

Turning Placement into Practice

Irvin Peckham

The history of writing assessment begins with composition: the first exams to weed out the rabble were oral; then came quizzes with questions like this one from an 1870 entrance exam: “He said that that that that that pupil parsed was not that that he should have parsed. Parse the *that’s* in this sentence” (qtd. in Crowley 65).¹ Our professional predecessors seem first to have moved from indirect to direct assessment of writing in the academic year 1872-1873 with writing tasks such as “Write a short composition on. [. . .] The Style of ‘Henry Esmond’” (qtd. in Crowley 67–68). Although replacing with short essays the early equivalent of bubbled-in test answers, the examiners’ responses (see Crowley 68) make clear they were looking for evidence of literary knowledge and mastery of spelling, punctuation, and grammar—and that the vast majority of high school graduates were sadly lacking in both.

We have come a way since then but hardly by leaps and bounds. I was teaching high school in California in 1974 when the minimum competency movement swept through the nation’s high schools. In Santa Clara County, the early response to demands for testing sixth-, eighth-, and twelfth-grade mastery of writing was to adopt one of several versions of an objective assessment of editing skills—not far from the clever parsing question in 1870. I was a member of the Santa Clara County Committee on Writing that eventually persuaded most school districts in our county to recommend versions of the holistic model for assessing writing in place of the objective exams. The effort sparked acrimonious arguments about our translating a formative assessment model, popularized by the Bay Area Writing Project, into an exit exam.

Involvement in this controversy was my introduction to the concept of assessment and its link with instruction. For ten years, I directed the minimum competency exams in the Morgan Hill Unified School District, where we discovered ways to turn the requirement for minimum competency exams to our advantage. Through our scoring sessions, teachers at the elementary, middle, and high-school levels talked with each other about the elements of good writing, the kind of writing tasks we should set for the next year, and how we might reshape our classes to help failing students, for whom we then developed an intensive individualized writing program. Through our common experience focused on serious discussions about writing, we all learned much about teaching writing and significantly reshaped writing instruction in our district. Most notably, it helped us resist the five-paragraph and power-writing models that were becoming the rage in California.

I have been involved in assessment since then, mostly as a consequence of three assumptions that emerged from my experience in Morgan Hill: first, it's better to have writing teachers take charge of assessment than to let outside forces tell them how to do it (see White, "Language"); second, if teachers are clear about what they are teaching, they should be able to develop some way of assessing performance;² and third, good assessment practices improve teaching performance.

Assessment theory and practice in colleges have largely been connected with placement: deciding whether the student is placed in a developmental, first semester, or second semester course or exempted. To the uninitiated, sorting students into poor, adequate, good, and excellent writers might not seem difficult, assuming that the sorters acknowledge the possibility of error when the essays lie near the boundaries of the categories. But anyone who has thought seriously about the link between writing and writers knows how complicated such sorting can be. We know that students who write well in one genre may not be equally successful in another (one only need look at Henry James' career as a dramatist to understand that). We know that writing situations and methods of scoring affect who will be placed where. In sum, placement is a messy business. We do not look for perfect answers; we look for better practices.

Although our deans may disagree, we know that placing students on the basis of their ACT or SAT scores is notoriously unreliable. Research on the correlation between indirect and direct methods of assessment is consistently negative, ranging from a low of .2 (Bamberg) to a high of .63 (Breland and Gaynor, cited in Stiggins). Richard Haswell, one of our field's most indefatigable collectors of information, summarized this research in the WPA listserv *Archives of WPA-L@ASU.EDU* posting ("RE: SAT Quote about Writing").³ Edward White summarized on WPA-L his unpublished report of the

correlation between the SAT and the California State University English Placement Test; he also confirmed these weak correlations, with an exception that at the lower and higher score levels, White found, the SAT was a reasonably accurate predictor of writing performance (“RE: SAT/ACT Very-High/low Score”), an exception that seems predictable.

A common alternative to the objective assessment for placement purposes has been the so-called “diagnostic,” usually administered during the first week of classes. Other systems are the single-draft timed essay, the multiple-draft timed essay, samples in different genres (Stitt-Bergh), portfolios (Condon and Hamp-Lyons; Willard-Traub, Decker, Reed, and Johnston; Yancey *Portfolios*), directed self-placement (Gilles and Royer), and informed self-placement (Bedore and Rossen-Knill). A few people have also experimented with untimed essays (Myers; Stitt-Bergh) and online assessments. There are, in addition, two general methods of evaluating essays for placement: variations of holistic scoring and the “expert reader” model (Calpus and Smith; Smith; Haswell and Wyche-Smith).

In “Toward a New Theory of Writing Assessment,” Brian Huot set these methods against each other, the former being marginalized as decontextualized, positivistic, and current-traditional, the latter being privileged as contextualized and postmodern. Huot’s argument for an “emergent” theory of assessment provides an important critique of assessment specialists: he sees them as people who get so bound up in statistics and terminology that they seem to forget they are assessing writing. But his argument is based on a false dichotomy. Those (like me) who have used variations of the holistic model are set up as strawpeople who hold beliefs that few would ascribe to; notably, that “student ability in writing, as in anything else, is a fixed, consistent, and a contextual human trait”(549–50). Huot imagines positivists who claim that if one can look carefully, one will discover a fixed point defining a person’s writing ability. Maybe some people in various industries believe this notion, but I doubt that anyone who has paid attention to genre and assessment theory would imagine that writing ability is a fixed point. Rather than use an “old-Europe/new-Europe” rhetoric to set the holistic model and the “emergent” theory of assessment against, it’s more useful to imagine ways the two models can amend and complement each other. Unlike Haswell and Wyche-Smith, Huot, and more lately, White (“Re: Synopsis”), I think writing teachers can look at writing samples written in situations comparable to the ones we create in our classrooms and, through comparing them and discussing their evaluations, arrive at some reasonable inferences about the writers’ abilities to meet other kinds of writing demands—always acknowledging that boundaries and immediately surrounding areas are at best fuzzy and that we are dealing with *probability*, not certainty. Rather than give

up on holistic models of assessment, we have to entertain broader notions of ways we can adapt those models in the process of generating responsible assessment practices. If we *must* sort (and that's always a question), our job is, then, to develop good means of doing it.

Citing the University of Pittsburg (Calpus and Smith; Smith) and Washington State University (Haswell and Wyche-Smith) placements as prototypes, Huot stands on stronger ground when he claims that assessment has to be developed locally and designed to provide answers to specific questions. Assessment needs to be tied to the local curriculum and always linked to the knowledge that when writing instructors examine specific pieces of writing produced in specific rhetorical situations and in specific genres, they are looking at only the tip of the iceberg, guessing about the vast geography that lies underneath. Haswell points out that there are many ways of looking at the tip, and that it is a serious mistake to rely on one glance, particularly when, as a result of a single score, students may be placed in the stigmatized developmental writing courses ("Post-secondary," par. 7). But the question is always one of time and money—at least for the WPA, who must negotiate competing interests. And one must always question the purpose of sorting, which has become a question so freighted with ideological overtones that it can provoke shouting matches between professors who claim to be teaching critical thinking in their classes.

If new WPAs are lucky enough to stumble into universities with sophisticated programs for placement and assessment, they won't have to worry about these issues, but I suspect many of us take positions in academic arenas where placement and assessment processes need revision. This article focuses on a placement strategy at Louisiana State University that had outgrown its use and on our attempts to revise it. I will describe the limits we faced while trying to imagine a responsible placement program and how we inadvertently turned the limits into assets through a hybrid strategy that combined direct assessment, indirect assessment, and self-placement in an online environment. The success of our Online Challenge led us to extend the procedure to assess student achievement at the end of the semester. Our purpose in the semester assessment was to improve course coherence, motivate student and teacher performance, and respond to the perennial charge of grade inflation.

PLACEMENT: OUR ONLINE CHALLENGE

In the writing program at Louisiana State University, the need for a required writing program is an article of faith. So is the assumption that we need to structure our sequence of writing tasks in two courses so that less accomplished writers will take both courses, the more accomplished will take only

the second course, and the most accomplished will take neither. Although I question these assumptions (see also Elbow 87–88), this is the belief system within which I operate as the WPA. Part of my job, then, is to determine the most effective method—within constraints particular to the institution—of classifying students on the basis of their writing ability and then linking that classification to what is being taught in the first and second courses.

The existing placement protocol at Louisiana State University had been developed in the early 1970s to operate as a check on the classification of writing ability by indirect measures. Students were initially placed in English 1000 (a no-credit course), 1001, 1002, and 1003 (the honors version of 1002), or students were exempted on the basis of their ACT, SAT, or AP scores. The former director of the program designed the “diagnostic”⁴ to check on placement via indirect assessment. During the first week of classes, students wrote an in-class essay and were either kept in the same class category or kicked up- or downstairs on the basis of these essays.

Several problems arose, however, in using the diagnostic as a check on the ACT/SAT/AP for placement. The procedure was clearly physically disruptive during the first two weeks of classes, creating chaos in our first-year writing office as essays and students were shuffled around. It certainly created an unsettling environment for students just entering the university. It wasted human resources as a consequence of the need to juggle class maximums to assure that “re-placed” students could then enter sections held at the same hours as their original class. In addition, the evaluation was uncontrolled, relying primarily on individual instructors’ sense of whether a student belonged in one of our four courses or should be exempted.⁵ And finally, although an improvement over indirect measures, timed, in-class, single-draft essays are the least valid of direct measures of a student’s ability to write.

White has suggested a hierarchy of testing methods according to what he calls “testing validity” based on the appropriateness of the testing instrument to what is being tested. In order from least to most valid, the methods are as follows: indirect measures (ACT or SAT); timed writing samples; un-timed writing samples; portfolios with samples of student writing in several genres (“Apologia”). The logic governing this hierarchy is commonsensical: the closer the structure of the assessment instrument to the structure of what is being assessed, the more valid the assessment. If we want to assess students’ editing skills, ACT or SAT tests might do the job. If we want to assess students’ abilities to respond to a generic topic within announced time limits, then the timed writing sample works. If we want to approximate the kind of writing situations most of us set up in our first-year writing classes—with time for invention, discussion, outlining, drafting, revision, peer responses,

and rewriting—then untimed writing samples are the way to go. If by “writing proficiency,” instructors mean the student writer’s ability to respond to a variety of writing situations in a variety of genres, then a portfolio assessment certainly makes more sense than a twenty-minute response to a generic topic like “creativity.”

In response to the problems with the diagnostic, we developed the Online Challenge for three reasons:

1. We wanted to have the placement completed by the time students entered classes, but we did not have enough money to pay readers to score 5,000 (the average number of students entering LSU) essays in the summer.
2. We wanted the writing situation to resemble as closely as possible the writing situations we created in our first-year writing courses.
3. We knew we could not ask students to come to LSU in the summer to write an essay. We might have managed to squeeze in a twenty-minute session for delivering such a writing sample, but we weren’t convinced that a twenty-minute impromptu would be significantly more reliable than indirect assessments.

We structured the Online Challenge like this: Students receive an initial placement on the basis of the ACT, SAT, or AP scores. After they are admitted, we send them a letter in which we invite them to challenge their placement by submitting an essay online.⁶ We run the assessment three times in the summer (in June, July, and August). The student chooses one of those assessments. The student goes online to download eight to ten articles we have selected on a particular subject. We give the student three days to read the articles and make notes. On the fourth day, the student goes back online to access the writing task. The student then has seventy-two hours to write the essay. He or she then returns to the online URL in order to upload the essay.⁷

Several experienced teachers and I score these essays in a controlled reading with explicit descriptions of score points,⁸ sample papers for each score point,⁹ and at least two readers for each essay. A third reader adjudicates when the two readers disagree by more than one point. We developed the descriptions by sorting 150 essays into low, medium, and high stacks, subdividing each stack into low and high to get six stacks. From each stack, we selected representative papers we thought were central examples of each stack, and then described them, developing our criteria along with our

descriptions. With every reading, we have followed this practice, building on our previous year's description, adapting the new description to the current year's specific writing task, and choosing sample papers for each score point from the new essays. Because we do not have a large number of essays to read (about 260), we read them carefully and break frequently to discuss problematic papers that we put on the LCD overhead, a procedure simplified by having all essays online. While acknowledging the probability of error, we know that through our collective reading and discussions we arrive at more valid judgments than any of us would make if we were to rank these essays alone.

Our assumption behind the Online Challenge is that students who think they should be in a different course or exempted are the students whose writing we should review. We preferred to look at everyone's writing, but scoring 5,000 essays was not economically feasible. In effect, we were creating a type of self-directed placement¹⁰ with the students determining whether they believed in their writing abilities and whether they were willing to do the extra work required to challenge their ACT/SAT/AP placements.

Assuming that most writers who participated in the Challenge would be on the upper end of the range of scores defining each category and would have a higher sense of self-efficacy than those in the middle or lower end of the categories (Reynolds), we expected a higher exemption rate than we had solely on the basis of the ACT/SAT/AP scores—in spite of the fact that the students with high ACT/SAT/AP scores had already been exempted. In line with prior research, we also predicted a low correlation between the ACT/SAT/AP scores and our scores for the students' essays.

Over the past two years, we have had 5% of the incoming students take up the Challenge. We have exempted between 7% to 8% of those (in contrast to the .8% that we exempt by virtue of ACT/SAT/AP scores). Our correlation coefficient with the ACT/SAT/AP scores was very low (.35). Below is a table of the reassignments on the basis of the Online Challenge:

Table 1.
Percentage of Reassignment in Each Category at LSU during Academic Years 2003–04 and 2004–05¹¹

Year, with Annual Student Reassignment	1001 Students Reassigned to 1002 ¹²	1001 to Exempt	1002 to Exempt	1003 to Exempt	1003 to 1001	1002 to 1001	Students Staying in Same Sections
2003–04	9.4	3.4	3.7	1.5	5.2	19.9	56.9
2004–05	10.6	1.9	1.5	2.7	5.7	16.7	61.8

NOTE: We had dropped English 1000 by 2003. In 2003–04, the program had 267 students reassigned; in 2004–05, 264 were reassigned.

Roughly, then, the students would have received the same placement with ACT/SAT/AP and essay scores a little over half of the time. That amounts to a flip of the coin. Two years of placement does not give us enough data to claim trends, but the percentages of “1001 to 1002” (9.4 and 10.6) and the percentages of “1003 to 1001” (5.2 and 5.7) and “1002 to 1001” (19.9 and 16.7) seem consistent. The percentages in “1001 to Exempt” (3.4 to 1.9) and “1002 to Exempt” (3.7 to 1.5) indicate that in 2004–05, the readers were less generous in their scoring, but the increase in the percentages of “1001 Students Reassigned to 1002” (9.4 to 10.6) and the lower percentages of “1002 to 1001” (19.9 to 16.7) argue against that interpretation. In any case, with the exception of “1001 to 1002,” the promotion between categories is between 1.5 (2003–04 “1003 to Exempt”) and 3.7% (2003–04 “1002 to Exempt”).

The dramatic bounce in the category “1001 to 1002” (9.4 and 10.6%) implies that the border between 1001 and 1002 is improperly drawn. But if we lowered the ACT/SAT/AP boundary to place more of the “1001 to 1002” students in 1002 on the basis of their ACT/SAT/AP scores, we would probably see an increase in the percentage of “1002 to 1001” students who would be demoted on the basis of their Online Challenge scores; the reverse might also true—if we raised the ACT/SAT/AP boundary between 1001 and 1002, we will increase the “1001 to 1002” category while decreasing the “1002 to 1001” category. Essentially, the ACT/SAT/AP is an unreliable indicator of a distinction between 1001 and 1002 students. This volatility at the mid-levels reinforces the results of White’s study cited above (“Re: SAT/ACT Very High/low Score”).

Two other statistics are striking: that more than 5 percent of the 1003 students should have been, according to my readers, placed in 1001 sections;¹³ and that 17 to 20 percent of 1002 students should have been placed in 1001. The demotion of student writers placed in 1003 was surprising because of the high ACT English threshold (31) for that group of students. Regardless of the verbal skills of student writers placed in 1003 on the basis of their ACT’s English exams, my readers thought that more than 5 percent of these writers should take *both* semesters of first-year writing! The disproportionate percentage of English 1002 students (those with a threshold score of 26 for ACT English) who, my readers thought, should have been placed in 1001 was equally surprising.

Initially, I thought my readers had been off-base in their scoring. But when I did some random checks on their scoring, I usually agreed within a one-point range on the assigned score. I found support for the disproportionate demotion of students in a essay by Russell Meyers. When comparing rankings based on take-home essays to rankings by indirect placement,

Meyers found that 44 percent of the students should have been assigned a lower placement. My readers, by comparison, had been too easy on student writers.

Two factors may account for this disproportionate downward movement: First, readers are notably more inclined to give lower scores to anonymous writers than to students in their own classes. Second, the standards that teachers insist on for exemption from courses are higher than the standards they use for assigning grades in their classes. This latter logic conflicts with the common assumption in proficiency testing that if the placement score indicates a student will receive a passing grade in a course, then the student should receive exemption from that course (see Calpus and Smith; Haswell, "Post-secondary"; McKendy). In fact, of the forty-one students in 1002 for whom we recommended demotion but who decided to stay in 1002, 51 percent of them received As, 43 percent received Bs, and 5 percent received Cs.¹⁴ Of the eight students who decided to follow our advice and took 1001, 63 percent of them received As and 37 percent received Bs.

One can reasonably conclude from these data that our bark is worse than our bite. More charitably, we can say that in placement, we assess writing ability to determine who obviously does not need one or two semesters of writing. We exempt those either from 1001 or from 1001 and 1002. But students who we think might benefit from further instruction are required to take either one or both courses. At the end of the semester, we graduate them from the course with some hope that they have learned about as much as they could, knowing that most students would not be able to come up to the level we had established as the threshold for exemption. Put another way: we place as idealists and grade as realists. Add to this some element of grade inflation.

SEMESTER ASSESSMENT

As a consequence of the many discussions we had in our scoring session, we decided to use the strategy we had established with the Online Challenge to assess student writing ability at the end of the semester. We had several purposes:

1. To address the issue of grade inflation. We believe there is some link between high grade distributions and high student evaluations, but we wanted to include in any assessment of grade distribution an evaluation of student achievement. We assume there will be some classes with higher student achievement and consequently a higher grade distribution; the converse is probably also true. We wanted to be able to give guidance to teachers for assigning grades on the basis of a controlled assessment of student performance.

2. To involve more teachers in workshops in which they would discuss evaluation practices and learn to evaluate student achievement with greater consistency.
3. To use the assessment to promote greater coherence in the program. We wanted some evidence that teachers were teaching a common curriculum in English 1001.
4. To assess overall student achievement so that we could measure any program changes against changes in student achievement.
5. To introduce in the classroom situation an additional motivation for teachers and students to improve performance. We wanted to have teachers strive for high scores for their students in the semester assessment; we also wanted to motivate student achievement by exempting them from English 1002 if they achieved high scores in the assessment.

For three years, we have been piloting the semester assessment in our English 1001 classes. We focused on English 1001 because we could expect greater course coherence among all the new GTAs, who follow a common curriculum, than in 1002, taught largely by instructors, who follow varied curricula. We could also use the assessment of student writing in 1001 to determine which students we could reasonably exempt from 1002 after a semester of instruction.

All new GTAs are required to participate in this project; in addition, I invite instructors to participate. For the last two fall semesters, an average of thirty teachers (including GTAs and instructors) and one thousand students have participated in each semester assessment.

Teachers reserve the last week of classes for this assessment. During that time, students participate in the semester assessment in exactly the same way as students have done in the Online Challenge. They have six days in which to complete the reading and the writing task. The writing task is in the same genre as the tasks with the Online Challenge.¹⁵ During finals week, teachers meet with me on Monday morning for four hours of training about scoring. We score for four hours on Tuesday morning and four hours on Wednesday morning. As with the Online Challenge, two readers read each essay with a third reading required if there is greater than a one-point difference between the scores the two readers assign.

By Wednesday afternoon, teachers can go online to access scores for their students. They receive individual scores, a graph showing students' score distribution, their students' mean scores, and the mean score for all students in the assessment. Figure 1 is a sample of data available for one teacher. The students in this class did well.

Figure 1.

Student Score Distribution in One Class for Semester Assessment, Louisiana State University Writing Program.

Department of English										
Overall mean: 5.76										
Instructor mean: 7.00										
				*						
				*		*		*		
		*		*		*		*		*
		*	*	*	*	*	*	*		*
2	3	4	5	6	7	8	9	10		12
Individual Student Scores										
Last Name		First Name		MI	ID Number			Score		
Daniels		Jessica		W	6320587			10		
Breaux		William		P	4559880			7		
Nichols		Mary		A	5784993			6		
Jameson		Elizabeth		N	4576822			8		

Note: Scores represent the sum of scores by two readers with less than two point difference on a 6 point scale. Scores for four students are shown as a sample. The teacher would see scores for the entire class. All names are pseudonyms.

These graphs and the comparison of the mean of the instructor’s class to the mean for all students participating in the assessment give instructors significant data on their students’ achievement. We do not claim this one score should definitively represent a student’s ability to write, but it does tell instructors, with some degree of objectivity, how well a student can write in a genre and in a writing situation that we feature in English 1001. In addition, the scores are contextualized by the instructors’ experiences in the scoring process. By having participated in the scoring, instructors know both the strengths and weaknesses of the process. They know the judgment is more objective than it would have been if they had evaluated their own students’ essays. They know we have discussed at length criteria that are central to the genre and essays that represent scores from one to six. They know that two readers had to agree on a score within one score point and that if there were

a disagreement, the essay would be read by a table leader, who would adjudicate the score, often by discussing the essay with the first two readers. But they also know that there are holes in the process—that in spite of our four-hour training sessions, two low-scoring readers or two high-scoring readers may have read a particular paper. To check on the possibility of egregiously inaccurate scores, we ask the instructors to read any of their students' essays that received scores not matching the instructors' expectations of the student's achievement (in fact, most instructors read all of them). If instructors challenge a score received in the assessment, they frequently ask one or two other instructors to read the essay as an additional check.

To be exempted from English 1002, students have to receive either a eleven or twelve on the assessment or a ten with their instructor's recommendation. Beyond those restrictions, the instructors decide how they will use the scores to determine final grades. But in order to mitigate erratic grade distribution patterns, we ask the instructors to pay attention to the distribution of scores and to the mean in their classes. If the mean of a particular class is higher than the overall mean, we can understand a higher than average grade distribution, but if a mean is significantly lower, we do not expect to see a disproportionate number of As and Bs. Because we know this method of evaluating writing ability is far from infallible, we expect exceptions, but we do not expect the exceptions to become the rule—e.g., a teacher who awards As to sixty percent of students receiving combined scores of 6 or lower.¹⁶

Table 2 shows the number and percentage of students at each score point in our Fall 2004 assessment at LSU.

Table 2.
Score Distribution of LSU Student Scores' for Semester Assessment, Fall 2004 (n = 1,056)

Score	2	3	4	5	6	7	8	9	10	11	12
No. of Essays per Score	29	37	70	136	165	200	160	140	68	45	11
Percentage	2.7	3.4	6.6	12.8	15.5	18.8	15.1	13.2	6.4	4.2	1.0

Note: Scores represent the sum of scores by two readers with less than two point difference.

There are several points of interest in this distribution: First, we exempt roughly 80 students out of the 1,056 who took part in this assessment. It is worth noting that when combined with the twenty or more students exempted from the Online Challenge, that exemption represents an instructional-cost savings of more than \$17,000 dollars. If this score distribution is

repeated when we assess all students in English 1001 in the fall of 2005, our program will save approximately \$51,000 in instructional cost (see Appendix).¹⁷

Second, approximately 30% of the students scored eight or better, with an additional 18.8% scoring at seven. In the Online Challenge, we consider a seven score to be on the border of capacity to give an acceptable performance in English 1002 courses. The 301 students who received scores of five or six may pass 1002 but not do well in the course. We could argue that students who received scores of two, three, and four should repeat 1001, but more than 130 students would have to repeat, which would dramatically increase instructor cost and obviously call teachers' grading practices into question. We would have to address quite a few other controversial issues if we were to switch this assessment into an exit exam, but this is a minefield we are not interested in crossing right now.

We have identified students who have taken both the Online Challenge and participated in the semester assessment. We tracked the difference in their scores in the challenge and the semester assessment to see whether overall growth in their skills existed, a phenomenon which we could then attribute to their 1001 experience. Figure 2 shows the difference in scores.

Figure 2.
Score Differences among Students Accepting the Online Challenge and Participating in LSU's Fall 2004 Semester Assessment. (n = 42 students)

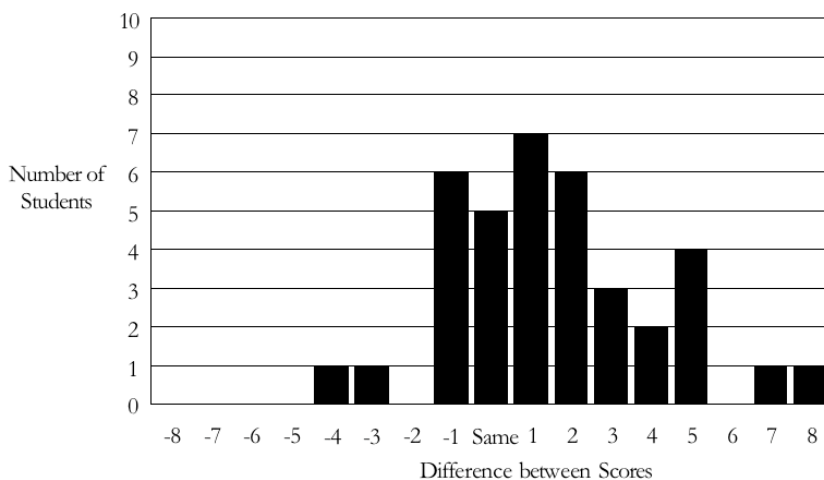


Figure 2 shows that five students had the same score in the Online Challenge and semester assessment. Six students' scores dropped by one point. One student's dropped by three points, and another, by four. The figure shows, however, that most students received higher scores at the end of the semester than they received in the Online Challenge before they began the semester. Admittedly, forty-two students is a small sample and therefore likely to be idiosyncratic; nevertheless, we see a pattern of growth.

We expect that this assessment will help us approach the problems of erratic grade distribution patterns as well as improve student writing. Teachers will have concrete evidence of student achievement, evidence that should guide them in their grading practices. In addition, we will provide them with a correlation coefficient linking their students' assessment scores with the grades assigned. We understand that factors other than writing achievement contribute to the course grade; nevertheless, an extremely low correlation, say of .6, should be cause for reviewing one's grading practices. By linking grade distribution to student achievement, we expect to improve both individual and class performance because students will know that by achieving a high class average, they will give their teachers a rationale for a higher grade distribution. For higher achievement, both teachers and students will have to pay attention to the explicit objectives for the course. And finally, teachers will learn how to evaluate essays using the same benchmarks and scoring criteria that their colleagues use. They will no longer be grading in the dark.

WHERE FROM HERE?

Although an improvement over placement by indirect measurement, the extant practice of a check by the diagnostic was clearly flawed, leading us to replace it with our Online Challenge. We originally designed this process as a concession to limitations of time and money, but we were able to turn the limitations to our advantage. By going online, we could approach the writing situations characteristic of our classroom assignments; by allowing students to challenge their placement by indirect measurement, we allowed them some sense of agency, dramatically cutting our scoring costs and focusing on students most likely to have been misplaced. The success of Online Challenge led us to adopt it as a method of assessing student achievement at the end of the semester. The Semester Assessment allowed us to focus more accurately on what we were teaching, normalize teacher evaluation, and respond to the overly simplistic linkage of higher grade averages with higher student evaluations.

These tandem processes have opened new possibilities for research we have yet to pursue. We need more data on student growth, measured by student scores on the Online Challenge and the semester assessment. We need to track student achievement, measured by grades, correlated with their scores on the Online Challenge. We can also develop a more coherent understanding of the relationship between placement and grading, which I have reframed as the idealist standards for placement measured against the realistic standards for grading. It's interesting, to put it mildly, that nearly 50% of the students that we pass with Cs or better in 1001 would not score high enough to be exempt from 1001 on the basis of our Challenge scores.

Finally, we believe we have barely opened the possibilities of online placement and assessment. By moving from the timed to the untimed writing sample, we have improved testing validity, but to get closer to an accurate assessment of students' writing abilities, we have to look at samples from several genres. We also know that this move increases writing time and scoring costs and decreases interrater reliability (Willard-Traub et al. 70–72). Making this move, however, is our direction; we will progress slowly, incrementally training our teachers to measure their scores against explicit criteria, and thereby normalizing their responses to student writing. The practical instructional effect will be our program's use of assessment as a reflective practice, measuring our pedagogy against student achievement, and then using both to move our assessment progressively toward a more accurate measure of what we claim to be teaching.

ACKNOWLEDGMENTS

I would like to thank Charles Cooper, who has been my mentor in many ways, but particularly in assessment. I have learned a good deal from the WPA listserv, which periodically erupts in assessment-centered conversations. I have also learned from specialists in the field—most notably, Edward White and Richard Haswell, who have generously and patiently responded to my inquiries. Finally, I have had the help of several colleagues around the country—Elizabeth Coughlin from DePaul University, Marlene Minor from the University of Cincinnati, and Steven Youra from California Institute of Technology—who have worked with me under the leadership of Les Perelman from Massachusetts Institute of Technology to develop a strategy for large-scale, online assessments, obviating the need for student presence. The five of us are deeply indebted to Shannon Larkin, from Massachusetts Institute of Technology, who kept us moving forward on the project.

APPENDIX

Table 3. Score Distribution of LSU Student Scores for Semester Assessment, Fall 2005 (n = 2,414)

Score	2	3	4	5	6	7	8	9	10	11	12
No. of Essays per Score	116	148	283	383	466	389	260	174	131	46	17
Percentage	4.8	6.1	11.7	15.9	19.3	16.1	10.8	7.2	5.4	1.9	0.7

While the score distributions in Fall 2005 followed a pattern similar to the one for Fall 2004, the overall distribution was shifted to the left—more low and middle scores and fewer higher scores. This shifted distribution resulted in a lower mean (6.94% for Fall 2004 versus 6.18% for Fall 2005). Our exemption rate was also consequently lower (7.5% for Fall 2004 versus 6.0% for Fall 2005). We suspect the lower exemption rate was the consequence of two factors: a more difficult writing task and a larger group of readers. The construction of the writing task was made more difficult by our decision to link the subject of the writing task to our summer reading program. The larger group of readers was not as well-trained as previous groups of readers who had volunteered to participate in the pilot program. A more careful analysis of scoring patterns will, we believe, at least partially answer the causes of the lower score distribution.

NOTES

1 I don't want to take the delight from this puzzle by naming the functions of the sequence of that's. The key lies in accenting the second, third, and fifth that's and pausing between the third and fourth. The examiner might have embellished the puzzle by writing three rather than two that's in the second sequence.

2 This is hardly my insight. Edward White has hammered this point home on the WPA listserv. We could consider it his mantra, without suggesting that he thinks linking assessment to instruction is easy to do.

3 With Haswell's permission, I have made this summary available at <http://www.english.lsu.edu/dept/programs/ugrad/firstyear/documents/indirectdirect.htm>.

4 I use scare quotes because calling a quick reading with a thumbs-up or a thumbs-down judgment a diagnostic seems mostly an act of rhetoric to professionalize the process.

5 On a superficial level, the system of scoring the diagnostic resembles the "expert reader" model pioneered by teachers at the University of Pittsburg (see Smith; Calpas and Smith) and later refined at Washington State University (Haswell and Wyche-Smith), but both of these practices relied on a small set of expert teachers reading in a semi-controlled environment with negotiations between read-

ers over disputed classifications. At LSU, we were relying on 60 instructors and 56 graduate students with significantly different degrees of expertise and experience to determine whether a writer belonged in one of the four categories. We had no reader training program, no sample papers, no genre-specific criteria.

6 For the letter, see <http://www.english.lsu.edu/dept/programs/ugrad/firstyear/documents/letter.doc>. For the description of the Online Challenge, see <http://www.english.lsu.edu/dept/programs/ugrad/firstyear/documents/description.doc>. Also, see <http://www.english.lsu.edu/dept/programs/ugrad/firstyear/placement>.

7 We are frequently asked how we know we are reading the student's writing. Admittedly, we can't guarantee that the essays represent the students' writing—no more than we can guarantee that out-of-class essays during the semester represent the students' writing. In designing this placement strategy, we wanted to reproduce as nearly as possible the writing situations we present in our first-year writing classes. A characteristic of those situations is that students might cheat. We do have students write a statement of authenticity. In addition, we make the writing situation demanding so that it will be a good friend indeed who will read all the articles and write an essay to get his or her friend out of English 1001 or English 1002.

8 For our matrix describing the characteristics of essays at each score point, see <http://www.english.lsu.edu/dept/programs/ugrad/firstyear/documents/matrix-homeless.htm>.

9 For essays representing each score point, see <http://www.english.lsu.edu/dept/programs/ugrad/firstyear/documents/anchors.htm>.

10 I was drawn to variations of guided self-placement (Gilles and Royer), but the belief system at Louisiana State University directly opposed notions of student agency underwriting any guided self-placement. As an incoming WPA, I thought that challenging the program's existing belief systems directly would be bad judgment.

11 Because percentages are rounded off, the totals may not equal 100.

12 We did not promote students to 1003 because we felt we would be distinguishing differences that were more hypothetical than real. Thus, if a 1001 student achieved a high score, we either put her in 1002 or exempted her, depending on how high that score was.

13 We don't have a "1003 to 1002" category because in translating scores to classes, we blurred the distinctions between 1002 and 1003. Scores of 8–9 were designated 1002. Scores of 10, 11, and 12 were "exempt." We could have designated 10 as the characterizing score for 1003, but such a distinction was pointless because we are not pretending that we can accurately tell the difference between essays separated by one point on a twelve-point scale.

14 We let students whom we recommended for a lower-level course decide for themselves whether they want to stay in 1002 or 1003, in order to encourage students to take the Challenge: obviously, we believed, far fewer would challenge if they thought they might lose rather than gain from the exercise.

15 We don't ask our faculty for more unremunerated work. Instead, we trade time off for hours spent: instructors are released from teaching duties during the last week of school while their students participate in the online assessment. In exchange, they work on the scoring for three half-days during finals.

16 For an online description of our process, see <http://www.english.lsu.edu/dept/programs/ugrad/firstyear/documents/outlinepilot.htm>. We make this description available to both students and instructors. The last item, "Score Reports," addresses the issue of grade distribution and provides explicit motivation for improved performance. Please note that we have hedged a bit under the item, "Exemption from 1002." To save the director and associate director from having to read all the 10s, we ask instructors to recommend whether we should exempt students in their classes who received 10s.

17 We have just completed the Fall 2005 assessment with seventy-four teachers and 2,414 students participating. Appendix, Table 3, shows the results.

WORKS CITED

- Bamberg, Betty. "Multiple-Choice and Holistic Essay Scores: What Are They Measuring?" *College Composition and Communication* 33 (1982): 404–06.
- Bedore, Pamela and Deborah F. Rossen-Knill. "Informed Self-Placement: Is a Choice Offered a Choice Received?" *WPA: Writing Program Administration* 28 (2004): 55–78.
- Calpus, Gary and William L. Smith. "The 'Expert Model' of Placement Testing." *Composition Chronicle* 8.1 (1995): 4–7.
- Condon, William and Liz Hamp-Lyons. "Questioning Assumptions about Portfolio-based Assessment." *College Composition and Communication* 44 (1993): 176–90.
- Crowley, Sharon. *Composition in the University*. Pittsburgh: U of Pittsburgh P, 1998.
- Elbow, Peter. "Writing Assessment in the 21st Century: A Utopian View." *Composition in the Twenty-first Century: Crisis and Change*. Eds. Lynn Bloom, Don Dai-ker, and Edward White. Carbondale, IL: Southern Illinois UP, 1997. 83–100.
- Gilles, Roger and Daniel J. Royer. "Directed Self-Placement: An Attitude of Orientation Special Section." *College Composition and Communication* 50.1 (1998): 54–70.
- , eds. *Directed Self-Placement: Principles and Practices*. Cresskill, N.J., Hampton Press, 2003.

- Haswell, Richard. "Re: SAT Quote about Writing." Online posting. 4 Nov. 2004. *Archives of WPA-L@ASU.EDU Writing Program Administration* 4 Nov. 2004. 16 Dec. 2004. <<http://lists.asu.edu/archives/wpa-l.html>>.
- . "Post-secondary Entry Writing Placement: A Brief Synopsis of Research." *CompPile*. Ed. Richard Haswell and Glenn Blalock. 2004. 12 Feb. 2005. <<http://comppile.tamucc.edu/writingplacementresearch.htm>>.
- Haswell, Richard and Susan Wyche-Smith. "Adventuring into Writing Assessment." *College Composition and Communication* 45.2 (1994): 220–36.
- Huot, Brian. "Toward a New Theory of Assessment." *College Composition and Communication* 47 (1996): 549–566.
- McKendy, Thomas. "Locally Developed Writing Tests and the Validity of Holistic Scoring." *Research in the Teaching of English* 26 (1992): 149–166.
- Meyers, Russell J. "'Take-Home' Placement Tests: A Preliminary Report." *College English* 44 (1982): 506–10.
- Reynolds, Erica. "The Role of Self-Efficacy in Writing and Directed Self-Placement." Gilles and Royer, *Self-Placement: Principles* 73–103.
- Smith, William. "Teachers Informing Placement Testing: A New Method Based on Teacher Expertise." Conference on College Composition and Communication. Cincinnati, April, 1992.
- Stiggins, Richard J. "A Comparison of Direct and Indirect Writing Assessment Methods." *Research in the Teaching of English* 16 (1982): 101–14.
- Stitt-Bergh, Monica. "Re: SAT or ACT Correlation with Direct Writing." Online Posting. 16 Dec. 2004. *Archives of WPA-L@ASU.EDU Writing Program Administration*. 16 Dec. 2004. <<http://lists.asu.edu/archives/wpa-l.html>>.
- Yancey, Katherine. *Portfolios in the Writing Classroom: An Introduction*. Urbana, IL: NCTE, 1992.
- White, Edward M. "An Apologia for the Timed Impromptu Essay Test." *College Composition and Communication* 46 (1995): 30–45.
- . "Language and Reality in Writing Assessment." *College Composition and Communication* 41 (1990): 187–200.
- . "Re: SAT/ACT Very High/low Score Placement Correlation." Online posting. 26 Oct. 2004. *Archives of WPA-L@ASU.EDU Writing Program Administration*. 26 Oct. 2004. <<http://lists.asu.edu/archives/wpa-l.html>>.
- . "Re: Synopsis of Research on Writing Placement." Online posting. 22 Jan. 2005. *Archives of WPA-L@ASU.EDU Writing Program Administration*.. 22 Jan. 2005. <<http://lists.asu.edu/archives/wpa-l.html>>.
- Willard-Traub, Margaret, Emily Decker, Rebecca Reed, and Jerome Johnston. "The Development of Large-Scale Portfolio Placement Assessment at the University of Michigan: 1992–1998." *Assessing Writing* 6 (1999): 41–84.

