Placement of Students into First-Year Writing Courses

Norbert Elliot, Perry Deess, Alex Rudiny, Kamal Joshi

New Jersey Institute of Technology

Author Note

Abstract

The purpose of the present study is to examine concurrent and predictive evidence used in the validation of ACCUPLACER, a purchased test designed to place first-year students into writing courses at an urban, public research university devoted to science and technology education. Concurrent evidence was determined by correlations between ACCUPLACER scores and scores on two other tests designed to measure writing ability: the New Jersey Basic Skills Placement Test and the SAT Writing Section. Predictive evidence was determined by coefficients of determination between ACCUPLACER scores and end-of-semester performance measures. A longitudinal study was also conducted to investigate the grade history of students placed into first-year writing by established and new methods. When analyzed in terms of gender and ethnicity impact, ACCUPLACER failed to achieve statistically significant prediction rates for student performance. The study reveals some limits of placement testing and the problems related to it.

Placement of Students into First-Year Writing Courses

Disjuncture between secondary and postsecondary writing proficiency expectations results in remediation for many first-year students admitted to community colleges, colleges, and universities (Gere, Aull, Green, & Porter, 2010; Mattern & Packham, 2009; Sullivan & Tinberg, 2006). The cost associated with this remediation is staggering. Michigan researchers estimate a cost of $601 million for that state (Greene, 2000), and the College Board has recently estimated that $17 billion is spent annually on remediation each year (Lee & Rawls, 2010). Strategies to align high school and college writing expectations—attention to coordinating constructs of writing (Applebee & Langer, 2006) and to aligning the content of college placement tests with K-12 exit level standards (Venezia, Bracco, & Nodine, 2010)—are known yet largely unrealized. As a result, information that would yield extrapolation about interpretations of placement test performance to claims about course performance is largely fuzzy. Failure to communicate extrapolation inferences—to explain what truly occurs in student learning assessment—has led to recent federal demands for accountability resulting from the 2008 reauthorization of the Higher Education Act of 1965 (H. R. 4137), including the requirement that graduation rates be posted prominently on all web sites and promotional materials, call attention to remedial placement patterns that delay graduation (Code of Federal Regulations, 2005; Federal Register, 2010). It is increasingly clear that the 2006 report of former Secretary of Education Margaret Spellings, with its identification of "shortcomings of postsecondary institutions in graduation rates and time to degree to learning outcomes and even core literacy skills," was an early signal that accountability have now become an enduring part of American higher education (U.S. Department of Education, 2006, p.3). In an environment of transparency and accountability, validation of placement test use—the accumulation of evidence to provide a sound scientific

basis for proposed score interpretations—is increasingly important (AERA, APA, & NCME, 1999). When a test is purchased from a vendor, the challenges are especially unique if the developer has not fully disclosed information on test construction, evaluation, documentation, fairness, and applications recommended by the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999).

As a report on one institution's efforts to validate the use of purchased writing tests to place students into composition courses that would be challenging but not daunting for them, the present study is a case documenting procedures used, evidence accumulated, interpretations proposed, conclusions drawn, and questions raised that are indicative of complexities encountered in writing placement testing. Our study is set within a context of student placement from 1998 to 2010 at an urban, public research university devoted to science and technology education.

Background: The Contemporary Placement System for Writing Ability

The most common type of research conducted by institutions participating in *Trends in College Admission 2000: A Report of a National Survey of Undergraduate Admission Policies, Practices, and Procedures* was the development of characteristics of incoming first-time, first-year students. Because four-year institutions are committed to developing profiles of incoming first-year students, validity studies designed to correlate placement test scores and high school grades with college achievement are common in public four-year institutions (Breland, Maxey, Gernand, Cumming, & Trapani, 2002). Of the 1,664 institutions participating in this national study, 42% reported that they had developed placement studies for particular courses or academic majors.

Because of construct domain demands—investigation of the relationship between the target domain of writing and its representation in the test—the widespread use of standardized testing raises questions about placement in post-secondary education. As Kane (2006) observes, when the measurement procedure is standardized, the target domain—the range of observations associated with the target behavior—is often severely narrowed. The behavior sampled may be legitimate, but it constitutes a "very narrow slice" of the target domain (p. 31). O'Neill, Moore, and Huot (2009) have argued persuasively that the assessment of writing should be site based, locally controlled, context sensitive, rhetorically informed, accessible, and consistent with contemporary theories about language, learning, literacy, and educational measurement. Yet this precondition for assessment—the demand for a robust representation of the target domain—is often viewed as burdensome as applicants increase. Between 1999 and 2009, post-secondary enrollment increased 38 percent, from 14.8 million to 20.4 million. From 2010 to 2019, NCES projects a rise of 9 percent in enrollments of students under 25, and a rise of 23 percent in enrollments of students 25 and over" (Snyder & Dillow, 2011, p. 281). Concurrent with this expansion, President Barak Obama (2009) has called for at least one year or more of higher education or career training for every American. Within this environment of rising enrollments, shifting demographics, and expanding expectations, there are exemplary responses to the call for increased construct representation through the use of portfolios (Hester, O'Neill, Neal, Edgington, & Huot, 2007), on-line assessment (Peckham, 2009, 2010), and directed self placement (Royer and Gilles, 1998). In these assessment contexts, the target domain is locally-developed; thus, assessment use at the specific institutional site is justified a priori as the validation process ensues. Yet the majority of 2-year and 4-year undergraduate programs—during the 2009–2010 academic year, 4,495 accredited institutions offered degrees at the

associate's degree level or above (Snyder & Dillow, 2011, p. 283)—will not achieve what has been accomplished in a limited segment of the higher education community. Indeed, 29% (n = 1,300) of Title IV postsecondary institutions used ACCUPLACER, the test we examine in the present study, to place students into English, reading, and mathematics courses (Mattern & Packman, 2009). By contrast, directed self placement appears to be used by less than one percent (n=35) of postsecondary institutions (Gere, Aull, Green, & Porter, 2010).

The typical form of purchased writing placements tests combines multiple choice items and a holistic score on a timed writing sample. Historically, the work of Educational Testing Service (ETS) researchers Godshalk, Swineford, and Coffman (1966) is often cited to represent the origin of holistic scoring in writing assessment. Equally significant is that their study ushered in the contemporary paradigm. While the direct method, holistic scoring, achieved acceptable levels of reliability (.7 for four readings of a single 20-minute topic), that finding was only one of three in the classic study. The second finding, that multiple choice questions designed to measure writing skills correlated with the essay at ranges from .7 to .75, led to the enduring third result of the study: The most efficient predictor of the direct measure of writing is one that includes essay questions in combination with objective questions (p. 41).

Building upon assumptions of the strength of a combined use of multiple-choice questions and a direct writing sample, the New Jersey College Basic Skills Placement Test (NJCBSPT) was launched in 1980 in a collaboration between the ETS and the New Jersey higher education community to design a state-wide placement test (New Jersey State Department of Higher Education, 1980). A similar program of assessment, the model for the New Jersey effort, was undertaken with ETS partnership in California (White, 2001). The multiple choice sections of the test (Reading Comprehension and Sentence Structure) and the direct assessment section

(the essay) were used to supply New Jersey colleges with placement scores established by a consensus method in which a passing score is computed on the basis of judges who assess the ability of a just-qualified student to answer correctly a test item (Cizek & Bunch, 2007; Nedelsky, 1954; Zieky & Perie, 2006). Under the persona of post-secondary instructors, the panelists establishing the placement scores were to answer this key question: "How many of the distracters on this item should the entering college freshman be able to eliminate as clearly incorrect? By definition, a college freshman does not need remediation" (p. 4). The persona itself reveals the disjuncture between secondary and post-secondary writing instruction and suggests the complexities involved in extrapolating test results. All New Jersey students admitted to two-year and four-year state institutions were required, by state law, to take the NJCBSPT language test, along with a mathematics test, for placement. The NJCBSPT was used to place students into first year writing at our university from 1980 to 2007.

In 2007, it became clear that our use of evidence derived from the NJBCBSPT could no longer be warranted to place students. In 1994 New Jersey Governor Christine Whitman created the Higher Education Restructuring Act (N.J. S. 18A:3b-1 et seq., 1994). The Act eliminated the Department of Higher Education and, with it, the NJCBSPT. There was no way to refresh the test with new multiple-choice items and writing prompts created with ETS and field-tested across the institutions. While field-testing of prompts was undertaken within our institution, the methodology was more impressionistic than empirical, and the statistics validating aging multiple choice questions were no longer satisfactory. At our institution, the NJBSPT scores in Reading Comprehension and Sentence Structure were combined with the SAT Critical Reading Section as an assurance of reading ability. This undesirable double use of our admissions test, combined with an emerging view that we were teaching a writing course in which reading,

though an integrated skill with writing, had little curricular implementation, led us to abandon the use of the SAT Critical Reading Section for placement purposes. We then turned, in the spring of 2008, to an effort to validate ACCUPLACER. The origin and development of that test is also significant.

Employing the multiple-choice and writing sample format, the College Board had contracted with ETS to establish the ACCUPLACER testing program in 1985. While a paper version of the test existed at first, ACCUPLACER was designed as a showcase vehicle for item response theory (IRT) methods that were then becoming popular (Yen & Fitzpatrick, 2006). Within a computerized system, items were selected rapidly to match the student's skill level as answers were submitted. This computer-adaptive technology was the engine that fuelled ACCUPLACER. In 1993, the College Board won a five year contract to supply placement tests in all Florida community colleges, and in 1998 Texas accepted the use of ACCUPLACER as an alternative to its own state tests. In that year, sole ownership of the test was transferred from ETS to the College Board, who would, over time, seek new vendors for the test. In 1999 the test was delivered in a Web-based format. At the time of the study reported here, ACCUPLACER was managed for the College Board by Vantage Learning. The multiple choice format had been retained in its IRT configuration, but a new technology had been added: a student essay scored by automated technology termed WritePlacer *Plus* essay (Elliott, 2003). Using semantic, syntactic, and discourse level features, Vantage Learning used automated essay scoring (AES) technology to provide a single score on a 12-point range. At the time of the present study, the test included three sections: Reading Comprehension, Sentence Skills, and the WritePlacer *Plus* essay assessment.

The origin of the dominant writing placement system, its techniques dating from the mid-1960s and its wide-spread use dating from the early 1970s, has been established by White (2001); Otte & Mlynarczyk (2010) have demonstrated the impact of this system and its relationship to definitions, pedagogies, and research on basic writing. So dominant is this system that Behizadeh and Engelhard (2011) have concluded that writing theory has had a minimal influence on writing assessment. The dominant form of testing and its established history constitute four characteristics of the contemporary placement system. The system is characterized, first, by an assumption that large scale testing will yield educational measurement benefits that locally-developed tests do not yield. In the case of multiple choice tests, the item development process and the analysis of item differentiating properties are conducted by the vendor supporting the test. In the case of the scored essay, field testing of prompts and reliable scoring are assumed to be the responsibility of the vendor. Second, the system is characterized by an assumption that the vendor will provide research support that includes item analysis studies leading to the establishment of placement scores. These placement scores, although sanctioned by the institution, are informed by standard setting practices and comparative analysis of results that lead to optimal placement decisions. Third, those who purchase the test assume that the vendor has designed the test based on a traditional view of a defined and articulated construct domain—a set of interrelated attributes that are included under a definition of the behavior under examination. This characteristic of the contemporary system is related to the fourth—an assumption that an institution itself will have a defined and articulated construct that is aligned, both theoretically and empirically, with the construct of the purchased test. This assumption allows the institution to validate the relationship between the construct as represented in the test and the construct as embodied in the curriculum into which the student is placed.

This ubiquitous system is reinforced by the organizational structure of the educational

measurement community (Reynolds and Miller, 2003), its research practices (Brennan, 2006),

and its standards (AERA APA, NCME, 1999). The writing assessment community has

questioned the characteristics and assumptions of the placement system (Haswell, 1998, 2005),

its research practices (Reinheimer, 2007), and its standards (Harrington, Rhodes, Fischer, &

Malenczyk, 2006). This contemporary placement system serves as the background to our study.

Study Design

We designed our study to include concurrent and predictive evidence that would inform

decisions at our institution. A study of concurrent evidence—the extent to which results are

correlated with criterion measures known to be valid—was designed to yield information about

the relationship of ACCUPLACER to scores on the NJCBSPT and the SAT Writing Section.

Obtained at the same time, this study of concurrent evidence was used to determine if

ACCUPLACER covered the same construct of writing as the NJCBSPT; or if the SAT Writing

Section, taken earlier, was related to the construct covered in ACCUPLACER and the NJBSPT.

If relationships among the tests, and to other established performance measures, could be

established, then we hypothesized that the new test could be as useful in placing students as the

old had been. A study of predictive evidence—the extent to which results can be used to estimate

future performance—was designed to establish the relationship of the ACCUPLACER and the

SAT Writing Section to end-of-semester performance measures, the final course grade, and the

portfolio score.

Portfolio scores are an important criterion measure in the study. Used at our institution

for program assessment after final grades are given each semester, portfolios have become an

especially rich source of information when centered on a set of goal statements—defined as core

competencies (Coppola & Elliot, 2010)—that focus reader observation. While our portfolios

focus on the core competencies of critical thinking, revising and editing, content and

organization, and sentence construction and mechanics, they also yield a holistic score (Elliot,

Briller, & Joshi, 2007). As a course-embedded method of writing assessment, student portfolios

represent the fullest representation of the construct of writing that we have instituted; thus, the

relationship between the limited representation of the construct of writing in the purchased tests

and the manifestation of that construct in student portfolios was of great significance to us.

During the fall of 2008, the portfolios (n = 181) were read with reasonable reliability considering

the complexity of the scoring model in which readers provided rubric-based, criterion scores for

critical thinking, writing and editing, content and organization, sentence construction and

mechanics, and an overall score. The adjudicated overall, holistic score—an important variable

used along with course grade in providing extrapolation evidence—achieving acceptable

correlations under both the Pearson correlation ($r = .67$, $\rho < .05$) and the weighted kappa statistic

$\kappa = .458$, $\rho < .01$). Similar levels of inter-reader reliability ($r = .67$, $\rho < .05$; $\kappa = .531$, $\rho < .01$)

were maintained during the spring 2009 reading (n = 103). The portfolios were collected within

a sampling plan designed to yield a sample of student that were representative of our NJIT

student population (Johnson & Elliot, 2010). While the criterion variable of final grades is

known to be problematic (Armstrong, 2000), the small sample size of the portfolio reading

would be expanded by use of course grades.

Tensions inherent in the design are related to critique of the dominant writing placement

system by the writing assessment community. The construct of writing operationally measured

by all three purchased tests was a narrow representation of the construct. Within our institution,

administrators and instructors held with Camp (1992) in her definition that writing was "a rich,

multifaceted, meaning-making activity that occurs over time and in a social context, an activity that varies with purpose, situation, and audience and is improved by reflection on the written product and on the strategies used in creating it" (p. 135). Based on our experiences with writing in digital environments—an emphasis central to the Outcomes Statement for First-Year Composition adopted by the Council of Writing Program Administrators (Harrington, Rhodes, Fischer, & Malenczyk, 2006)—during the period of the study we were conceptually re-defining writing as an activity that emphasizes elements of visual design in the service of aesthetic aims, an activity that—within a media-rich environment—actively constructs a role for the reader to play and is improved by reflection on the impact of the digitally-mediated communicative act. The limited view of writing presented within each of the three tests, and the item types used to capture writing performance, did not fully reflect our conceptual definition of writing.

The NJCBSPT, a test that was first introduced in 1978 and had not benefitted from new item development since the state discontinued its use in 1966, used a 20-minute writing sample produced in response to a prompt, 35 items within a Sentence Structure section administered in 35 minutes, and 40 multiple-choice items in a Reading Comprehension Section administered in 30 minutes (Hech, 1980). The NJCBSPT was understood to be a test of English, with Sentence Structure, Reading Comprehension, and the essay as elements of that construct. The essay is defined as "a direct measure of writing ability" (State of New Jersey, 1983, p. 3). The essay is evaluated by features of coherence, effectiveness, support, and effective language use.

ACCUPLACER, using modern IRT assumptions (Yen and Fitzpatrick, 2009), is an untimed test monitored under secure administrations; that test employs a writing sample produced in response to a prompt, 20 multiple-choice items in a Sentence Skills Section, and 20 multiple-choice items in a Reading Comprehension Section (Mattern & Packman, 2009). The

essay is defined as the way that "a piece of writing communicates a whole message" (College Board, 2007, p. 6). The essay section is evaluated by features of focus, organization, development and support, sentence structure, and mechanical conventions. The elements of the test may be used together for placement, or they may be used individually. Because its sections reflected those of the NJCBSPT, our concurrent and predictive study sought to investigate connections among the three parallel sections.

The SAT Writing Section, first introduced in 2005, includes a 25-minute essay produced in response to a prompt and 49 multiple-choice questions asking students to identify sentence errors, improve sentences, and improve paragraphs (Norris, Oppler, Kuang, Day & Adams, 2006). In the SAT, writing is defined as assessment "that takes into account both the student's ability to develop ideas in a thoughtful, coherent, and cogent essay, and his or her understanding of using the conventions of language to express ideas" (Morgan, 2006, p. 32). The essay section is evaluated by features of point of view, focus, language use, sentence structure, and correctness. Both the multiple choice section and the essay are combined into a single score. Although the SAT Writing Section has been studied for admissions purposed by the College Board, it has not been formally validated by the Board as a placement test (Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008; Norris, Oppler, Kuang, Day, and Adams, 2006). The SAT Writing Section was not used at our institution for admissions purposes and would therefore have to be validated for local use.

Boorsboom (2003) has shown that the concept of validation has evolved as researchers have grappled with a wide-variety of test-related problems. As Boorsboom has demonstrated, the question of validity has shifted from attention to score and construct alignment (Kane, 1927), to alignment between scores as manifestations of theoretical relationships within the target domain

(Crohbach and Meehl, 1955), to emphasis on score use and social consequence (Messick, 1989).

At the present writing, the evidence-centered approach of Mislevy (2007) is widely favored in its

ability to allow researchers to make decisions on test use. Our study design necessitated what

Kane (2006) has termed an interpretative argument for validity underlying the special case of

placement testing (Cronbach & Snow, 1977). Recognizing that validation as a process employing

arguments (Cronbach, 1988) affords a rhetorical understanding that subordinates traditional

positivistic approaches to the contingencies of language involved in selecting one test over

another (Huot, 2002). In similar fashion, as Yancey (1999) and Elliot (2005) have shown, the

concept of writing assessment has also evolved. The definition of a valid writing assessment has

emerged from scales emphasizing error identification (Thorndike, 1911), to deployment of

multiple choice tests (Palmer, 1961), to use of a holistically-scored essay (White, 1973), to

emphasis to portfolio assessment (Belanoff and Dickinson, 1991). Our study design was

informed by the emphasis of Huot (1996) for assessment to be site-based, locally-controlled,

context-sensitive, rhetorically-based, and assessable, as well as by emphasis on meaningfulness

and ethical conduct stressed by Lynne (2004). While our long-standing tradition of portfolio

assessment for assessment of student learning adhered to these principles, validation of a

placement test also had to be attended to as it impacted admitted students at our specific

institutional site

  Our design was thus based on its potential to allow us to develop four interpretative

arguments. At their essences, these interpretative arguments follow the Toulmin (1958) model in

which claims are made that are supported with information. Warrants—generalizations made to

justify links between particular claims and the data that support them—are offered when explicit

connections of proof are necessary. Qualifications are used in the model to express contingency (Miselvy, 2006, 2007).

Operationally, this plan for an interpretative argument for our study allowed us to design the following four studies:

- **Study 1: Construct Evidence**. First, we believed (and therefore would seek to claim) that proficiency in focus, support, and effective language use—common elements among the tests—was both a precondition to and a part of the conceptual definition of writing offered by Camp (1996) and the WPA Outcomes Statement (Harrington, Rhodes, Fischer, & Malenczyk, 2006) in our traditional and honors first-year writing curriculum. We also believed that the ability to read critically—in terms of ACCUPLACER, to identify explicit statements in a reading passage related to ideas, application, and inference—was related to competency in writing ability and was, as well, a precondition to course success. Problematic here was the inclusion of the SAT Writing Section that was solely a test of writing ability. Nevertheless, we hypothesized that ACCUPLACER and the SAT Writing had the potential to yield a construct representation argument regarding student writing performance. In order to establish a construct representation argument, we hoped to discover correlations of a medium effect size among the three tests: NJCBSPT, ACCUPLACER, and the SAT Writing Section. Following statistical power analysis effect size indexes established by Cohen (1992), for all cases in which $\rho < .05$ correlations (r) would have to be at least .3 for a medium effect size and at least .5 for a large effect size. This concurrent evidence study would allow us to make generalization arguments that all three tests captured the construct under examination.

- **Study 2: Scoring Evidence**. Second, we believed we would be able to make a scoring argument that rules could be developed to yield a placement score that would be applied accurately and consistently. In order to establish a scoring argument, we determined to employ a consensus method for establishing placement scores (Cizek & Bunch, 2007). Establishing the placement score would be based on information that the test maker furnished about test design and validation studies, as well as by scores formally derived by a panel of our instructors.

- **Study 3: Extrapolation Evidence**. Third, we believed that placement test scores would allow us to make predictions about student competencies beyond test performance; this extrapolation argument would thus allow us to estimate how well the student would do in basic, traditional, and honors courses. In order to establish an extrapolation argument, we determined to establish the predictive power of the ACCUPLACER and the SAT Writing Section once the NJCBSPT had been retired. Following Cohen (1992), for all cases in which $\rho < .05$ the analysis of variance (F) would have to be at least .25 for a medium effect size and at least .4 for a large effect size in relation to the holistic portfolio score and course grades. We would also have to demonstrate that students achieved grades of C or above, our definition of student success, at rates that were historically congruent with past success rates.

- **Study 4: Decision Evidence**. Fourth, we believed that a longitudinal study would allow us to establish grade patterns that could be used in an interpretative argument to justify the use of the decision to recommend a new placement test. In order to establish a decision argument, we decided to augment our investigation to see if a consistent level of admitted students had been maintained based on SAT Critical Reading scores for a 10

year period. If that pattern was consistent, we would then be able to determine if we had

placed students according to consistent patterns, and we would also be able to investigate

final course grades consistency patterns. We also used in-place student surveys to

determine if students felt they were placed in the appropriate course level.

These four studies would allow us to better discuss our findings to administrators, instructors,

and students. The evidence-based recommendations that we would ultimately propose—that

ACCUPLACER was not the best writing placement test for our students and that the SAT

Writing Section was a preferable placement test—would have to link concepts, evidence, social

and personal consequences, and values (Cronbach, 1988). Based on construct representation, test

scoring, and result extrapolation, our study design would allow us to communicate gathered

evidence.

<div align="center">**Results**</div>

Our study began in the spring of 2008 and continues today. While there are many levels

of information that led to our decision to reject ACCUPLACER and use the SAT Writing

Section for placement purposes, we present here critical analyses that will allow replication of

this study.

Study 1: Construct Evidence

In anticipation of using ACCUPLACER in the fall of 2008, we begin our work in the

spring of 2008, a semester before we would have to make our decision to use ACCUPLACER.

Because the NJCBSPT was the test in place, we wanted to determine its relationship to

ACCUPLACER and the SAT Writing Section. We used the SAT Writing Section to establish a

base-line relationship. Four sections of students (n=97) took all three tests, just over the sample

size of 85 established by Cohen for a medium effect size (1992). The results are given in Table1.

Internally, the NJCBSPT showed large statistically significant correlations between its Reading and Sentence Structure sections. The correlations with the essay were lower. This same pattern appeared in the ACCUPLACER Reading and Sentence Skills scores, which showed large correlations with each other but a lower correlation with WritePlacer *Plus*. The correlations between NJBSPT Reading Comprehension Test and ACCUPLACER Reading Comprehension Test were large, as were the correlations between NJBSPT Sentence Structure Test and ACCUPLACER Sentence Skills Test. The relationship between the NJCBSPT human-scored essay and the AES of WritePlacer *Plus* was moderate. There was no statistically significant relationship between the SAT Writing Section and the NJCBSPT essay, and there was a moderate relationship between the SAT Writing Section and WritePlacer *Plus*.

We drew three conclusions from this analysis. First, the statistically significant correlations between the multiple choice sections of NJCBSPT and ACCUPLACER were evidence that the tests appeared to tap the same constructs of reading ability and sentence control. Second, the statistically significant correlations between the NJCBSPT essay and WritePlacer *Plus* suggested that they addressed the same construct of writing. Third, the statistically significant correlations between WritePlacer *Plus* and the SAT Writing Section suggested that they captured the same construct of writing. We concluded that there was enough preliminary evidence to establish placement scores for students who would be admitted in the fall of 2008 using ACCUPLACER. We also concluded that we would continue to study the placement potential of the SAT Writing Section.

Study 2: Scoring Evidence

When we began our study, we knew that we could not employ the Nedelsky (1954) method that had been used to set placement scores with the NJBSPT because of the IRT design

of the ACCUPLACER's Reading Comprehension and Sentence Skills tests. Popular methods developed by Angoff (1971) and those seeking to extend his work (Hambleton & Plake, 1995) would also be closed to us. Designed to ask judges to evaluate items based on the index of difficulty supplied by item characteristic curves of IRT, the bookmark method (Mitzel, Lewis, Parz, & Green, 2001) seemed most promising. However, it soon became clear that both the College Board and its vendor, Vantage Learning, considered the items and all difficulty measures associated with them as proprietary. Refusal to release information obtained in development ACCUPLACER—evidence of test validity required by Standards 1.1 - 1.24 and supporting documentation required by Standard 6.4 of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999)—has also been noted by Ericsson and Haswell (2006). We then realized that we knew nothing about how the test had performed with other students; indeed, we did not even have an item key. After a series of attempts to exhaust the test bank by snagging the items with screen shots and creating a key proved futile, it became clear that none of the standard setting methods described in Zieky and Perie (2006) or Cizek and Bunch (2007) would be available to us to establish placement scores.

Scaffolding a qualitative solution, we asked five experienced instructors to take ACCUPLACER three times—under the persona of an honors student, a traditional student, and a basic writing student. To establish the persona for the instructor to use, we used a two step process. In the first, we referred the instructor to the *Accuplacer Coordinator's Guide* (College Board, 2007) and the scores associated there with certain performance levels. A score of about 86, for example, on the Sentence Skills test described a skill set that appeared to reflect the skills of students who would be placed in basic writing, and we referred our instructors to the pages in the *Guide* describing those skills. We then asked the instructors to use their own unique

practitioner's experience to answer each item as would a student at each of the three performance levels. While there is no formal tradition for our *ad hoc* procedure, the persona method we developed is related to the Simulated Minimally Competent Candidate method described by O'Neill, Tannenbaum, and Tiffen (2005).

Table 2 presents the results of our expert review under persona. Because the WritePlacer *Plus* score descriptions were similar to those used in the NJBSPT, we were easily able to establish a score of 6 or below for placement in basic writing, a score of 7 to 8 for traditional writers, and a score of 9 or above for honors writers. The scores of each of the five panelists for the multiple choice sections of ACCUPLACER are, in general, within the range of the proficiency score statements of the College Board (2007).

Understanding our qualitative study as a benchmark, we then used basic equipercentile equating methods (Livingston, 2004) in which a score on a new form of a test (ACCUPLACER) and a score on the reference form of the test (NJBSPT) are equivalent if the test takers have the same percentile rank in a group. We thus established placement scores for our three cohorts of students.

To be placed in an honors section of writing, students would have to earn a score of 105 or above on the Reading Comprehension test, a score in line with those recommended by our expert panel (103) and the Board proficiency score (108). On the Sentence Skills test, students would have to earn a score of 112 or above, a score slightly below those recommendations by our expert panel (117) and in line with the Board proficiency score (110). On the WritePlacer *Plus* section, students would have to earn a score of 9 or above, a score aligned to the same score on the NJBSPT, the recommendations of the expert panel, and the score of 9 on the Board

proficiency score descriptors. Our placement scores would allow between 10 and 11% of students to be placed into honors classes.

The situation was less clear regarding the placement scores for basic writing students. On the ACCUPLACER Reading Comprehension test, the recommended Board score of 51 and expert faculty score of 52 would have placed only 16% of our students into basic writing. However, to place 37% of students in basic writing—the pattern from the fall of 2007—we would need a score of 72 or below. On the Sentence Skills test, the recommended Board score of 53 and expert score of 57 would have placed only 5% of our students into basic writing. However, to place students into basic writing in patterns similar to those from the fall of 2007, we would need a score of 80 or below. Without contradictory evidence, we decided to adopt the more rigorous provisional scores for the admitted class of 2008. If students achieved a score below 7 on the WritePlacer *Plus* scored essay, the student would be placed into basic writing. This scoring rule was consistent with our NCJBSPT placement score.

We therefore concluded that we had created a scoring method were justified in asking students beginning in the fall of 2008 to drive to our campus for two hours of ACCUPLACER testing on all three sections of the test.

Study 3: Extrapolation Evidence

As Matter and Packham (2009) have written in their meta-analysis of ACCUPLACER, the predictive power of a placement test is a significant part of the validation process. A thought experiment will allow us to see why this is so.

It might be assumed that a high correlation between a placement test and an outcome score means that the intervening teaching positively impacted student writing. A brief examination of an idealized data set demonstrates why this conclusion is so. Imagine that 60

students have been placed into basic writing because their timed essay scores were below 7. The placement test scores reveal that 20 of the students achieved a score of 4, 20 a score of 5, and 20 a score of 6. Let us then assume that, under expert instruction in a consistent learning environment, at mid-semester each of the scoring groups had increased performance by one point: The bottom students now have a score of 5, the mid-range students a score of 6, and the highest students a score of 7. At the end of the semester, say that the scores continue to increase according to the same pattern: There are 20 scores of 6, 20 or 7, and 20 of 8. Say that a *t*-test reveals that, between the beginning of the semester and mid-semester, the scores increased at a statistically significant level (t[df=59]= 47.04; ρ < .01), while a second t-test reveals that, between mid-semester and the end of the semester, the scores continued to increase at a statistically significant level (t[df=59] = 65.85; ρ < .01). Now that the statistically significant difference is established, we would look at the correlation. Imagine that the correlation between the placement test and the mid-semester test is perfect (*r* = 1.0, ρ < .01). So, too, is the correlation between the mid-semester test and the end-of-semester test. A high correlation, in this case a perfect one, would suggest that effective teaching was one of several variables in a complex learning environment in which student improvement occurred (Sternberg & Subotnik, 2006).

In practice, the relationships are much lower and, as such, often fuzzier. The meta-analysis by Matter and Packham (2009), published after our ACCUPLACER study was completed, demonstrates that the correlations between the test and final grades are low to modest in range. In individual studies, correlations between ACCUPLACER Reading Comprehension and final grades ranged from *r* = .18 and .38, and correlations between ACCUPLACER Sentence Skills and final grades ranged from *r* = .15 and .34. Related to course grades of C or higher, the

correlation between ACCUPLACER Reading Comprehension and final grades (n = 12,699) was

.1. The correlation between ACCUPLACER Sentence Sense and final grades (n = 12,485) was

.13. The correlation between ACCUPLACER *WritePlacer* and final grades (n = 3,408) was .13.

When the three sections of the test were combined (n = 1,693), the correlation increased to .24.

To gain a sense of the entire range of relationships, we performed a regression analysis,

shown in Table 3, of ACCUPLACER and the SAT Writing Section with the performance

variables of holistic portfolio scores and course grades during the fall and spring semesters. At

the end of the first semester, the ACCUPLACER scores captured slightly more of the variance

(27%) in predicting the holistic portfolio scores than did the SAT Writing Section (23%). These

coefficients of determination are nearly identical to that observed by Norris, Oppler, Kuang,

Day, and Adams (2006) in their study of English composition grade point average and its

relationship to the SAT Writing Section. The prediction falls dramatically when the tests are

compared to course grade: ACCUPLCER captures only 7% of the variance and the SAT Writing

Section captures 6%. However, in terms of the holistic score, at the end of the second semester,

the ACCUPLACER prediction falls demonstrably (to 15%) while the SAT Writing remains

approximately the same. In terms of course grades at the end of the spring semester, the

predictive power of ACCUPLACER increases slightly (9%) while the SAT Writing Sections

substantially increases its predictive power (to 16%). Based on Cohen's (1992) guidelines, the

tests demonstrate statistically significant small-to-medium effect sizes in their relationship to

holistic scores; the coefficients of determination constitute statistically significant small effect

sizes for both in their relationship to course grades. These findings of low-to-moderate predictive

power parallel recent research on admissions tests (Bowen, Chingos, and McPherson, 2009).

Of special interest is the ability of a test to predict scores for gender and ethnic groups assigned to basic writing courses in which issues of equity are especially relevant (Camili, 2006). In terms of gender and ethnicity, the predictive power of ACCUPLACER and the SAT Writing Section are shown in Table 4. Fall 2008 holistic portfolio scores, obtained by a sampling plan (Elliot, Briller, Joshi, 2008; Johnson & Elliot, 2010) based on a 95% confidence interval that does not include the total number of enrolled students, allowed a sufficient sample to analyze prediction for all except Black students. In all cases, both ACCUPLACER and the SAT Writing Section demonstrated statistically significant low-to-medium correlations with the holistic score. This finding was not the case, however, when the two tests were examined in their ability to predict course grades. Both ACCUPLACER and the SAT Writing Section did not achieve statistical significance for Hispanic and Black students. At the end of the spring semester, regression analysis demonstrated that ACCUPLACER's ability to predict the holistic portfolio scores for men decreased from 30% to 17% yet its ability to predict holistic scores for white students increased from 26% to 35%. For the spring semester, the SAT Writing Section increased slightly in its ability to predict holistic portfolio scores for males and increased dramatically in its ability to predict scores for white students, exceeding the .4 specification for a large effect size (Cohen, 1992). While the SAT Writing Section yielded statistically significant low correlations for all groups in predicting the spring writing course grade, ACCUPLACER failed to achieve statistical significance for females, Asian, Hispanic, and Black students.

Table 5 demonstrates the performance differences across gender and ethnic groups on all measures. Each of the purchased tests shows statistically significant differences in performance across groups. Of interest is that the holistic portfolio scores show no statistically significant difference in either the fall or spring semesters. While statistically significant differences in

writing course grades appear among white, Hispanic, and Black students in the fall of 2008, there are no differences in the spring of 2009.

Study 3 reveals the complexities involved in making an extrapolation argument. Desired were claims about student competencies beyond test performance. That is, justification of the use of a test—its ability to tell us something about student performance in subsequent courses—was paramount. Simply put, if such justification could not be warranted, the test could not be used. While the construct representation argument and the scoring argument were able to be justified, it was becoming clear that, for the diverse population at hand, ACCUPLACER did not promise the prediction we had anticipated.

Study 4: Decision Evidence

Table 6 presents a 13 year history of admitted students. In this period, between 19% and 39% of the admitted class was placed in basic writing before the placement study began in 2008. While the placement rate varied, the level of admitted students did not. At the 25th percentile, for example, the range of student scores on the SAT Critical Reading Section were equivalent in 1999 and 2006; based on level of admitted student, a 20 % increase in 2006 of students placed into basic writing does not appear warranted. The logic of placement becomes even more difficult to establish when Critical Reading scores are compared to national and state scores. For each year, the mean scores at our institution exceed national and state averages.

While the placement rates varied greatly during this period, the grades did not. Except for a period in which the honors students, small in number, may have had low grades than typical in 2000 and 2001, these students earned grades of C or better between 96% and 80% of the time. Traditional composition students earned grades of C or above between 77% and 90% of the time; on average, only 13% of the students received grades of D or F. No more than 17% of the

students had ever failed basic writing in a single semester; from 1998 to 2007, of the 2, 225

students who had been placed into basic writing, only 9 % had received grades of D or F. And, as

Table 6 demonstrates, the failure rates of our ESL basic writers were even lower.

Students reported increasing satisfaction with their placement. The 2007 enrolling student

survey recorded that 82% of the students believed they had been placed correctly. Under the new

placement system begun in the fall of 2008, student beliefs changed. Only 10% the responding

students (n = 364) believed they had placed incorrectly. In the fall of 2009, the second year of

our placement study, 98% of the students (n = 484) believed they had been placed correctly. In

the fall of 2010, 88% (n = 411) of the surveyed students reported that they were placed in the

appropriate level course.

Discussion

In our efforts to place students into first-year courses, we were able to make a construct

representation argument that the scores on the NJCBSPT and ACCUPLACER addressed the

same construct of student performance, a construct that was, in turn, related to our conceptual

definition of writing. The scoring argument, however, was difficult to support. Background on

test validity and supporting documentation were not provided by the vendor, and we knew little

about the test we intended to use. While we did adopt ACCUPLACER to place students into

classes in the fall of 2008, subsequent extrapolation studies—designed to investigate the impact

of the test on a diverse student population—revealed that ACCUPLACER had limited predictive

power, especially at the end of the second semester, in relation to gender and ethnicity of our

university's diverse student body. While we hold with Reynolds (1982) that sole focus on

differences in predictive power or mean scores constitutes invalid definitions of test bias, we also

with Cleary (1968) that a connotation of "unfair" may arise when the score predicted from a

regression line is too low. Rather an impulse to be empirically neat, attention to low prediction rates suggests that populations of diverse studies may be disenfranchised by tests that fail to capture complexities attendant to discourse features in essays submitted by diverse student writers. While researchers such as Smitherman (2000) have demonstrated that Black English Vernacular did not affect rater score on the National Assessment of Educational Progress in papers written by 17-year-old Black students from 1960 to 1989, no such study had been offered for ACCUPLACER. As Table 4 demonstrates, both ACCUPLACER and the SAT Writing Section failed to achieve score prediction of course grades for Hispanic and Black students in the fall of 2008; nevertheless, studies by the College Board of the SAT Writing Section attended to definition of groups and their performance (Breland, Kubota, Nickerson, Traipani, Walker, 2004; Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008). Because no such analysis existed for ACCUPLACER, we concluded that we could not have legally defended our continued use of the test under disparate impact challenges stemming from Title VI and Title VII of the federal Civil Rights Act. Taken at the time of admissions testing and supported by detailed impact studies, the SAT Writing Section was the preferable placement test. We ethically feared tests that would exploit the communicative divisiveness explored by Gilyard (1991), potentially awarding low scores for diverse ways of approaching a writing task, when we had worked hard for decades to develop curricula that encouraged and celebrated diversity as integral to the university mission.

The decision argument, based on an assumption that we had placed students of similar skill levels according to consistent patterns, was equally problematic. A post-hoc investigation of our placement system revealed that, without the use of the SAT Critical Reading Section as part of our former placement process, an additional 10% of the students admitted in 2007 (n = 62) would have been placed into traditional writing, and an additional 19% of the students (n = 115)

would have been placed into honors writing. The archival study revealed that our use of the SAT Critical Reading Section had a substantial impact on student placement. In combination with the success rates of basic writers—an indication that they could handle more challenging work—there was evidence that we were over-remediating our students. As well, as we investigated our definitions of the construct of writing, we had to acknowledge that we were teachers of writing and, as such, had no instructional plan in place that reflected current theories of reading instruction (O'Reily & Sheehan, 2009). Both statistically and conceptually, our placement was inconsistent. Yet our grading patterns were not. A return to Table 6 reveals the placement system to date in the two years following our validation study. Discarding the NJCBSPT placement system and re-conceptualizing the placement process allowed us to remediate far fewer students, as the decrease from a 30% remediation rate in 2007 to an 11% remediation rate in 2007 reveals. Overall, correct placement and course failure patterns from 2008 to 2010 are consistent with patterns before our study, evidence that we were over-remediating our students from 1997 to 2007.

However, variations in grade patterns remain. For example, from 1998 to 2007, an average of 89% of our students earned grades of C or above in basic writing; these average grade levels declined slightly in 2008, declined more sharply in 2009, and then rose in 2010. These variations remind us that instructor and student interactions are always complex and may never achieve complete stability, even if the admitted student population remains stable. Equally complex under conditions of variation is the decision for adjusting a placement score. Had we raised the placement score following the rise of failing grades among our basic writers in 2009, we would have again been over-remediating these students, especially when we realize that in 2010 course failure for basic writers declined to 11%, only one percent higher than the average

rate from 1998 to 2007. In writing placement research, monitoring student performance in the classroom remains one of the most complex aspects of establishing consequential validity.

While the system will remain in need of careful monitoring, three reflections are in order as we began the fourth year of our on-going writing placement research. As illustrative of the complexities involved in writing placement, our research suggests that a program of writing placement will benefit by attention to the societal context of placement, the pursuit of an idiographic orientation, and adoption of a methodological approach.

First, out study illustrates the presence of a national culture of remediation and the role that tests such as ACCUPLACER play in that culture. A fundamental assumption underpinning writing assessment undertaken for placement purposes is that the nation's high schools are not doing their jobs in the teaching of writing. The most recent national study, completed by the National Center for Educational Statistics, supports this point: In fall 2000, 68% of the surveyed institutions—Title IV degree-granting 2- and 4-year institutions that enrolled first-year students—offered at least one remedial writing course, and 14 percent of entering freshmen enrolled in at least one of those courses. The data further show that, on average, surveyed institutions offered two remedial writing courses (Parsad and Lewis, 2003). However, student performance does not necessarily support these rates of remediation. When 2007 NAEP national results are compared with results from the 2002 and 1998 assessments of writing ability, we see that student scores are rising. At grades 8 and 12, average writing scores and the percentages of students performing at or above basis levels were higher than in both previous assessments (Salahu-din, Persky, and Miller, 2008). Tensions that arise between in the transition from high school to college give rise to debate about the ability of purchased tests to cover the construct of writing. Capitalizing on these debates, both for-profit and non-profit organizations have

developed tests for writing placement designed to sort students into highly prepared, prepared, and unprepared students, with a special emphasis on the last group fuelled by controversy. Expansion of these tests depends on a perceived increase in the need for the services they provide. Against the growth of these tests there is no defense save the development of rigorous assessment research designed to allow campus leaders to make hard decisions about when remediation benefits students and when it unnecessarily consumes resources. The worship of efficacy, as Williamson (1994) has warned, may result in a failure to acknowledge institutional complexity and variety and its preservation within the classroom.

Second, our study illustrates the need for attention to ideographic analysis rather than pursuit of nomothetic span (Embrestson, 1983). Rather than engage hoary arguments over effectiveness in secondary education and construct representation in purchased tests—debates that provide only limited answers for those undertaking writing assessment research for placement purposes—institutions will benefit by embracing the ideographic orientation of validation that is common to both the writing assessment community (Huot, 1996; Lynne, 2004) and the educational measurement community (Cizek, 2008). As Cronbach (1975) long-ago observed, the general nature of Aptitude X Treatment interactions (ATIs) are most fruitfully researched in local use. Though enduring systematic theories about individuals in society are not likely to be achieved, Cronbach proposed, systematic inquiry can realistically hope to make two contributions. One is to assess local events accurately, to improve short-run control. The other reasonable aspiration is to develop explanatory concepts relevant in local contexts and adding valuable heuristic value in nomothetic span.

Third, even though practice may be local, our study suggests that a systematic approach is needed for placement testing. It is no coincidence that Kane (2006) takes placement testing as

his example of validation in the fourth edition of *Educational Measurement*. His example of an interpretative argument for placement testing articulates procedures that can be most meaningful on the local level. Within a local context, decisions can be made to communicate to all shareholders that the framework offered by Willingham (1974) for nature of placement (and exemption), the system of decisions for placement, the reason or assignment of students to traditional and alternate pedagogies, and the findings from longitudinal study of student performance. Because validation of a test is the shared responsibility of both the test developer and the test user, new technologies such as automated assessment must be increasingly transparent in their development, providing a broad base of validity evidence for score meaning (Bennett, 2011). Indeed, our study is not a renunciation of automated essay scoring; rather it is an example of how innovative assessment systems must be fully explained by the test developer—whether local or commercial—and validated by reference to institutional mission before they can be used. When writing assessment is used for any high stakes decisions that may interrupt or facilitate student progress, the design and preliminary results from the system are prerequisite to score interpretation; without information on the impact of a test on diverse student populations in field testing, scores cannot be interpreted. Whether the test is developed locally or commercially, disaggregated information on student performance, understood in terms of gender and race, must be provided before the test can be meaningfully used. An example of such recent research is automated scoring has been provided by Bridgeman, Trapani, and Attali (in press) on the performance of gender, ethnic, language, or country groups on essays scores by humans and machines. A research agenda can be imagined for automated essay scoring by using the *Framework for Success in Postsecondary Writing* (CWPA, NCTE, WPA, 2011) to study how this new technology fosters the desired habits of mind, as well as writing, reading, and critical

analysis experiences for diverse groups of students; indeed the *Framework* can be a meaningful heuristic for a program of research to engage all forms of writing assessment as they are used in placement contexts. While there can be no one handbook for how students should be placed into writing courses, there can be systematic, innovative approaches—yet to be developed for new assessment methods—that will sufficiently attend to the complexities of how writing assessment is used to support distinct institutional missions.

References

American Educational Research Association, American Psychological Association, & National

    Council on Measurement in Education. (1999). *Standards for educational and*

    psychological testing. Washington, DC: American Educational Research Association.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.),

    Educational measurement (2nd ed., pp. 508-600). Washington, DC: American Council on

    Education.

Applebee, A. N., & Langer, J. A. (2006). *The state of writing instruction in America's Schools:*

    What existing data tell us? Albany, NY: Center on English Learning & Achievement.

    Retrieved from http://www.albany.edu/aire/

Armstrong, W. B. (2000). The association among student success in courses, placement test

    scores, student background data, and instructor grading practices. *Community College*

    Journal of Research and Practice, *21*, 681-695.

Belanoff, P., & Dickson, M. (Eds.). *Portfolios: Process and product*. Portsmouth, NH: Boynton,

    1991.

Behizadeh, N., & Englahard, G. (2011). Historical view of the influences of measurement and

    writing theories on the practice of writing assessment in the United States. *Assessing*

    Writing, *16*, 189-211.

Bennett, R.E. (2011). *Automated scoring of constructed-response literacy and mathematics*

    items. Washington, DC: Arabella Philanthropic Investment Advisors. Retrieved from

    http://www.ets.org/research/

Black, R. W. (2009). Online fan fiction, global identifies, and imagination. *Research in the*

    Teaching of English, *43*, 397- 425.

Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the finish line: Completing college at America's public universities*. Princeton, NJ: Princeton University Press.

Breland, H., Kubota, M., Nickerson, K., Trapani, C., & Walker, M. (2004). *New SAT® writing prompt study* (College Board Report 2004-1). New York, NY: College Board. Retrieved from http://professionals.collegeboard.com/gateway

Breland, H., Maxey, J., Gernand, R., Cumming, T., & Tripani, C. (2002). *Trends in college admission 2000: A report of a survey of undergraduate admissions policies, practices, and procedures*. (Sponsored by ACT, Inc., Association for Institutional Research, The College Board, Educational Testing Service, and National Association for College Admission Counseling). Retrieved from http://www.airweb.org/images/trendsreport.pdf

Brennan, R. L. (2006). Perspectives on the evolution and future of educational measurement. In R. L. Brennan (Ed.). *Educational measurement* (4th ed., pp. 1-16). Westport, CT: American Council on Education and Praeger.

Bridgeman, B., Trapani, C., & Attali, Y. (in press). Comparison of human and machine scoring of essays: Differences by gender, ethnicity, and country. *Applied Measurement in Education*.

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.). *Educational measurement* (4th ed., pp. 221-256). Westport, CT: American Council on Education and Praeger.

Camp, R. (1996). New views of measurement and new models for writing assessment. In E. M. White, W. D. Lutz, & S. Kamusikiri (Eds.), *Assessment of writing: Politics, policies, practices* (pp. 135-157). New York, NY: Modern Language Association.

Cizek, G. J. (2008). Assessing *Educational measurement*: Ovations, omissions, opportunities. [Review of *Educational Measurement*, 4th ed., by R. L. Brennan (Ed.).] *Educational Researcher*, *37*, 96-100.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Press.

Cleary, T. A. (1965). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement*, *5*, 115-124.

Code of Federal Regulations. (2005). Title 34, 668.14(b)(22)(ii)(A) through (L). 2005. Retrieved from http://ecfr.gpoaccess.gov/

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

College Board. (2007). *ACCUPLACER Coordinator's guide*. New York, NY: College Board.

College Board (2010). *State profile report: New Jersey*. New York: NY: College Board

Coppola, N. &, Elliot, N. (2010). Assessment of graduate programs in technical communication. In M. Hundleby & J. Allen (Eds.), *Assessment in technical and professional communication* (pp. 127-160). Amityville, NY: Baywood.

Council of Writing Program Administrators, National Council of Teachers of English, & National Writing Project. (2011). *Framework for success in postsecondary writing*. Retrieved from http://wpacouncil.org/files/framework-for-success-postsecondary-writing.pdf

Cronbach, L. J. (1975, February). Beyond the two disciplines of scientific psychology. *American Psychologist*, 116-127.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test Validity* (pp. 3-17). Hillsdale, NJ: Lawrence Earlbaum.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

Cronbach, L. J, & Snow, R. E. (1997). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington Publishers.

Dewsbury, D. A. (Ed.). (1996-2000). *Unification through division: Histories of the divisions of the American Psychological Association* (Vols. 1-5). Washington, DC. American Psychological Association.

Elliot, N. (2005). *On a scale: A social history of writing assessment*. New York, NY: Peter Lang.

Elliot, N., Briller, V., & Joshi, K. (2007). Portfolio assessment: Quantification and community. *Journal of Writing Assessment, 3*, 5–30.

Elliott, S. (2003). IntelliMetric™: From here to validity. J. C. Burstein (Eds.). *Automated essay scoring: A cross-disciplinary perspective* (pp. 71-86). Mahwah, NJ: Lawrence Erlbaum.

Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, *93*, 179-97.

Ericcson, P. F., & Haswell, R. (2006). Introduction. In P. F. Ericcson & R. H. Haswell (Eds.), *Machine scoring of student essays: Truth and consequences*. Logan, UT: Utah State University Press.

Federal Register (2010, October). *Program Integrity Issues: Final Rule*. Retrieved from http://www2.ed.gov/legislation/FedRegister/

Gere, A. R., Aull, L., Green, T., & Porter, A. (2010). Assessing the validity of directed self-placement at a large university. *Assessing Writing*, *15*, 154-176.

Gilyard, K. (1991). *Voices of the self: A study of language competence*. Detroit, MI: Wayne State University Press.

Greene, J. P. (2000). *The cost of remedial education*. Midland, MI: Mackinac Center for Public

    Policy.

Godshalk, F., Swineford, F., & Coffman, W.E. (1966). *The measurement of writing ability*.

    Princeton, NJ: Educational Testing Service.

Hambleton, R. K., & Plake, B. S. (1995). Using an extended Angoff procedure set standards on

    complex performance assessments. *Applied Measurement in Education*, *8*, 41-56.

Harrington, S., Rhodes, K., Fischer, R.O., & Malenczyk, R. (Eds). (2006). *The outcomes book:*

    Debate and consensus after the WPA outcomes statement. Logan, UT: Utah State

    University Press.

Haswell, R. H. (1998). Multiple inquiry in the validation of writing tests. *Assessing Writing*, *5*,

    89-101.

Haswell, R. H. (2005). Post-secondary entrance writing placement. Retrieved from

    http://comppile.tamucc.edu/placement.htm

Hech, L. W. (1980). *Validation of the New Jersey college basic skills placement test* (ED 204

    945). Retrieved from http://www.eric.ed.gov/PDFS/ED214945.pdf

Hester, V., O'Neill, V., Neal, M., Eddington, A, & Huot, B. (2007). Adding portfolios to the

    placement process: A longitudinal perspective. In P. O'Neill (Ed.), *Blurring boundaries:*

    Developing writers, researchers, and teachers (pp. 61-90). Creskill, NJ: Hampton Press.

H. R. 4137. *Extension of Higher Education Act of 1965*. 110th Cong. (2008). Retrieved from

    http://www.govtrack.us/congress/

Huot, B. (1996). Towards a new theory of writing assessment. *College Composition and*

    Communication, *47*(4), 549-566.

Huot, B. (2002). *(Re)Articulating writing assessment in teaching and learning*. Logan, UT: Utah
State University Press.

Issacs, E., & Molloy, S. A. (2010). SATs for writing placement: A critique and counterproposal.
College English, *72*, 518-538.

Johnson, C., & Elliot. N. (2010). Undergraduate technical writing assessment. *Programmatic
Perspectives*, *2*, 110-151. Retrieved from http://www.cptsc.org/

Kane, M. T. (2006) Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp.
17-64). Westport, CT: American Council on Education/Praeger.

Kobrin, J. L., Patterson, B. F., Shaw, E. J, Mattern, K. D., & Barbuti, S. M. (2008). *Validity of
the SAT® for predicting first-year college grade point average* (College Board Report
2008-5). New York, NY: College Board. New York: College Board. Retrieved from
http://professionals.collegeboard.com/gateway

Lee, J. M., & Rawles, A. (2010). *The college completion agenda: 2010 progress report*. New
York, NY: College Board. Retrieved from http://professionals.collegeboard.com/gateway

Livingston, S. (2004). *Equating test scores (without IRT)*. Princeton, NJ: Educational Testing
Service. Retrieved from http://www.ets.org/research/

Lynne, P. (2004). *Coming to terms: A theory of writing assessment*. Logan, Utah: Utah State
University Press.

Mattern, K., & Packham, S. (2009). *Predictive validity of ACCUPLACER® scores for course
placement: A meta-analysis* (Research Report 2009-2). New York, NY: College Board.
Retrieved from http://professionals.collegeboard.com/gateway

Messick. S (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 134-
103). New York, NY: American Council on Education and Macmillan.

Messick, S. (1998). Test validity: A matter of consequences. *Social Indicators Research*, *45*, 35-
    44.

Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In R. L. Brennan (Ed.),
    Educational measurement (4th ed., pp. 257-305). Westport, CT: American Council on
    Education/Praeger.

Mislevy, R. J. (2007). Validity by design. *Educational Researcher 36,* 463–469.

Mitzel, H. C., Lewis, D. M., Parz, R. J., & Green, D. R. (2001). The bookmark procedure:
    Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards:*
    Concepts, methods, and perspectives (pp. 249-281). Mahwah, NJ: Lawrence Erlbaum.

Morgan, D. L. (2006). *Setting local cut scores on the SAT Reasoning Test™ Writing Section for*
    use in college placement and admission decisions. New York, NY: College Board.
    Retrieved from http://professionals.collegeboard.com/gateway

Mutnick, D. (2011). Review: Basic writing and the future of higher education. *College English*,
    73, 322-336.

Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and*
    Psychological Measurement, *14*, 3-19.

New Jersey State Department of Higher Education (1980). *Establishing cut scores on the New*
    Jersey college basic skills placement test. Trenton, NJ: New Jersey State Department of
    Higher Education. Retrieved from http://www.eric.ed.gov/

N.J. S. 18A:3b-1 et seq., 1994. *Higher Education Restructuring Act of 1994*. Retrieved from
    http://www.state.nj.us/highereducation/

Norris, D., Oppler, S., Kuang, D., Day, R., & Adams, K. (2006). *The College Board SAT®*
    Writing validation study: An assessment of predictive and incremental validity (Research

Report 2006-2). New York, NY: College Board. Retrieved from

http://professionals.collegeboard.com/gateway

Obama, Barak. (2009). Remarks prepared for delivery address to joint session of Congress,

Tuesday, February 24, 2009. Retrieved from http://www.whitehouse.gov/

Otte, G., & Mlynarczyk, W. (2010). *Basic writing*. West Lafayette, IN: Parlor Press and WAS

Clearninghouse.

O'Neill, P., Moore, C., & Huot, B. (2009). *A guide to college writing assessment*. Logan, UT:

Utah State University Press.

O'Neill, T. R., Tannenbaum, R. J., & Tiffen, J. (2005). Recommending a minimum English

proficiency standard for entry-level nursing. *Journal of Nursing Measurement*, *13*, 129-

146.

O'Reilly, T., & Sheehan, K. (2009). *Cognitively based assessment of, for, and as Learning: A

framework for assessing reading competency* (ETS RR-09-26). Princeton, NJ:

Educational Testing Service.

Palmer, O. (1961). Sense or nonsense? The objective testing of English composition. *English

Journal*, *50*, 314-320.

Parsad, B., & Lewis, L. (2003). *Remedial education in degree-granting postsecondary

institutions in fall 2000*. (NCES 2004-010). Washington, DC: U.S. Department of

Education, National Center for Educational Statistics.

Peckham, I. (2009). Online placement in first year writing. *College Composition and

Communication*, *60*, 517-540.

Peckham, I. (2010). Online challenge versus offline ACT. *College Composition and

Communication*, *61*, 718-745.

Perelman, L. (2005, May 29). New SAT: Write long, badly, and prosper. *Los Angeles Times*, p. 5. Retrieved from

Reinheimer, D. A. (2007). Validating placement: Local means, multiple measures. *Assessing Writing*, *12*, 170-179.

Reynolds, C. R. (1982). Methods of detecting construct and predictive bias. In R. A. Berk (Ed.), *Handbook of methods for detecting bias* (pp. 199-227). Baltimore, MD: Johns Hopkins University Press.

Reynolds, W. R., & Miller. G. E. (Eds.). (2003). *Handbook of psychology: Vo. 7. Educational psychology*. Hoboken, NJ: John Wiley.

Royer, D.J., & Gilles, R. (1998) Directed self-placement: An attitude of orientation. *College Composition and Communication*, *50*, 54-70.

Salahu-Din, D., Persky, H., & Miller, J. (2008). *The nation's report card: Writing 2007*. (NCES 2008–468). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C. Retrieved from http://nces.ed.gov/

Smitherman, G (2000). African American student writers in the NAEP, 1969-1988/89. In G. Smitherman (Ed.), *Talk the Talk: Language, culture and education in African America* (pp. 163-176). New York: Routlegde.

Snyder, T. D., & Dillow, S.A. (2010). *Digest of educational statistics, 2010* (NCES 2011-015). Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/

State of New Jersey. *New Jersey college basic skills placement test: Scoring the essays*. Trenton, NJ: Department of Higher Education. Carl Campbell Brigham Library, ETS, Princeton, NJ.

Sternberg, R., & Subotnik, R. F. (2006). *Optimizing student success in school with the other three Rs: Reasoning, resilience, and responsibility*. Charlotte, NC: Information Age Publishing.

Sullivan. P., & Tinberg, H. (Eds.). (2006). *What is college level writing?* Urbana, IL: National Council of Teachers of English.

Thorndike, E. L. (1911). A scale for merit in English writing by young people. *Journal of Educational Psychology*, *2*, 361-368.

Title VI of the Civil Rights Act, 42 U. S. C. § 2000d (1964). Retrieved from http://www.justice.gov/

Title VII of the Civil Rights Act, 42 U. S. C. § 2000e *et*. *seq*. (1964). Retrieved from http://www.justice.gov/

Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

Trainor, J. (2005). 'My ancestors didn't own slaves': Understanding white talk about race. *Research in the Teaching of English*, *40*, 140-167.

Trainor, J. (2008). *Rethinking racism: Emotion, persuasion, and literacy education in an all-white high school*. Carbondale, IL: Southern Illinois University Press.

U.S. Department of Education (2006) *A Test of Leadership: Charting the Future of U.S. Higher Education.* Washington, D.C: U.S. Department of Education. Retrieved from http://www2.ed.gov/

Venezia, A., Bracco, K. R., & Nodine, T. (2010). *One-shot deal? Students' perceptions of assessment and course placement in California's community colleges*. San Francisco: WestEd.

White, E. M. (1973). (Ed.). *Comparison and contrast: The 1973 California State University and College Freshman English Equivalency Examination*. (ERIC ED 114825). Retrieved from http://www.eric.ed.gov/

White, E. M. (2001). The opening of the modern era of writing assessment: A narrative. *College English*, *63*, 306-320.

Williamson, M. (1994). The worship of efficiency: Untangling theoretical and practical considerations in writing assessment. *Assessing Writing*, *1*, 147-174.

Willingham. W. W. (1974). *College placement and exemption*. New York: College Board, 1974. Retrieved from http://www.eric.ed.gov/

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, *39*, 1-97.

Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment as a rhetorical act. *College Composition and Communication*, *50*, 483-503.

Yen, W. M., & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 111-153). Westport, CT: American Council on Education/Praeger.

Zieky, M,. & Perie, M. (2006). *A primer on setting cut scores on tests of educational achievement*. Princeton, NJ Educational Testing Service. Retrieved from http://www.ets.org/research/

Table 1

Summary of Intercorrelations, NJCBSPT, ACCUPLACER, and SAT Writing

	1	2	3	4	5	6	7
1. NJCBSPT Reading	—	.659**	.382**	.669**	.595**	.395**	.541**
2. NJCBSPT Sentence		—	.48**	.639**	.688**	.335**	.5**
3. NJCBSPT Essay			—	.408**	.378**	.342**	.159
4. ACCUP Reading				—	.631**	.213**	.555**
5. ACCUP Sentence					—	.365**	.546**
6. ACCUP *Plus*						—	.359*
7. SAT Writing							—

*p < .05, ** p < .01

Table 2

Results of Standard Setting under Persona

College Board Proficiency Score Statements	Expert Review under Persona				
Reading Comprehension	Reading Comprehension				
The University Basic Writer Total Score of about 51	Reviewer 1 = 60	Reviewer 2 = 56	Reviewer 3 = 42	Reviewer 4 = 41	Reviewer 5 = 59
The University Traditional Writer Total Score of about 80	Reviewer 1 = 93	Reviewer 2 = 91	Reviewer 3 = 66	Reviewer 4 = 87	Reviewer 5 = 94
The University Honors Writer Total Score of about 103	Reviewer 1 = 108	Reviewer 2 = 109	Reviewer 3 = 110	Reviewer 4 = 100	Reviewer 5 = 117
	Reading Comprehension Persona Scores: Basic Writer: Mean = 51.6 Range = 19 SD = 9.34 Traditional Writer: Mean = 86.2 Range = 28 SD = 11.6 Honors Writer: Mean = 108 Range = 17 SD = 6.05				
Sentence Skills	Sentence Skills				
The University Basic Writer Total Score of about 53	Reviewer 1 = 59	Reviewer 2 = 60	Reviewer 3 = 38	Reviewer 4 = 70	Reviewer 5 = 60
The University Traditional Writer Total Score of about 86	Reviewer 1 = 92	Reviewer 2 = 98	Reviewer 3 = 64	Reviewer 4 = 76	Reviewer 5 = 97
The University Honors Writer Total Score of about 110	Reviewer 1 = 120	Reviewer 2 = 115	Reviewer 3 = 115	Reviewer 4 = 117	Reviewer 5 = 120
	Sentence Skills Persona Scores: Basic Writer: Mean = 57 Range = 32 SD = 11.73 Traditional Writer: Mean = 85.4 Range = 34 SD = 14.8 Honors Writer: Mean = 117 Range = 8 SD = 2.5				

Table 3

Regression Analysis of ACCUPLACER and SAT Writing as Predictors of Holistic Portfolio Score and Course Grade, Fall 2008 – Spring 2009

	Sem. 1 Holistic Score (n = 181)		Sem. 1 Course Grades (n = 774)	
		Fall 2008		
	R^2	F	R^2	F
ACCUPLACER	.267**	(3,177) = 21.45	.066**	(3,770) = 18.24
SAT Writing	.234**	(1,163) = 49.89	.063**	(1,748) = 49.92

	Sem. 2 Holistic Score (n = 103)		Sem. 2 Course Grades (n = 454)	
		Spring 2009		
	R^2	F	R^2	F
ACCUPLACER	.15**	(3,99) = 5.84	.092**	(3,450) = 15.12
SAT Writing	.235**	(1,99) = 30.4	.164**	(1,438) = 85.93

* $\rho < .05$, ** $\rho < .01$

Table 4

Regression Analysis of ACCUPLACER and SAT Writing as Predictors of Holistic Portfolio Score and Course Grade, Fall 2008 – Spring 200: Gender and Ethnicity Analysis

Sem. 1 Holistic Score (n = 178)			Sem. 1 Course Grade (n = 774)		
			Fall 2008		
	R^2	F		R^2	F
ACCUPLACER			ACCUPLACER		
Males (n = 140)	.304**	(3, 136) = 19.79	Males (n = 621)	.064**	(3, 617) = 14.04
Females (n = 38)	.219*	(3, 34) = 3.18	Females (n = 153)	.132**	(3, 149) = 7.56
White (n = 57)	.255**	(3, 54) = 7.5	White (n = 318)	.079**	(3, 314) = 8.99
Asian (n = 58)	.292**	(3, 55) = 17.54	Asian (n = 188)	.08**	(3, 184) = 5.3
Hispanic (n = 33)	.236*	(3, 30) = 3.09	Hispanic (n = 161)	.032 *ns*	(3, 157) = 1.75
Black (n = 15)*qns*	—	—	Black (n = 70)	.089 *ns*	(3, 66) = 2.16
SAT Writing			SAT Writing		
Males (n = 134)	.219**	(1, 132) = 37.1	Males (n = 605)	.049**	(1, 603) = 30.74
Females (n = 31)	.296**	(1, 29) = 12.21	Females (n = 145)	.128**	(1, 143) = 20.9
White (n = 56)	.207**	(1, 54) = 14.1	White (n = 308)	.088**	(1, 306) = 29.45
Asian (n = 54)	.251**	(1, 52) = 17.46	Asian (n = 183)	.068**	(1. 181) – 13.3
Hispanic (n = 31)	.231**	(1, 29) = 8.73	Hispanic (n = 154)	.021 *ns*	(1, 152) = 3.28
Black (n = 13)*qns*	—	—	Black (n = 68)	.052 *ns*	(1 , 66) = 3.65

*ρ < .05, ** ρ < .01

Note: Sample sizes under 30 are designated *qns* (quantity not sufficient).

Table 4 Continued

	Sem. 2 Holistic Score (N = 103)			Sem. 2 Course Grade	
			Spring 2009		
	R^2	F		R^2	F
ACCUPLACER			ACCUPLACER		
Males (n = 84)	.169**	(3, 80) = 5.43	Males (n = 359)	.096**	(3, 355) = 12.6
Females (n = 19)*qns*	—	—	Females (n = 95)	.063 *ns*	(3, 91) = 2.04
White (n = 49)	.348**	(3, 45) = 8.0	White (n = 205)	.125**	(3, 201) = 9.61
Asian (n = 16) *qns*	—	—	Asian (n = 94)	.07 *ns*	(3, 90) = 2.24
Hispanic (n = 22)*qns*	—	—	Hispanic (n = 97)	.071 *ns*	(3, 93) = 2.35
Black (n = 10)*qns*	—	—	Black (n = 39)	.073 *ns*	(3, 35) = .951
SAT Writing			SAT Writing		
Males (n = 82)	.234**	(1, 80) = 24.46	Males (n = 349)	.171**	(1, 347) = 71.54
Females (n = 19)*qns*	—	—	Females (n = 91)	.123**	(1, 88) = 12.47
White (n = 48)	.561**	(1, 46) = 58.67	White (n = 197)	.232**	(1, 195) = 58.9
Asian (n = 16)*qns*	—	—	Asian (n = 92)	.088**	(1, 90) = 8.64
Hispanic (n = 22)*qns*	—	—	Hispanic (n = 95)	.134**	(1, 93) = 14.44
Black (n = 10)*qns*	—	—	Black (n = 37)	.174**	(1, 35) = 7.38

* $\rho < .05$, ** $\rho < .01$

Note: Sample sizes under 30 are designated *qns* (quantity not sufficient).

Table 5

Means, Standard Deviations, and Significant Differences on All Measures, Fall 2008 - Spring 2009

	Total	Male	Female	White	Asian	Hispanic	Black
Fall 2008							
SAT Writing	n = 783	n = 628	n = 155	n = 321	n = 190	n = 160	n = 74
	518.7 (83.85)	513.74$_a$ (83.52)	538.77$_a$(83.5)	534.67$_a$ (82.29)	519.05$_a$ (92.84)	492.88$_a$ (74.17)	496.22$_a$ (73.1)
ACCUP RC	n = 885	n = 695	n = 189	n = 361	n = 212	n = 181	n = 82
	85.21 (17.89)	85.74 (17.93)	84.04 (17.8)	89.64$_a$ (12.73)	79.14$_a$ (20.8)	82.93$_a$ (16.34)	83.09$_a$ (17.1)
ACCUP SS	n = 885	n = 695	n = 189	n = 361	n = 212	n = 181	n = 82
	94.82 (16.64)	93.77$_a$ (17.13)	98.68$_a$ (14.15)	96.89$_a$ (14.98)	91.12$_a$ (17.78)	93.88 (15.76)	93.91 (16.15)
ACCUP Plus	n = 885	n = 695	n = 189	n = 361	n = 212	n = 181	n = 82
	7.75 (1.39)	7.66$_a$ (1.39)	8.1$_a$ (1.37)	7.88 (1.36)	7.73 (1.48)	7.71 (1.21)	7.37 (1.57)
Holistic Score	n = 181	n = 140	n = 38	n = 58	n = 59	n = 34	n = 15*qns*
	8.04 (1.63)	7.96 (1.63)	8.24 (1.6)	8.28 (1.49)	7.86 (1.76)	8.09 (1.52)	7.4 (1.35)
Course Grade	n = 774	n = 621	n = 153	n = 318	n = 188	n = 161	n = 70
	2.91(1.1)	2.89 (1.10)	3.0 (1.11)	2.98$_a$ (1.06)	2.89 (1.1)	2.96$_a$ (1.21)	2.53$_a$ (1.17)
Spring 2009							
Holistic Score	n = 103	n = 84	n = 19*qns*	n = 48	n = 17*qns*	n = 22*qns*	n = 10*qns*
	7.82 (1.8)	7.86 (1.86)	7.63 (1.6)	7.82 (1.87)	8.24 (1.09)	7.5 (1.92)	7.1 (2.13)
Course Grade	n = 454	n = 382	n = 95	n = 205	n = 94	n = 97	n = 39
	2.92 (1.11)	2.88 (1.11)	3.08 (1.1)	3.03 (.993)	2.92 (1.3)	2.71 (1.18)	2.74 (1.02)

Note. Different subscripts ($_a$) within a row represent means different at the 0.05 level by Tukey's honestly significant difference (HSD) test. Sample sizes under 30, too small for HSD analysis, are designated *qns* (quantity not sufficient).

Table 6.

Overview of Admission and Placement, 1998-2010

	1998 (N=678)	1999 (N=710)	2000 (N=729)	2001 (N=740)	2002 (N=685)	2003 (N=770)	2004 (N=713)	2005 (N=809)	2006 (N=907)	2007 (N=858)	2008 (N=939)	2009 (N=1021)	2010 (N=1006)
Admissions Test Scores													
National SAT Critical Reading: Mean Score	505	505	505	506	504	507	508	508	503	502	502	501	501
NJ SAT Critical Reading: Mean Score	497	498	498	499	498	501	501	503	496	495	495	496	495
Class SAT Critical Reading: Mean Score	546	534	533	533	545	540	544	529	520	535	523	537	526
Class SAT Critical Reading: 75th Percentile	600	590	589	590	590	590	600	590	570	580	580	590	587
Class SAT Critical Reading: 25th Percentile	490	480	480	490	490	480	490	470	470	480	460	490	460
First Year Writing Placement													
Honors	102 (15%)	87 (12%)	78 (11%)	111 (15%)	82 (12%)	87 (11%)	81 (11%)	74 (9%)	102 (11%)	92 (11%)	88 (9%)	91 (9%)	62 (6%)
Traditional	357 (53%)	419 (59%)	417 (57%)	409 (55%)	385 (56%)	421 (55%)	394 (55%)	397 (49%)	408 (45%)	453 (53%)	704 (75%)	818 (80%)	766 (76%)
Basic Writing	161 (24%)	134 (19%)	178 (24%)	177 (24%)	179 (26%)	216 (28%)	200 (28%)	298 (37%)	355 (39%)	254 (30%)	103 (11%)	92 (9%)	157 (16%)
ESL Basic Writing	58 (9%)	70 (10%)	56 (8%)	43 (6%)	39 (6%)	46 (6%)	38 (5%)	40 (5%)	42 (5%)	59 (7%)	44 (5%)	20 (2%)	21 (2%)
Correct Placement: Grade C or above													
Honors	92 (90%)	78 (90%)	59 (76%)	81 (73%)	74 (90%)	73 (84%)	78 (96%)	59 (80%)	86 (84%)	84 (91%)	81 (92%)	84 (92%)	60 (97%)
Traditional	298 (83%)	341 (81%)	342 (82%)	352 (86%)	346 (90%)	338 (80%)	329 (84%)	309 (78%)	316 (77%)	393 (87%)	602 (86%)	663 (81%)	631 (82%)
Basic Writing	150 (93%)	118 (88%)	155 (87%)	146 (82%)	163 (91%)	194 (90%)	178 (89%)	270 (91%)	311 (88%)	225 (89%)	88 (85%)	71 (77%)	141 (90%)
ESL Basic Writing	55 (95%)	70 (100%)	54 (96%)	37 (86%)	34 (87%)	39 (85%)	33 (87%)	38 (95%)	39 (93%)	54 (92%)	40 (91%)	17 (85%)	19 (90%)
Course Failure: Grades of D or F													
Honors	5 (5%)	6 (7%)	16 (21%)	27 (24%)	8 (10%)	14 (16%)	3 (4%)	13 (18%)	12 (12%)	6 (7%)	5 (6%)	6 (7%)	2 (3%)
Traditional	37 (10%)	65 (16%)	60 (14%)	44 (11%)	31 (8%)	60 (14%)	54 (14%)	71 (18%)	79 (19%)	45 (10%)	83 (12%)	124 (15%)	117 (15%)
Basic Writing	5 (3%)	10 (7%)	17 (10%)	30 (17%)	10 (6%)	20 (9%)	17 (9%)	23 (8%)	39 (11%)	26 (10%)	13 (13%)	17 (18%)	14 (9%)
ESL Basic Writing	2 (3%)	0 (0%)	2 (4%)	5 (12%)	2 (5%)	6 (13%)	5 (13%)	0 (0%)	2 (5%)	4 (7%)	3 (7%)	2 (10%)	1 (5%)
Course Withdrawals and Incompletes													
Honors	5 (5%)	3 (3%)	3 (4%)	3 (3%)	0 (0%)	0 (0%)	0 (0%)	2 (3%)	4 (4%)	2 (2%)	2 (2%)	1 (1%)	0 (0%)
Traditional	22 (6%)	13 (3%)	15 (4%)	13 (3%)	8 (2%)	23 (5%)	11 (3%)	17 (4%)	13 (3%)	15 (3%)	19 (3%)	31 (4%)	18 (2%)
Basic Writing	6 (4%)	6 (4%)	6 (3%)	1 (1%)	6 (3%)	2 (1%)	5 (3%)	5 (2%)	5 (1%)	3 (1%)	2 (2%)	4 (4%)	2 (1%)
ESL Basic Writing	1 (2%)	0 (0%)	0 (0%)	1 (2%)	3 (8%)	1 (2%)	0 (0%)	2 (5%)	1 (2%)	1 (2%)	1 (2%)	1 (5%)	1 (5%)