WRITING SAMPLE RELIABLY EVALUATES NON-NATIVE SPEAKER’S SKILLS

INTRODUCTION: MEASURING STUDENT WRITING SKILLS
The number of immigrant students and other non-native speakers of English enrolling at the College of San Mateo (CSM) has steadily increased during the past five years. This change in the student population poses new challenges and opportunities. In order to provide the appropriate academic support, the College needs accurate placement instruments which profile student abilities to speak, listen, read, and write English. CSM’s Non-Native Speaker’s (NNS) program uses three placement instruments to recommend appropriate courses at five skill levels: (1) a reading diagnostic; (2) a listening comprehension test; and (3) a writing sample diagnostic. The NNS program offers students a course of study fitted to individual abilities at various skill areas.

Special problems occur whenever “direct” measures of writing performance (i.e., writing samples) are employed in testing programs. First, student performance on writing sample prompts should be consistent—regardless of which prompt a student receives [Inter-Prompt Reliability]. Second, because scoring of student writing samples involves variability such as multiple readers, essay scoring criteria must be consistently employed [Inter-Rater Reliability]. Third, placement recommendations should reflect the skills necessary for success in courses into which students are placed [Criterion-Related Validity].

The following presents the results of studies into three identified issues of “reliability” and “validity” associated with the NNS writing sample diagnostic. Overall, the purpose of these studies is to assess the extent to which CSM’s writing sample yields sound advice for both students and counselors/advisors regarding writing competencies.

RESEARCH FINDINGS: INTER-PROMPT RELIABILITY
For reasons of test security, the NNS writing diagnostic employs six different prompts. Ideally, student performance should be the same regardless of the writing prompt employed. Prompt reliability was assessed using a study design with 249 students enrolled at various NNS course levels. Students were tested/re-tested [with different prompts] in class during the Fall 1994 semester. A total of 498 essays were evaluated.

<table>
<thead>
<tr>
<th>Prompts</th>
<th>Consistent Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>85.2%</td>
<td>89.6%</td>
</tr>
<tr>
<td>84.6%</td>
<td>95.0%</td>
</tr>
</tbody>
</table>

Figure 1: CSM NNS Writing Diagnostic Inter-Prompt Reliability
Student essays were holistically evaluated by CSM faculty readers. Scoring guidelines were adapted from developmental writing guidelines in use at other community colleges, universities, and Educational Testing Service rubrics. Readers evaluated student competencies for fluency in English, types of errors, sophistication of thought, vocabulary, reading level, mechanics, and the amount of written text in a 45 minute period. Two readers evaluate student writing samples; if both readers cannot agree on the evaluation, a third reader is consulted.

Study results are presented in Figure 1. This data indicates an extremely high level of prompt reliability for the NNS writing sample diagnostic: ranging from 84.6% to 95.0% consistency.

**RESEARCH FINDINGS: INTER-RATER RELIABILITY**

Inter-rater reliability reflects the extent to which different readers or scorers reach the same decision when evaluating student writing samples. As noted, scorers undergo thorough training in standardized procedures for test scoring. In addition to inter-prompt reliability, the study design examined the consistency of scores between two raters on a total of 498 essays. Study results indicate very high inter-rater reliability and reflect positively on the training of the readers of writing samples: the extent of evaluative agreement between multiple readers is 86.5%.

**RESEARCH FINDINGS: CRITERION-RELATED VALIDITY**

Beyond inter-prompt and inter-rater reliability, an educationally sound writing diagnostic should yield accurate course placement recommendations. Student academic performance data (Fall 1988 - Spring 1993) were analyzed to assess course placement accuracy. Data presented in Figure 2 indicates that students who follow NNS writing placement recommendations are significantly more likely to successfully complete their writing coursework (80.1% vs. 54.8%). Conversely, students who self-placed, contrary to writing placement recommendations, were nearly three times more likely to fail or withdraw from their coursework (45.2% vs. 19.9%).

Analysis of student records shows approximately the same proportional gain in academic success for followers of placement advice—regardless of student ethnicity, gender, or age. In other words, students of all ages, gender, or ethnicity are more likely to successfully complete their NNS writing coursework if they follow the recommendations of CSM’s NNS placement program.

**CONCLUSION**

This data strongly supports the accuracy and reliability of CSM’s writing assessment tests for non-native speakers of English. Moreover, student outcome data have shown that individual assessment test scores are powerful and unbiased predictors of performance in the classroom. To this extent, CSM’s placement program offers both the student and the college a way of effectively identifying the appropriate level at which students should be studying. Moreover, study results show that CSM’s assessment program does exceedingly well in enhancing student achievement and increasing retention.

For further information, please contact the Office of Articulation and Research. Telephone: (415) 574-6196 Fax: (415) 574-6680 Internet: sewart@smcccd.cc.ca.us