative of what a student will actually encounter during the course of the semester in order to better predict classroom performance.

In determining how many topics to use, one must consider how much time and money are available for placement programs. If one uses twenty-twenty-minute topics, the raters will need more time, but they can get a fairly good idea of what a writer's average work is like. Once again, one must decide whether it is an average writing sample that the program elicits or whether it is the students' best writing or worst writing that is desired. When evaluating the essays, raters must decide what it is they are to use as criteria for their evaluation—production of words, gut feelings, coherence, or something else. In any case, their criteria must be stated so that someone else can use them to help the students.

The actual conditions under which students take a placement test, of course, make a difference in their performance. After extensive evaluation, it appears that students do best on a summer placement test if it seems as much like a classroom test as possible. When given two hours to complete their writing, 74 percent of the students in Smith's study took between sixty and ninety minutes. Only 6 percent took more than ninety minutes. However, those who took less than sixty minutes actually wrote more than the other students. Other findings were that smaller testing groups produced better writing, as did the presence of a teacher and the use of a warm-up activity. In determining the reliability of the raters, Smith and his colleagues followed up on their students. They found that students who were simultaneously rated very high and very low by different raters tended to get better grades as they progressed in their writing courses than did those who were consistently rated as "middling" by various raters. As a result, this particular phenomenon of wide-ranging scores is also taken into consideration when placing students.

Smith concluded by reminding the audience not to buy anyone else's system, for if it doesn't fit, a school may be buying someone else's mistakes. He encouraged each institution to consider its particular needs, time, money, and faculty commitment. If an assessment program is theoretically sound and its directors can sleep at night, it will probably be quite satisfactory, given enough time, patience, and imaginative hard work.

MATCHING ASSESSMENT PURPOSES AND INSTRUMENTS


Introduction: Ellen Kibo Summy, University of Akron, Ohio

Gretchen Glick discussed four interrelated options that testers and educators must consider in order to match a writing test to the testers' purposes: formats, prompts, scoring, and reporting. When choosing a test, testers must first consider the information they want to obtain, the purpose of the testing, and their resources—both financial and temporal. Indirect assessment of writing, done through an objective instrument, may be adequate and desirable in some circumstances, but most programs use direct measures (a writing sample) for purposes of placement and admissions, research, and proficiency.

The ultimate purpose of direct writing assessment is, of course, to improve instruction and to benefit students.

Once the testing purpose is established, testers must develop an assignment or prompt designed to produce the kind of information that they want. Mechanical skills, such as punctuation, spelling, and vocabulary, can be evaluated from a diversity of prompts. However, if the purpose is to evaluate content or organization, all students must be given one prompt, and that prompt must have directions about the topic and the type of writing, whether descriptive, narrative, expository, or persuasive. Then, all papers written in response to a prompt can be compared to a standard—either a model or a set of guidelines. When all students write to the same prompt, their papers can be compared to each other and ranked. In choosing the prompt to be used, testers must consider their program, its texts, and its writing instruction. They should also consider what type of writing is appropriate for the students' ages and abilities. Finally, the topic used should interest students, and the prompt should be general enough to give them something to think about, but specific enough to focus the content of their writing. Sometimes, more than one prompt will be used to provide more information to evaluators; however, the testers' rule is to do the minimum amount of testing that will provide the information needed.

Next, testers must consider scoring options, including the method, the scale, and the location. Although there is no nationwide consensus on the best method for scoring writing samples, the three used most are holistic, analytic, and primary-trait (but even these labels have different meanings to different groups of educators). In choosing a method of scoring, testers must focus on two issues: the testing purpose and the kind of information desired. For instance, holistic scores can be used to evaluate programs and rank students; analytic scoring can give teachers diagnostic information on individual weaknesses so that they can plan classroom instruction. The point scale for rating is also based on the testing purpose. To separate writers into ability levels, a four-point scale might be used: good, acceptable, below average, unacceptable. But to give a more detailed evaluation, an eight- or ten-point scale might be used. The choice of a location for scoring involves the choice of evaluators. With purpose as well as resources in mind, testers could choose professional evaluation with experienced evaluators and computer-monitored, bias-free scoring.