VALIDITY OF THE IUPUI PLACEMENT TEST SCORES FOR COURSE PLACEMENT: 1995-1996

Howard R. Mzumara, Ph.D.
IUPUI Testing Center and School of Education

Mark D. Shermis, Ph.D.
IUPUI Testing Center and Department of Psychology

Deborah Wimer, B.A.
IUPUI Testing Center and Department of Psychology

620 Union Drive,
Suite G003
Indianapolis, IN 46202-5167
317-274-2620

November 1996

INDEX

Introduction

The present study was designed to assess the validity of placement test scores for Fall 1995 in making placement decisions at Indiana University Purdue University Indianapolis (IUPUI). Placing new students into appropriate first year courses has become an increasingly challenging task for colleges and universities. Also, the percentage of postsecondary institutions with some form of placement and developmental instruction has steadily increased in the past decade and is now about 90% (Sawyer, 1996). In a recent survey on *Remedial Education in Higher Education Institutions*, conducted by the National Center for Education Statistics (NCES), about three-quarters (78 percent) of higher education institutions that enrolled freshmen offered at least one developmental reading, writing, or mathematics course in Fall 1995 (NCES, October 1996).

A cursory review of the literature (Sawyer, 1996, and the references cited therein) offers two potential explanations for the increase in placement and developmental instruction. "One suggested explanation is that American high schools have become less effective in preparing students for college" (Carnegie Foundation for the Advancement of Teaching, 1988; National Commission on Excellence in Education, 1983; Singal, 1991, as cited in Sawyer, 1996, p. 271). A second explanation is that more students from disadvantaged backgrounds are attending higher education institutions (College Entrance Examination Board, 1977; Munday, 1976, as cited in Sawyer, 1996).

In placement decisions, the concern of the institution is to create learning environments in which all students will learn. Hills, Hirsch, and Subhiyah (1990) define placement as a process by which students are assigned to courses commensurate with their past achievements in order to facilitate expeditious further learning. "The underlying idea is that students differ. They may differ in their level of preparation, in their adeptness at learning, in their interests, in their ability to organize for themselves, and so on. As a result, for efficient instruction, one arranges for different approaches for the different students or groups of students. Ideally, a student is placed in the learning situation which is best for him" (Hills, Hirsch, & Subhiyah, 1990, p.5). Accordingly, IUPUI has implemented a mandatory placement testing program for all undergraduate students in order to facilitate the academic success of students at the University.

The IUPUI placement tests were developed for the purpose of course placement (i.e., matching students with instruction appropriate to their academic preparation) in English (writing), mathematics, and reading. Thus, like most other higher education institutions (NCES, October 1996), IUPUI provides developmental courses in reading, writing, and mathematics. Generally speaking, developmental courses are provided to those college students lacking in academic skills necessary to perform college-level work at the level required by the institution (NCES, October 1996). Although what constitutes developmental courses varies from institution to institution, often developmental courses do not carry credit toward satisfying degree requirements.

In general, the rationale for placement testing is threefold: First, students who enroll in appropriate university courses should have a more positive experience than those who enroll in courses that are either too difficult or too easy. They should be more satisfied with their university experience, and thus more likely to be retained. Second, because students are more likely to be retained in classes that are appropriate to their ability level, departmental administrators can more carefully plan how best to allocate faculty resources to respective class sections. Finally, the placement tests might serve as a basis for assessing the contributions of the University to the development of general educational skills. If placement tests perform the three vital functions well, then University funds spent on these assessments are wisely expended. A brief description of the major IUPUI placement tests follows.

Mathematics Placement Exam. The mathematics placement examination was designed to assess students' college-level mathematics readiness skills upon admission into IUPUI. The mathematics exam used in the present study was a linearly computerized test which consisted of 40 objective items that assessed skills ranging from basic arithmetic operations to introductory calculus. The total raw score on the mathematics test ranged from 0 to 40, and represents the number of items a student answers correctly. For placement purposes, students who score 6 or less on the test are advised to take Mathematics M010 (Pre-Algebra), a developmental mathematics course. Students scoring between 7 and 14 on the mathematics placement test are advised to take Mathematics 001 (Introduction to Algebra). Students who score between 15 and 23 on the mathematics test receive a placement recommendation of Mathematics 111 (Algebra). Students scoring at 24 or higher are advised into a variety of course offerings, depending on their academic major. Based on this assessment, therefore, placement is made into an appropriate mathematics class. Mathematics placement test scores are valid for one year from the test date.

Reading Placement Exam. The reading placement exam comprised the *Nelson-DennyForm E*, an objective reading test published by the Riverside Publishing Company (Brown, Bennett, & Hanna, 1981). The purpose of this test was to assess students' vocabulary and reading comprehension skills upon entry into IUPUI. The exam consisted of three parts. The first part (vocabulary) covered definitions or synonyms for college-level words. The second part assessed the reading speed of the examinee. In the third part, students were asked to read several short-timed passages, and then asked questions designed to assess their comprehension of each passage. Based on this test, a student's eligibility for university reading requirements or the need for reading improvement was determined. The total reading score ranged from 0 to 170, and indicates a weighted raw score composite. Students who scored less than 68 were advised to enroll in Education X150, and a score of 68 to 79 resulted in a placement recommendation for Education

enroll in Education X150, and a score of 68 to 79 resulted in a placement recommendation for Education X152. Students who obtained a score equal to or greater than 80 were exempt from further requirements in reading skill development. Thus, students who read at college level were exempt from taking college reading classes. Reading placement test scores are valid for one year from the test date.

English Placement Exam. The English placement exam is a one-hour exam that asks students to write an essay that explains and supports their opinion on a current social issue. The test provides a brief explanation of the issue or the context in which the issue is posed. Students are also asked to evaluate their answer and explain what changes they might make, had they the time to do so. When readers assess the English placement tests, they look for the presence or absence of organization, support, development, and position. Students who need extra help focusing their essays around a major theme, or students who need extra help understanding the relationship between assertion and support, are placed into an appropriate developmental course.

The purpose of the English placement exam is to assess students' ability to write an essay that explains and supports one's viewpoint or opinion on a given issue. Examinees have a choice of two questions, each of which allows the students to use their personal experiences and/or observations in writing the essay. It is important that the test conveys not only the examinee's viewpoint on the selected topic, but also the reasons for taking a particular position. The test, however, does not require any specialized knowledge or research, only an ability to discuss an individual's opinion and reasons. Within the one hour time allotted to the English placement test, students are expected to (a) think seriously about the topic selected, (b) state an opinion clearly, (c) present examples or details that support an opinion, and (d) organize the essay clearly. The English test score is valid for two years from the test date. Students are required to register for the respective courses into which they are placed [i.e., E010 - Access to Writing (a developmental course for students judged not yet ready for W001); W001 - Fundamentals of English; W131 - Elementary Composition; or W140 - an honors version of W131].

Changes Made in Past Two Years

Several changes were implemented based on the recommendations outlined in the previous placement validity reports. First, it was recommended that both mathematics and reading should use alternate outcome variables. Note that the traditional variable of "grade" was problematical because mathematics employs a truncated scale that excludes the grade of "D". Similarly, the grades in reading were based on a dichotomous outcome of "Pass/Fail" and were further complicated because they included non-reading performance criteria (e.g., attendance). The collapsing of course grades made this outcome less variable and tended to depress the true linear relationship. Partly because of this, correlation coefficients in mathematics were in the low .20s; and reading correlations were close to zero. Because mathematics uses a common final exam, this criterion seemed to be an appropriate supplement to course grades. Moreover, since reading instructors administer *Form F* of the *Nelson-Denny Reading Test* at the end of each semester, this reading score provided a convenient outcome replacement.

The second recommendation was that the Mathematics Department consider moving towards an adaptive test that was both configured to the department's curriculum and also drew from a larger item bank than the linearly computerized test. Consequently, full implementation of the computerized adaptive mathematics placement test took place in late October 1995. A cursory review of the literature suggests that the adaptive procedure should produce a more precise estimate of ability, take a shorter time to complete, and be perceived as easier for most students. Preliminary results from the adaptive mathematics placement test are quite promising. So far, it seems the current adaptive mathematics test is working properly in that the new testing procedure is relatively efficient and yields higher coefficients with the mathematics outcome measures.

The third recommendation was that the reading faculty consider developing an in-house computerized test that would eventually incorporate both adaptive and diagnostic features. While there was no widespread dissatisfaction with the current test (*Nelson-Denny Reading Test, Form E*) among counselors, near zero predictive validity coefficients coupled with the cumbersome test administration suggested that an assessment more closely linked to the curriculum was appropriate. The vision for the test was that it would perform as a placement test for all students, but would transform into a diagnostic procedure for those who were recommended for developmental coursework. Consequently, the linearly computerized reading placement test was implemented in late June of this year (Shermis, Wolting, & Lombard, 1996). The adaptive version of the test should be completed early next year.

The final suggestion had to do with modifications to the scoring procedure for the English test. The

The final suggestion had to do with modifications to the scoring procedure for the English test. The problem with the previous procedure was that only one score was associated with each course placement. This meant that there was no variance in the predictor and consequently no correlation coefficient could be calculated. The recommendation was that each class placement take on a range of predictor scores (ideally, 5 to 7 categories for each class). The English Department developed a new scale during Fall 1994 and incorporated the new rating system in January, 1995. Thus, the current English placement test scores are based on the new scoring procedure.

Method

Sample

The target population comprised all students who took either an English, mathematics, or reading placement test from January, 1995, through August, 1995, **and** enrolled in an English writing, mathematics or reading course during the Fall 1995 session at IUPUI's Indianapolis campus. (Note that withdrawals or students with incomplete course grades were excluded from the present study.)

Procedure for Obtaining the Data

Students' raw data were obtained through a FOCUS query (and/or students' academic records provided by the respective course coordinators) from the Fall 1995 cohort of students who took both placement exams and subsequently enrolled in a mathematics, reading or English course. Because the IUPUI placement tests are seen as advisory rather than prescriptive, in some cases a student enrolled in a course that was not recommended by the placement test score. Consequently, for the purpose of statistical analyses, students were divided into two categories, namely, "compliant" and "non-compliant." The compliant group comprised students who took the recommended courses based on the placement test scores. The non-compliant group consisted of students who chose (sometimes in consultation with their advisors) not to take the recommended course. The current extract resulted in a pool of 2,675 students for mathematics, 718 students for reading (note that students exempt from taking reading classes were excluded from the present study), and 2,356 students for the English written essay. With respect to compliance, there were 2,025 students who took the mathematics placement exam and then enrolled in a recommended mathematics M010, math 001, math 111 or a higher class during the Fall of 1995. Likewise, 676 students took the *Nelson-Denny Reading Test Form E* and then enrolled appropriately in either Education X150 or Education X152. Also, there were 1,387 students who took the English placement exam and then enrolled in one of the following recommended courses of interest: E010, W001, W131, or W140. Overall, the total compliance rates were 76% for mathematics, 94% for reading, and 96% for English.

The present study employed simple correlation and regression analyses to demonstrate the relationship between predictor scores and outcome variables. Most college placement exams have correlations that run between .20 to .40 (cf. Hills, Hirsch, & Subhiyah, 1990). Previous studies conducted at IUPUI regarding the relationship between SAT/CAT scores and course outcomes have yielded correlations of near zero.

The predictor variables consisted of students' placement exam scores on mathematics, reading, and English, respectively. The outcome measures were the corresponding exam scores and/or course grades in the respective content areas. The rationale for the validation/research design was as follows. To the extent that the usefulness of a placement test depends on the existence of statistical relationships, such evidence is clearly essential to validation. Thus, by measuring the strength of this statistical relationship, we obtained evidence on the validity of the placement test scores for making course placement recommendations.

The primary criterion measure for mathematics comprised final examination scores based on a common final exam. The secondary outcome variable for mathematics consisted of the course grade obtained by the student at the end of the Fall semester. The mathematics grades ranged from "A+" to "F". For data analysis purposes, these letter grades were converted to a numeric scale ranging from 4.33 for an "A+" to 0.33 for an "F".

The primary outcome variable for reading consisted of the total reading score received by a student on Form *F* of the *Nelson-Denny Reading Test.* The predictor variable for reading comprised students' placement test scores based on the *Nelson-Denny Reading Test Form E.* In reading courses, an "S/F" grading scheme is employed. The "S" grade indicates that the course was completed satisfactorily, whereas, the "F" grade means that the student failed the course. For the purpose of statistical analysis, an

"S" grade was converted to a score of 2.00, and an "F" grade received a score of 0. As mentioned earlier, students who withdrew from either Education X150 or X152 course were excluded from the present study.

The outcome variable for English was the course grade obtained at the end of the fall semester. English grades ranged from "A+" to "F". The letter grades were converted to a numeric scale ranging from 4.33 for "A+" to 0.33 for "F". Students who withdrew from courses of interest were excluded from the present study.

The present validity study attempted to provide three major types of information. The first was the correlation among all the variables studied (i.e., predictor and outcome measures). The intercorrelations (validity coefficients) between the predictors and each outcome measure indicated how the predictors were working, but did not provide a means of making specific predictions for individuals. This was accomplished with the second type of information, prediction equations, which were the product of regression analyses. In essence, the prediction equations used one variable (i.e., the placement test scores) to predict an outcome. This information, however, did not indicate how effective the predictions were. Thus, the third type of information attempted in this study was measurement of the likely error in prediction (i.e., standard errors of estimate). Because the simple regression analyses did not seem to yield remarkably different results than those obtained with correlation analyses, correlation coefficients are presented in the next section.

Results

The basic placement procedure applied at IUPUI is mostly dependent upon students' academic achievement, as measured by the respective placement tests. The manner in which placement decisions are made has been described in the Introduction section. Thus, it is imperative that the effectiveness of the existing placement procedures be known. To provide some information, correlation (validity) coefficients were calculated between IUPUI placement test scores and subsequent achievement in the respective courses in which students enrolled during the Fall 1995 semester. Pearson product-moment correlation coefficients were generated for both compliant and non-compliant groups for the mathematics, reading and English courses. [Note: The correlation coefficient (r) reflects the nature and strength of the linear association between two variables. The *coefficient of determination* (r^2) and the *coefficient of nondetermination* ($1 - r^2$) are useful for getting an intuitive feel of the strength of association represented by r. These two statistics provide an interpretation of r in terms of explained and unexplained variation. That is, for a given linear relationship between predictor and outcome measures, we would know how much of the outcome-score variance is accounted for by the predictor-score variance and how much is not accounted for. Details regarding the interpretation of r are presented in most introductory statistics texts (e.g., Glass & Hopkins, 1984; Kirk, 1990).]

Table 1 provides a summary of the descriptive statistics for the compliant and non-compliant groups based on the respective placement test scores. The correlation matrices for the respective groups by course are provided subsequently as Tables 2 to 17. Based on the tables, a summary of the results per course is presented in turn.

Mathematics

The following is a summary of the descriptive statistics based on compliant and non-compliant groups for mathematics (see Table 1). The compliant group for mathematics M010 (n = 434) had a mean placement test score of 4.23 with a standard deviation of 1.33. In contrast, the non-compliant group (n = 131) had a mean placement test score of 4.50 and a standard deviation of 1.43. The compliant group for Mathematics 001 (n = 1154) obtained a mean mathematics placement test score of 10.26, with a standard deviation of 2.15. The corresponding non-compliant group for Mathematics 001 had a mean placement test score of 10.91 and a standard deviation of 2.26. For Mathematics 111, the number of compliant students was 437, with a mean placement score of 17.34 and a standard deviation of 2.21. In contrast, there were 141 non-compliant students in Mathematics 111 course with a mean placement score of 18.33 and a standard deviation of 2.55.

Table 1

Descriptive Statistics for the Compliant and Non-Compliant Groups Based on the Placement Exam Scores (PES) for Fall 1995

SUBJECT	COURSE	GROUP	PES MEAN	PES*SD*	*N*
Mathematics	M010	Compliant	4.23	1.33	434
		Non-compliant	4.50	1.43	131
	001	Compliant	10.26	2.15	1154
		Non-compliant	10.91	2.26	378
	111	Compliant	17.34	2.21	437
		Non-compliant	18.33	2.55	141
Reading	X150	Compliant	58.63	6.11	306
		Non-compliant	61.21	5.65	19
	X152	Compliant	73.85	3.34	370
		Non-compliant	72.74	3.33	23
English	E010	Compliant	1.73	.65	11
		Non-compliant	1.43	.79	7
	W001	Compliant	8.72	.97	580
		Non-compliant	8.62	1.15	439
	W131	Compliant	14.03	.97	786
		Non-compliant	14.02	.97	512
	W140	Compliant	20.80	1.03	10
		Non-compliant	19.91	.54	11

(*Note*. The non-compliant groups for English include students who did not register for any English course in Summer I, Summer II, or Fall, 1995.)

Correlation matrices for the various mathematics compliant groups are provided in Tables 2 to 4. Considering mathematics exam scores as the primary criterion measure, the correlation coefficients for the compliant students ranged from r = .17 for Mathematics M010 (n = 282) to r = .25 for Mathematics 001 (n = 817). The average correlation coefficient was r = .23 for the compliant students. Using course grades as the outcome measures, however, the correlation coefficients for compliant students ranged from r = .16 for mathematics M010 (n = 391) to r = .20 for Math 111 (n = 393); and the average r = .19. Correlation matrices for the three mathematics non-compliant groups are presented in Tables 6 to 8. With mathematics exam scores used as the main outcome variable, the correlation coefficients for the non-compliant students ranged from r = -.02 for math M010 (n = 35) to r = .39 for math 111 (n = 3); and the average r = .25. Similarly, the correlation coefficients for the non-compliant students ranged from r = -.08 for math M010 (n = 46) to r = .31 for math 001 (n = 176), with mathematics course grades used as the outcome measure. The non-compliant students had an average r = .17 for the respective courses in which they were enrolled.

A somewhat interesting situation can be seen in Table 2, which presents the correlation coefficients for the Math M010 compliant group. A statistically significant correlation was observed between the placement test scores and course grades for mathematics (r = .16, p < .05). In spite of the increased variability in the mathematics examination grades,

1. The average correlation coefficients were calculated by taking the correlation for each class, converting it to a z score using Fisher's r, weighting the z scores by the number of students in each class, dividing the weighted z scores by the total number of students, and then converting back using Fisher's formula.

Table 2

Correlation Matrix for the Math M010 Compliant Group

	MA100	MATHGRD1	EXAMSCR
MA100	1.000		
	(434)		
MATHGRD1	.156*	1.000	
	(391)	(391)	
EXAMSCR	.167	.739***	1.000
	(282)	(282)	(282)

Note:
* $p < .05$; ** $p < .01$; *** $p < .001$ (nondirectional test)
MA100: Math Placement Test Score; EXAMSCR: Math Exam Score;
MATHGRD1: Math Course Grade.
The numbers in parentheses indicate the respective sample sizes.

Table 3

Correlation Matrix for Math 001 Compliant Group

	MA100	MATHGRD1	EXAMSCR
MA100	1.000		
	(817)		
MATHGRD1	.200***	1.000	
	(1004)	(1004)	
EXAMSCR	.253**	.778***	1.000
	(817)	(817)	(817)

Note:
* $p < .05$; ** $p < .01$; *** $p < .001$ (nondirectional test)
MA100: Math Placement Test Score; EXAMSCR: Math Exam Score;
MATHGRD1: Math Course Grade.
The numbers in parentheses indicate the respective sample sizes.

Table 4

Correlation Matrix for Mathematics 111 Compliant Group

	MA100	MATHGRD1	EXAMSCR
MA100	1.000		
	(487)		
MATHGRD1	.204***	1.000	
	(393)	(393)	
EXAMSCR	.230***	.787***	1.000
	(315)	(315)	(315)

Note:
* $p < .05$; ** $p < .01$; *** $p < .001$ (nondirectional test)
MA100: Math Placement Test Score; EXAMSCR: Math Exam Score;
MATHGRD1: Math Course Grade.
The numbers in parentheses indicate the respective sample sizes.

Table 5

Correlation Matrix for Math M010 Non-compliant Group

	MA100	MATHGRD1	EXAMSCR
MA100	1.000		
	(131)		
MATHGRD1	-.082	1.000	
	(46)	(46)	
EXAMSCR	-.021	.705***	1.000
	(35)	(35)	(35)

Note:
* $p < .05$; ** $p < .01$; *** $p < .001$ (nondirectional test)
MA100: Math Placement Test Score; EXAMSCR: Math Exam Score;
MATHGRD1: Math Course Grade.
The numbers in parentheses indicate the respective sample sizes.

the validity coefficient between mathematics placement test scores and examination scores was not statistically significant. Generally speaking, if the placement test is related to instruction and both are reflected in the final exam, one would expect higher validity coefficients for variables with increased variability. The current lack of statistical significance might be an artifact of a smaller sample size for the exam variable, hence less statistical power.

Table 6, which provides the correlation coefficients for the Math 001 non-compliant group, shows another surprising situation. A statistically significant correlation was obtained between the placement test scores and course grades for mathematics ($r = .31$, $p < .001$). It would be interesting to find out the individual characteristics and performance of students who did not comply with the placement test recommendation. The sample sizes given in Table 7 are too small to yield stable or statistically significant coefficients.

Table 6

Correlation Matrix for Math 001 Non-compliant Group

	MA100	MATHGR1	EXAMSCR
MA100	1.000		
	(378)		
MATHGRD1	.312***	1.000	
	(176)	(176)	
EXAMSCR	.330	.638***	1.000
	(113)	(110)	(113)

Note:
* $p < .05$; ** $p < .01$; *** $p < .001$ (nondirectional test)
MA100: Math Placement Test Score; EXAMSCR: Math Exam Score;
MATHGRD1: Math Course Grade.
The numbers in parentheses indicate the respective sample sizes.

Table 7

Correlation Matrix for Mathematics 111 Non-compliant Group

	MA100	MATHGRD1	EXAMSCR
MA100	1.000		
	(137)		
MATHGRD1	-.060	1.000	
	(69)	(69)	
EXAMSCR	.394	.999	1.000
	(3)	(3)	(3)

Note:
* $p < .05$; ** $p < .01$; *** $p < .001$ (nondirectional test)
MA100: Math Placement Test Score; EXAMSCR: Math Exam Score;
MATHGRD1: Math Course Grade.
The numbers in parentheses indicate the respective sample sizes.

Reading

Tables 8 and 9 show the correlation coefficients for the reading compliant groups. The validity coefficients between the *Nelson-Denny Reading Form E* and *Form F* test scores range approximately from .10 for Education X150 to .18 for Education X152. Only the latter correlation coefficient is statistically significant at the .01 level. The average correlation coefficient was .14 for the reading compliant groups. Note that none of the other obtained validity coefficients between the placement test scores and course grades for reading were statistically significant at the .05 alpha level. The relatively small sample sizes and range restriction are most likely the two most important reasons for the present statistically non-significant results for reading. The null results for the non-compliant groups for Education X150 and X152 are presented in Tables 10 and 11.

Table 8

Correlation Matrix for Reading X150 Compliant Group

	READINGE	READINGF	X150GRD
READINGE	1.000		
	(306)		
READINGF	.070	1.000	
	(177)	(177)	
X150GRD,/TD>	.192	-.028	1.000
	(12)	(6)	(12)

Note:
None of the above correlation coefficients are statistically significant at $p < .05$.
READINGE:*Nelson-Denny Reading Test Form E*;
READINGF:*Nelson-Denny Reading Test Form F*;
X150GRD: EDUC X150 Course Grade for Fall 1995 semester
The numbers in parentheses indicate the respective sample sizes.

Table 9

Correlation Matrix for Education X152 Compliant Group

	READINGE	READINGF	X152GRD
READINGE	1.000		
	(370)		
READINGF	.177**	1.000	

	(272)	(272)	
X152GRD	-.070	n/a	1.000
	(46)		(46)

Note:
*p < .05; ** p < .01; *** p < .001 (nondirectional test)
READINGE:*Nelson-Denny Reading Test Form E*;
READINGF:*Nelson-Denny Reading Test Form F*;
X152GRD: EDUC X152 Course Grade for Fall 1995 semester
The numbers in parentheses indicate the respective sample sizes.
n/a: data not available.

Table 10

Correlation Matrix for Education X150 Non-compliant Group

	READINGE	READINGF	X150GRD
READINGE	1.000		
	(19)		
READINGF	-.081	1.000	
	(7)	(7)	
X150GRD	-.149	n/a	1.000
	(12)		(12)

Note:
None of the above correlation coefficients are statistically significant at p < .05.
READINGE: *Nelson-Denny Reading Test Form* E;
READINGF: *Nelson-Denny Reading Test Form* F;
X150GRD: EDUC X150 Course Grade for Fall 1995 semester
The numbers in parentheses indicate the respective sample sizes.
n/a: data not available

Table 11

Correlation Matrix for Education X152 Non-compliant Group

	READINGE	READINGF	X152GRD
READINGE	1.000		
	(23)		
READINGF	.507	1.000	
	(4)	(4)	
X152GRD	--	--	--
	(0)	(0)	(0)

Note:
None of the above correlation coefficients are statistically significant at$p$ < .05.
READINGE:*Nelson-Denny Reading Test Form E*;
READINGF:*Nelson-Denny Reading Test Form F*;
X152GRD: EDUC X152 Course Grade for Fall 1995 semester
The numbers in parentheses indicate the respective sample sizes.

English

The descriptive statistics for English are summarized in Table 1. The compliant students for W001

The descriptive statistics for English are summarized in Table 1. The compliant students for W001 writing course (n = 580) obtained a mean English placement test score of 8.72 and a standard deviation of .97. In comparison, there were 439 non-compliant students for W001 who had a mean English placement test score of 8.62 and a standard deviation of 1.15. The compliant group for W131 course comprised 786 students who had a mean English placement score of 14.03 with a standard deviation of .97. Similarly, there were 512 students with a mean English placement score of 14.02 and standard deviation of .97 for the W131 non-compliant group. For the W140 compliant group (n = 10), the mean placement test score was 20.80 with a standard deviation of 1.03. Likewise for the W140 non-compliant group (n = 11), the mean placement test score was 19.91 and a standard deviation of .54.

Tables 12 to 14 show the correlation matrices for the respective compliant groups for English. The correlation coefficients between the English placement exam scores and the writing course grades for the compliant groups ranged from r = -.01 for W140 to r = .09 for W131. Partly due to the small sample sizes and range restriction in the English placement test scores, most of the obtained correlation coefficients for English were not statistically significant at the .05 level. Surprisingly, however, a statistically significant correlation coefficient was obtained for the W001 non-compliant group (r = .58, p < .05).

(Note: The sample sizes shown in Tables 15 to 17 are too small to yield reliable results.)

Table 12

Correlation Matrix for English W001 Compliant Group

	EN100	ENGRD1
EN100	1.000	
	(580)	
ENGRD1	.079	1.000
	(536)	(536)

Note: *p< .05; ** p < .01; *** p < .00(nondirectional test)
EN100: English Placement Test Scores
ENGRD1: English Course Grade for Fall1995 semester
The numbers in parentheses indicate the respective sample sizes.

Table 13

Correlation Matrix for English W131 Compliant Group

	EN100	ENGRD1
EN100	1.000	
	(786)	
ENGRD1	.089*	1.000
	(732)	(732)

Note:
* p < .05; ** p < .01; *** p < .001 (nondirectional test)
EN100: English Placement Test Scores
ENGRD1: English Course Grade for Fall 1995 semester
The numbers in parentheses indicate the respective sample sizes.

Table 14

Correlation Matrix for English W140 Compliant Group

	EN100	ENGRD1

EN100	1.000	
	(10)	
ENGRD1	-.009	1.000
	(8)	(8)

Note:
* $p < .05$; ** $p < .01$; *** $p < .001$ (nondirectional test)
EN100: English Placement Test Scores
ENGRD1: English Course Grade for Fall 1995 semester
The numbers in parentheses indicate the respective sample sizes.

Table 15

Correlation Matrix for English W001 Non-compliant Group

	EN100	ENGRD1
EN100	1.000	
(14)		
ENGRD1	.584*	1.000
	(14)	(14)

Note:
* $p < .05$; ** $p < .01$; ***$p < .001$ (nondirectional test)
EN100: English Placement Test Scores
ENGRD1: English Course Grade for Fall 1995 semester
The numbers in parentheses indicate the respective sample sizes.

Table 16

Correlation Matrix for English W131 Non-compliant Group

	EN100	ENGRD1
EN100	1.000	
	(18)	
ENGGRD1	-.019	1.000
	(18)	(18)

Note:
* $p < .05$; ** $p < .01$; ***$p < .001$ (nondirectional test)
EN100: English Placement Test Scores
ENGRD1: English Course Grade for Fall 1995 semester
The numbers in parentheses indicate the respective sample sizes.

Table 17

Correlation Matrix for English W140 Non-compliant Group

	EN100	ENGRD1
EN100	1.000	
	(11)	
ENGRD1	.000	1.000
	(4)	(4)

Note:
* $p < .05$; ** $p < .01$; ***$p < .001$ (nondirectional test)

EN100: English Placement Test Scores
ENGRD1: English Course Grade for Fall 1995 semester
The numbers in parentheses indicate the respective sample sizes.

Results of Gender and Ethnic Analyses

The issue of gender differences in academic achievement was explored using the Student *t*-test for independent samples. Preliminary results based on gender analysis indicated that there were some statistically significant gender differences in achievement as reflected in performance on the respective placement tests. Specifically, statistically significant differences were obtained on mathematics placement test scores [*M*(females) = 11.05, *M*(males) = 12.25; *t*(2817) = -4.68, $p < .000$] with males scoring higher. However, females tended to obtain better course grades in mathematics [*M*(females) = 2.53), *M*(males) = 2.20; *t*(2335) = 5.66, $p < .000$]. There were no statistically significant gender differences on the mathematics final exams [*M*(females) = 59.33), *M*(males) = 60.65; *t*(837) = -1.24, $p = .214$]. Female students performed significantly better than males on both the English placement tests [*M*(females) = 11.89, *M*(males) =11.57; *t*(2366) = 2.40, $p < .017$] and English courses [*M*(females) = 2.79, *M*(males) = 2.44; *t*(1629) =5.86, $p < .000$]. In contrast to last year's observation, there were no statistically significant gender differences in both reading placement test scores [*M*(females) = 95.63, *M*(males) = 97.32; *t*(2455) = -1.68, $p = .093$] and reading post-test (*N-D Form F)* scores [*M*(females) = 85.33, *M*(males) = 89.01; *t*(228) = -1.83, $p = .069$].

The aforementioned results indicated that males performed significantly better than females on the mathematics placement test. In contrast, female students had significantly better performance on both the English placement tests and courses. No gender-related differences were obtained in reading.

The analyses based on the ethnic status of students did not yield statistically significant results. This finding is similar to those obtained in previous validity studies at IUPUI. However, given the exploratory nature of the present investigations on gender and ethnic bias, it will be important to investigate further the link between gender or ethnicity and academic achievement at IUPUI.

Discussion and Conclusion

The primary purpose of the present study was to investigate the validity of the IUPUI placement exam scores for course placement in mathematics, reading, and English. In comparison with the last two years' findings, the overall validity coefficients for mathematics were quite similar, and generally lower for reading. Some of the factors for explaining the present results are outlined below. As noted in last year's report, however, a modest increase in the validity coefficients is observed when the *Nelson-Denny Reading Test Form F* scores are employed as a primary criterion. Probably due to the influence of score range restriction and small sample size on validity coefficients, the correlations obtained for English were very low and mostly not statistically significant. Another important factor to consider for English is that a majority of high school graduates are poorly prepared in writing. Specifically, recent research indicates that most high school students neither write well nor write much in school or outside of school [see the Education Research Report titled *What's Wrong with Writing and What Can We Do Right Now?* by the Office of Educational Research and Improvement (OERI), April 1993].

The overall findings indicate the need to implement an alternative to the conventional approach to placement testing. One alternative that has recently been implemented for the mathematics placement test is the conversion to computerized adaptive testing (CAT). Details regarding the procedures and advantages of CAT are presented elsewhere in the literature (cf. Wainer et al., 1990).

Computerized Adaptive Testing is a set of techniques developed to increase the efficiency of testing. A full description of adaptive testing is well beyond the scope of this paper, but the interested reader is referred to *CAT Primer* (Wainer et al., 1990). The many advantages (and the relatively few disadvantages) of CAT are discussed elsewhere in the literature (see e.g., Wainer, 1993, Wainer et al., 1990). However, the choice to convert a computerized test into a CAT is mainly to achieve increased efficiency in the use of both students' time and test items. By "efficiency" we mean

increased efficiency in the use of both students' time and test items. By "efficiency" we mean obtaining the same accuracy as a linear test while using fewer items.

Given the overall weak relationships between the predictor and criterion measures, several points need to be considered when interpreting the present validity coefficients. For this reason, it is important to review the following points:

- It is unusual for a validity coefficient to rise above 0.60 (Cronbach, 1970). Although it is desirable to obtain higher validity coefficients, any positive correlation indicates that predictions from the test will be more accurate than guesses (Cronbach, 1970). Cronbach further states that "if a criterion can be predicted only with a validity of 0.20, the test may still make an appreciable practical contribution" (p. 135).
    - The typical range of validity coefficients (i.e., correlations between placement test scores and final course grades) reported in the literature is from about 0.20 to 0.50 (see CEEB, 1982; Ebel, 1972; Nitko, 1983). Even in excellent placement programs, correlations between placement test scores and course grades are weak. Nevertheless, coefficients as low as 0.30 are of definite practical value. Placement tests with validity coefficients in the range from about 0.30 to 0.50 make a considerable contribution to the efficiency of an institution, though they can forecast wrongly for many individuals (see Cronbach, 1970).
    - Correlation does not imply causality. If two measures correlate, the underlying attributes are not necessarily causally related. A theoretical understanding of the processes involved, or data from controlled experiments, permit us to draw conclusions about causes that underlie a correlation. Without such information, the only safe conclusion is that correlated measures are influenced by a common factor.

In addition to the aforementioned points, several constraints and/or extraneous factors might have affected the results of the present correlation study. The following extraneous factors have been noted to lower (or raise) the observed correlation coefficient, thereby masking the true nature of the relationship between the variables being measured. The factors mentioned below are, therefore, crucial in interpreting the results of the present validity study.

Range of the group. In interpreting a correlation coefficient, the range of the group studied must be considered. In general, the correlation is smaller in a select group than in a group comprising a wide range of ability. For instance, in the present study, students in each course were divided into two groups (compliant and non-compliant) for the purpose of conducting statistical analyses. Therefore, the correlation coefficients reported in this study were based on selected groups of students. Also, in several cases, the resulting sample sizes per group were too small to obtain stable results and/or decent statistical power. Note that with small sample sizes, correlations may vary widely from year to year. In other words, validity coefficients based on small samples are extremely sensitive to chance fluctuations. Thus, it is difficult to make strong recommendations based on results obtained from relatively small sample sizes.

Sample heterogeneity. In general, tests predict individual differences poorly within a homogeneous group. Validity coefficients are largest in a group with a wide range of ability, and tend to be small in a restricted, preselected group. Failure to recognize a range effect sometimes leads to discarding useful tests (Cronbach, 1970). In the present study, both mathematics and reading had score range restrictions on the criterion variables. Mathematics developmental courses eliminated a grade of "D"; and reading courses employed a dichotomous ("S/F") grading scheme. These restrictions in range of ability limited the amount of variability that could have been observed in the criterion variable. Consequently, the validity coefficients were generally reduced. To boost the correlation for reading, however, the scores from the *Nelson-Denny Reading Test Form F* were obtained and used as the main criterion variables. As expected, there was some modest increase in the validity coefficients between the reading placement test scores and the *Nelson-Denny Reading Test Form F* scores.

With respect to English placement, the statistical problem involved the lack of sufficient score variance in the predictor variable. Although the English Department currently employs a wide range of scores, the score variance in the current English placement exam is still restricted. Although the statistical analyses for English were based on the new grading system, most of the obtained validity coefficients were extremely low and not statistically significant. This

situation indicates that score range restriction and sample size are important practical issues (or problematical factors) for course placement, particularly in English.

Random error in the outcome or predictor measure. Intuition suggests that both outcome and predictor measures were subject to sampling error; this tended to lower the validity coefficient. (See Allen and Yen, 1979.) Unfortunately, the*correction for attenuation* (i.e., estimates of the correlations that would be found if predictor and outcome measures were perfectly reliable) was not attempted due to lack of reliability data for the outcome measures. An acceptable degree of reliability would be a necessary condition for a test score to be valid. However, we could not realistically expect that any examination would yield perfectly consistent or reliable scores. "Nevertheless, as the degree of reliability diminishes, so does their degree of validity" (Nitko, 1983, p. 388). The lack of correction for attenuation is a serious limitation of the present validity study.

Notwithstanding the aforementioned factors for consideration, several recommendations are outlined in the next section.

Recommendations

Based on this study and our experience, the following advice should be helpful for course placement and test construction. Recommendations are presented separately for each content domain under investigation.

Mathematics

- To improve the validity coefficients for mathematics, a new or improved Mathematics placement exam, utilizing Computerized Adaptive Testing (CAT) procedures, was implemented to replace the current conventional mathematics test. The conversion to adaptive testing in mathematics was actually accomplished in October 1995.

Since the switch-over to CAT, the Testing Center has been collecting information to evaluate the effectiveness of computerized adaptive testing in mathematics, and the preliminary findings have been very encouraging. For instance, we have obtained a positive validity coefficient as high as .55 ($p < .001$) between the computerized adaptive mathematics test scores and the final exam scores for the math 001 compliant group ($n = 36$). A similar high positive correlation ($r = .54$, $p < 001$) was observed between the mathematics CAT scores and the course grades for the math 001 compliant group ($n = 36$). The overall correlation coefficients of .37 ($p < .01$) with the exam scores as criteria, and .38 ($p < .01$) with course grades as the outcome were obtained for the total compliant group ($n = 45$) in mathematics. Note that the preliminary results should be interpreted cautiously because of the small sample size employed.

- Mathematics placement cutoff scores may need to be modified and should take into account the standard error of measurement and other practical issues. These cutoff points represent the standards upon which placement recommendations are made. Thus, cutoff scores should be chosen judiciously to ensure that a student enters only those courses for which he/she is prepared. Various methods for setting cutoff scores are discussed elsewhere in the literature (e.g., see Hills, Hirsch, & Subhiyah, 1990; Nitko, 1983).

- In addition to modifying the current cutoff scores, the Testing Center staff (in conjunction with the mathematics faculty) should consider adding appropriate items to the current item bank or periodically revise the local mathematics placement test to improve its content validity or curricular relevance.

Reading

Because of the low correlations between the *Nelson-Denny Reading Test Form E* and *FormF*, a new IUPUI Reading Placement Test has now replaced the *Nelson-Denny Reading Test FormE* (Shermis, Wolting, & Lombard, 1996). The non-adaptive version of the reading test was implemented in late June of this year. This linearly computerized reading test will soon be converted to an adaptive format. Previous pilot information on the reading test designed to work in a CAT environment has yielded promising results.

English

Because the current report provides the initial comprehensive analyses for English, relevant validity data should be collected over an extended period from large data samples in order to establish an English placement test with good predictive validity. However, we recommend that the English Department should review the current scoring procedure and also carefully construct appropriate placement test items. Questions on a placement exam should test for or measure the prerequisites a student needs in order to be successful in a course. Alternatively, the English faculty might want to adopt alternative (portfolio) assessment, in conjunction with traditional forms of assessment, to obtain a wider range of scores.

Overall recommendation. One of the continuing problems has to do with enforcing compliance in course placement. Although the current compliance rates in English, mathematics, and reading are generally higher than for the past two years, overall compliance rates could be increased if the University implements an enforcement mechanism for class enrollment other than relying on post-registration audits. Thus, we recommend that the Registrar create a "barring mechanism" akin to the financial bars that are already in place. The bar would prevent a student from enrolling in a class that is inappropriate for him/her based on placement test scores unless special consideration is granted by the respective academic departments. Such a system is currently being planned.

General Remarks

Future placement validity studies should include an analysis of the nature and content of the outcome measures. It is likely that any one year's examination is different in its content coverage of the curriculum, standards or difficulty. Such year-to-year differences in skills tested may account for a considerable amount of any change in performance on the examinations. A content validity analysis would provide some useful and/or supplementary information that would help explain some of the observed inconsistency in the relationships between predictor and outcome scores.

It is noteworthy that placement tests are designed to be used in conjunction with counselors' recommendations, and provide one source of information about the student's current ability level. Decisions regarding course placement ought to include a student's previous academic record and not be made on the basis of placement test scores alone. Because the IUPUI placement decisions are advisory, the placement tests are designed to be used in conjunction with counselors' recommendations. Thus, the use of multiple predictors for course placement is desirable (see Sawyer, 1996). Rather than use placement test scores alone, additional predictors may include one or more of the following variables: high school (HS) overall grade point average (GPA), HS subject GPA, HS (percentile) rank, and HS courses completed. It seems likely that the more information counselors have about a student, the better the chance of their making an appropriate placement decision.

From an exploratory investigation of the non-compliant groups, it was evident that quite a considerable number of students, perhaps in conjunction with their academic advisors, ultimately selected either more challenging or easier courses than those recommended by the placement test scores. Perhaps some returning adult students enrolled in an easier course as a means of developing confidence about their ability to succeed in college studies Probably due to academic major requirements, career choices, and/or the inadequacy of content coverage or cutoff points for the respective placement tests, some students opted for more challenging courses and apparently performed well on the criterion measures for some courses. This situation might warrant a review of the content coverage and/or cutoff points for the respective IUPUI placement tests.

After several years of placement testing in English, mathematics and reading, the respective stakeholders should be fairly convinced that the IUPUI placement testing program provides an efficient, practical, and workable method of placing students in appropriate courses which give them the best chance for academic success. The IUPUI placement tests, however, are a guide, based on the past performances of other similar students, of potential success in specific sets of courses in English, mathematics, and reading. Although the placement tests cannot measure motivation or other affective variables in academic learning, they are quite useful in providing

motivation or other affective variables in academic learning, they are quite useful in providing our students with a healthier and more valid learning experience in the respective courses. Placement testing, therefore, offers the University a practical way to accommodate the more diverse academic needs and talents of the entering students.

References

Akst, G., & Hirsch, L. (1991). Selected studies on mathematics placement. *Review of Research in Developmental Education, 8*(4), 1-3.

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.*Monterey, California: Brooks/Cole.

Brown, J. I., Bennett, J. M., & Hanna, G. (1981). *The Nelson-Denny Reading Test: Examiner's Manual Forms E and F.*Chicago, IL: Riverside.

Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items.*Journal of Educational Psychology, 78*, 141-146.

College Entrance Examination Board (1982). *Guide to the College Board Validity Study Service.* New York: Author.

Cronbach, L. J. (1970) *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.

Cronbach, L .J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

Ebel R. L. (1972). *Essentials of educational measurement* (2nd ed.). New Jersey: Prentice-Hall.

Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology* (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall.

Hills, J. R., Hirsch, T. M., & Subhiyah, R. G. (1990). *Issues in placement.* Washington, DC: ERIC Clearinghouse on Tests, Measurement and Evaluation/American Institutes for Research.

Kerlinger, F. N., & Pedhazur, E. J. (1973).*Multiple regression in behavioral research.* New York: Holt, Rinehart & Winston.

Kirk, R. E. (1990).*Statistics: An introduction* (3rd ed.). Fort Worth: Holt, Rinehart & Winston.

National Center for Education Statistics (October 1996). Developmental education at higher education institutions in fall 1995. *Statistical Analysis Report (NCES 97-584).* (WWW: http://www.ed.gov/NCES/pubs/97584.html)

Nitko, A. J. (1983).*Educational tests and measurement: An introduction.*New York: Harcourt Brace Jovanovich.

Office of Educational Research and Improvement (April 1993). What's wrong with writing and what can we do right now?*OERI Research Report*. Washington, DC: Author. (WWW: http://www.ed.gov/pubs/OR/ResearchRpts/writing.html)

Sawyer, R. (1996). Decision theory models for validating course placement tests.*Journal of Educational Measurement, 33*, 271-290.

Shermis, M. D., Wolting, M., & Lombard, D. (1996). Computerized adaptive testing for reading placement and diagnostic assessment.*Journal of Developmental Education, 20*(2), 18-24.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment.*Journal of Educational Measurement, 22*, 271-286.

Wainer, H. (1993). Some practical considerations when converting a linearly administered test to an adaptive format.*Educational Measurement: Issues and Practice, 12,* 15-20.

Wiseman, S. (1967). The effect of restriction of range upon correlation coefficient.*British*

Wiseman, S. (1967). The effect of restriction of range upon correlation coefficient. *British Journal of Educational Psychology, 37,* 248-252.