ACT Research Report Series 2000-7

Posttesting Students to Assess the Effectiveness of Remedial Instruction in College

Richard Sawyer

Jeff Schiel

ACT

June 2000

For additional copies write:
ACT Research Report Series
PO Box 168
Iowa City, Iowa 52243-0168

Posttesting Students
to Assess the Effectiveness of Remedial Instruction in College

Richard Sawyer and Jeff Schiel

Abstract

At many postsecondary institutions, there are two levels of first-year courses: a "standard" course in which most students enroll, and a "remedial" course for academically underprepared students. This paper is concerned with determining whether taking a remedial course increases the cognitive skills that students need to succeed in a standard course. The paper describes some indicators based on data from posttesting students (i.e., testing them after they have completed a remedial course). The paper also contains a discussion of how prior selection and measurement error in the initial placement test and the posttest affect the indicators. An example is provided to illustrate the indicators and the effects of prior selection and measurement error.

Acknowledgements

The authors thank David Woodruff for his many helpful comments on earlier drafts of this paper, and James Sconing for his advice on statistical issues.

Posttesting Students
to Assess the Effectiveness of Remedial Instruction in College

A typical and important use of college entrance tests is course placement (i.e., matching students with instruction appropriate to their academic preparation). For example, students whose academic skills are insufficient for them to be successful in a standard first-year mathematics course (e.g., college algebra) might, on the basis of their test scores and other characteristics, be advised or required to enroll in a lower-level mathematics course (e.g., elementary algebra). On the other hand, unusually well prepared students might be encouraged to enroll in a higher-level course (e.g., calculus). Of course, what constitutes standard, lower-level, and higher-level courses varies from institution to institution.

At many postsecondary institutions, there are two levels of first-year courses: a "standard" course in which most students enroll, and a "remedial" course for students who are not academically prepared for the standard course. Often, "remedial" courses do not carry credit toward satisfying degree requirements. At many institutions, the lower-level course is given other names, such as "college-preparatory," "compensatory," "developmental," or "review," and may include important supplemental content, such as instruction in study skills and personal counseling. Carriuolo (1994) articulated differences in the meanings of the terms "remedial" and "developmental." McCabe (in press) pointed out that the term "remedial," while commonly used by policy makers and the general public, may have negative connotations to students and faculty. Furthermore, some institutions offer courses that require more knowledge and skills than the lowest-level courses, but less than the standard courses. For simplicity in this discussion, however, only a single lower-level course is considered, and it is designated "remedial."

The percentage of postsecondary institutions with some form of placement and remedial instruction is about 90% ("Colleges and Universities Offering Remedial Instruction," 1994). A survey by the American Council on Education (1996) found that about 17% of students in community colleges and about 11% of students in public four-year institutions take remedial courses. Another survey, by the National Center for Education Statistics (1996), found that 29% of all first-year students take remedial courses. Whichever result is more accurate, it is clear that a significant percentage of college students are involved in remedial course work, according to the standards of the institutions in which they are enrolled.

Evaluating Remedial Courses

Before remedial courses can be designed and implemented at an institution, administrators must decide to allocate resources to these tasks. This decision is often difficult, because the required resources may be substantial and could be allocated to other worthy programs or projects. From an institution's perspective, remedial instruction must improve students' academic skills and knowledge sufficiently for them to succeed in standard courses; otherwise, the institution's resources will have been used ineffectively.

Students are similarly concerned with the effectiveness of remedial instruction. For example, a student who is placed in a remedial course may incur additional tuition expense beyond what he or she initially anticipated, and may not complete her or his degree as quickly as planned. Delayed degree completion can also have negative financial consequences; if the student does not begin full-time employment when planned, then potential income will be lost. If the student does not later successfully complete the standard courses (or, at least, the remedial courses), then the investment of time and money will have been wasted.

Aspects of Evaluating Remedial Courses

Noncognitive variables are important when evaluating the effectiveness of a course placement system. Administrative data (e.g., the number of students who are tested, exempted from testing, or who file appeals of placement decisions) can, when monitored over time, signal changes in how well the system is working. Data on affective characteristics (e.g., do students believe the advice they have been given is appropriate? Do students think that they have been treated well by the faculty and staff who operate the system? Do the faculty and staff themselves believe that their needs are considered and that their skills are effectively used?) can also, when monitored over time, alert staff to important changes in the system. Using standardized survey forms, administrators can also compare their students' opinions to those of students at similar institutions (ACT, 2000a).

Financial considerations are another important characteristic. Murtuza and Ketkar (1995) studied a course placement and advising program at an urban university for its effect on retention and for its cost-effectiveness. Cost-effectiveness was determined by a *break-even analysis*: Does the program increase retention enough so that the resulting extra tuition income offsets the program's cost? When they compared recent retention rates to those observed before the program began, Murtuza and Ketkar found that the program was cost-effective, but their analysis of data from only recent years produced an inconclusive result. Murtuza and Ketkar also found that a centralized program (in which staff were hired and assigned to work specifically on course placement and advising) was more cost-effective than a decentralized program (in which these functions were assigned to faculty members).

Prediction plays a significant role in evaluating the effectiveness of remedial courses. Do students who successfully complete a remedial course eventually succeed in the standard course? Do they stay in school and complete their programs? What is their academic achievement, as measured by course grades, overall GPA, or other criteria? Does taking a remedial course improve students' chances of success with respect to these criteria, as compared to their chances of success if they did not take the remedial course? Hodges (1998) showed how data from ACT's Underprepared Student Follow-Up Report (ACT, 2000b) can be used to monitor the academic success of students who take remedial courses.

Posttesting

All the preceding issues need to be addressed in evaluating the overall effectiveness of remedial course placement systems, but they are beyond the scope of this paper. This paper is concerned with the narrower issue of determining whether a remedial course achieves its basic goal, to teach the cognitive skills students need to succeed in the standard course. One method for studying students' educational growth is to posttest them with an equated alternate form of the same test used to place them into the remedial course. If:

- the placement test score is a valid measure of the knowledge and skills required for success in the standard course;
- the remedial course is effective in teaching students the required knowledge and skills; and
- an alternate form of the placement test is administered at the end of the remedial course,

then students' test scores obtained at the end of the remedial course should exceed their scores obtained at the beginning of the course. The purpose of this paper is to illustrate how this design can be used to assess students' increase in cognitive skills, and to consider some of its limitations.

The term *posttesting* can be distinguished from the term *retesting*: *Retesting* involves repeating a placement test (the pretest), because of some reason that would lead one to believe that the test score obtained initially was not a valid measurement of a student's knowledge and skills. Students often retest to achieve a score sufficient for placement into a standard course. In retesting, students have not yet taken a course, and short time intervals occur between administrations of the test. *Posttesting*, on the other hand, refers to testing that occurs after the remedial course has been taken, the pretesting having been used to place the student in the remedial course in the first place. Posttesting may also be repeated until the cutoff score for placement into the standard course is achieved. Using assumed utilities in a decision theory model, van der Linden (1998) derived optimal cutoff scores for the placement test and the posttest.

Educational growth is often measured by subtracting each student's pretest score from the posttest score. The distribution of the resulting difference scores can then be summarized (e.g., by the mean and variance). Difference scores are often negatively correlated with pretest scores (Rogosa, Brandt, and Zimowski, 1982). This phenomenon could, in principle, result from a weak or even negative relationship between pretest true scores and difference true scores (caused, for example, by ceiling effects on the test or by differential instruction to students with different pretest scores). More typically, this phenomenon occurs because of measurement error, irrespective of relationships among true scores. Consider, for example, a particular individual with a given pretest true score and a given posttest true score: Whatever these true scores may be, measurement error in the observed pretest score will be negatively correlated with measurement error in the observed difference score. This phenomenon can be expressed statistically in the classical measurement model $X_i = T_i + \varepsilon_i$, where X_i is the observed score, T_i is the true score, and ε_i is the measurement error on testing occasion *i*, (*i*= 1,2): $Corr[X_2 - X_1, X_1 \mid T_1, T_2] = -1/\sqrt{2}$. (The appendix contains a derivation of this result.)

When students are explicitly selected for a treatment on the basis of their low pretest scores (e.g., through a cutoff on a placement test), they are being selected partly on the basis of low true scores, but also partly on the basis of negative measurement errors. Given the previous discussion, even if all these students had zero difference in their pretest and posttest true scores (i.e., no true growth), they would likely have positive observed difference scores. (In other contexts, this phenomenon is sometimes referred to as a "regression effect" or "regression toward the mean.") Thus, a positive observed score difference for students selected on the basis of their low pretest scores needs to be interpreted with respect to prior selection and measurement error.

This paper contains a discussion of indicators that are based on data collected from posttesting students. The discussion includes an analysis of the effects of measurement error due to selection using a cutoff score on the pretest. By taking into account these effects, one can interpret the indicators more accurately.

The indicators discussed here pertain only to the *apparent* effectiveness of remedial instruction: Do students who take a remedial course increase the skills they need to succeed in the standard course? The indicators do not pertain to causation (i.e., whether the increase in skills is, in fact, due to the remedial instruction). It is conceivable that students could acquire the necessary skills even if they did not take a remedial course. To investigate causation, one would need to obtain posttest data from a control group of students who do not receive remedial instruction, compute indicators for them, and compare the indicators to those of the students who did receive remedial instruction. Ideally, the control group would consist of students who needed remedial instruction (as indicated by their pretest scores), but did not receive it. Clearly, this sort of experimental research is rarely done. Alternatively (but less plausibly), one could assume that students who need remedial instruction, but do not receive it, do not experience any growth in their skills.

Other Ways to Measure Growth

Authors have proposed and studied other kinds of measures of growth. Rogosa, Brandt, and Zimowski (1982), for example, argued that deficiencies with difference scores are related to the amount of data collected, not necessarily to the method of measurement. They advocated collecting measurements at more than two points in time, and fitting growth curves to the resulting data.

Maris (1998) described "covariance adjustment" methods for making inferences about the effectiveness of remedial instruction. One application of covariance adjustment involves attempting to mitigate biases due to nonrandom selection by adjusting the mean difference scores of the treatment and control groups on the basis of other variables (the "covariates"). Of course, the benefit of this approach depends very much on being able to identify and collect data on the right covariates.

In one important application of covariance adjustment, the pretest score is used as a covariate. The statistical relationship between the posttest score and the pretest score among subjects in the control group is extrapolated to the treatment group. The difference between the average observed posttest score of the treatment group and the average extrapolated posttest score from the control group is taken as an indicator of the average treatment effect for the treatment group. Analogously, one could estimate the average treatment effect for the control group as the difference between the average extrapolated posttest score from the treatment group and the average observed posttest score of the control group. One could proceed further to estimate an average treatment effect for all students, thereby obtaining an indicator related to causation. We have chosen to study difference scores, rather than use the pretest scores as covariates, because difference scores are more easily understood by many users of standardized tests and by policy makers.

The dependent variables ("posttest scores") in covariance models need not actually be on the same scale as the pretest scores. For example, one could use successful completion of relevant standard courses as dependent variables, and study the relationship between students' probability of success in these courses, their pretest scores, and their enrollment in (or completion of) a remedial course.

A Basic Indicator of Remedial Effectiveness

One basic indicator of the effectiveness of the remedial course is the proportion of students who complete it:

$$I^{(1)} = \frac{N_1}{N_0} , \tag{1}$$

where N_0 is the number of students who enroll in the remedial course, and N_1 is the number of students who complete the remedial course. In the following discussion, we assume that $N_1 > 0$, and so $I^{(1)} > 0$.

McCabe (in press) argued that successfully completing a remedial course is a positive outcome, irrespective of any follow-up course work. The reason is that even if students do not continue their postsecondary education, those who have successfully completed remedial courses tend to find employment in occupations that pay substantially more than the minimum wage.

Because students drop out of a remedial course for academic as well as for non-academic reasons, $I^{(1)}$ is not purely an indicator of educational achievement. When remedial instruction is viewed as one component of a larger course placement system, however, $I^{(1)}$ can be thought of as one indicator of the apparent effectiveness of the system.

Note that N_1 (and, therefore, $I^{(1)}$) is measured without error. With respect to other sources of variation, however, the statistical distribution of N_1 likely is related to the pretest scores, because students with higher pretest scores are more likely to complete the remedial course. Standard techniques, such as logistic regression, can be used to model the conditional distribution of N_1, given the pretest scores.

Indicators Based on Posttesting

We now propose to describe various indicators based on posttesting and to investigate the effects of prior selection and measurement error on their statistical properties. With respect to prior selection, we make the simplifying assumption that students are assigned to the remedial course if their pretest scores are less than some cutoff, K. With respect to the measurement error issue, *the only random quantities considered in this paper are measurement errors.* Furthermore, all probabilistic statements are conditioned *on the particular students who take the pretest and on their pretest true scores.* The conditional properties could be extended to unconditional properties, using additional assumptions about the distribution of true scores, but that is beyond the scope of the paper.

One indicator of the educational effectiveness of the remedial course is the proportion of students who complete the remedial course and whose posttest scores meet or exceed the cutoff, among all students who enroll in the remedial course:

$$I^{(2)} = \frac{N_2}{N_0} , \tag{2}$$

where (as before) N_0 is the number of students who enroll in the remedial course, and where N_2 is the number of students who complete the remedial course and who obtain a posttest (observed) score greater than or equal to K.

One can also calculate the proportion of students whose posttest scores equal or exceed K, *among those who complete the remedial course:*

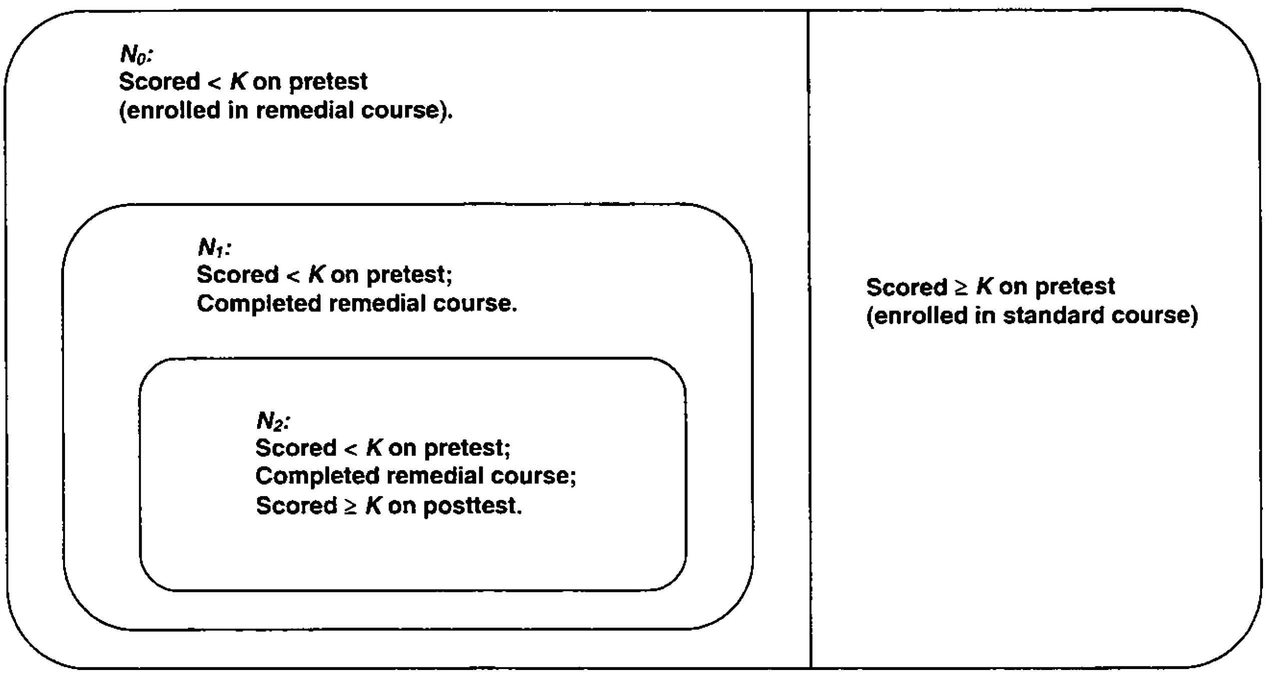
$$I^{(3)} = \frac{N_2}{N_1}\ , \tag{3}$$

where N_1 is the number of students who complete the remedial course, and where N_2 is as defined in the previous paragraph.

The counts $N_0, N_1,$ and N_2 are illustrated in Figure 1 below. Recall that $I^{(1)} = \frac{N_1}{N_0}$, $I^{(2)} = \frac{N_2}{N_0}$, and $I^{(3)} = \frac{N_2}{N_1}$. Note that $0 \le I^{(2)} \le I^{(3)} \le 1$ and $0 \le I^{(2)} \le I^{(1)} \le 1$.

FIGURE 1

Classification of Students According to Score on Pretest/Posttest and Completion of Remedial Course



Ideally, all students who complete the remedial course obtain posttest scores of K or higher, indicating that they all have a reasonable chance of success in the standard course. In this ideal situation, $I^{(3)} = 1$; short of this ideal, the statistical distribution of $I^{(3)}$ depends on the posttest true scores and on the statistical distribution of the measurement errors. With simple assumptions, properties of the distribution of $I^{(3)}$, given that there has been no increase in true scores (and, therefore, no real gain in educational achievement) can be estimated.

Suppose student j completes the remedial course and, therefore, has a posttest score. Consider the classical measurement model $X_{i,j} = T_{i,j} + \varepsilon_{i,j}$, where $X_{i,j}$ is the observed score, $T_{i,j}$ is the true score, and $\varepsilon_{i,j}$ is the measurement error for student j on testing occasion i; and where the measurement errors $\varepsilon_{i,j}$ are independent with mean 0 and variance σ_ε^2. If we make the additional assumption that the measurement errors are normally distributed, then it is possible to estimate the expected value of $I^{(3)}$, given that the true scores have not changed:

$$\hat{E}\left[I^{(3)} \mid T_{2,j} = T_{1,j} \right] = \frac{1}{N_1} \sum_j \left\{ 1 - \Phi\left(\frac{K - \hat{T}_{1,j}}{\sigma_\varepsilon} \right) \right\} \tag{4}$$

where Φ is the standard normal distribution function, σ_ε is the standard error of measurement, and $\hat{T}_{1,j} = \overline{X}\,(1 - r_{XX}) + r_{XX} X_{1,j}$. In the expression for $\hat{T}_{1,j}$, $\overline{X}$ is the mean pretest score in the entire pretested group, and r_{XX} is the reliability of the pretest score in the entire pretested group. A derivation of this result is contained in the appendix.

Of course, in practical applications (where scores are bounded), the normality assumption about the error terms can not be true. Nevertheless, this assumption is commonly made (for example, in relating standard errors of measurement or conditional standard errors of measurement to confidence intervals for true scores).

A value of $I^{(3)} \geq \frac{1}{N_1} \sum_j \left\{ 1 - \Phi\left(\frac{K - \hat{T}_{1,j}}{\sigma_\varepsilon} \right) \right\}$ would suggest that the proportion of remedial course completers who achieved scores high enough for them to enroll in the standard course was greater than would be expected from the effects of random measurement error alone. The summation is over students who complete the remedial course (and who, by assumption, have posttest scores).

Note that if all students' true scores on the pretest were equal to K (indicating that they were all minimally adequately prepared for the standard course) and if they nonetheless enrolled in and completed the remedial course, then we would require $I^{(3)} \geq 0.50$. If the students' pretest scores were all much lower than K, then the probability that their posttest scores meet or exceed K due solely to measurement error would be much less than 0.50. In this case, the conditional expected value of $I^{(3)}$, given no gain in true scores, would also be much less than 0.50.

By a similar argument, $\frac{1}{N_0} \sum_j \left\{ 1 - \Phi\left(\frac{K - \hat{T}_{1,j}}{\sigma_\varepsilon} \right) \right\}$ is an estimate of the conditional expected value of $I^{(2)}$, given that there has been no increase in true scores. Again, the summation is over students who complete the remedial course (and who, by assumption, have posttest scores).

Average Gain

A more modest indicator of effectiveness is whether, on average, students' test scores improved at all after they finished the remedial course:

$$I^{(4)} = \bar{X}_2 - \bar{X}_1 , \qquad (5)$$

where $\overline{X}_1$ is the average score on pretesting, and $\overline{X}_2$ is the average score on posttesting, of the N_1 students who completed the remedial course. Interpreting $I^{(4)}$ is more complicated than interpreting $I^{(2)}$ or $I^{(3)}$, because $I^{(4)}$ depends on the pretest scores themselves, as well as on the fact that students were selected on the basis of their pretest scores. Given the particular examinees tested, their true scores, and their selection on the basis of observed pretest scores, the conditional expected value of $I^{(4)}$ is:

$$E\left[\overline{X}_2 - \overline{X}_1 \,\middle|\, T_{i,j};\quad X_{1,j} < K;\quad (i=1,2;\ j=1,\ldots,N)\right]$$

$$= \overline{T}_2 - \overline{T}_1 + \frac{\sigma_\varepsilon}{N\sqrt{2\pi}} \sum_j \frac{\exp\left[-(K - T_{1,j})^2 / (2\sigma_\varepsilon^{\,2})\right]}{\Phi[(K - T_{1,j})/\sigma_\varepsilon]} \qquad (6)$$

$$= \overline{T}_2 - \overline{T}_1 + G(K, \sigma_\varepsilon, T_{1,1}, \ldots, T_{1,N})$$

where $\overline{T}_i = \frac{1}{N}\sum_j T_{i,j}$, and where G is a function of the cutoff score K, the standard error of measurement σ_ε, and the true scores of the students on the first administration of the test. For simplicity, we have deleted the subscript on the sample size N_1; (i.e., $N = N_1$). A derivation of this result can be found in the appendix.

Although G is a function of the latent variables $T_{1,1}, \ldots, T_{1,N}$, an estimator $\hat{G}$ can be constructed from the reliability and mean pretest score in the unselected group, as described previously. To be indicative of effective remedial instruction, an observed value of $I^{(4)}$ should exceed $\hat{G}$.

Given the particular examinees tested, their true scores, and their selection on the basis of observed pretest scores, the conditional variance of $I^{(4)}$ is:

$$Var\left[\bar{X}_2 - \bar{X}_1 \mid T_{i,j};\ X_{1,j} < K;\ (i = 1,2;\ j = 1,\ldots,N)\right] = \frac{3\sigma_\varepsilon^2}{2N} +$$

$$\frac{1}{N^2}\sum_j \left\{ \frac{\frac{\sigma_\varepsilon^2}{\sqrt{\pi}} Sign\left(K - T_{1,j}\right) H\left[\left(K - T_{1,j}\right)^2 / (2\sigma_\varepsilon^2)\right]}{\Phi\left[\left(K - T_{1,j}\right)/\sigma_\varepsilon\right]} - \frac{\frac{\sigma_\varepsilon^2}{2\pi} exp\left[-\left(K - T_{1,j}\right)^2 / \sigma_\varepsilon^2\right]}{\Phi^2\left[\left(K - T_{1,j}\right)/\sigma_\varepsilon\right]} \right\}, \qquad (7)$$

where $Sign(x) = 1$, if $x \geq 0$ and $Sign(x) = -1$, if $x < 0$; where $H(x) = \int_0^x z^{1/2} e^{-z} dz$ is the "incomplete gamma function"; and where the other symbols are as defined previously. The appendix contains a derivation of this result.

Note that as $K \to \infty$, $H\left[\left(K - T_{1,j}\right)^2 / (2\sigma_\varepsilon^2)\right] \to \sqrt{\pi}/2$, and so the conditional variance approaches $2\sigma_\varepsilon^2/N$. With an appropriate estimate of the variance (e.g., by estimating the true scores), one could construct an approximate confidence interval for the conditional mean (6). The confidence interval would, like the other quantities, be conditional on the true scores of the particular sample of students.

More Complex Investigations

The indicators described here are all simply averages of various types. One could undertake more sophisticated investigations by modeling individual students' completion of the remedial course, successful posttesting, or gain scores as functions of relevant covariates (e.g., age, work responsibilities, other remedial courses taken). Such analyses could provide an institution with the capability to identify particular types of students who are more or less successful in the remedial course. The information from these analyses could suggest modifications in the remedial courses that would benefit particular groups of students.

Example

Pre- and posttest data were obtained from students enrolled in 9 two-year institutions and 10 four-year institutions in a state postsecondary education system. In this system, placement decisions are made using one of two screening tests, followed, in certain cases, by the administration of a placement test. The two screening tests are the ACT Assessment and the SAT, both of which are commonly used in making postsecondary admissions decisions. The ACT Assessment has four subject area tests (English, Mathematics, Reading, and Science Reasoning) and is used by postsecondary institutions in making both admissions and placement decisions. Scores on each of the four subject area tests range from 1 to 36. The SAT, in comparison, has Mathematical and Verbal tests; scores on each test range from 200 to 800. Its use as a placement instrument is not as widespread as is that of the ACT Assessment.

Students whose scores on the ACT Assessment or on the SAT meet or exceed certain cutoffs are placed directly in standard English and mathematics courses. Those scoring below the cutoffs, on the other hand, are administered the COMPASS placement tests. COMPASS, a computer adaptive testing system developed by ACT, measures students' academic skills and knowledge in mathematics, reading, and writing. Scores are reported on a scale that ranges from 1 to 99, and are interpreted as estimates of the percentage of items in a subject area item pool that a student can answer correctly. In the placement system in this example, placement decisions pertaining to standard and remedial mathematics courses are made using the COMPASS Algebra test. The Reading test is used to make placement decisions for courses requiring a substantial amount of reading (e.g., history and political science courses), and the Writing Skills test is used to make placement decisions for writing courses. In this system, students are usually permitted to take the COMPASS pretest only once, although individual institutions may make exceptions to this rule. When retesting does occur, the time interval between testings may be very brief, due to the administration mode of the test. In this example, the numbers of pretested students were: 4,434 (Algebra), 8,563 (Reading), and 6,281 (Writing Skills).

Students who score at or above the COMPASS cutoff are placed into a standard-level course, and are not posttested. Students who score below the COMPASS cutoff are placed into a remedial course. Some of these students do not complete the remedial course. Those who do complete the remedial course must take COMPASS as a posttest and meet or exceed the cutoff before they are permitted to enroll in the standard course.

It is possible that some students, after learning that they scored below the COMPASS pretest cutoff, delayed their enrollment in the remedial course for a term. Students delaying enrollment in a remedial course beyond one term could have other educational experiences, whether occurring in a classroom setting or elsewhere, that make it hard to isolate the effect of the remedial course itself. Moreover, the minimum length of remedial instruction was two months (during summer school). Therefore, the data in this example were restricted to students with at least two months, but no more than eight months, between pre- and posttesting.

Because a few students took the COMPASS pretest more than once (in order to earn a score of K or higher), the highest score of those who pretested multiple times was retained for analysis. Some students also took the COMPASS posttest more than once. We elected to use the first posttest scores of such students, thereby constructing indicators of the initial effectiveness of the remedial courses.

Some pretested students might not have enrolled in the institution at which they pretested. This might occur, for example, if a student's pretest score was lower than the cutoff, and if the student decided not to enroll at all in the institution, rather than take remedial courses. Such students would not, of course, have any posttest scores. In the data available for this analysis, there was no formal indication that a student completed a remedial course; we could only infer that a student completed a remedial course by the presence of a posttest score. Students who pretested, but never enrolled in the institution are therefore counted as not having completed the remedial course (which would artificially lower the indicators $I^{(1)}$ and $I^{(2)}$).

The accuracy of the results also depends in another way on the representativeness of the data within institutions. If some institutions, for whatever reason, did not send some of their COMPASS posttest data to ACT, then the indicators $I^{(1)}$ and $I^{(2)}$ would be artificially lowered. In this example, we eliminated from the analysis institutions that did not report any posttest data within the two- to eight-month window. It is possible, however, that some institutions reported only some of their posttest data and therefore, the data from these institutions are incomplete.

Results

Table 1 on the following page shows summary statistics (mean and standard deviation of COMPASS scores, and sample size) for the data. The statistics are presented separately for students whose pretest scores exceeded the cutoff established for the subject area (and who therefore enrolled in a standard course); for students whose pretest scores were lower than the cutoff (and who therefore enrolled in a remedial course); and for the total group of students.

TABLE 1

Mean and Standard Deviation of COMPASS Scores (and Number of Students), by Performance Category

		COMPASS test		
Student group		**Algebra**	**Reading Skills**	**Writing Skills**
Scored ≥ *K* on pretest; enrolled in standard course	Mean	46	86	79
	Std. dev.	15	7	14
		(N=1,857)	(N=5,979)	(N=4,704)
Scored < K on pretest; enrolled in remedial course	Mean	21	61	30
	Std. dev.	4	11	13
		(N=2,577)	(N=2,584)	(N=1,577)
All pretested students	Mean	31	79	66
	Std. dev.	16	15	25
		(N=4,434)	(N=8,563)	(N=6,281)

The mean scores for Algebra, Reading Skills, and Writing Skills were 31, 79, and 66, respectively. The corresponding national mean scores for all COMPASS-tested students were 37, 77, and 61, respectively (ACT, 1998). The differences between the two sets of means are not large in a practical sense, given the corresponding national standard deviations of 20, 17, and 28.

As the total group sample sizes in the bottom row of Table 1 suggest, most students took more than one test. Of 10,591 students administered any test, about 37% took all three tests, and about 27% took two tests.

TABLE 2

Summary of Indicators of Effectiveness of Remedial Instruction, By Remedial Course (COMPASS test)

	Remedial course (COMPASS test)		
Indicator	**Mathematics (Algebra)**	**Reading (Reading Skills)**	**Writing (Writing Skills)**
$I^{(1)} = N_1 / N_0$ Percentage of students who completed remedial course	22	45	27
$I^{(2)} = N_2 / N_0$ Percentage of students who completed remedial course and who scored $\geq K$ on posttest	21 (17)	32 (13)	24 (6)
$I^{(3)} = N_2 / N_1$ Percentage of remedial course completers who scored $\geq K$ on posttest	97 (17)	71 (9)	90 (7)
$I^{(4)} = \overline{X}_2 - \overline{X}_1$ Mean score gain of remedial course completers (Scale=1 to 99)	33 (3)	17 (1)	41 (1)

Note: The numbers in parentheses are estimated expected values of the indices assuming no change in true scores from pretest to posttest. The estimates incorporate effects due to selection and measurement error only.

Table 2 summarizes the effectiveness indicators, by remedial course and the COMPASS test used to place students in the course. (Indicators $I^{(1)} - I^{(3)}$ are reported as percentages.) The percentage of these students who completed the remedial course (indicator $I^{(1)}$) ranged from 22% in remedial mathematics courses to 45% in remedial reading courses. Such results could occur for a number of reasons. For example, the comparatively low completion rate for remedial mathematics and writing courses could indicate that these courses are more difficult than the other courses. Alternatively, this result could occur if the students who took the remedial mathematics and writing courses had other characteristics (e.g., a native language other than English in the case of remedial writing) that were related to lower retention rates.

There is a wide divergence of findings on the percentage of students nationally who successfully complete remedial courses (the indicator $I^{(1)}$). According to a study by the National Center for Education Statistics (1996), about 75% of students complete remedial courses of all types. According to a recent study by McCabe (in press), however, only about 43% of students who enroll in remedial courses successfully complete them. According to H. R. Boylan (personal communication, July 25, 1999), the completion rate can sometimes be as low as 14% in remedial mathematics courses. The divergence of these findings may be due partly to differing definitions for successfully completing a remedial course.

The percentage of students who completed the remedial course and who scored at or above K (indicator $I^{(2)}$) ranged from 21% (mathematics) to 32% (reading). When only the scores of the completers are considered, the percentage of students scoring at or above K (indicator $I^{(3)}$) increases considerably, ranging from 71% (reading) to 97% (mathematics).

It is interesting to note that remedial mathematics courses were much more successful with respect to $I^{(3)}$ (97%) than with respect to $I^{(2)}$ (21%). This suggests that although some students may, for whatever reasons, decide to drop out of remedial mathematics courses, the students who do complete the courses benefit considerably. Similar, but less dramatic, differences occurred for the other two courses.

Table 2 also contains estimates of the expected percentages of students whose posttest scores would meet or exceed the cutoffs even if there were no change in their true scores. These estimates are shown in parentheses for the rows corresponding to $I^{(2)}$ (all students in remedial course) and to $I^{(3)}$ (remedial course completers). For both indicators, the observed percentage of students whose posttest scores exceeded the cutoffs substantially exceeded the estimated expected percentage under the assumption of no change in true scores.

The final row of Table 2 contains mean difference scores for students who completed a remedial course. The mean differences ranged from 17 (Reading Skills) to 41 (Writing Skills). The estimated expected mean differences assuming no change in true scores ranged from 1 (Reading Skills and Writing Skills) to 3 (Algebra). Ninety-five percent confidence interval half-widths (based, like all other probabilistic quantities in this paper, on measurement error only) for all the estimates were less than 1. The estimated mean differences are far less than the observed mean differences, which suggests that the joint effects of measurement error and selection were small.

One can compute "adjusted mean difference scores" by subtracting from each observed difference score the estimated expected mean differences associated with measurement error. The adjusted mean difference scores were 30, 16, and 40 for the Algebra, Reading Skills, and Writing Skills tests, respectively. Dividing the means of the adjusted mean difference scores by the standard deviations of the estimated pretest true scores for the total group (see Table 1) provides another way to interpret score gains. The resulting adjusted mean difference scores, expressed in standard deviation units, were 1.9 (Algebra), 1.1 (Reading Skills), and 1.6 (Writing Skills).

Discussion

Although it has limitations, posttesting can provide useful information about the effectiveness of remedial courses. We have not surveyed postsecondary institutions to determine their activities in evaluating the effectiveness of their remedial courses, but we believe that the example presented in this paper describes one of the more concerted efforts in this area. In this example, pretest data were available for a reasonably large segment of students, but posttest data were available only for students who completed a remedial course. Although the resulting lack of data limits causal inferences about remedial instruction, one can construct indicators of its effectiveness.

The results suggest that students who completed the remedial courses offered by this group of institutions increased their academic skills. Mean difference scores, even when adjusted for the effects of selection and measurement error, ranged from about 1 to 2 COMPASS standard deviation units. In addition, the results suggest that students completing remedial mathematics and writing courses have a relatively high probability of scoring at or above the posttest cutoffs and, therefore, of being permitted to enroll in the corresponding standard-level courses. These results are consistent with those of McCabe (in press), who found that students who successfully complete remedial courses in community colleges are typically well prepared to take standard-level courses.

Because the data were subject to selection on the pretest, the indicators proposed here potentially can be inflated because of measurement error in the pretest. Fortunately, the results of the example suggest that the indicators were influenced much more by real improvement in students' skills than by artifacts due to prior selection and measurement error. Although the results are based on data from postsecondary institutions in only one state, they should give confidence to practitioners at other institutions that the indicators are useful in practical situations involving operational course placement systems. Moreover, for individuals who want to estimate the size of prior selection/measurement error artifacts in their particular situations, the equations in this paper provide tools to do so.

Of course, to compute the indicators discussed in this paper, an institution must first collect detailed and complete follow-up data on the students who receive remedial instruction. An institution needs to know whether or when each student completes a remedial course, as well as the student's posttest score. Furthermore, the institution must be able to match all these data elements into a single unified record for each student. Ideally, the institution would also posttest a sample of students who did not take remedial courses, thereby permitting estimates of what students learned *because* they took remedial courses.

References

ACT (1998). *COMPASS National Composite Report.* Iowa City, IA: Author.

ACT (2000a). *Evaluation/Survey Services: Assessing Attitudes and Opinions of Postsecondary Students and Alumni.* (Available at www.act.org/ess /index.html.)

ACT (2000b). *ACT Research and Information Services. Underprepared Student Follow-Up Report.* (Available at www.act.org/research/services/upsfur/index.html.)

American Council on Education (1996). *Remedial education: An undergraduate student profile.* Washington D.C.: Author.

Carriuolo, N. (1994, April 13). Why developmental education is such a hot potato. *The Chronicle of Higher Education,* Sec. 2, pp. 1-2.

Colleges and universities offering remedial instruction and tutoring. (1994, April 13). *Education Week,* XIII, No. 29, p. 6.

Hodges, D. Z. (1998). Evaluating placement and developmental studies programs at a technical institute: Using ACT's Underprepared Student Follow-up Report. *Community College Review, 26*(2), 57-66.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, Mass.: Addison Wesley.

Maris, E. (1998). Covariance adjustment versus gain scores--revisited. *Psychological Methods, 3*(3), 309-327.

McCabe, R. H. (in press). *No one to waste.* Washington, D. C.: American Association of Community Colleges.

Murtuza, A., & Ketkar, K. (1995). Evaluating the cost-effectiveness of a freshman studies program on an urban campus. *Journal of The Freshman Year Experience, 7*(1), 7-26.

National Center for Education Statistics (1996). *Remedial education and higher education institutions in fall 1995.* (NCES 97-584). Washington, D.C.: U.S. *Government* Printing Office. (Also available at the NCES WebSite *www.ed.gov/NCES/pubs).*

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92*(3), 726-748.

van der Linden, W. J. (1998). A decision theory model for course placement. *Journal of Educational and Behavioral Statistics, 23*(1), 18-34.

Appendix

Proposition 1: Suppose $X_i = T_i + \varepsilon_i$, $(i = 1, 2)$, where T_i are constants and ε_i are independent, normally distributed measurement errors with mean 0 and variance $\sigma_\varepsilon^2 > 0$. Then

$Corr[X_2 - X_1, X_1 \mid T_1, T_2] \;=\; -1/\sqrt{2}$.

Proof: This result is a consequence of the following relationships:

$Cov[X_2 - X_1, X_1 \mid T_1, T_2] \;=\; Cov[X_1, X_2 \mid T_1, T_2] - Var[X_1 \mid T_1, T_2] \;=\; -\sigma_\varepsilon^2$;

$Var[X_2 - X_1 \mid T_1, T_2] = 2\sigma_\varepsilon^2$; and $Var[X_1 \mid T_1, T_2] = \sigma_\varepsilon^2$.

Proposition 2: Suppose $X_{i,j} = T_{i,j} + \varepsilon_{i,j}, (i = 1, 2; j = 1, \ldots, N_0)$, where $T_{i,j}$ are constants and $\varepsilon_{i,j}$ are independent, normally distributed measurement errors with mean 0 and known variance $\sigma_\varepsilon^2 > 0$. Let $N_2 = \sum_j D_j$, where

$D_j = 1$, if student j completes the remedial course and if $X_{2,j} \geq K$;
$= 0$, otherwise.

(N_0 is the number of students who enroll in the remedial course. N_2 is the number of students who complete the remedial course, and whose posttest scores exceed the cutoff, K.) Then, the maximum likelihood estimator of the conditional mean $E\left[N_2 \mid T_{2,j} = T_{1,j}\right]$ is:

$$\hat{E}\left[N_2 \mid T_{2,j} = T_{1,j}\right] \;=\; \sum_j \left\{ 1 - \Phi\left(\frac{K - \hat{T}_{1,j}}{\sigma_\varepsilon} \right) \right\}, \qquad \text{(A-1)}$$

where Φ is the standard normal distribution function, σ_ε is the standard error of measurement, and $\hat{T}_{1,j} = \overline{X}\left(1 - r_{XX}\right) + r_{XX} X_{1,j}$. In the expression for $\hat{T}_{1,j}$, $\overline{X}$ is the mean pretest score in the entire pretested group, and r_{XX} is the known reliability of the pretest score in the entire pretested group.

Proof: Note that $E\left[N_2 \mid T_{2,j} = T_{1,j}\right] = \sum_j P\left[X_{2,j} \geq K \mid T_{2,j} = T_{1,j}\right] = \sum_j \left\{1 - \Phi\left(\frac{K - T_{1,j}}{\sigma_\varepsilon}\right)\right\}$,

where the summation is over students who complete the remedial course. (For students who do not complete the remedial course, $D_j = 0$, and this event is observed without measurement error.) A maximum likelihood estimator for the true score $T_{1,j}$ is $\hat{T}_{1,j} = \bar{X}\left(1 - r_{XX}\right) + r_{XX} X_{1,j}$, where r_{XX} is the reliability of the pretest score in the entire pretested group, and $\bar{X}$ is the mean pretest score in the pretested group (Lord & Novick, 1968). Therefore, $1 - \Phi\left(\frac{K - \hat{T}_{1,j}}{\sigma_\varepsilon}\right)$ is a maximum likelihood estimator for $1 - \Phi\left(\frac{K - T_{1,j}}{\sigma_\varepsilon}\right)$.

Estimates of the expected values of the indicators $I^{(2)} = \frac{N_2}{N_0}$ and $I^{(3)} = \frac{N_2}{N_1}$ can then be obtained by dividing (A-1) by N_0 and N_1, respectively.

Proposition 3: Suppose $X_{i,j} = T_{i,j} + \varepsilon_{i,j}, (i = 1, 2; j = 1, \ldots, N)$, where $T_{i,j}$ are constants and $\varepsilon_{i,j}$ are independent, normally distributed measurement errors with mean 0 and variance $\sigma_\varepsilon^2 > 0$.

Let $\overline{X}_i = \frac{1}{N}\sum_{j=1}^{N} X_{i,j}$ and $\overline{T}_i = \frac{1}{N}\sum_{j=1}^{N} T_{i,j}$, $(i = 1,2)$, and let K be a cutoff score. Then

$$E\left[\overline{X}_2 - \overline{X}_1 \mid T_{i,j};\quad X_{1,j} < K;\quad (i = 1,2;\ j = 1,\ldots,N)\right] =$$

$$\overline{T}_2 - \overline{T}_1 + \frac{\sigma_\varepsilon}{N\sqrt{2\pi}}\sum_{j=1}^{N}\frac{\exp\left[-(K - T_{1,j})^2/(2\sigma_\varepsilon^2)\right]}{\Phi[(K - T_{1,j})/\sigma_\varepsilon]} ,$$

where Φ is the standard normal distribution function.

Proof: Suppose W ~ $n(0,\sigma^2)$, and let h < 0. Then

$$E[W \mid W < h] = \frac{\int_{-\infty}^{h}\frac{w}{\sigma\sqrt{2\pi}}\exp\left[-w^2/(2\sigma^2)\right]dw}{\Phi[h/\sigma]}$$

. Substituting $z = w^2/(2\sigma^2)$ in the numerator,

we obtain:

$$E[W \mid W < h] = \frac{-\left(\sigma/\sqrt{2\pi}\right)\int_{h^2/(2\sigma^2)}^{\infty} e^{-z}dz}{\Phi[h/\sigma]}$$

$$= \frac{-\left(\sigma/\sqrt{2\pi}\right)\exp\left[-h^2/(2\sigma^2)\right]}{\Phi[h/\sigma]} . \qquad \text{(A-2)}$$

Note that if $g(u) = u\exp\left[-u^2\right]$, then $g(-u) = -g(u)$. Therefore, result (A - 2) is also true if $h > 0$.

Now,

$$E[\overline{X}_2 - \overline{X}_1 \mid T_{i,j};\quad X_{1,j} < K;\quad (i = 1,2;\ j = 1,\ldots,N)]$$

$$= \overline{T}_2 - \overline{T}_1 - E\left[\frac{1}{N}\sum_{j=1}^{N} e_{1,j} \,\middle|\, T_{1,m};\ X_{1,m} < K;\ (m = 1,\ldots,N)\right]$$

$$= \overline{T}_2 - \overline{T}_1 - \frac{1}{N}\sum_{j=1}^{N} E[e_{1,j} \mid e_{1,j} < K - T_{1,j}] . \qquad \text{(A-3)}$$

Substituting (A-2) into (A-3), with $h = K - T_{1,j}$, and $\sigma^2 = \sigma_\varepsilon^2$, we obtain

$$E[\overline{X}_2 - \overline{X}_1 | \; T_{i,j}; \quad X_{1,j} < K; \quad (i = 1,2;\ j = 1,\ldots, N)]$$

$$= \overline{T}_2 - \overline{T}_1 + \frac{\sigma_\varepsilon}{N\sqrt{2\pi}} \sum_{j=1}^{N} \frac{\exp\left[-(K - T_{1,j})^2 / (2\sigma_\varepsilon^2)\right]}{\Phi[(K - T_{1,j})/\sigma_\varepsilon]}. \qquad \text{(A-4)}$$

Proposition 4: Suppose $X_{i,j} = T_{i,j} + \varepsilon_{i,j}, (i = 1, 2;\ j = 1,\ldots, N)$, where $T_{i,j}$ are constants and $\varepsilon_{i,j}$ are independent, normally distributed measurement errors with mean 0 and variance $\sigma_\varepsilon^2 > 0$. Then:

$$Var\left[\overline{X}_2 - \overline{X}_1 \middle| T_{i,j};\ X_{1,j} < K;\ (i = 1,2;\ j = 1,\ldots,N)\right] = \frac{3\sigma_\varepsilon^2}{2N} +$$

$$\frac{1}{N^2}\sum_j \left\{ \frac{\frac{\sigma_\varepsilon^2}{\sqrt{\pi}} Sign\left(K - T_{1,j}\right) H\left[\left(K - T_{1,j}\right)^2 \Big/ (2\sigma_\varepsilon^2)\right]}{\Phi\left[\left(K - T_{1,j}\right)\Big/\sigma_\varepsilon\right]} - \frac{\frac{\sigma_\varepsilon^2}{2\pi} \exp\left[-\left(K - T_{1,j}\right)^2 \Big/ \sigma_\varepsilon^2\right]}{\Phi^2\left[\left(K - T_{1,j}\right)\Big/\sigma_\varepsilon\right]} \right\},$$

where:

$Sign(x) = 1$, if $x \geq 0$ and $Sign(x) = -1$, if $x < 0$;

$H(x) = \int_0^x z^{1/2} e^{-z} dz$ is the incomplete gamma function; and

Φ is the standard normal distribution function.

Proof:

$$Var\left[\overline{X}_2 - \overline{X}_1 \middle| T_{i,j};\ X_{1,j} < K\right] = Var\left[\overline{X}_2 \middle| T_{i,j};\ X_{1,j} < K\right] + Var\left[\overline{X}_1 \middle| T_{i,j};\ X_{1,j} < K\right]$$

$$= \frac{1}{N}\sigma_\varepsilon^2 + \frac{1}{N^2}\sum_j Var\left[e_{1,j} \middle| T_{1,j};\ X_{1,j} < K\right]. \qquad \text{(A-5)}$$

Furthermore,

$$Var\left[e_{1,j} \middle| T_{1,j};\ X_{1,j} < K\right] \;=\; E\left[e_{1,j}^2 \middle| T_{1,j};\ X_{1,j} < K\right] \;-\; \left(\frac{\sigma^2}{2\pi}\right) \frac{\exp\left[-(K - T_{1,j})^2 / \sigma_\varepsilon^2\right]}{\Phi^2[(K - T_{1,j}) / \sigma_\varepsilon]} \qquad \text{(A-6)}$$

with the last term a consequence of (A-2).

Let $W \sim n(0, \sigma^2)$. Then, $E\left[W^2 \,\middle|\, W < h\right] \;=\; \dfrac{\int_{-\infty}^{h} w^2 \phi(w)\, dw}{\Phi[h/\sigma]}$. If $h < 0$, then

$$\int_{-\infty}^{h} w^2 \phi(w)\, dw \;=\; \frac{\sigma^2}{\sqrt{\pi}} \int_{h^2/(2\sigma^2)}^{\infty} z^{1/2}\, e^{-z}\, dz$$

$$= \frac{\sigma^2}{\sqrt{\pi}} \left[\frac{\sqrt{\pi}}{2} \;-\; H\left(\frac{h^2}{2\sigma^2}\right)\right],$$

where $H(t) = \int_0^t z^{1/2}\, e^{-z}\, dz$ is the incomplete gamma function, as previously defined. (Note that $H(\infty) \;=\; \sqrt{\pi}/2$.) If $h > 0$, then

$$\int_{-\infty}^{h} w^2 \phi(w)\, dw \;=\; \frac{\sigma^2}{2} \;+\; \frac{\sigma^2}{\sqrt{\pi}} \int_0^{h^2/(2\sigma^2)} z^{1/2}\, e^{-z}\, dz$$

$$= \frac{\sigma^2}{\sqrt{\pi}} \left[\frac{\sqrt{\pi}}{2} \;+\; H\left(\frac{h^2}{2\sigma^2}\right)\right].$$

Therefore,

$$E\left[W^2 \,\middle|\, W < h\right] \;=\; \frac{\sigma^2/2 \;+\; (\sigma^2/\sqrt{\pi})\, Sign(h)\, H\left(h^2/2\sigma^2\right)}{\Phi\left[h/\sigma\right]}, \qquad \text{(A-7)}$$

where $Sign(x) = 1$, if $x \geq 0$ and $Sign(x) = -1$, if $x < 0$. On substituting the result (A-7) into equation (A-6), with $h = K - T_{1,j}$ and $\sigma^2 = \sigma_\varepsilon^2$, we obtain:

$$Var\left[e_{1,j} \middle| T_{i,j};\ X_{1,j} < K\right] = \frac{\frac{\sigma_\varepsilon^2}{2} + \frac{\sigma_\varepsilon^2}{\sqrt{\pi}} Sign(K - T_{1,j})\, H\left[(K - T_{1,j})^2 / (2\sigma_\varepsilon^2)\right]}{\Phi\left[(K - T_{1,j}) / \sigma_\varepsilon\right]} - \frac{\frac{\sigma_\varepsilon^2}{2\pi} exp\left[\frac{-(K - T_{1,j})^2}{\sigma_\varepsilon^2}\right]}{\Phi^2\left[(K - T_{1,j}) / \sigma_\varepsilon\right]} \quad \text{(A-8)}$$

The formula for the variance follows from substituting (A-8) into (A-5).