PLACEMENT OF ESL STUDENTS IN
WRITING-ACROSS-THE-CURRICULUM PROGRAMS

J. D. BROWN
University of Hawai‘i

Four years ago, a writing-across-the-curriculum (WAC) program was established at the University of Hawai‘i. As part of that program, new English composition courses were implemented and placement testing for writing courses was mandated for all incoming undergraduate students on the campus—native speakers and ESL students alike. As a result, all new freshmen are assigned by placement tests to one of six composition courses: accelerated composition, regular composition, regular composition with required laboratory, remedial composition, regular ESL composition or preparatory ESL composition. The five-hour Mānoa Writing Placement Examination (MWPE) requires students to write on two topics and revise both essays later in the day. ESL students were also required to sit the three-hour English Language Institute Placement Test (ELIPT), which has subtests for ESL listening, reading and writing skills. This paper explores the place of ESL students in such a WAC placement testing system.

The data for this study are based on an entire year’s administrations of the MWPE (N = 1769 incoming freshmen), and of the ELIPT (N = 470 incoming international students). Not surprisingly, the position of the ESL students is low in the overall distribution of MWPE scores. The ways in which foreign students are identified at UHM are discussed in light of these results. In addition, the relative reliability and validity of the MWPE and the ELIPT are examined for appropriateness in making decisions about the placement of ESL students. The results are further examined in terms of how ESL testing and decision making were affected by these university-wide policies.

INTRODUCTION

An extensive literature exists on the teaching of writing in ESL/EFL contexts (see Brown & Bailey 1984, Connor 1987, Zamel 1987 and Piper 1989 for overviews of this work). Unfortunately, within this literature, there is a striking, yet important, area that seems to remain largely unexplored by ESL professionals: writing-across-the-curriculum programs. A writing-across-the-curriculum (WAC) program can be roughly defined as any writing program that is institution-wide and fosters writing as a means of meaningful communication and learning within the differing contexts of specific
disciplines. Spack (1988) briefly touches on WAC as a part of her discussion of the ESL teachers' responsibilities vis-à-vis readiness international students for writing in the academic discourse community. However, she must draw heavily on the first language literature to do so. Swales (1987) describes means for teaching ESL students techniques for writing discipline specific term papers, but never refers directly to the existence of WAC programs. In short, there is very little information to be had on the topic of writing-across-the-curriculum programs—particularly with regard to how ESL programs and students are affected by them.

As a result, when ESL programs are confronted with emerging WAC programs, they will find very little to draw on in the ESL literature. Because WAC programs are growing increasingly numerous at major universities across the United States, because the population of foreign students on such campuses is also growing, and because of the institution-wide nature of WAC programs, the place of ESL students in WAC programs will inevitably demand more prominent attention from the ESL/EFL field. Hopefully, this article will stimulate discussion of the issues that arise with regard to ESL students when a WAC program is established.

Background

In 1986, a proposal was made and approved for the development of a writing-across-the-curriculum program at the University of Hawai’i at Mānoa (UHM). The main purpose of this WAC program (later named the Mānoa Writing Program) was to improve the quality of writing at UHM by teaching students “to communicate clearly and effectively in standard English”* and “to reason clearly and effectively” through writing (UHM 1984).

The Program provides intensive training in writing beginning in the freshman year and continuing throughout undergraduate studies by requiring writing intensive courses within the student’s major or other related fields. Clearly, such a WAC program depends for its survival on the cooperation of faculty members across the entire campus.

The English Language Institute (ELI) first became directly involved when the Director and Assistant Director of the ELI, and the Chair of the ESL

* The phrase “standard English” has itself been an important political issue. Amongst linguists and ESL professors, it has largely been replaced with “Hawai’i standard English and mainland dialects.”
Department were invited to meet with the Dean of our college, the Chair of the English Department and the Director of Composition to discuss "the ESL problem." As in many ESL programs, we faced the contention that ESL students somehow lack the ability to communicate in writing. At the meeting, anecdotal evidence was presented for the particular weakness of foreign students' writing abilities. The ESL students in the ELI finish their training in writing with a course, ESL 100, which is treated at UHM as an exact equivalent to the English 100 freshman composition course which is offered by the English Department. It was suggested that the ESL students should be tested at the end of their training to determine whether they were up to the same "standard" as the students in the English Department. It was eventually decided that it would be useful to test our ESL students at the end of instruction and that students in English 100 could be tested at the same time for comparison. A formal study was set up and funded by the University.

That study (Brown 1990b; also see Brown & Durst 1987) investigated the degree to which significant differences exist between the mean writing scores for ESL and English Department students at the end of their respective composition courses. Eight ELI teachers and eight English Department teachers were paid to rate randomly assigned compositions without knowing which type of students had written each. A holistic six point rating scale (0-5) (described below under MWPE scoring materials) was used by all raters. Analysis of variance (ANOVA) procedures revealed no significant differences for main effects due to the type of student (ESL 100 vs. English 100), type of rater (ESL or English Departments), the order in which the compositions were read, or any interaction of these three factors.

Such studies can help substantially in formulating policies across departments. For instance, in addition to quelling discussion of the "ESL problem" and ending the need for the committee involved, this cooperative study opened up new avenues of communication and cooperation between the ELI and English department. Since the study was conducted, there has been a noticeable increase in consultations between departments on many policy and testing issues, and the Director of ELI has been appointed a permanent ex officio member of the Mānoa Writing Board which governs the Mānoa Writing Program. Thus, from early in its development, the ELI has been involved in the Mānoa Writing Program and in the formulation of policies that rather
dramatically affect the academic lives of native speakers and ESL students alike.

The remainder of this paper will focus on the campus-wide writing placement policies that resulted from this sweeping set of decisions, and the effect of those policies on the English Language Institute and the international students that it serves.

Placement Decisions

As one component of the Mānoa Writing Program, the Mānoa Writing Board was charged with insuring that all students demonstrate their ability in expository writing within two semesters of admission to UHM by passing English 100. All entering freshmen and transfer students without English 100 were to take the Mānoa Writing Placement Examination (MWPE) administered periodically by the Board. On the basis of their writing performance, students were to be assigned to one of the following options:

1) English 100A (accelerated);
2) English 100 ("regular");
3) English 101 (with a one credit supplemental lab);
4) English 22 (special course of instruction for those students whose writing shows that they are unprepared to meet the demands of English 100).

A similar, but separate, system of writing placement existed for international students. On the basis of their writing performance on the English Language Institute Placement Test (ELIPT), international students have traditionally been assigned to one of the following options:

1) ESL 100 (the exact equivalent of English 100, but tailored to foreign students' needs)
2) ESL 73 (intermediate ESL writing, preparatory for ESL 100)

Both the native speakers and foreign students were required to take the required course(s) within their first two semesters. After completing that requirement, students would also be required to take five writing-intensive courses as part of their other degree requirements (at least two of these in
upper-division courses). The Mānoa Writing Program was to be phased in, with two writing-intensive courses required for those students matriculating in 1987, three in 1988, four in 1989, and the full five course requirement in 1990. It soon became clear that the 1300 foreign students on campus at the time might have different needs and options under rules operating at the time. Unfortunately, it was not clear in the state of Hawai‘i just how a “foreign” student should be defined. Thus it was decided by the Mānoa Writing Board that foreign students would temporarily be subject to both the ELIPT and the MWPE and that the actual placement of foreign students into writing courses would be handled jointly by the Director of the ELI and the Director of the Mānoa Writing Program.

Purpose

The overall purpose of this research was to investigate the degree to which definitions of “foreign” students are adequate for purposes of distinguishing the students who can most benefit from ESL training from those students who will most profit from regular English Department composition courses. This not only opens the question of what constitutes a “foreign” student, but also the degree to which “foreign” students actually form a separate population with regard to their ability to write in English. In addition, the relative appropriateness of ESL and English Department placement tests for measuring the writing ability of such students was investigated.

To help clarify the above issues, this study was designed to investigate the following research questions:

1) What were the descriptive statistics for those students who took both the ELIPT and the MWPE?
2) On the MWPE, did “foreign” students form a separate population from the population of native speakers of various dialects of American English? What characteristics should be used to distinguish these two populations?
3) To what degree were the ELIPT and MWPE rating scales reliable as applied at UHM? How did the two tests compare in reliability for placing “foreign” students consistently into appropriate writing courses?
4) To what degree were the ELIPT and MWPE rating scales valid as applied at UHM? How did the two tests compare in validity for purposes of placing “foreign” students into appropriate writing courses?

**METHOD**

**Subjects**

As mentioned above, all incoming Freshman and transfer students without English 100 credit were required to take the Mānoa Writing Placement Examination (MWPE). The MWPE was administered on four dates in the Spring semester 1988 and on three during Fall 1988. These examinations were conducted primarily at the Mānoa campus of the UH but were also offered at other sites on Maui, Kaua‘i and the island of Hawai‘i for the convenience of potential students from all parts of the state. A total of 1769 students took the MWPE—1063 in Spring semester and 706 in Fall. They included 57.8 percent females and 42.2 percent males. All of these students were undergraduates, mostly entering university for the first time, i.e., transfer students typically had already satisfied the composition course requirement and therefore were not required to take either writing examination.

In addition to the MWPE, the foreign students were told that they were required to sit for the three-hour ELIPT. At UHM, “foreign” students are defined as those who:

1) were born in countries other than the United States, Canada, United Kingdom, Australia, and New Zealand, or
2) claim a native language other than English, and
3) have been in an English speaking country for 11 years or less.

Automatic exceptions from that definition include students who have:

1) 12 or more years of schooling in any of the English speaking countries listed above,
2) received a baccalaureate degree from an accredited university in any of those English speaking countries,
3) attained a score of 600 or higher on the Test of English as a Foreign Language.
There were approximately 1300 foreign students (out of about 21,000) on the UHM campus at any given time during the year. Of these, 470 took the ELIPT during the calendar year. Among students taking the test, 30.2 percent were graduate students, 61.2 percent were undergraduates and 8.6 percent were in non-degree status. In addition, they were 51.4 percent female and 48.6 percent male in a wide variety of majors. These students came predominantly from Asian countries (85 percent), but there were some students from elsewhere in the Pacific Basin (5 percent), Latin America (4 percent), Europe (4 percent), Africa (1 percent), and the Middle East (1 percent). Of the Asian countries, Hong Kong, the People’s Republic of China, Japan and Singapore had the largest contingents, but all Asian countries were represented. The mean TOEFL score for this group of students was 544.45 with a standard deviation of 29.88.

All international students were told to take the MWPE when they registered to take the ELIPT. This was done systematically, but orally, by the department secretaries when students came to the ESL Department to register for the ELIPT. It turned out that only 43 students actually did as they were instructed. Unfortunately, this means that the results are difficult to interpret because this group of students is self-selected, i.e., the results cannot be generalized beyond the particular group involved because the students themselves appear to have decided whether or not to take the MWPE based on reasons of their own. Nevertheless, these 43 students form an important subgroup because their results on the two tests can be directly compared.

Materials

MWPE testing materials. Students were notified of the placement examination requirements in their acceptance letters. This notice included detailed information about what they should expect to happen at the actual testing session. This was done so that they would be as relaxed and comfortable as possible under such testing conditions.

During the examination, students wrote in response to two essay topics. The questions in the first part required the students to write an analytic essay after reading approximately one and one-half pages of prose (e.g., one passage presents two points of view on genetic engineering and another gives an account of the present state of the women’s movement). The second part asked
the students to write an essay based on personal experience (e.g., one question asked for an analytic discussion of the effects of television on contemporary society and another required a discussion of prejudice in education). Each question had lined paper attached to it upon which the students were to write their responses. Seven different sets of questions were used. These sets were randomly distributed throughout the entire population.

**MWPE scoring materials.** Prior to the scoring session, raters were sent the holistic rubric which was used for scoring essays. The rubric contained sample test questions, example responses which had been scored according to the five point rubric, and a justification for those holistic scores. Raters were told to read the rubric so that they would be thoroughly acquainted with the scoring procedures. The rubric was a six point scale with descriptors for typical writing at each point on the scale (see Brown 1988, 1989, or 1990a for a copy of the scale).

**ELIPT testing materials.** ESL students were also notified of placement examination requirements in their acceptance letters. This notice included detailed information about where and when they should sign up for the ELIPT. The ELIPT included six subtests, one of which was the Writing Sample (the others test the students abilities in listening, reading and grammar). During the Writing Sample portion of the examination, students chose one of four possible essay topics to write on. These topics generally required personal experience types of responses. The directions and questions had lined paper attached upon which the students were to write their responses.

**ELIPT Writing Sample scoring materials.** An analytic scoring scale (see Jacobs et al 1981) was used in the ELI to score the writing samples. This scale required separate (weighted) scores in each of five categories: content (30 percent), organization (20 percent), vocabulary (20 percent), language use (25 percent) and mechanics (5 percent). Within each of these separate categories, four brief descriptors were provided to help determine the separate scores. These five subscores were then added up to determine the total score for each student.

**Procedures**

**MWPE test administration.** The MWPE took a total of about five hours to administer. The first 75 minutes were devoted to drafting a response to the first question. Then there was a 15 minute break, followed by an additional 75
minute period for drafting the second response. There was an additional two hour period after the lunch break that students could use to revise their responses to both topics. Students were required to return for the revising session but were allowed to leave as soon as they had completed their revisions. They had access to dictionaries throughout the test.

**MWPE scoring.** The examinations were scored holistically. Scoring sessions were three hours long, with the first hour devoted to training. The sessions began with the raters reviewing a one-page explanation of the purposes and problems in holistic rating procedures. Typical scoring sessions were devoted to only one topic. Raters then practice-scored three essays on the day's topic. Raters' practice scores were compiled and discussed until a consensus emerged on the appropriate ratings for the practice essays. If no consensus was reached, raters repeated the practice procedure. A two-hour scoring session followed this hour of training. Raters did not know the name, gender or ethnicity of the writers of any essay.

Each essay was rated by a minimum of two readers. If there was a two-point or greater discrepancy between scores, the essay was read by a third scorer. Thus for the two essays written by each student, there were from four to six ratings.

**ELIPT test administration.** The entire ELIPT took about three hours and thirty minutes to administer. The battery of tests included six subtests: the Academic Listening Test, Dictation, Reading Comprehension Test, Cloze, Writing Sample and Academic Writing Test. These subtests were all administered in the morning and were followed in the afternoon by personal placement interviews with each student. During the Writing Sample subtest, students were allowed thirty minutes to plan and write their essays.

**ELIPT scoring.** Immediately prior to the scoring session raters received a brief orientation to help familiarize them with the scoring grid and its use. Because of serious problems with the reliability of ratings for the Writing Sample (first noted in 1986), each composition was read by a minimum of three readers. This policy much improved the reliability of this rating procedure in the ELI (from estimates in the .55 to .59 range to much higher estimates in the .80 to .85 range).
Analyses

The analyses included descriptive statistics for central tendency (mean), as well as indicators of the dispersion of scores (range and standard deviation) around that central tendency. These statistics were used to indicate the degree to which the test was doing a sound job of centering the students’ scores and spreading them out along a continuum of writing abilities to aid in the placement decisions that would be based on the scores. Frequencies and percentages were cross tabulated to determine how many students in the various groupings being studied were placed into each of the courses.

Since there were a minimum of four scores for each student on the MWPE and three scores each on the ELIPT, the reliability of these tests was estimated using a version of the K-R20 formula specifically adapted for use with rating scales (after Ebel 1979). This approach was felt to be superior to the more traditional interrater approaches to composition rater reliability in that it allowed for estimating the overall reliability of the procedure while considering the contribution of all three of four ratings simultaneously. This procedure is related to the intraclass correlation coefficient discussed by Krzanowski and Woods (1984).

A one-way ANOVA procedure and a single sample t test were used to explore differences among the means for various groupings of students. Chi-square analysis was also employed to investigate the significance of differences among the frequencies for placement of students in various groups into the four English courses. All analyses were performed using the Quattro (Borland 1987) spreadsheet program or the ABSTAT (Anderson-Bell 1989) statistical program on an IBM personal computer. The alpha decision level for all statistical decisions was set at $\alpha < .05$.

RESULTS

Descriptive Statistics

The overall results of the 1988 MWPE and the ELIPT Writing Sample are described in Table 1. The first row shows the number of students (N) who took the examination. The means, standard deviations (STD), minimum score (MIN) and maximum score (MAX) are in the rows that follow. Starting with central tendency, notice that the mean on the MWPE is very well centered as indicated by the fact that it is approximately halfway between the possible
extreme scores of 0 and 20 on the scale. This 20 point scale is based on the sum of the four ratings (described above in the Materials section) for each student. The ELIPT Writing Sample has a mean that is not quite so neatly centered (the mean represents a score of about 70 percent of the total possible points). However, this does not appear to be a problem because there is no apparent skewing. The scores on both tests seem to be reasonably well dispersed about the mean as indicated by the standard deviations (which are relatively large as a proportion of the total number of possible points) and ranges. Remember that these results are based on the entire populations that took each of the two tests.

<table>
<thead>
<tr>
<th>STATISTIC</th>
<th>MWPE 1988</th>
<th>ELIPT 1988 WRITING SAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1769</td>
<td>314</td>
</tr>
<tr>
<td>MEAN</td>
<td>10.86</td>
<td>69.46</td>
</tr>
<tr>
<td>STD</td>
<td>2.54</td>
<td>9.54</td>
</tr>
<tr>
<td>MIN</td>
<td>0.00</td>
<td>34.00</td>
</tr>
<tr>
<td>MAX</td>
<td>19.00</td>
<td>92.00</td>
</tr>
<tr>
<td>RANGE</td>
<td>20.00</td>
<td>58.00</td>
</tr>
</tbody>
</table>

Table 1: Summary Descriptive Statistics (MWPE)

Are “Foreign” Students Really Different?

As mentioned above, there is some question as to the adequacy of definition of “foreign” students at UHM. This issue becomes particularly important when it comes to distinguishing those students who would most benefit from ELI training in writing from those who would most profit from English Department composition courses.
<table>
<thead>
<tr>
<th>TEST</th>
<th>Grouping</th>
<th>STATISTIC</th>
<th>N</th>
<th>MEAN</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MWPE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>Born Anglophone country</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L1 = English</td>
<td></td>
<td>1502</td>
<td>10.99</td>
<td>2.46</td>
</tr>
<tr>
<td></td>
<td>L1 = Other language</td>
<td>More than 12 yrs.</td>
<td>23</td>
<td>10.52</td>
<td>2.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11 yrs. or less</td>
<td>3</td>
<td>8.67</td>
<td>2.08</td>
</tr>
<tr>
<td>*</td>
<td>Born other country</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>L1 = English</td>
<td></td>
<td>102</td>
<td>10.89</td>
<td>2.52</td>
</tr>
<tr>
<td></td>
<td>L1 = Other language</td>
<td>More than 12 yrs.</td>
<td>49</td>
<td>10.93</td>
<td>2.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11 yrs. or less</td>
<td>90</td>
<td>8.77</td>
<td>3.04</td>
</tr>
<tr>
<td>*</td>
<td>Students who took both tests</td>
<td></td>
<td>43</td>
<td>7.82</td>
<td>2.54</td>
</tr>
<tr>
<td>*</td>
<td>Total</td>
<td></td>
<td>1769</td>
<td>10.86</td>
<td>2.54</td>
</tr>
</tbody>
</table>

| ELIPT |          |           |    |      |     |
|       |          |           |    |      |     |
| *    | Undergrads. | (non-transfers) | 173 | 69.22 | 8.83 |
|       |               | 47.79% |   |      |     |
|       |               | (i.e., 173/363 total UGs) |   |      |     |
| *    | Students who took both tests |           | 43 | 72.99 | 6.86 |
| *    | Total |           | 314 | 69.46 | 9.54 |

Table 2: Group Statistics (MWPE & ELIPT)

To examine the degree to which the foreign students are different from the mainstream population, all students who took the MWPE were asked their country of birth, their native language and the amount of time that they had spent in Anglophone countries (narrowly defined as the United States, Canada, United Kingdom, Australia or New Zealand). Our decisions in identifying
foreign students have hitherto been based on the notion that "foreigners" are those born in other than Anglophone countries, or claim a native language other than English, and have spent 11 years or less in an Anglophone country. The top portion of Table 2 shows the results on the MWPE for the various combinations of those three variables. Notice that the results are first categorized by whether or not the students were born in an Anglophone country. Under each of these categories, the students are grouped into those who claimed English as their L1 and those who claimed some other language as L1. Since none of the students who claimed English as L1 had been in an Anglophone country for 11 years or less, only the students claiming some other language as L1 were further subdivided into those who had spent 12 or more years in an Anglophone country and those who had spent 11 years or less.

Notice that the means for all students claiming English as L1 are remarkably similar whether they were born in an Anglophone county or not (10.99 and 10.89, respectively). The means for students claiming some other language as L1 and 12 years or more in an Anglophone country are also remarkably similar to each other and to the English L1 students regardless of place of birth (10.52 and 10.93, respectively). One-way analysis of variance procedures indicate that there is no significant difference among these four groups of students (F = .2406; df = 3, 1673; p = .7865). The groups that do stand out are those students who claim some other language as L1 and have been in Anglophone countries for 11 years or less. A single sample t test indicated that these students are statistically different from the population as estimated by the performance of all students on the 1988 MWPE (t = 6.59; df = 89; p < .001; two-tailed). This result indicates that these students are indeed statistically different from the population as a whole and therefore probably form a separate population.

Another way to approach this issue was to separately examine the percentage of students in each group that was placed into each of the composition courses offered at UHM. Table 3 presents the percentages (for rows) of students in each group who fell within the score ranges for each of the four courses. In this table, one pattern seems clear: the students who claim that L1 is other than English (regardless of where they were born) who have been in an Anglophone country for 11 years or less were much more likely than all other groups to be placed into English 22 (remedial) and much less likely to be placed into English 100 (regular) or English 100A (accelerated). Even more
striking differences were found for those students who took both the ELIPT and the MWPE.

<table>
<thead>
<tr>
<th>* Grouping</th>
<th>COURSE</th>
<th>ENG22</th>
<th>ENG101</th>
<th>ENG100</th>
<th>ENG100A</th>
<th>TOTAL N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0−7</td>
<td>8−9</td>
<td>10−14</td>
<td>15−19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Born Anglophone country</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1 = English (84.9%)</td>
<td>7.9%</td>
<td>16.1%</td>
<td>68.7%</td>
<td>7.3%</td>
<td>1502</td>
<td></td>
</tr>
<tr>
<td>L1 = Other language</td>
<td>15.4%</td>
<td>23.1%</td>
<td>57.7%</td>
<td>3.8%</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>12 yrs. or more</td>
<td>(13.0%)</td>
<td>(21.7%)</td>
<td>(60.9%)</td>
<td>(4.3%)</td>
<td>(23)</td>
<td></td>
</tr>
<tr>
<td>11 yrs. or less</td>
<td>(33.3%)</td>
<td>(33.3%)</td>
<td>(33.3%)</td>
<td>(0.0%)</td>
<td>(3)</td>
<td></td>
</tr>
<tr>
<td>* Born other country</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1 = English</td>
<td>6.9%</td>
<td>21.6%</td>
<td>62.7%</td>
<td>8.8%</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td>L1 = Other language</td>
<td>25.9%</td>
<td>18.7%</td>
<td>51.8%</td>
<td>3.6%</td>
<td>139</td>
<td></td>
</tr>
<tr>
<td>12 yrs. or more</td>
<td>(6.1%)</td>
<td>(14.3%)</td>
<td>(75.5%)</td>
<td>(4.1%)</td>
<td>(49)</td>
<td></td>
</tr>
<tr>
<td>11 yrs. or less</td>
<td>(36.7%)</td>
<td>(21.1%)</td>
<td>(38.9%)</td>
<td>(3.3%)</td>
<td>(90)</td>
<td></td>
</tr>
<tr>
<td>* Students who took both tests</td>
<td>(46.5%)</td>
<td>(27.9%)</td>
<td>(25.6%)</td>
<td>(0.0%)</td>
<td>(43)</td>
<td></td>
</tr>
<tr>
<td>TOTAL 1988</td>
<td>9.3%</td>
<td>16.7%</td>
<td>66.9%</td>
<td>7.1%</td>
<td>1769</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Percentages Placed in Each Course by MWPE (for Different Groupings by Rows)

Another implication for our decision making was that the students' country of birth appears to be a revealing variable. Chi-square analysis of the frequencies underlying the percentages presented in Table 3 indicates that there were significant differences in placement due to whether a student was born in an Anglophone country or not ($\chi^2 = 28.07$; df = 3). Further analysis indicated that the differences from expected frequencies occurred principally in the numbers of students in each group placed into English 22.

In short, the following were found to be important identifiers for a "foreign" student, in terms of distinguishing them from the main body of students (at least with regard to writing ability): they were born in a non-
Anglophone country, they claim a language other than English as L1, and they have been in an Anglophone country for 11 years or less.

Relative Reliability

The reliability of a test is defined as the degree to which it is consistent. This section will be devoted to examining the relative reliability of the ELIPT and the MWPE. The term reliability will be used here to refer to the degree to which scores were assigned consistently using the administration and scoring practices of the MWPE and the ELIPT Writing subtest. In order to understand the implications of this reliability information, it is useful to assess the results in two different ways. First, reliability coefficients should be examined and then the standard error of measurement (SEM) should be considered along with its implications for the decision making processes.

Traditionally, writing scales such as these are evaluated for reliability in terms of interrater reliability, i.e., the degree to which the raters agree in their assignment of scores. Since correlation coefficients estimate the degree to which two sets of numbers are related, Pearson product-moment coefficients can be calculated between the first and second ratings assigned for each composition to estimate the reliability of either set. For instance, the correlation coefficient between the first and second ratings of the MWPE was .53 (N = 1769). The correlation between the third and fourth ratings was .54. Both of these correlations indicate that there was a certain amount of agreement (53 and 54 percent, respectively) but also a good deal of disagreement (47 and 46 percent, respectively) in the sets of ratings being analyzed. The apparently high degree of unreliable variation in scores might be worrisome if it were not true that the actual placement of students is based of four scores rather than one or two.

Reliability coefficients. Application of Ebel's (1979) version of K-R20 indicates that the MWPE was producing scores that were reliable at .67 overall for all 1769 scores taken together (see Table 4). This can be interpreted directly as the proportion of the variance among the scores that is consistent. In other words, approximately 67 percent of the score variance can be considered true score variance, while the remaining 33 percent must be viewed as random variation which cannot be systematically accounted for. This is useful information in the sense that it helps understand the degree to which the students' writing abilities are being assessed in a consistent manner.
Table 4: Summary Reliability Statistics

Table 4 also gives the equivalent figures for the ELIPT when calculated for the total population of students (.88) and when calculated for the undergraduate students only (.85). In both cases, the reliability of ELIPT analytic scoring procedures appears to be moderately higher than that for the MWPE—even though four ratings are consistently used for each student on the MWPE, while only three ratings are normally given on the ELIPT. A similar, though less striking, difference is reflected in the reliability of the ratings on both tests when they are calculated for those students who took both tests.

Standard error of measurement. Another way of looking at the reliability of a test is to consider the standard error of measurement (SEM)—especially as it relates to the decision making processes. In brief, the SEM is a statistic which expresses (in probability terms) the degree to which scores are likely to fluctuate due to unreliable variance. For example, the SEM of 1.46 found for the MWPE (see Table 4) indicates that students’ scores on the MWPE would fluctuate ± 1.46 points with 68 percent certainty if they were to take the test again. This has important implications for decision making especially for those students who are close to the cut points between levels in the placement decisions. Consider a student who has a score of seven and is therefore placed into the English 22 course. The SEM indicates that, with 68 percent certainty, a student with a seven might score as high as 8.46 (7 + 1.46) and be placed into the next level higher, or as low as 5.54 (7 – 1.46) and still be placed into English 22. If the probability level is raised to say 95 percent certainty, the band of potential fluctuation becomes even wider, i.e., ± 2 SEM = ± 2.92.
From a practical point of view, educators usually wish to protect the interests of the students and be fair by not placing them too low. As such, it is sometimes important that additional pertinent information be considered in placing those students who fall below such a cut point, but within one SEM of it. A narrow band is desirable for decision making. Hence, a low SEM is a good SEM. In absolute terms, the SEMs reported in Table 4 for the MWPE are lower than they are for the ELIPT Writing Sample. However, it is important to remember that the ELIPT scale has 100 possible points while the MWPE scale only has 20. For this reason, the SEM is also reported as a percent (in parentheses) of the total possible on each scale. In these percent terms, the SEM is about half as large for the ELIPT Writing Sample as it is for the MWPE regardless of the group that is considered.

The reliability and standard error of measurement estimates are useful for assessing how consistently raters are scoring and how much variation can be expected in scores with regard to placement decisions. But this does not help much in deciding how appropriate a test is for a particular set of placement decisions. After all, a test can produce scores that vary consistently while testing the wrong things.

Relative Validity

Thus validity, though related to reliability, is also important as a separate issue. Validity will be defined here as the degree to which a test is measuring what it claims to measure. Since the MWPE claims to measure the incoming students' ability to write academic prose for purposes of placement into the English composition courses, and since the ELIPT Writing subtest is used for purposes of placement into the four ELI writing courses, the validity question is slightly different for each. One way to examine the validity of each measure for the corresponding placement decisions is to gauge the degree of successful placement as judged by the teachers of the courses to which the students were assigned. This task was approached in two slightly different ways by the ELI and the Mānoa Writing Board.

The Mānoa Writing Board sent out a survey questionnaire to relevant teachers to ask them about the success of placement decisions. As explained in Brown (1989), the survey was answered by 23 teachers who were collectively responsible for thirty sections of the composition courses. A majority (78 percent) felt that the placement was "very effective" or "effective" in sending
students to the correct level of instruction. Four instructors felt that the placement was "somewhat effective." These instructors generally felt that those students who had been misplaced should have been placed lower. One instructor rated the placement as having "room for improvement." This instructor felt that half of the students should have been placed higher. In short, there appears to have been at least a consensus, among the English composition teachers who responded to the survey, that the placement decisions based on the 1988 MWPE were effective.

The ELI required teachers to test the students once again with an in-class writing assignment during the first week of class and consider whether any misplaced students should be moved up to the next level. Thus the inaccuracy of placement decisions can be gauged to some degree by the number of students that teachers recommended moving to another level of study. There were 24 such requests out of nearly 700 ELI course enrollments during 1988; 13 of these misplacements were in writing courses.

On the whole, both tests appear to be satisfactory with regard to the teachers' views of how accurate the placement decisions are. It would be pointless to argue that one test is better, or more accurate, than the other.

However, other issues related to validity do merit comment. Recall that the MWPE was administered in two parts, which required responses to two different topic types. The two parts together were designed to test how students respond to reading-based and personal-experience-based writing assignments. The scoring for the MWPE was done holistically. Alternatively, the ELIPT used a writing subtest that was based on a composition written in 30 minutes. The questions were all designed to force the student into writing a comparison or contrast based on personal experience. The scoring on the ELIPT was based on an analytic scoring grid, which provided separate scores for content, organization, vocabulary, language use and mechanics.

On the whole, the test administration procedures for the MWPE appear to be more thorough than those for the ELIPT in that the MWPE allows five hours for writing on two topics, while the ELIPT allows only 30 minutes on one topic. In contrast, the ELIPT appears to be more thorough than the MWPE in terms of scoring procedures; the ELIPT requires analytic scoring, which seems to focus the raters on more detailed analysis of each student's writing performance, while the MWPE involves a holistic scoring procedure that ends in a single "gut reaction" decision. Clearly there is room for much more
research on the relative validity of these contrasting administration and scoring variables.

The fact that only a non-significant \( (p > .05) \) correlation of .2278 (or .3145 corrected for attenuation in both measures) was found between the two sets of scores for the 43 students who took both tests suggests that these tests are not very similar in what they assess. It is important to note that this group of students (the ones who took both tests) is self-selected and may not be typical of students who took either the MWPE or the ELIPT.

**DISCUSSION**

To sum up, the MWPE and ELIPT both appear to have functioned reasonably well during all of the 1988 administrations as indicated by the descriptive statistics, reliability estimates and validity. Thus they seem to be reasonably sound as tools for making placement decisions about the two groups of students involved.

The descriptive statistics indicate that both tests are producing scores that are reasonably well-centered and are dispersing the students with adequate efficiency within the range of possible scores. In addition, examination of the frequency distributions indicates that the scores on both tests are normally distributed about the mean. In other words, the MWPE and ELIPT appear to be efficient tests for the types of norm-referenced placement decisions that are being made.

In addition, it appears that, on the MWPE, "foreign" students form a separate population from the population of native speakers. Decisions about who belongs in this group can apparently be productively based on the self-reported place of birth and L1 of the students—particularly for those who have been in an Anglophone country for 11 years or less. Within that group, there was no meaningful correlation between number of years in Anglophone country and scores on the test.

All of this is important information in the ongoing search for ways to distinguish those students who can best be served by the English and ESL departments. It also emphasizes the need for ongoing cooperation in serving the needs of the differing groups of students on our campus. This is particularly important for those students on the borderline between the courses
offered in the ELI and English department. Perhaps, those students who wish to take both tests should continue to be allowed that right, especially since it appears that those who would choose to do so tend to be the stronger writers on the ELIPT (see Table 2). Under those conditions, the results from both tests could be considered jointly by the two organizations who would then issue a joint placement decision.

The reliability estimates in this study indicate that the scores on both the MWPE and ELIPT are reasonably consistent when calculated using reliability estimates that consider all raters simultaneously—though the ELIPT scoring procedures appear to be marginally more reliable. Naturally, these results must be viewed as lower-bound estimates of the reliability of each test because they would probably all be higher if the extra readings (beyond the four, or three that all students received) could be factored into the analysis.

Although the tests already appear to be reasonably reliable, the Mānoa Writing Board and English Language Institute should make every effort to improve the reliability. Generalizability studies (Brown 1988, 1989, 1990a) of the MWPE indicated that increasing the number of topics would have a greater effect on improving the reliability (G coefficients) than increasing the number of raters. In other words, it would probably be more profitable to increase the number of topics that students must address than to increase the number of raters who read their compositions. Other measures that might be taken to help increase the reliability of both tests would be to 1) strengthen the training that the raters receive, 2) clarify the rubrics used to score the compositions, and 3) improve the testing conditions.

The results here suggest that both tests were reasonably accurate in terms of the placement decisions for which they were designed as determined in two different ways: 1) 78 percent of the English department instructors rated the placement “effective”; and 2) there were only 13 misplacements noted by ELI teachers in first week testing. The relative validity of these two approaches to placement on the UHM campus certainly remains an open question. Indeed, much more research should be done on all aspects of both tests.

Further Research

As is often the case, the results of this study have raised more questions than they have answered. For that reason, the following suggestions for further research are made:
1) How would the results have changed if the dividing line had been set to consider students as "foreign" when they have lived in Anglophone countries for 12 years or less, or 13 years or less, or 7 or 9 years?

2) How would the writing performances of students ("foreign" and native speakers alike) who transfer composition courses into UHM differ from students who do not transfer such courses? Upon arrival? At graduation?

3) What is the relative validity of the contrasting administration and scoring variables that differentiate the MWPE and ELIPT?

4) What alternative and/or additional sources of information (e.g., ACT Verbal scores, GPA, TSWE scores, portfolios, etc.) should be used in the placement of students into courses in the ELI and English department?

5) Will broadening the TOEFL requirement to include all "foreign" students (transfer or not) help determine which students should take each test?

6) Why did many of the "foreign" students who were told to take the MWPE in addition to their ELIPT requirement fail to do so?

CONCLUSION

There is no doubt that the writing-across-the-curriculum program and the resulting testing systems have had a particular impact on the foreign students and on the ELI that serves them. For instance, we see increasing numbers of students taking their English composition courses at community or private two-year institutions and transferring those credits into UHM (yet placing into our lower level courses in listening and reading). There is also little doubt that the ELIPT and MWPE have had an impact on the public schools in the state of Hawai‘i. It is natural that secondary school teachers who have many students bound for UHM will want their students to attain high scores.

In short, the advent of the Mānoa Writing Project and the coordinated testing between the English and ESL departments have had a number of dramatic effects on writing instruction in the state of Hawai‘i and the strategies that students use to enter the university. Some of these effects have been
salutary and should be encouraged. To that end, open communication about the purposes and effectiveness of the ELIPT and MWPE is important. If teachers, administrators and the general public have correct and clear information about the tests, positive effects should be maximized. Colleges and universities across the United States have much to offer their respective communities in terms of providing models of sound testing and teaching of academic writing, but only if they constantly examine and upgrade their writing curriculum policies and testing procedures.

Received March 14, 1990

Author's address for correspondence:
J.D. Brown
Department of English as a Second Language
University of Hawai'i
1890 East-West Road
Honolulu, HI 96822
REFERENCES

