Valid Assessment of Writing and Access to Academic Discourse

John A. Albertini
Joseph H. Bochner
Fred Dowaliby
Janette B. Henderson
National Technical Institute for the Deaf
Rochester Institute of Technology

One way to improve students' access to and retention in post-secondary degree programs is to assess their readiness for such programs accurately. To place deaf and hard-of-hearing students in preparatory courses and to determine their readiness for degree programs more accurately, a direct measure of writing was developed for deaf and hard-of-hearing students at a large technical university. The purpose of this study was to estimate the concurrent and predictive validity of this measure. The Test of Written English (Educational Testing Service, 1992) served as the criterion in the concurrent validity study, and student success in the university's gateway freshman composition course served as the criterion in the predictive validity study. Results provide evidence of the concurrent and predictive validity of the measure, supporting its use for course placement and early planning purposes.

Like hearing students who belong to ethnic and linguistic minorities, many deaf students in the United States have limited access to academic discourse. At one large technical university, deaf students' access to associate and baccalaureate degree programs is often delayed because of their skills in English language and literacy. On average, reading comprehension of these students is approximately 7.5 grade equivalents (NTID Annual Report, 1995). While this is higher than the national norm for college-age deaf students (approximately 4.0 grade equivalents as reported by Schildroth, Rawlings, & Allen, 1991), most deaf students at this university are placed in developmental English courses until they are eligible to enter the required freshman composition course. Students must pass this course in order to enter most degree programs. For many, preparation for degree programs takes two years; for those unable to get into this course, the options are limited.

Currently, students are placed in developmental English courses based on their performance on the Michigan Test of English Language Proficiency (English Language Institute, 1977) and the reading comprehension portion of the California Achievement Tests (Tiegs & Clark, 1957). Previous research with college-age hearing students has indicated that measurements of English language proficiency and reading comprehension are both related to writing ability (for correlations of English language proficiency with writing ability, see English Language Institute, 1977 and Pitcher & Ra, 1967; and for correlations of reading comprehension with writing ability, see Educational Testing Service, 1995). Readiness for the freshman composition course is determined, in part, on the basis of cut-off scores achieved on the Michigan and California tests. A student who achieves a score of 80 or better on the language proficiency test and a grade equivalent score of 10.0 or above on the reading test is considered...
“proficient” and allowed to write an essay for placement into “the writing sequence,” which consists of two preparatory writing courses and the freshman composition course.

Placement of students in developmental writing courses based on language proficiency and reading scores has been problematic; increasingly, students enter the university with the goals of obtaining associate, baccalaureate, and graduate degrees. Thus, to provide information about a student’s writing skill upon entry (to aid them in planning a course of study) and to provide information for placement in developmental courses, a direct measure of writing, the NTID Writing Test, was developed. In some programs, indirect measures of writing (multiple-choice tests) have been used in place of actual samples of student writing. Recent research by Berent et al. (1996) indicates that the validity of such tests for purposes of placement of deaf and hard-of-hearing baccalaureate-level students is weak, and they recommend using direct writing tests “to ensure fair and accurate assessment of writing competence.”

Results of previous studies provide evidence of the external validity and rater reliability of the NTID Writing Test (Albertini, Bochner, Cuneo, Hunt, Nelson, Seago, & Shannon, 1986; Bochner, Samar, Kelly, Bochner, Albertini, & Sacken, 1992). To estimate external validity, ratings of experienced, professional judges (English teachers) were compared with those of inexperienced, nonprofessional judges (graduate students). The nonprofessional judges employed a different rating procedure requiring no training since the training of nonprofessional judges would impart certain biases regarding the skill dimensions considered significant in writing assessment. The ratings assigned by the inexperienced group were strongly associated with those of the experienced raters ($R^2 = .94$). Interrater reliabilities for the NTID Writing Test ranged from .61 to .87, with the average reliability for a single rater estimated at .75 with the use of Fisher’s $r$ to $z$ transformation. Alpha reliabilities for the writing test summary scores, which represent the average of the scores assigned by three raters, ranged from .83 to .91 for all combinations of three raters.

The design of the NTID Writing Test is similar to the design of tests developed for adults learning English as a second language (see for example, Jacobs, Zinkgraf, Wormuth, Hartfiel, & Hughey, 1981). The use of English-as-second-language methods and materials in the education of students who are deaf or hard of hearing has long been advocated (Bochner, 1977, 1982; Goldberg & Boardman, 1974; Ragosta & Nelson, 1986; Swisher, 1989). One reason for this is that deaf students’ writing resembles that of hearing students of English as a second language (Langston & Maxwell, 1988; Swisher, Butler-Wall, & Stavans, 1988).

This study was undertaken to estimate the concurrent and predictive validity of the NTID Writing Test. We take concurrent validity to be the measure of the extent to which the NTID Writing Test produces the same rank order of individual performances as another previously validated test (Hamp-Lyons, 1991, p. 10). To estimate concurrent validity, we compared scores on the NTID Writing Test with those on the Test of Written English (TWE, Educational Testing Service, 1992), a direct measure of writing typically administered to hearing foreign students wishing to study at universities in the United States and Canada. Research by Traxler (1990) has indicated that the TWE can provide valid estimates of the writing abilities of deaf college students.

To estimate predictive validity, one uses a criterion measure at a future point in time (Hamp-Lyons, 1991, p. 10). We chose a performance criterion: the number of quarters necessary to successfully complete the freshman composition course, gateway to the university's degree programs. If the test provides useful information about deaf students’ writing skills upon entering college and about their probability of success in the writing sequence, this information may be used to help faculty plan curriculum and students to plan their courses of study.

Method

The Writing Test

Students are given thirty minutes to write an essay on an assigned topic. For the concurrent validity study, students were asked to write on the following topic: Some college students live in dormitories. What do you like about dorm life? What don’t you like about dorm
life? Explain. For the predictive validity study, students wrote on the following topic: You are in a new place. What do you like about NTID and the people here? What don’t you like about NTID and the people here? Explain.

Each essay is read by three raters. Each rater is an English teacher who has been trained in the scoring protocol. Training consists of scoring a selected sample of 25 papers and discussion of the ratings. In addition, all raters periodically participate in calibration sessions to maintain consistency. Raters may assign a maximum of 25 points to each of four categories: content, organization, language, and vocabulary (for descriptors of each category, see Appendix) with the total possible score being 100. The reported score represents an average of the three raters’ scores.

Subjects and Procedure

The subjects for the two validity studies reported here were deaf and hard-of-hearing college students from a variety of cultural and language backgrounds. Most had grown up in the United States. The students in both studies were selected from the population of students entering NTID from 1988 to 1992. A representative sample of 65 students volunteered to participate in the concurrent validity study, which meant taking two tests, the TWE and the NTID Writing Test. To encourage participation in the study, seven instructors gave up scheduled class time for the testing. The classes represented the full range of ability at the college, from basic English to the freshman composition classes. The order of the tests was counterbalanced across groups, and actual testing time was 30 minutes for each test. At the conclusion of the testing, students were paid for their participation and refreshments were provided. The TWE papers were scored by the Educational Testing Service according to their procedures. Scores on the TWE may range from 1 to 6 points in increments of 0.5 points. The NTID Writing Test papers were scored by NTID faculty members trained in the scoring protocol. Scores on this test may range from 0 to 100 points.

For the predictive validity study, the course histories of 1,198 deaf and hard-of-hearing students entering the university from the Fall 1988 through Fall 1992 were examined to determine whether entry writing test scores would predict the number of quarters needed to complete the required freshman composition course. Based on results of a placement essay administered to eligible students, 582 students from this group were placed into “the writing sequence” (the two preparatory courses and the freshman composition course). Of this number, 180 students did not complete the freshman composition course, and 375 successfully completed the course with a grade of D or better. Results for 27 students were missing. In this study, entry scores on the NTID Writing Test and course histories for the successful students were analyzed. Entry scores on the reading comprehension subtest of the California Achievement Tests and the Michigan Test of English Language Proficiency were also included in the analysis.

Results

Concurrent Validity

For the sample of 65 students participating in the concurrent validity study, the mean score on the NTID Writing Test was 53.26 (standard deviation [SD] = 10.34, range = 34–81). The mean score on the TWE was 3.31 (SD = 0.93, range = 1.5–6.0). The correlation between scores on the NTID Writing Test and the TWE was observed to be .83.

Predictive Validity

For the population of 1,198 students entering NTID from 1988 to 1992, the mean score on the NTID Writing Test was 51.66 (SD = 11.12, range = 21–86), and the scores were distributed normally as shown in Figure 1.

For the sample of 375 students participating in the predictive validity study, the mean score on the NTID Writing Test was 59.19 (SD = 9.68, range = 34–86). The mean score on the reading comprehension subtest of the California Achievement Tests was 9.74 (SD = 1.36) and the mean score on the Michigan Test of English Language Proficiency was 77.97 (SD = 10.60). The mean number of academic quarters needed to successfully complete the freshman composition course
upon entry to the writing sequence (i.e., the two preparatory courses and the freshman composition course) was 2.33 (SD = 1.20, range = 1–8). Correlations involving students' performance on these variables are displayed in Table 1. All correlations exceed the .01 level of significance (two-tailed), and probably represent conservative estimates of true, underlying relationships because of the reduction in variances resulting from the elimination from the analysis of students who did not successfully complete the freshman composition course. As shown in Table 1, the correlation between the NTID Writing Test and number of quarters was observed to be -.48. This means that students scoring low on the Writing Test required more quarters to successfully complete the freshman composition course. Conversely, those scoring high on the Writing Test required fewer quarters. While this relationship is statistically significant, the magnitude is modest and indicates that approximately 23% of the variance in successful completion of the sequence is accounted for by performance on the Writing Test.

The number of quarters required for successful completion of the freshman composition course was used as the dependent variable in a stepwise, multiple regression analysis in order to assess the predictive validity of the NTID Writing Test within the context of the reading and English language proficiency measures currently used for placement purposes at this university. The .05 level of significance was employed as the criterion for variables entering the equation, and .10 was employed as the criterion for the removal of variables. Both the writing and English language proficiency (ELP) measures satisfied the entry criterion and, in combination, accounted for 27.2% of the variance in the number of quarters needed to successfully complete the Freshman composition course ($R = .52$, $p < .001$). The partial correlation describing the relationship between writing and quarters was $- .20$, and the partial correlation describing the relationship between ELP and quarters was $- .22$. Reading did not satisfy the criterion for entry into the regression equation.

A separate stepwise, multiple regression analysis
was performed to evaluate the relative contributions of writing and reading as determinants of ELP and to provide evidence of the construct validity of the NTID Writing Test. A partial correlation of .62 was obtained between writing and ELP, and a partial correlation of .53 was obtained between reading and ELP. Writing and reading combined accounted for 69% of the variance in ELP \( (R = .83) \).

**Discussion**

The results of this study demonstrate that scores on the NTID Writing Test are strongly associated with scores on the TWE \( (r = .83) \) and provide evidence of the concurrent validity of the NTID Writing Test. NTID Writing Test scores are also related to time required for successful completion of the freshman composition course \( (r = -.48) \), supporting the test’s predictive validity. Students scoring high on the NTID Writing Test tend to take fewer quarters to complete the freshman composition course upon entry to the writing sequence than those scoring low on the NTID Writing Test. The multiple regression results and corresponding significant, partial correlations between the number of quarters needed to successfully complete the freshman composition course and performance on the writing \( (r = -.20) \) and English language proficiency \( (r = -.22) \) tests reinforce the results of the simple correlation analysis, demonstrating the unique contributions of each measure. These multivariate relationships show that writing and ELP correlate uniquely and independently of each other with the length of time needed to successfully complete the writing course sequence. These findings combine with data reported by Bochner et al. (1992) to corroborate and extend the validity of the NTID Writing Test.

The conventional measures of student success, grade point average (GPA) and degree attainment, tend to be restricted in range and limited in utility. The criterion of success used here, the number of academic quarters needed to successfully complete the freshman composition course upon entry to the writing sequence, is a more direct indicator of student success than GPA and is associated with a range of student performance. Simply stated, more capable students are able to complete the freshman composition course in fewer academic quarters than less capable students. Thus, number of quarters required to pass the course appears to be a valid and useful measure within the context of the writing curriculum at this university.

Scores on the NTID Writing Test are distributed normally with the mean located in the center of the test-score range, as are scores on the TWE (Educational Testing Service, 1992). Presumably, these tests of writing ability have similar distributions because each population of examinees is normally distributed with respect to the construct of writing ability. The similarity of these distributions, therefore, may be a result of defining the construct in terms of the population. However, there may also be a tendency for raters to rely heavily on scores in the middle range of the distribution when evaluating papers. Previous research has shown that raters are not able to consistently evaluate separate dimensions of writing quality, such as organization, content, language use, and vocabulary, apart from one another (Bochner et al., 1992). Similarly, if raters are not able to consistently discern fine distinctions in writing ability, their ratings may tend to cluster in the central region of the distribution because of relative uncertainty associated with the evaluation of midrange papers. The shape of the test-score distribution, therefore, appears to be attributable to both population (writer) characteristics and rater (reader) capabilities.

The results of this study indicate that scores on measures of writing, reading, and English language proficiency are interrelated. Consistent with the results of studies conducted on students of English as a second language (Educational Testing Service, 1992; Pitcher & Ra, 1967), the writing ability of deaf students is more strongly associated with English language proficiency \( (r = .76) \) than with reading ability \( (r = .55) \). English language proficiency, however, is strongly associated with both writing and reading ability \( (r = .76 \text{ and } r = .71, \text{ respectively}) \). The relationship among these variables is asymmetrical in the sense that writing and reading are not associated with each other in quite the same way as they are associated with English language proficiency. English language proficiency is comprised of subsidiary skills pertaining to knowledge of grammar and vocabulary, thought to underlie writing and reading performance. Therefore, the correlations among writing, reading, and English language proficiency suggest that knowledge of grammar and vocabulary represents an underlying competency common to
both writing and reading performance.

Although writing and reading are related to one another and to English language proficiency, they may nevertheless be considered separate and distinct constructs. The results of the partial correlation analysis provide strong evidence of the distinctiveness of writing and reading. The partial correlations indicate that distinctive attributes of writing ability are uniquely associated with a larger share of the variance in English language proficiency ($r = .62$) than are distinctive attributes of reading ability ($r = .53$), despite the fact that reading comprehension questions constitute 20% of the items on the Michigan Test. This finding suggests that knowledge of grammar and vocabulary, which is assessed by 80% of the items on the Michigan Test, may be more strongly associated with features of writing ability than with features of reading ability. The asymmetrical relationship among these variables strongly supports treating writing and reading as separate but related constructs and also provides evidence of the construct validity of the NTID Writing Test (cf. Educational Testing Service, 1992).

The distinction between writing and reading ability is also reflected in the results of the multiple regression/correlation analysis used to evaluate how well scores on tests of writing, reading, and English language proficiency are able to predict success in freshman composition. Scores on measures of writing and English language proficiency were shown to be uniquely and independently related to a measure of students’ success in the writing course sequence ($r = -.20$ and $r = -.22$). Scores on the reading test, however, did not contribute in a unique and independent manner to predicting students’ success in the writing course sequence. This asymmetrical pattern of results, once again, demonstrates the distinction between writing and reading and provides evidence of the construct validity of the NTID Writing Test.

The results of these studies have implications for future testing and curriculum planning and for the assessment of college students’ writing in general. For deaf and hard-of-hearing students, scores on the NTID Writing Test and the English language proficiency test may be used as rough, early predictors of readiness for degree programs. Further research is necessary to determine whether more than one writing test sample or samples of other kinds of writing (such as samples from a writing portfolio) will improve the accuracy of such predictions. Research on the actual placement and performance of students in writing courses is also necessary. More accurate determination of writing skills would allow more accurate placement of students. More homogeneous classes, that is classes where students had more comparable writing skills, would allow instructors to focus on skill areas of benefit to the class as a whole. More focus in turn should lead to more efficient and effective instruction.

Improving access to academic discourse means among other things providing students with a variety of degree options. Recently, Associate in Occupational Studies (AOS) degree programs have been introduced. Compared to existing degree programs, AOS programs place less emphasis on traditional academic writing and hence are accessible to a broader range of students. To make it possible for more students to pursue degrees in the existing Associate in Applied Science (AAS) programs, the writing sequence consisting of two preparatory courses and the freshman composition course was created. The analyses reported here suggest that more needs to be done for those wishing to pursue AAS degrees. Of the 1,198 students whose records were analyzed for the predictive validity study, 582 (49%) gained entry into the sequence. Roughly two thirds of those who entered the sequence (375 or 64%) completed it successfully, and one third (180 or 31%) did not. For the unsuccessful students in this sample, improving access to academic discourse implies encouraging some to pursue an AOS degree and increasing the chances for others to complete the AAS degree writing sequence successfully in a reasonable period of time.

Appendix

NTID Writing Test Scoring Categories and Descriptors

Organization (25 pts.): This includes such features as:
- clear statement of topic placed appropriately;
- intent is evident to readers;
- plan of paper could be outlined by reader (i.e., paper is unified and coherent);
• appropriate transitions (i.e., transitional markers and clear paragraphing).

Content (25 pts.): This includes such features as:

• paper addresses the assigned topic;
• generalizations are supported by examples;
• no extraneous material;
• pertinence and noteworthiness of ideas.

Language Use (25 pts.): This includes such features as:

• correct use of grammatical structures (sentence and discourse level) and punctuation;
• correct use of complex structures;
• intelligible spelling;
• clarity of style and expression;
• clarity of reference.

Vocabulary Use (25 pts.): This includes such features as:

• appropriate semantic use of vocabulary;
• consistent register;
• sophisticated choice of vocabulary;
• appropriate use of figurative and idiomatic expressions.

References


