WPA: Writing Program Administration Volume 10, Number, 3 Spring 1987
© Council of Writing Program Administrators

Another Look: Standardized Tests for Placement in College Composition Courses

Barbara L. Gordon

According to a recent survey conducted by the National Testing Network in Writing, 84 percent of all colleges use a placement test to exempt or place students in composition courses. The majority of institutions decide placement using a single writing sample.[1] Though the National Testing Network in Writing and the National Council of Teachers of English recommend placing students using a writing sample, past research and the research I have recently completed, indicate that standardized tests are more accurate than a single writing sample for placing students. In fact, with regard to validity and reliability, a single writing sample is among the most unacceptable means to place students. There are compelling pedagogical reasons for using a writing sample for placement; however, writing program administrators should be aware of the inadequacies of using a single sample and, if persuaded to use standardized test scores, administrators must know how to use standardized test scores for the most accurate placement.

Standardized Tests vs. a Writing Sample

To conduct valid and reliable placement using a writing sample requires a large investment of time, energy, and money. First, finding appropriate essay topics is, as Gertrude Conlan put it, like panning for gold. She states, "The Writer of the essay question is at the mercy of not only the vagaries of students' thinking (often better described as reasonableness, imagination, and inability to read teachers' minds), but also of the diversity of experiences, cultures and environments represented in a large testing population." Conlan cites as one example how the essay topic "Waste is a necessary part of the American way of life" had to be changed to "Wastefulness is a part of the American way of life" after the pilot study revealed students, pardon the upcoming pun, were regularly writing about bodily functions.[2]

Such an unexpected task misinterpretation is relatively harmless, but not so harmless is a misinterpretation that immediately puts certain groups of students at a disadvantage, a topic, for example, that unknow-

WPA: Writing Program Administration, Volume 10, Number, 3 Spring 1987
© Council of Writing Program Administrators

ingly depends on white middle class values. Designing fair and appropriate writing tasks is a demanding endeavor. The pitfalls have been discussed by many including Leo Ruth, Sandra Murphy, James Hoetker, Gordon Brossell, Edward White, Barbara Ash, and Karen Greenberg.[3]

After designing the task, evaluating the resulting essays requires equally careful and painstaking procedures. A minimum of care would require that writers remain anonymous during the evaluation, and that at least two raters rate each paper. Edward White points out further criteria to reduce scoring variability such as controlling the circumstances under which the papers are scored, developing and familiarizing raters with a scoring rubric, selecting sample papers to establish consistent ratings using rater consensus, checking ratings during the scoring, using multiple independent scoring, and recording raters' scores.[4] Following these careful procedures is not merely advantageous but essential if a college is to claim a fair evaluation of student essays.

The research John Dick, Howard Scheiber, and I recently completed confirms that a great amount of time is required to conscientiously create tasks and evaluate essays. We carefully devised and piloted tasks and followed White's criteria in assessing approximately 1,500 papers for a large assessment at the University of Texas at El Paso. A conservative estimate of the time we each spent is one hundred twenty hours. This estimate does not take into account students' time, test givers' time, and raters' time. But, even with this large investment of resources, using a single sample of a student's writing is an inappropriate method to place students. Over and above the time and cost necessary to devise tasks and evaluate papers, the most important reason for not using a single writing sample as a measure of a person's writing ability is that it is not reliable or valid. The task may have been reliably rated, but students' writing performance is not reliable; students' performance is quite variable depending on the task and test circumstances. The writing sample may be valid for assessing students' writing ability for that task—but not for assessing students' overall writing ability. It is not established just how many writing samples would be needed from each student to accurately assess writing ability, but current research makes it clear that one is not enough.

In our assessment approximately 700 students each responded to two tasks in the Fall of 1983. The first task asked the students to pick a good quality in themselves and relate an incident that displayed their good quality. The second task was a comparison of the advertising techniques in two cigarette ads. The students' papers were holistically scored using ETS guidelines. In a similar manner these same papers were scored with primary trait scoring guides. Using only cases where students completed both tasks, I ran Pearson correlations to see if

students who had high or low ratings on one essay would score similarly on their second essay (see Table 1). The correlation between the holistic ratings on the two tasks was .35 and the correlation between the primary trait ratings was .20. Statistics texts suggest that psychological tests given within two-week intervals have test-retest reliabilities between .7 and .1.0.[5] The two tests of writing ability are not reliable by this criterion. These weak correlations supported by the experience of assessment specialists such as Gertrude Conlan and Edward White confirm that writers' performance can vary substantially from task to task.[6] It is likely, therefore, that many students are misplaced as a result of using a single writing sample.

TABLE 1. Correlations Between Holistic and Primary Trait Ratings on Two Tasks

	Holistic Rating Task 1	Holistic Rating Task 2	Primary Trait Rating Task 1	Primary Trait Rating Task 2
Holistic Rating Task 1	1	.35	.60	.34
Holistic Rating Task 2		1	.21	.73
Primary Trait Rating Task 1			1	.20
Primary Trait Rating Task 2				1

n = 497
All correlations are significant at the 0.000 level.

Though a single writing sample is not a valid placement measure, many find it preferable to a standardized test because a sample has face validity, meaning the measure being taken appears valid. If you want to measure writing ability what better way than having someone write. Face validity is the least sophisticated kind of validity and psychologists are not satisfied with a measure if it meets this criterion alone.[7] Construct validity is the type of validity which best assures that the measure being taken is a valid one. Construct validity has to do with how accurately a measure reflects the attribute or ability at the theoretical level. Multiple measures of a construct are used to determine if a measure has construct validity. Construct validity is established when a consistent pattern of relationships is found between a measure of a construct and other measures for the same construct.[8]

WPA: Writing Program Administration Volume 10, Number, 3 Spring 1987
© Council of Writing Program Administrators

Writing ability, like intelligence, is a complex theoretical construct, and like intelligence it is difficult to define. Does a single sample of a person's writing rated holistically reveal enough of a person's overall writing ability, enough that decisions concerning placement can be made? It appears not, given the likelihood that people perform differently depending on the task. Although a standardized test does not face validity, it is a more valid measure of a person's writing ability than a single writing sample rated holistically. It somehow captures to a fuller extent what constitutes writing ability.

Hunter Breland's findings substantiate this view. His correlational analyses showed that objective measures correlated more highly when multiple samples of a student's writing were used rather than a single writing sample. The objective measure captures the factors that underlie writing ability better than a single rating of a single essay. It was found that a student's score on an objective measure grows stronger as it is related to more samples of his or her writing. Breland concludes that a single brief writing sample is probably a less useful indicator of writing ability than a brief objective measure.[9]

Though standardized scores lack face validity, Breland's research shows that standardized scores give a more accurate idea of a person's writing ability since they have construct validity. In addition, standardized test scores are known to have strong and consistent relationships with various means to measure writing ability, such as essays and success in composition courses. Numerous studies document the strong relationship of standardized scores, particularly the English ACT and, to a somewhat lesser degree, the TSWE (the Standardized Test of Written English) with both writing samples and success in composition classes.

After completing a two-year study at Ferris State College, John Alexander found that writing samples did not predict student success in composition courses more accurately than ACT scores. Using criterion-referenced analysis on writing samples of 1,200 incoming freshmen, he found the standardized test scores correlated well with composition grades and samples of writing, though there were some discrepancies at the high end of the ACT scores.[10]

In past issues of *Education and Psychological Measurement* Jack Snowman et al. report the ACT accurately predicted grades in writing courses,[11] and in other research, Bill Fowler and Dale Ross found a correlation of .56 between ACT scores and students' composition grades.[12] Recently in *Research in the Teaching of English* Donna Gorrell reported the English ACT had correlations of .65 and .61 with holistically rated essays.[13]

The correlations derived from the assessment at UT El Paso were nearly identical. Writing ability was defined as a student's total score compiled by adding together holistic and primary trait scores for the student's two essays (see Table 2). The Pearson correlation between writing ability and the English ACT was .61, between writing ability and the TSWE it was .39. Though the TSWE correlation is less than stunning, the English ACT is highly correlated with writing ability and confirms that standardized tests can be devised to be strongly related to writing ability.

TABLE 2. Correlations Between Standardized Tests Scores and Writing Ability—Defined as a Person's Cumulative Holistic and Primary Trait Scores on Two Essays

	ACT English	TSWE	Verbal SAT
Writing Ability	.61	.39	.39
N =	102	296	300

All correlations are significant at the 0.000 level.

Apparently, then, some standardized tests are better measures of writing ability than others. Clearly, objective tests must be carefully designed to assure they are reliable and valid measures of writing ability. Some institutions operate under the uninformed assumption that nearly any objective test with some face validity is an appropriate measure. To devise a reliable and, at least to some degree, valid objective test would necessitate piloting the test, conducting an item analysis, collecting longitudinal data to determine test reliability, and correlating the test results with writing samples and course grades. Unless this is done the "in house" objective test is unscientific and unacceptable. Designing a test takes considerable time, money, and most importantly, it requires a great deal of expertise. Unless institutions are willing to make a substantial investment, they should not consider an "in house" objective test to place students.

Using Standardized Tests for Placement

Given the likelihood of misplacing many students using a single writing sample or "in house" objective test, many colleges may consider placing students with a standardized test. However, using standardized scores for placement is not as easy as looking at scores and slotting students in courses, initially anyway. The crucial decision is deciding appropriate cutoff scores. Consider Russell Meyer's findings. Using 854 students attending the University of Missouri-Columbia, Meyer examined

WPA: Writing Program Administration - Volume 10, Number, 3 Spring 1987
© Council of Writing Program Administrators

whether the exemption score on the University's objective writing test matched well with the students' writing ability. Based on the University's established test cutoff, 57 percent of the incoming students exempted the freshman level writing course, while an examination of students' writing indicated that only 18 percent should be exempt.[14] Any college that decides to use standardized tests must check their cutoff scores against writing samples to assure that they have selected an appropriate score.

Our composition committee at UT El Paso was pleased to discover that our cutoff scores conformed with the committee's idea of the type of writing that would be expected of students in our developmental and freshman composition courses. To determine this, the students' essays from the assessment were compared with students' standardized test scores. As stated previously, each student in our assessment wrote two essays and each essay was rated twice using holistic scoring, and twice using primary trait scoring. The scoring ranged from a low of 1 to a high of 4; therefore, the lowest possible score a student could receive was an 8, representing 1's from all raters on both essays, and the highest possible score a 32, representing 4's from all rates on both essays. After examining numerous papers, including anchor papers, we determined that any student who received a score of 24 (an equivalent of 3's from all raters) was too advanced for our developmental class. The next step was to see how many students with a total writing ability score of 24 or higher had been misplaced into the developmental composition course on the basis of our current English ACT and TSWE cutoff scores.

At UT El Paso any student who scores 18 or below on the English ACT, or scores 39 or below on the TSWE, is placed in our developmental course. I looked to see how many students with an ACT English score of 18 or below had received a writing ability score of 24 or higher. I found 6 percent of the students misplaced by the English ACT cutoff and 10 percent misplaced by the TSWE cutoff (see Table 3).

TABLE 3. Percent of Misplaced Students (As Determined by Writing Ability Score) in Developmenal Composition Courses at Selected Standardized Test Scores

	ACT English Score 18 or Below	**TSWE Score 39 or Below**
Writing Ability Score 24 or Higher	6%	10%
N=	68	137

I followed the same procedure to see what percentage of students was not advanced enough for our regular freshman level composition class (see Table 4). An ACT score of 19 or higher or a TSWE score of 40 or above is required for a student to enroll in the freshman composition class. After reviewing papers from the writing sample, our committee decided that any student who received a writing ability score of 16 or less (an equivalent of 2's from all raters) wrote too poorly to be enrolled in the freshman composition course. After sorting out the incoming students with English ACT scores of 19 or higher, we discovered that approximately 4 percent of the students had writing scores of 16 or less. In other words, about 4 percent of the students were misplaced by our ACT cutoff. The TSWE was surprisingly more accurate. No students were found to have writing scores of 16 or less who had a TSWE score of 40 or higher, or, in other words, no students in our sample were misplaced.

TABLE 4. Percent of Misplaced Students (As Determined by Writing Ability Score) in Regular Composition Courses at Selected Standardized Test Scores

	ACT English Score 19 or Above	**TSWE Score 40 or Above**
Writing Ability Score 16 or Lower	4%	0%
N=	31	118

Our placement could be more accurate. We now have an estimate of the percentage of students misplaced by our current standardized cutoff scores. What is needed, however, are frequency distributions of students' writing ability for scores on the English ACT and the TSWE. We need to examine the frequency distribution of writing ability for students who scored, for example, a 20 on the English ACT so that we will know the percentage of students who received a writing score of 8 (1's from all raters) and 9's, up through scores of 32. By examining sample papers that exemplify the writing ability score of 8 through the highest, 32, we can decide more accurately what our standardized cutoff scores should be, and simultaneously approximate the percentage of misplaced students. We are currently determining this.

If a college decides to collect its own writing sample, it should understand that this procedure for determining cutoff scores is a large undertaking, larger even than using more than one writing sample for

WPA: Writing Program Administration Volume 10, Number, 3 Spring 1987
© Council of Writing Program Administrators

placement—at least initially. To determine the cutoff scores, a minimum of two writing samples must be collected using careful and painstaking procedures in selecting topics and evaluating the papers. In addition, frequency distributions must be obtained from the data, and lastly a committee must decide cutoff scores using sample papers. However, unlike placement with writing samples, establishing cutoffs need only be conducted once, or perhaps infrequently. The short-term investment had long-term gains.

It may be possible to dispense with the need for individual colleges to collect writing samples and produce frequency distributions in order to determine their cutoff scores. With the exception of certain ethnic and socioeconomic groups, it is unlikely that the frequency distribution of students' writing ability will vary greatly nationally. The exceptions are writers from minority cultures, since these writers are known to produce a different frequency distribution from "white" participants. Generally the standardized tests render a more negative view of minority students' writing ability than their writing samples indicate.[15] Nationally as colleges conduct their own assessments, share their frequency distributions, and characterize their student populations, a reliable database will be founded. Colleges can then dispense with their own data collection and form committees to determine local cutoff scores by examining distributions and sample papers from colleges with similar student populations.

Conclusion

The most valid and reliable means of assessing writing ability remains a thorough examination of carefully obtained multiple samples of an individual's writing. Since this is an unrealistic alternative for most institutions, the majority have opted to use a single writing sample. Perhaps many administrators are unaware that a single sample is not very reliable or valid. Others, knowing this, may continue to use a writing sample because they fear the message standardized tests send to students and educators. The negative message some educators fear students will receive is: "The university considers writing a collection of editorial skills," since this is the emphasis of standardized tests. But students who take standardized tests do so with college entrance in mind, not composition placement, and the time between the test and entrance to a university is generally long enough that students would not think the content on standardized tests is the content of college composition classes. A writing sample for placement sends out a negative message as well: "The university considers writing a way to assess certain abilities to place you in a course. This composition is a test. What you have to say will not be responded to, rather how you say it will be evaluated." Writing for the purpose of placement turns writing into a vehicle for assessment rather than a means of communication.

What is cause for concern is the message sent to educators. In the recent past the shift to placement by a writing sample helped usher in an era of more reasoned composition instruction. Is it possible to keep the public and educators from interpreting the use of standardized tests as a call for instruction in editing skills? Until educators distinguish testing from instruction there will remain compelling reasons to include a writing sample in conjunction with standardized test scores. The standardized test is necessary for valid, reliable placement. A writing sample is, for the time being, the costly and perhaps necessary insurance to preserve the current strides in composition pedagogy.

Notes

[1]Karen L. Greenberg, Harvey S. Wiener, Richard A. Donovan, eds., *Writing Assessment* (White Plains: Longman, Inc., 1986), xiii-xiv.

[2]Gertrude Conlan, "Panning for Gold: Finding the Best Essay Topics," *Notes from the National Testing Network in Writing* (October 1982): 11.

[3]See Leo Ruth and Sandra Murphy,"Designing Topics for Writing Assessment: Problems of Meaning," *College Composition and Communication* 35 (December 1984): 410-422; Edward M. White, *Teaching and Assessing Writing* (San Francisco: Jossey Bass, 1985), 100-119; James Hoetker and Gordon Brossell, "A Procedure for Writing Content-Fair Essay Examination Topics for Large-Scale Writing Assessments," *College Composition and Communication* 37 (October 1986): 328-335; Gordon Brossell, "Rhetorical Specification in Essay Examination Topics," *College English* 45 (February 1983): 165-173; Gordon Brossell and Barbara Hoetker Ash, "An Experiment with the Wording of Essay Topics," *College Composition and Communication* 35 (December 1984): 423-425; Karen Greenberg, *The Effects of Variations in Essay Questions on the Writing of CUNY Freshman* (New York: City University of New York, Instructional Research Center, 1981).

[4]Edward M. White, "Holisticism," *College Composition and Communication* 35 (December 1984): 400-409.

[5]George S. Howard, *Basic Research Methods in the Social Sciences* (Glenview, IL: Scott, Foresman and Company, 1985), 26.

[6]See Gertrude Conlan, "'Objective Measures of Writing Ability," in *Writing Assessment*, edited by Karen L. Greenberg, Harvey S. Wiener and Richard A. Donovan (White Plains: Longman, Inc., 1986), 112-113; Edward M. White, "Pitfalls in the Testing of Writing," in *Writing Assessment*, edited by Karen L. Greenberg, Harvey S. Wiener and Richard A. Donovan (White Plains: Longman, Inc., 1986), 98-107.

[7]Howard, p. 100.

WPA: Writing Program Administration, Volume 10, Number, 3 Spring 1987
© Council of Writing Program Administrators

[8]H. Gleitman, *Psychology* (New York: W. W. Norton, 1981), 107.

[9]Hunter M. Breland, "Can Multiple-choice Tests Measure Writing Skills?" *The College Review Board*, no. 103 (Spring 1977): 11-15.

[10]John Alexander, "Policy and Assessment," *Notes from the National Testing Network in Writing* (December 1983): 19.

[11]Jack Snowman, Dennis W. Leitner, Vivian Snyder, and Lillie Lockhart, "A Comparison of the Predictive Validities of Elected Academic Tests of the American College Test (ACT) Assessment Program and the Descriptive Tests of Language Skills for College Freshmen in a Basic Skills Program," *Educational and Psychological Measurement* 40 (1980): 1159-1166.

[12]Bill F. Fowler and Dale H. Ross, "The Comparative Validities of Differential Placement Measures for College Composition Courses," *Education and Psychological Measurement* 42 (1982): 1107-1115.

[13]Donna Gorrell, "Toward Determining a Minimal Competency Entrance Examination for Freshman Composition," *Research in the Teaching of English* 17 (October 1983): 263-274.

[14]Russell J. Meyer, "Take-Home Placement Tests: A Preliminary Report," *College English* 44 (September 1982): 506-510.

[15]Edward M. White and Leon L. Thomas, "Racial Minorities and Writing Skills Assessment in the California State University and Colleges," *College English* 43 (March 1981): 276-282; Roscoe C. Brown, Jr., "Testing Black Student Writers," in *Writing Assessment*, edited by Karen L. Greenberg, Harvey S. Wiener and Richard A. Donovan (White Plains: Longman, Inc., 1986), 98-107.