

**FROM BLACK BOX TO
PANDORA'S BOX:
EVALUATING REMEDIAL/
DEVELOPMENTAL EDUCATION**

W. Norton Grubb
David Gardner Chair in Higher Education
School of Education
University of California, Berkeley

February 2001

Supported by the Alfred T. Sloan Foundation

For additional copies please contact-
Community College Research Center
Teachers College, Columbia University
439 Thorndike Building
525 W. 120th Street, Box 174
New York, New York 10027
212-678-3091 (telephone)
212-678-3699 (fax)

Abstract

Our conceptions of institutional responsibility for learning have been changing. An older conception is that educational institutions provide a curriculum imparted by teachers through traditional didactic methods monitored by conventional assessments—for example, multiple-choice tests covering the facts and skills of a body of content. Students are responsible for learning this content, and too bad if they fail. A high rate of non-completion is if anything a badge of the institution's honor.

Over time, this conception has been modified by one that places greater responsibility for student success on the institution itself. Accepting this responsibility means identifying and then correcting the many possible reasons for non-completion or failure to learn—in short, providing remedial/developmental education. No longer is it possible to be complacent about high non-completion rates, particularly since open access in community colleges has brought to postsecondary education more under-prepared students—and more lower-income, minority, and immigrant students—whose high dropout rates are both personal tragedies and institutional embarrassments. Thus, the roster of student services has expanded, especially remedial education.

Relatively few evaluations of remedial programs have been conducted, and many existing evaluations are useless because, failing to recognize what the program does, they provide little information about what should be changed to make it more effective. In place of this kind of “black box” evaluation, I recommend a variety of evaluation approaches that can improve information about many different aspects of remediation, including not only its effects but also the instructional methods used, the progress of students, and the ways students are

assigned to remedial programs. I call this a “Pandora’s box” approach because it is designed to open up the black box, to reveal the problems with existing programs, including the potential reasons for their effectiveness or ineffectiveness—and then to improve them.

Adult education is such a vast and varied world—adult basic ed, job training, welfare-to-work, community college—that it is difficult to characterize what happens in this sphere. Although some teachers have developed student-centered approaches to teaching and although some community colleges have established “learning communities” that show considerable promise, by far the most common approach to remedial/developmental education is the approach I have labeled “skills and drills.” This tends to focus on arithmetic procedures, punctuation and vocabulary, math problems of the most contrived sort, and passages from texts that have been simplified for low reading levels. This approach takes place not only in classes identified as remedial; it also emerges in college-level classes that become remedial if the majority of students are not ready for what the instructor considers college-level work. Conventional skills and drills approaches violate all the maxims for good teaching in adult education.

Does remedial/developmental education work? The evidence is sparse: most states and colleges have not yet evaluated their remedial programs. No one knows much about what works and what does not—or why. In this vacuum, it is not helpful to recommend one particular approach to evaluation over others. Too many dimensions of remedial education are poorly understood; investigating them requires several different methods. Each of the following evaluation approaches has the potential to illuminate a different aspect of this difficult problem.

1. Dropout rates from remedial courses need more investigation.

Complex combinations of reasons are responsible, and students themselves cannot articulate why they stay with or leave a particular

program. A combination of qualitative, interview-based studies and quantitative studies might begin to provide evidence for improving remedial courses.

2. We need a more systematic collection of outcome measures, but these measures need to include more than test scores of basic skills. Such measures should include persistence in college and completion of degrees, writing portfolios, and completion of occupational courses.
3. It is important in institution- or state- or national-level studies to have control or comparison groups. One instance when this would be especially useful is where some students are thought to need remediation but do not take such courses.
4. Classroom practices in remedial courses must be observed and described. Otherwise it is difficult to know what might have generated a particular set of outcomes—and therefore what might be changed.
5. If evaluation is to have any influence on classroom practice, it needs to compare different approaches to teaching. Some successes may be replicable and others may not, but understanding them better is a necessary first step to improving the quality of instruction.
6. The “assignment” of students to remedial courses needs to be better understood. The question is whether some students who might benefit do not attend remedial courses—either because the assignment test fails to identify those in need of remediation, or because enrolling in such courses is voluntary. Some consideration of alternative assignment procedures is appropriate—either different basic skills tests, or procedures that incorporate other information and counseling as well as testing.

The expansion of postsecondary education since the 1960s, especially the growth of open-access community colleges, has provided opportunities for some students where none existed before, and the dedication of many colleges and most instructors to their non-traditional students is unmistakable. But dedication and student-centeredness, while necessary, may not be sufficient, so a program of evaluation and improvement is central to improving the performance of students.

TABLE OF CONTENTS

Introduction.....2

Multiple Approaches to Remedial/Developmental Education.....5

Multiple Approaches to Evaluation.....17

An Eclectic Approach to Evaluation.....30

Appendix: The "Assignment Problem" in Developmental Education.....36

Endnotes.....42

References.....46

INTRODUCTION

Our conceptions of institutional responsibility for learning have been changing, albeit slowly and incompletely. An older conception is that educational institutions provide a specified curriculum, imparted by teachers to students through traditional didactic methods monitored by conventional assessments — for example, multiple-choice tests covering the facts and skills of a body of content. Students are responsible for learning this content, and too bad if they fail. A high rate of non-completion is if anything a badge of the institution's honor, and in any event not the responsibility of the institution or its instructors. Like the caricature of the college with a dropout rate of two thirds ("Look to your right; look to your left . . ."), dropout rates are expected to be high — and in many two- and four-year colleges they are, often atrociously so.

But over time this conception of student responsibility has been modified by one that places greater responsibility for success on the institution itself. Accepting such responsibility means identifying and then correcting the many possible reasons for non-completion or failure to learn: providing remedial/developmental education,ⁱ tutoring, counseling, and other forms of student services; providing financial aid for low-income students who might otherwise drop out for financial reasons; and providing child care, transportation, and other social services as necessary.

The existence and growth of remedial/developmental education in both two- and four-year colleges are testimony to a shift toward a greater institutional responsibility for learning and completion. No longer is it possible to be

complacent about high rates of non-completion, particularly since open access (in community colleges) and the expansion of higher education have brought to postsecondary education more under-prepared students — and more lower-income, minority, and immigrant students — whose high dropout rates are both personal tragedies and institutional embarrassments. And so the roster of student services has expanded, especially remedial education.

At the same time, colleges remain wedded to older ideals about their responsibilities. Various factions still emphasize student responsibility — sometimes older faculty, sometimes trustees and administrators wedded to reputations gained through high standards, sometimes the forces urging community colleges to remain "collegiate" rather than more varied in their purposes (e.g., Eaton, 1994), certainly critics deploring the dreadful state of higher education (e.g., Traub, 1994; McGrath and Spear, 1991), and certainly policymakers wanting to reduce funding for remediation. These groups find the expansion of remedial education somewhere between worrisome and abominable. And so in practice most institutions are effectively hybrid in their approach, taking greater responsibility for their students' success but often ambivalent about their efforts to do so. Most postsecondary institutions now provide some form of developmental education or basic skills, but remedial education itself has remained marginal in most institutions, squeezed into the back pages of college catalogues, usually under-funded, taught by part-timers, provided as an afterthought, segregated from the "regular" offerings. In every way, colleges signal that this is not real education.

But remedial/developmental education *is* real education, of the most difficult sort. Under the best circumstances, it tries to do more than simply filter accomplished students from the others; it tries to educate all of them, including those who seem not to have learned much in ten or twelve years of conventional schooling. It requires the *most* skilled instructors — not simply part-timers pressed into service, nor individuals untrained in its special challenges.ⁱⁱ The task is self-evidently difficult, and given a lack of certainty about "what works" in teaching of any sort, the approaches to remedial/developmental education vary enormously — as I illustrate in Section I. Remediation is not simple, and it certainly is not a single kind of program.

Because remedial education has developed as a solution to a particular problem — the lack of educational progress of many students — almost no one views it as valuable in its own right. Instead, it is usually considered instrumental to achieving other goals including increased learning (of basic reading, writing, or math) increased retention, and other measures of educational progress.ⁱⁱⁱ Therefore remedial education should be easy to evaluate because — unlike other forms of education freighted with multiple purposes (Labaree, 1997) — its goals are relatively clear. But there have been relatively few evaluations of remedial programs, and many existing evaluations are quite useless because, as I point out in Section II, they fail to recognize what the program does — and therefore they provide little information about what should be changed to make it more effective. In place of this kind of "black box" evaluation, I recommend in Section III a variety of evaluation approaches that can improve information about many different aspects of remediation, including not only its effects but also the

instructional methods used, the progress of students, and the ways students find themselves in remedial programs. We might call this a "Pandora's Box" approach because it is designed to open up the black box, to reveal the problems with existing programs including the potential reasons for their effectiveness or ineffectiveness — and then to improve them.

In the end, debates about remedial/developmental education involve the most central issues in American schooling, particularly those about equity and about student versus institutional responsibility for outcomes. Is it possible to have comprehensive institutions that include a broad variety of students, coming with very different levels of preparation, and teach them all? Or are our schools and colleges merely sorting mechanisms, sifting the competent from the incompetent through batteries of tests and placements, coursework and assignments, promoting one group and "cooling out" the other? The answers to these questions may be partly ideological, but they are empirical as well — and here lies a role for evaluation.

I. MULTIPLE APPROACHES TO REMEDIAL/DEVELOPMENTAL EDUCATION

For all the debate over remedial education, there is almost no discussion about what it looks like — what goes on in classrooms, whether it appears to be educative in any sense of the word, whether it stands any chance of bringing students up to "college level."^{iv} It is important to understand the variety of activities that march under the banner of remedial education. Otherwise it is easy

to fall into the trap of assuming that developmental education is well defined, and can be readily evaluated like any other program.

There are different approaches to remediation in the various institutions and programs of postsecondary education. These institutions are almost completely independent of one another, with instructors in one unaware of what their peers — who are often literally down the street — are doing, even though their instructional tasks are similar. In community colleges, what is often termed developmental education is usually placed in departments separate from English (or reading and writing) departments. Although the variety of remedial or developmental education is staggering, there is at least some sense of this as a distinct field of instruction, with a journal, an association, conferences, and from time to time some efforts at state-level reform.^v In four-year colleges, the same field is generally referred to as "basic English" or basic instruction, and usually takes Shaughnessy's (1997) *Errors and Expectations* as its starting point; she argued that the errors in the writing of basic writers follow patterns rather than simply being random mistakes, and that a skilled instructor could use these errors to reconstruct writing — an approach that requires a certain kind of student-centeredness, rather than blindly plowing ahead with a standardized program. Since then the field of compositional studies has further elaborated various approaches to basic reading and writing courses (Hull, undated). However, educators in two- and four-year colleges have virtually no contact with one another; even though there are journals and associations to which the two groups might contribute, like *College Composition and Communication*, in practice these are dominated by four-year colleges.

Adult education is a vast chaotic world of programs funded with state and federal money, as well as a bewildering array of local funds (for example, for library-based programs) and charitable donations. It is very difficult to characterize what happens in this sphere since there is so much variation. Because there is so little institutional oversight, small innovative programs coexist with much larger and more conventional programs. I have visited a number of community-based programs that look promising, and it is possible that library-based programs are more constructivist and student-centered in their teaching than the rest of adult education^{vi} — but these cells of innovation often do not know what the rest of adult education is like or how different they are, and there is not enough contact among adult programs to facilitate discussion about different approaches. Conventionally, however, adult education programs at several levels (adult *basic* education distinct from adult *secondary* education) prepare students to pass the GED, a credential of dubious value^{vii} that, because of its conventional multiple-choice format, encourages a "skills and drills" approach to instruction. Often, because of its belief in flexible enrollment and open-entry/open-exit, students in remedial adult ed work independently on programmed texts, working through vast arrays of homonyms and synonyms, of grammar exercises and sentence completion exercises, of short passages read to answer simple factual questions (Grubb and Kalman, 1994). In an era of insistent concern with the higher-order skills for a flexible labor force, the complex area of literacy and communication has been shriveled to grammar and punctuation.

Finally, the equally vast and complex area of job training programs — including welfare-to-work programs — provides remedial education too,

sometimes as a prerequisite to vocational skills training. It is impossible to know how much of this goes on, because providers are not required to distinguish remedial education from other services and because remedial education is usually a local option; but there is general agreement that programs are forced to provide more remediation than they would like. Often, these programs subcontract with adult education to provide remediation, so it is back to skills and drills and programmed workbooks for such clients (Grubb and Kalman, 1994). Sometimes job training programs create remedial labs with computer-based programs — but these programs are invariably just skills and drills conveyed to the computer, with even shorter reading passages because of the small size of the screen, a rigid progression through topics, and a lab "manager" (rather than a teacher) whose jobs is to turn the machines on and off and monitor progress but who is not trained to provide any instruction. Finally, some adult education and job training programs have become enamored with functional context literacy training (Sticht et al., 1987), an approach that uses the materials from a "functional context" (like employment or the military) to teach multiple literacy skills (e.g., 186 "reading to do" skills and 143 "reading to learn" skills, in Sticht, 1979). While the functional context approach can be used in constructivist and student-centered ways, it also lends itself to the most didactic and skills-oriented teaching. Functional context instruction has become orthodoxy in some circles, including workplace literacy programs (Schultz, 1997; Gowan, 1992).

These areas of remediation are remarkably different from one another, with different histories, different students, and different goals. Certain pedagogies emerge in all of them, particularly the behaviorist, didactic, teacher-centered (or,

more often, text-centered) approach I often call skills and drills. But because programs are not in communication with one another, examples of innovation and good practice cannot readily spread from one area to another, so the prospects for reform often look gloomy — particularly in the spheres of adult education and job training.

But there is substantial variation *within* each of these areas as well, and I will illustrate this with examples from community colleges.^{viii} These institutions have certain advantages over the others that provide remedial education. As open access institutions, many of their students come unprepared for college-level work, and so (unlike four-year colleges) the necessity for remedial education is built into their basic structure. Community colleges also pride themselves on being "teaching colleges"; even though this ideal is "honored more in the breach . . . the tradition is there and can be called upon when warranted," as one English instructor described it. Unlike adult education and job training, with their reliance on untrained instructors hired in casual ways for part-time work, instructors in community colleges generally have master's degrees and are hired through painstaking procedures (even though these usually have little to do with the quality of teaching). Although these colleges have come to rely too much on part-time instructors, there is still a commitment to teaching as a career. Of all the postsecondary institutions that offer remedial education, community colleges may have the greatest chance of doing it well.

By far the most common approach to remedial/developmental education within community colleges is the approach I have labeled "skills and drills" (Grubb and Associates, 1999, Ch. 5). This tends to focus on sub-skills — on

arithmetic procedures like multiplication and percentages, on grammar and punctuation and vocabulary, on math "problems" of the most contrived sort and reading passages from texts that have been simplified for low reading levels ("there is nothing to read in these texts" complained one instructor). Occasionally instructors will bring in reading from outside the class — from newspapers, for example — but in a typical heterogeneous remedial class there are few common experiences to use as the basis of more contextualized instruction. Students rarely know one another, because of the common pattern of taking courses almost randomly, and therefore do not serve as resources for one another; mastering "literacy" is an individual responsibility with the teacher as the sole authority, rather than a collective and social activity (Worthen, 1997). This approach to remediation takes place not only in classes identified as remedial, to which students are referred if they score below a certain score on a basic skills test; it also emerges in covert or hidden remediation, which takes place in some "college-level" classes — particularly in English, Business English, or Technical Math — that are converted into remedial classes because the majority of students are not ready for what the instructor considers "college level work" (Grubb and Associates, 1999, Ch. 5). Thus the amount of remediation in most community colleges almost surely exceeds the count of official remedial courses, and is therefore difficult to estimate; according to conventional estimates, the proportion of students needing remediation varies among colleges from 25 percent to 50 percent to 78 percent in Tennessee (Grubb and Kalman, 1994).

Often, remedial instructors use computer programs to supplement their instruction. Invariably, the programs are simply drills transferred to the screen,

with short reading passages followed by questions of fact, fill-in-the-blank exercises, arithmetic drills, conversions of fractions and percentages and simple word problems.^{ix} They typically allow students to move to the next level only when they have passed a short "test" on one subject, so they manage the student's progress carefully — and often create records for the instructor. Often, students work on these programs in large labs — perhaps 60 students — overseen by an instructor or "manager," but this individual typically has neither the time nor the training for instruction: if a student gets stuck, he or she has to go back in the computer program to try to work out the problem, but there is no teaching in the conventional sense of the term. Some of these programs are quite elaborate, covering many different topics, and some are quite expensive; they are often promoted with elaborate claims about teaching "higher-order skills." Unfortunately, the majority of these are simply repackaging and peddling the "skills and drills" model.

The problem with this approach is not just that these classes are deadly, with low levels of student engagement. They also violate all the maxims for good teaching in adult education (Grubb and Kalman, 1994). And their tactic is simply "more of the same": they take students who have not learned well in ten or twelve years of standard didactic instruction, and then put them through an additional 15 weeks of similar instruction. There may be some success stories,^x but overall there is little chance that this dominant pedagogy can be very effective. It is foolish to think that students who have never learned to read for meaning, or who have no real understanding of numerals, can suddenly learn quickly from another round of skills and drills.

In our observations, substantial numbers of community college instructors have come to see didactic and behaviorist methods as unsuccessful, so they develop approaches to teaching — largely through trial and error — that are more constructivist, student-centered, and interpretive (Grubb and Associates, 1999, Ch. 5). They are quite aware that community college students have suffered a great deal of humiliation in their earlier education, as well as a remarkable amount of poor teaching; they are likely to blame urban school districts for the low levels of their students. These instructors are more likely to bring in reading materials from work, or from newspapers and political debates; they tend to spend considerable time probing the interests of their students and their purposes for attending college, so they can mold reading and writing to these interests. Here is a description from one such instructor:

I operate a student-centered classroom, so that means a kind of teaching from the sidelines in the early part of the semester. I kind of try to have people doing activities which I direct, but sort of low key, to give me a chance to see how people perform in all different situations . . . Then toward the middle third of the semester, I try to begin giving information in whatever ways I've seen people's interest. Like this semester I have one student who is particularly interested in myth, the Indian guy. So I did a lot of stuff to kind of relate this European and Native American tradition. . . Then by the last third of the semester I really move back into the early format, but this time folks are much more independent. They have much more — they have the tools and then it's mostly just fine-tuning . . .

Everybody is so different, you can't assume they're like you; you can't assume they're like each other. So you really have to spend that first month doing what I call four for nothing. You're just beating; you're just finding the rhythm, one, two, three, four, go; one, two, three, seeing what they do. Then after you kind of know how people do things, then you can

begin to teach what they are interested in, what they need, what makes sense to them.

These instructors also foster work in groups rather than individual drill, partly in recognition of language as a communicative and therefore social process rather than an individual struggle with an unresponsive text. They do not avoid conventional grammar and punctuation drills, but they stress above all having students learn to create meaning from and with texts, and they subordinate drills to that kind of reading. Here, for example, is one instructor talking about her approach, where drill is subordinated to meaning-making:

It's very student-centered — it focuses on what students need to be able to do to succeed . . . they need to be able to write in ways that let their papers be read with respect . . . more bottom-up than top-down, because I'm trying to get them to have the meaning — I try to have meaning drive what they're doing. Although we may need to do a drill, time is so precious that I'd rather that they do more writing and talking than doing worksheets. And I expect *them* to take responsibility for a lot of it themselves — I'm not the error police.

The topics these classes cover are idiosyncratic, because a student-centered class invariably proceeds in different ways depending on the backgrounds and interests of students. This creates problems for evaluation, because the outcomes are not necessarily well defined — in fact, they are partially *student*-defined — but it does mean that such classes are livelier than skills-oriented remediation. Students are much more engaged, with each other as well as the instructor; the activities and materials of the class are generally adult, rather than the childish drills of the behaviorist classroom; and there seems much greater chance that this approach

can finally teach students about the complexities of language and mathematical thinking.

In most colleges, the appearance of more student- and meaning-centered teaching seems random and idiosyncratic, because the odyssey from didactic to constructivist teaching is usually one that instructors make on their own, through trial and error, with at best a little help from their friends. In most community colleges, there are few institutional resources to help instructors make this transition — though there are a few. We discovered a developmental studies division, in an institution we call North County Community College, which developed a coherent philosophy about remedial/developmental education, codified in two enormous volumes referred to as the "basic writing curriculum book." This is a self-consciously hybrid approach to instruction; the head of the division complained that existing basal readers generally follow either a "phonics" or a "comprehension" strategy, and that debates about remediation are similarly polarized: "We're back to the same old thing — top-down or bottom-up; and that's ridiculous." Instead, the philosophy of this department follows "transactional theory," in which language including writing is a "dialectic or interchange among writer, audience, and reality." Writing is a "recursive activity" incorporating prewriting, rewriting, and revision, and includes "strategies for invention and discovery whereby instructors help students to generate content and purpose." The approach includes grammar, spelling, and other mechanics, but only in the final stages of writing since its use early in writing has been so counter-productive.

The introduction to the "basic writing curriculum book" is an elegant approach to meaning-centered teaching, replete with examples from the latest

research and practice. The rest of the volumes contain examples, applications, and syllabuses in great profusion. The purpose of the "basic writing curriculum book" was partly to prepare new and part-time instructors to the division's methods. The division had developed a hiring process selecting individuals for their commitment to constructivist practice, and then assigned them a mentor to help them with their initial stages of teaching. In this division — and in a small number of other community colleges that have developed institution-wide support for teaching — the appearance of student-centered and constructivist approaches to remediation is deliberate and planned, rather than idiosyncratic.

A third major approach to developmental education in community colleges is the use of learning communities (LCs). Generically, LCs develop when students take two or more classes jointly; then, *if* instructors spend sufficient time planning together, each course can support and complement the others. LCs are infinitely flexible and can be used for a variety of purposes including multi-disciplinary approaches to general education, the integration of academic and vocational education, and the presentation of complementary subjects like science and math, or history and literature.^{xi}

The use of learning communities for remedial purposes has several distinctive features.^{xii} Typically, a "lead" course — an occupational course, or a central academic course — is matched with an English and/or a math course. For example, one institution found that a particular biology course was blocking the progress of students who wanted to go into health occupations, partly because of their problems in reading and math. The biology course was then joined with supportive math and English courses, which in turn modified their content to

provide the kinds of academic competencies necessary in biology. An automotive instructor who discovered the problems his students had with reading devised a learning community with an English instructor called "Reading, Writing, and Wrenches." At LaGuardia Community College in New York, all programs for welfare recipients are taught in learning communities. The Bridge program at Laney College, Oakland, and the Puente program in numerous colleges in California (specifically for Hispanic students, emphasizing bilingual education and some multi-cultural subjects) are other examples of learning communities devised specifically for remedial/developmental education.

The benefits of LCs are multiple, at least when they work well. Most obviously, students find themselves making progress in subjects that they care about — biology for health occupations, or an automotive program, for example — rather than simply being in drill-oriented remedial classes with no apparent relation to their future goals. The combination of classes allows instructors to contextualize their teaching: examples and applications in English and math can come from the "lead" course, and the lead instructor can develop writing exercises and problem sets that are used in the other courses. Students within learning communities get to know one another much better than most community college students do, and they universally report forming study groups as a result.^{xiii} And instructors report benefits too, since learning communities break down the isolation of instructors and allow them to create communities of like-minded teachers. With all these apparent benefits, the evidence so far indicates that students in learning communities tend to persist longer and earn higher grades than do similar students in conventional classes.^{xiv}

There are a number of other approaches to improving developmental instruction that have been suggested, though their prevalence is unknown. Weinstein and her colleagues (Weinstein et al., 1998) have suggested strategic learning as one approach; Cross and Angelo's (1993) classroom assessment techniques can be used in developmental education as well as other classes, and are sometimes used as the basis for staff development for all faculty in a community college.

By now my point should be clear: Remedial education in community colleges can vary enormously. The student-centered teaching from constructivist instructors (and typical of many LCs) looks completely different from conventional skills and drills; the combinations of courses in learning communities are vastly different from the conventional tendency to fragment the curriculum into stand-alone courses. Some of these approaches to developmental education stand virtually no chance, on *a priori* grounds, of helping students who have come to college with many years of formal schooling but without adequate command of language and math; others promise new approaches that might correct these problems. Even within one institution like the community college, remediation is not just one thing.

II. MULTIPLE APPROACHES TO EVALUATION

But does remedial/developmental education work? Do any of the approaches described in the previous section enable students to make further progress in their education, to complete community college programs and either

move out into the world of employment or onto further education? The evidence is sparse, and partly it is for lack of trying: most states and most colleges that provide remediation have not yet started to evaluate their programs in any way.^{xv}

Evaluation serves multiple purposes, of course. At the level of policy, decisions to expand or abolish programs might be based on such evidence: if, for example, there were many years of evidence that no form of remediation benefits students, then we might be tempted to eliminate all funding for such programs. But given the educational imperatives behind remediation and the lack of evidence, very few educators would be willing to take such a step — so I presume the purpose of evaluation is both to learn more about the conditions of success and to improve remedial programs.

One common form of evaluation is to examine the completion rates in developmental courses. However, such an approach fails to see whether there are any long-run effects from completion — for example, effects on subsequent retention and completion of credentials. While completion of such courses may be a good thing, it cannot by itself help people find jobs or help them vote, or provide them transcendent experiences of literature or art; such information provides no evidence about whether remedial programs have helped individuals get along with their lives.

Another common way of evaluating remedial efforts has been the comparison of pre-tests and post-tests, usually on some test of basic skills like the

TABE or the CASAS. For example, the Learning Assessment and Retention Consortium (LARC) of the California community colleges used to publish volumes of such figures (e.g., LARC, 1989a and b); they enable one to determine, for example, that increases in reading scores were higher in one college than another. But such results are almost useless, for a number of reasons. Most obviously, pre-test/post-test comparisons are available only for students who have stayed with a course until the end, when the post-test is given; if weaker students drop out, or if only "brush-up" students survive until the end, then the test increases will badly overstate the results for the average or random student. In addition, without knowing about the backgrounds of students in different colleges, the comparisons among institutions are impossible to interpret. At the extremes it might be possible to draw some conclusions, either if gains were close to zero or if they seem high, by some unarticulated and possibly idiosyncratic standard; but even that is risky. The meaning of such results improves a little when they include the proportion of students who complete the course. For example, recent figures from the CUNY system reveal that in 1990, 64.7 percent of post-tested students gained at least one year on the TABE; since only 61.7 percent of students were post-tested, this means that — under certain shaky assumptions — perhaps as few as 40 percent of students entering these adult education programs gained at least a year.^{xvi}

But other problems with pre-test/post-test results cannot be resolved simply by collecting more data. From some perspectives, the tests themselves are objectionable: they ask for the kinds of responses, about grammar and vocabulary, arithmetic operations and simple word problems, that fuel skills and drills approaches to remediation. No constructivist teacher, and few of those in learning communities, would accept the results of such tests; the vitriolic debate between McKenna, Miller, and Robinson (1990) and Edelsky (1990) about the evaluation of whole language is an indication of the debates over alternative outcome measures. In addition, like the completion of remedial courses, increasing test scores may be better than the opposite, but it may still not lead to further progress, the completion of meaningful degrees, or other outcomes. Finally, and worst of all from the perspective of improving programs, these statistical results say nothing about why test scores are what they are: they provide literally no information about why students fail to complete courses in such large numbers, about whether some approaches are better than others, about whether stand-alone remedial courses are effective compared to those embedded in other programs (like learning communities). Such figures, if they are discouraging enough, may lend some urgency to the problem of reform, but they do not provide any guide about what to do next.

Pre-test/post-test comparisons are close to the simplest and weakest evaluation designs, but many of the problems with such results are repeated in the

most sophisticated approach to evaluation — exemplified by the random-assignment evaluation of the remedial program in the San Diego GAIN program (Martinson and Friedlander, 1994). This evaluation assigned welfare recipients in five California counties randomly to a welfare-to-work program (GAIN), which included an initial assessment of whether individuals needed remedial education; a control group of welfare recipients was not assigned to such a program, though 8.4 percent of them participated in adult basic education or some kind of GED program on their own (compared to 43.6 percent of the experimental group enrolled in GAIN). Participating in the GAIN program increased the rate at which individuals received the GED — since 9.1 percent of the GAIN group but only 2.0 percent of the control group received a GED. However, the GAIN group as a whole did not improve their scores on the Test of Adult Literacy Skills (TALS),^{xvii} though individuals who scored the highest on an initial screening test (the CASAS) did increase their TALS scores overall.

Only in San Diego County was there a statistically significant increase — from 454 among the experimental group to 488 among the GAIN group, on a scale ranging up to 1,000. There are a few clues about the distinctiveness of the San Diego program: it was designed specifically for the GAIN program, "built on the premise that existing adult education services were not appropriate for the GAIN population because of their previous negative experiences in school"; key features included "up-to-date computer assisted learning combined with

classroom instruction, integrated academic and life-skills instruction," and "a new teaching staff." The results are not particularly encouraging even for San Diego, then, because even there the increase in TALS is trivial.^{xviii} But the TALS is not a self-evidently meaningful measure of outcomes (though elsewhere it has been linked to higher earnings^{xix}). Aside from an implication that effective remedial programs need to be different from conventional "school-like" adult education programs, there is not much guidance about how to reform the remedial programs that were part of GAIN — and there is no information whatsoever on the nature of the remedial programs that caused even more dismal results in other counties.

Another approach to evaluation fixes some of these problems but not others. For a number of years Miami-Dade Community College has evaluated its remedial programs in the format of Table 1: completion rates are calculated for students who are judged "below standard" in one, two, or three subjects, and who have successfully completed all appropriate remedial courses versus those who have not.^{xx} (Table 2 provides some earlier results, in a slightly more detailed format.) The amount of information in this table is substantial. It indicates, for example, that of the 6,324 students who entered, 59 percent were judged to need some kind of remedial education. Students with three deficiencies had a much harder time than students with one deficiency: only 42 percent of the former group corrected all three deficiencies and only 9 percent of these students graduated within three years, while 63 percent of students with one deficiency

corrected it and 28 percent of these graduated. (There are no surprises here, but there is a substantial warning to high school students who think they can easily make up during college the learning they have failed to do in high school.^{xxi}) And even students who take the full complement of remedial courses they presumably need, graduated at much lower rates than those who entered needing no remediation. Other conclusions are possible, of course, depending on the outcomes for those students still enrolled after five years: some of these will graduate, though the probability of doing so surely decreases with time. A reasonable conclusion from these results is that remedial courses help a great deal, but they cannot eliminate the gap between students with and without some need for remedial education, and a substantial fraction of students judged to need remediation fail to complete these courses.

It is easy to critique these results: they fail to control for variation in academic achievement and other characteristics (like family background) among groups; they neglect maturation effects, test effects, regression to the mean (which describes some "brush-up" students), and selection effects — particularly greater motivation among students who complete all remedial classes necessary. No doubt these results overstate the effects of remedial courses, and more sophisticated statistical analysis could improve the results. But they are a vast improvement over some of the evaluation results already presented: the outcome measure is one of intrinsic value,^{xxii} they clarify that the amount of remedial

education completed matters a great deal, and they compare many different groups of students with varying needs for remediation. However, they still fail to investigate what remediation is: they provide no clue about why so many students fail to complete remedial courses, they have not examined what about them attracts and repels students, and they do not investigate what these courses are like and whether some of them are more effective than others. Interesting as they are, it is hard to know what to do next.

Table 1
Five-Year Outcomes (Graduated or Still Enrolled)
For First-Time Degree-seeking Students Entering Fall 1989
Miami-Dade Community College

| | | Successfully Completed Remedial Courses in: | |
|---------------------------|----------------|---|---|
| Below standard in: | | Completed all remedial courses | Did not complete all remedial courses |
| <hr/> | | | |
| No Subject (N=2581) | N | 2581 | (these students did not need remediation) |
| | Graduated | 45% | |
| | Still enrolled | 14% | |
| | Total | 59% | |
| One Subject (N=1735) | N | 1097 | 638 |
| | Graduated | 28% | 7% |
| | Still enrolled | 24% | 10% |
| | Total | 52% | 17% |
| Two Subjects (N=1118) | N | 485 | 633 |
| | Graduated | 16% | 5% |
| | Still enrolled | 34% | 12% |
| | Total | 50% | 17% |
| Three Subjects (N=890) | N= | 218 | 672 |
| | Graduated | 9% | 2% |
| | Still enrolled | 40% | 11% |
| | Total | 49% | 13% |

"Still enrolled" refers to those still enrolled with a GPA of 2.00 or better.
Source: Morris (1994), Table 6.

Table 2

**Three-Year Persistence Rates (Graduated or Re-Enrolled)
For Tested First-Time Students Entering Fall Term 1982
Miami-Dade Community College**

| Below Standard in: | | Successfully Completed Remedial Courses in: | | | |
|----------------------------|----------------|--|----------------|-----------------|-------------------|
| | | No Subject | One Subject | Two Subjects | Three Subjects |
| No Subject (N=2021) | N | 2021 (these students did not need remediation) | | | |
| | Graduated | 26% | | | |
| | Still enrolled | 21% | | | |
| | Total | 47% | | | |
| One Subject (N=1524) | N | 873 | 651 | | |
| | Graduated | 11% | 21% | | |
| | Still enrolled | 17% | 25% | | |
| | Total | 28% | 46% | | |
| Two Subjects (N=1360) | N | 530 | 509 | 321 | |
| | Graduated | 5% | 11% | 15% | |
| | Still enrolled | 9% | 26% | 33% | |
| | Total | 14% | 37% | 48% | |
| Three Subjects (N=1457) | N= | 641 | 357 | 303 | 156 |
| | Graduated | 1% | 4% | 8% | 9% |
| | Still enrolled | 9% | 19% | 29% | 37% |
| | Total | 10% | 23% | 37% | 46% |

"Still enrolled" refers to those still enrolled with a GPA of 2.00 or better.
Source: Losak and Morris (1985), Table 1.

I know of no outcome evaluations that try to compare different approaches to remediation except for the investigations of learning communities cited above (endnote 15). And a large testimonial literature indicates that students like learning communities more than conventional classes (MacGregor, 1991) — a result that is significant if only because of the evident dreariness of and disengagement within most remedial classes.

A final issue important to remediation is rarely mentioned in program evaluation, perhaps because the dominant focus has been on measuring the effects of a program on those enrolled in it. But what I will call, for want of a better term, the "assignment problem" arises because students are typically assigned to remediation based on an assessment of some sort — usually a basic skills test like the TABE or the CASAS, sometimes with a writing sample, sometimes with some counseling. Then the question becomes whether students enrolled in remediation fare better than those who did not. It is possible that some students assigned to remediation do not need it — for example, students only needing brush up — and of course if remediation programs are completely ineffective then everyone assigned to them is misassigned. But a different question is what happens to those students who did not "fail" the assignment test — who scored just above the cut-off point, and did not enroll in remedial programs. It is quite possible that these students would still benefit from remediation, or that some students score well on basic skills tests but still cannot write or reason well — because these tests do not measure such higher-level competencies. Any assignment procedure runs the risk of false assignment — Type I errors, or assigning students to remediation who do not need it — as well as allowing students who need some kind of remedial work

to progress to college-level work (Type II errors). Both Type I and Type II errors may reduce rates of completing programs, and Type II errors have the added cost of putting unprepared students in regular classes — a process that ends up generating a good deal of hidden remediation as "college-level" courses are converted into remedial courses. The assignment problem is a difficult one, and I have further explored its difficulties in an appendix.

These issues about errors in assignment are at the heart of debates about community colleges and whether they are egalitarian, advancing students who otherwise would have no access to postsecondary education, or whether instead they "cool out" students who otherwise would go further in four-year colleges.^{xxiii} A particular incident illustrates this problem clearly. California instituted a process known as "matriculation," intended to help entering students be placed correctly in regular and remedial/developmental classes. However, some colleges implemented matriculation poorly, using tests that had not been validated for the purposes used, preventing many individuals from enrolling on the basis of irrelevant tests, and using test results by themselves whereas the education code stated that they were to be advisory only. To improve the quality of matriculation, MALDEF (the Mexican-American Legal Defense and Education Fund) successfully sued a college on equal protection grounds, claiming that the requirement discriminated against Mexican-American students. In effect, MALDEF claimed that there were too many Type I errors, of Mexican-American students incorrectly assigned to remedial education whose progress through college was thereby impeded. As a result, enrolling in remedial courses became voluntary; what's more, the Office of the Chancellor in Sacramento imposed

regulations requiring that *any* prerequisites for *any* courses be justified through a validation study, a burdensome procedure that has all but eliminated prerequisites in California community colleges. Now there may be fewer Type I errors — and if there are, they are presumably the choice of students rather than the result of college assignment — but there may be more Type II errors, of students who need more remedial/developmental education (or other prerequisites) than they get. Indeed, the lawyer involved in the case admitted that the remedy was imprecise and "broad brush"; MALDEF intended to improve the sensitivity of the matriculation process, but instead got a series of crude rulings and bureaucratic procedures, as is typical in legal cases.^{xxiv} But the result of the lawsuit is unclear: no one knows whether the MALDEF case helped Mexican-American students or hurt them, or whether it benefited or harmed other students who have been equally affected by the ruling and the response of the Chancellor's Office. Indeed, no one has even thought to ask the question, never mind answer it.

It is clear, then, that the evaluation of remedial education is still in its infancy, and no one knows much about what works and what does not.^{xxv} What little evidence there is indicates that completion rates in remedial courses are low, that the amount of remediation does matter to important outcomes like persistence in and completion of college programs, and that learning communities are probably more effective than stand-alone classes. There is some suggestion (from the GAIN evaluation) that "school-like" programs are less effective, but it is not entirely clear what this means. The observational evidence (such as that in Section D) indicates how much remedial courses vary, though my interpretation of these classes — that many of them provide virtually no possibility for significant

learning — might not be widely accepted and might not even be correct; there is no particular reason to think that the remedial courses at Miami-Dade are particularly innovative and yet they have substantial effects on graduation and retention (Table 1). But, aside from the possible recommendation to teach all remediation in LC formats, there is not much evidence to suggest how to improve the state of remedial/developmental education.

III. AN ECLECTIC APPROACH TO EVALUATION

In this vacuum, it is not helpful to recommend one particular approach to evaluation over others. The orthodoxy in the evaluation literature — random-assignment studies, to rule out as many selection and self-selection effects as possible — is no more useful than the most basic pre-test/post-test designs, as the GAIN results illustrate. There are too many dimensions of remedial education that are poorly understood, and investigating them requires several different methods.

Furthermore, there are at least two levels at which evaluation is useful. One is the program level, where information about a particular course and a specific instructor would be useful in diagnosing what is going well and badly. This kind of information might include completion rates, assessments of academic progress including locally developed measures (writing portfolios, for example), and subsequent progress through the college, but it should also include peer observation so that instructors can begin to create communities of discussion and practice around remedial education.^{xxvi} If such program-level evaluations are to be

useful to instructors themselves, they cannot be particularly complex — they surely cannot use comparison groups, or follow students over long periods of time, or introduce special assessments unrelated to normal teaching, for example. A second level includes more formal evaluations carried out at the institution level — for example, like the evaluations of Miami-Dade's programs in Table 1 — or at the state or national level. These can be more complex, with control or comparison groups, and can follow students over longer periods of time; their purpose is not to improve the practice of specific instructors, but rather to assess institutional and state policy, the overall effects of remediation, and the effectiveness of different approaches.

Instead of a single approach, therefore, I recommend a number of different approaches, each of which has the potential for illuminating a different aspect of this difficult problem:

1. The dropout rates from remedial courses need more investigation.

While it is plausible that dreary teaching is the reason, the difficult lives of many community college students — including financial problems, child care problems, transportation problems, other family problems including abusive spouses and boyfriends, and the pervasive indecision of experimenting and uncommitted students — surely play important roles. It is also possible that complex combinations of reasons are responsible, and that even students themselves cannot articulate why they stay with or leave a particular program. As one student commented on his leaving the community college,^{xxvii}

It was not even a decision. I just did not go. Sometimes you decide on certain things. It was not a decision at all. Just like you go home, tired

from work, you don't decide about "Oh, I'm just going to go to sleep now." You just doze off and go to sleep. It wasn't a plan. That's the way [dropping] the class was: it wasn't a plan.

A combination of qualitative, interview-based studies and quantitative studies might begin to provide evidence for improving remedial courses.^{xxviii}

2. Outcome measures need to include more than test scores of basic skills.

(Indeed, it is an open question whether such test scores mean anything at all.)

Persistence in college and completion of degrees are obvious measures, because completion is particularly important to the economic benefits of community colleges.^{xxix} However, constructivist teachers have their own measures of success like writing portfolios, and other measures emerge from a college's intentions for remedial programs — for example, completion of occupational programs may be the most valuable outcomes in some cases. More systematic collection of outcome measures could build up better understanding of the different outcomes that remedial courses can achieve. Some students may declare other purposes — political or familial goals, for example — to be more important, and qualitative studies can clarify these goals and the contribution of courses toward achieving them.

3. It is important, at least in institution- or state- or national-level studies, to have some kinds of comparison or control groups, to see if completing remedial courses produces benefits compared to students who do not take such courses.^{xxx} But while it might be possible to design a random-assignment study under certain

conditions — for example, comparing the effects of learning communities to conventional formats, or comparing one pedagogical approach with another^{xxxii} — it would not be ethical or feasible to compare the effects of remediation to its absence through random assignment. Instead, the kinds of comparison groups in Table 1 — where some students are thought to need remediation, but do not take such classes^{xxxii} — is the only feasible way to construct comparison groups. Evaluation studies can still collect other information like prior test scores and grades to use as regression controls, to improve comparability somewhat.

4. No outcome evaluation should ever fail to understand the program it is evaluating — and this means observing and describing the classroom practices in remedial courses. The conventional "black box" evaluation, in which the nature of the program being evaluated is never described, should be replaced with a "Pandora's box" approach that clarifies both the triumphs and the troubles of classroom practices.^{xxxiii} Otherwise it becomes difficult to know what might have generated a particular set of outcomes, and therefore what might be changed.

5. If evaluation is ever to have any influence on classroom practice, it needs to compare different approaches to teaching. This in turn requires some conceptualization of different approaches. The differences between behaviorist and didactic practices on the one hand, and constructivist and student-centered practices on the other, are dimensions of teaching that emerge over and over again, both in the comments of instructors themselves and in various theories of teaching and learning. Although there is endless debate about what dimensions of teaching are important, these two polar approaches can be operationalized for purposes of evaluation^{xxxiv} and compared along their many dimensions.

However, there may be other ways to think about the power of different approaches, particularly since student motivation arises for many complex reasons external as well as internal to classrooms. While skeptical about "skills and drills," we have observed drill-oriented remedial classes where students seemed to be attentive and engaged, possibly because the class was followed by an occupational class where the academic material would be applied. Some teachers following behaviorist approaches develop idiosyncratic methods,^{xxxv} or a special rapport with students, that overcome the limits of drill, and some students — particularly ESL students, who seem to be able to sit through anything, and older students with clear and passionate goals — are able to learn from even the most dreary teaching. Some of these possible successes may be replicable and others may not, but understanding them better is a necessary first step to improving the quality of instruction.

6. The "assignment" problem needs to be better understood. (See the appendix for an initial effort.) Understanding this issue depends first on ascertaining whether remedial programs themselves are effective: if they are ineffective, then every student assigned to them is misassigned. However, in the case of programs judged to be effective, the question of Type II errors is whether some students who might benefit do not attend them — either because the assignment test fails to identify those in need of remediation, or because enrolling in such courses is voluntary (as in post-MALDEF California). Examining this problem requires looking at the subsequent experiences of several groups: (a) students judged in need of remediation who did not enroll in such courses — like some of the groups whose progress is measured in Table 1; (b) the "near misses,"

or those who barely passed the assignment test, compared both to those who enrolled in remediation and those who clearly do not need remediation, at least based on the initial basic skills test.^{xxxvi} Finally, some consideration of alternative assignment procedures is appropriate — either different basic skills tests, or procedures that incorporate other information and counseling as well as testing.

In the end, many questions about remedial/developmental education are empirical issues of this kind. The expansion of postsecondary education since the 1960s, and especially the expansion of open-access community colleges, has provided opportunities for some students where none existed before, and the dedication of many colleges and most instructors to their non-traditional students is unmistakable. The shift toward viewing institutions as responsible for learning and advancement is a move in the right direction, certainly for proponents of equity. But dedication and student-centeredness, while necessary, may not be sufficient, so a program of evaluation and improvement is central to improving the performance of students.

APPENDIX:

The "Assignment Problem" in Developmental Education

The "assignment problem" arises in education whenever a student is assigned to one form of education rather than another, based on an assessment of some kind. In remedial education, the assessment is usually a basic skills exam, though a few colleges add a holistically graded writing sample; the assessment process could be a more complex procedure in which multiple tests (and more sophisticated tests) are used along with interviews, an examination of prior education, and the like. Then a person is assigned to remedial education based on this assessment; the assignment may be mandatory or voluntary, and in either event some students enroll in the remedial course and others do not. Of those who enroll, some fraction complete and others (often a very large percent) do not. Sometimes there is an exit exam to move to the next level of education (for example, the first college-level course, or the next remedial course in a sequence), and sometimes course completion is sufficient. While this is a relatively familiar sequence in remedial education, the assignment problem arises in many other contexts including the assignment of students to special education and to various tracks in K-12 education, admission to college, and to various majors within a college.

The question is whether assignment to a remedial program benefits the student or not, or conversely whether those not assigned to remediation would have benefited had they been assigned (Type II errors). To see the complexity of the issue, it is helpful to describe a simplified world. Imagine that facility in

reading, writing, math, or any other subject can be measured on a 100-point scale, ascertained by a conventional test, and that (arbitrarily, perhaps, but with the weight of tradition) a certain point on this scale (say 70) is considered necessary for college-level work. Then anyone with a score below 70 is assigned to remediation; otherwise the college-level course would have to go over sub-college material, which is the ubiquitous problem of "hidden" remediation. Students stay in remediation until they achieve 70 on an exit exam; there may be different levels of remedial courses depending on the scores students attain (e.g., one for those scoring 50-60, another for those scoring 60-70). The effectiveness of remedial programs is simply measured by the absolute increase in the score for the average student in the class, which is the conventional pre-test/post-test comparison. Finally, there might be a different standard (say 85) for graduation from this institution, perhaps a standard established by demands in employment; an exit exam for graduation (as Florida has in its rising junior exam, or high schools now have in many states) is sufficient to prevent under-prepared students from graduating.

In this simple world, there are not any special problems once the test has been created and the various cut-points established: the test used for initial assignment is highly valid, pre-test/post-test evaluations of remedial courses are adequate, and the required exit exam is obvious. Indeed, the conventional pattern of developmental education seems to *assume* that the world of education follows this simple model. But of course facility in various subjects does not follow this model at all because there are multiple dimensions to reading, writing, or math. The specific dimensions of reading necessary in an automotive course are

different from those in a standard English course; the dimensions tested in a diagnostic basic skills test may be different from those required in any subsequent college-level courses; and the content of any specific remedial course may be different from both the diagnostic exam and subsequent courses. Thus the initial assessment may not be aligned with the remedial program (which is a problem of predictive validity); the remedial program may not be aligned with the exit exam, if the exit exam is established independently of the program (though in this case one might assume that instructors would start teaching to the test); and neither the program nor the exit exam may be aligned with subsequent "college-level" education, in which case successful completion of developmental courses may not enhance subsequent outcomes. Of course, the college program as a whole may not be aligned with the competencies necessary for employment, generating complaints from employers about under-prepared employees. So there are many points at which the multiple dimensions of any particular subject can create problems, all of them leading to students assigned to remediation who do not benefit (Type I errors). When a college has a series of remedial courses, the problems of alignment are simply compounded.

A further problem, of course, is that attending remedial courses requires time and (usually) money. The additional time required to complete a remedial sequence may itself lead to non-completion. Distinguishing the time and money dimensions of remediation from the alignment issues is important because the remedies are different.

By the same token, students may not enroll in remedial courses even if they would benefit — the problem of Type II errors. This may happen (as in

California and some other states) because a placement test is advisory only, because students dislike the additional time and money costs of remediation, or because college-level courses stress competencies that are not measured by the simple initial assessment. If, for example, developmental instructors understand that subsequent college-level work requires facility with analytic thought and try to teach that, while the initial basic skills assessment measures facility with grammar and vocabulary, then students will pass the initial assessment but may still lack the analytic abilities necessary to succeed subsequently.

Now we can see a little more clearly where conventional evaluations fail to incorporate the complexities of the assignment problem. In the first place, it is necessary to have some intrinsically valuable outcome measure, like the graduation and re-enrollments rates for Miami-Dade in Table 1. Then equations describing the probability of graduation as a function of initial assessment scores, completion (or non-completion) of remedial courses, plus the many other variables that help explain graduation (gender, family background, race/ethnicity, family support or family responsibilities, etc.) can identify the increase in the probability of graduation due to completing remedial courses for those judged in need of them. Then one minus this probability is the probability of Type I error; for example, using the simplified figures in Table 1 (instead of the logit or probit equations that could be estimated), 55 percent of those below standard in one subject benefited from completing the appropriate remedial course, in the sense that they graduated rather than not graduating, but 45 percent did not.^{xxxvii} Using the figures on those who graduated *or* were still enrolled, the results are much better: 83 percent benefited and only 17 percent did not. However, the likelihood

of finally graduating surely decreases as the period of time enrolled increases, and so the equations predicting graduation should incorporate a measure of time, or should use event history methods to examine the probability of completion as a function of time, where the time necessary to graduate is increased (and the likelihood of graduation reduced) by the need to take remedial courses as well as many other variables including the demands on students' time. So it is necessary to estimate a system of equations, some describing outcomes and others describing the time enrolled, with remediation affecting both of these.

However, when there is a statistical finding of large Type I errors, this kind of statistical analysis cannot distinguish among potential explanations. While the nature of alignment or misalignment among the different aspects of the overall assignment problem could in theory be resolved by having sub-tests for different dimensions of reading, writing, and math, the difficulty of doing this is overwhelming. The alternative is a careful content analysis of the diagnostic exam, the various remedial courses, exit exams (if any) and the courses students subsequently enroll in. However, while community college instructors often complain about misalignment (Grubb and Associates, 1999, Ch. 5), careful analysis of the problem is rare.

The examination of Type II errors is made virtually impossible because of the lack of a particular kind of information. When students are judged in need of remediation according to some assessment procedure, some of them fail to enroll in (or complete) remedial courses, so the effectiveness of completing remediation can be determined by comparing the two groups (as in Tables 1 and 2). But if students are judged not in need of remediation, then none of these students enroll

in remedial courses — so it is impossible to tell if some of them would have benefited from remediation. It might be possible to simulate the experience of this group: if we assume that students within one standard deviation of the critical cut score are statistically indistinguishable from one another, then analysis of this restricted group and the effects of remediation might provide an estimate of Type II errors. Some understanding of Type II errors now comes from instructors complaining that students are not prepared for the specific uses of reading, writing, or math in their courses, as occupational instructors often do; the solutions sometimes include learning communities or applied academics courses with the basic skills necessary to particular occupational areas. But in the absence of such solutions, the consequence is a large number of under-prepared students in conventional classes, leading either to hidden remediation, to non-completion, or to some of both.

My contention is that the "assignment problem" needs to be much better understood before there can be much progress on the quality of remediation. Many developmental instructors are quite aware of many of these problems, and they complain about several dimensions of misalignment (see Grubb and Associates, 1999, Ch. 5, especially section 1). But most writing about remedial/developmental education fails to address this question, and the policy debates (e.g., at CUNY) have failed to address any dimensions of quality whatsoever. Until these issues can be more carefully examined and understood, the effectiveness of remedial education will continue to be haphazard and its evaluation be incomplete.

ENDNOTES

ⁱ I will not repeat the various debates over the terms "remedial", "developmental", and "basic skills" education; see Goto (1995) for an excellent review of these issues and the compromise of "remedial/developmental" education.

ⁱⁱ A full appreciation of the difficulty of remedial/developmental teaching can be found in Goto (1998), who followed a number of students in two community college classrooms. Such understanding can come only from examining the lives of students as well as activities within the classroom and pedagogical strategies.

ⁱⁱⁱ In reality, most instructors and institutions interpret remediation in instrumental terms, but students may not. A more student-centered conception could accept education that looks to be relatively basic to be valuable in its own right, for students whose purposes may not include completing advanced degrees. See Goto's (1998) description of basic writing students in a community college, some of whom view it as valuable regardless of its instrumental purposes. This way of looking at remedial education is more common in certain branches of adult education; see for example, Gowen and Bartlett's (1997) description of several women able to confront domestic abuse through a remedial writing program.

^{iv} Traub (1994) includes some descriptions of one College Skills class at CCNY with many comments from the instructor about the lack of academic preparation among his students. These descriptions convey an example of a disequilibrium between instructor and students — where the instructor has an expectation of what students should be able to do that is not matched by their preparation. For other descriptions of remedial/developmental classes in community colleges, see Grubb and Associates (1999), Ch. 5, and Goto (1998). But most examinations of remedial or developmental education contain no analysis of classrooms whatsoever; see, for example, Roueche and Roueche (1993) and McCabe and Day (1998).

^v See the National Center for Developmental Education at Appalachian State University in Boone, NC, which publishes the *Journal of Developmental Education* and a newsletter, the *Review of Research in Developmental Education*. The national association is the National Association for Developmental Education.

^{vi} I based this statement on recent observations in several library and adult programs by Caleb Paull.

^{vii} Cameron and Heckman (1993) found no employment value to the GED, using sophisticated statistical techniques; reworking the same data, Murnane, Willett, and Boudett (1995) found a small effect, though they noted that it might not be enough to overcome the pedagogical disadvantages of the test. Nor does the GED appear to enhance subsequent education attainment; see Quinn and Haberman (1986).

^{viii} This section draws heavily on a book about teaching in community colleges (Grubb and Associates, 1999) based on observations of and interviews with about 280 instructors (including 27 English instructors and 36 remedial/developmental instructors) and about 60 administrators. See also Worthen (1997), drawn from the same data. This is, amazing to say, almost the only empirical work on teaching in community colleges since Richardson, Fisk, and Okun (1983).

^{ix} See Grubb and Kalman (1994), and the earlier review of computer programs for job training programs by Weisberg (1988). While the latter review is by now several computer generations old, we saw only drill-oriented computer programs in our observations during 1993 - 1997. There are some interesting constructivist uses of computers by a few community college instructors, but they are all individual efforts by instructors developing computer applications on their own; see Grubb and Associates (1999), Ch. 7.

^x Community college and adult instructors sometimes tell stories of students, invariably older, who breeze through a programmed text or workbook. I interpret some of these as "brush-up" students, who have been out of school for a decade or more and have failed an initial placement exam because they have forgotten the trivia involved in such tests. If they have learned basic English and math in their earlier schooling, one additional exposure is sufficient to brush up on these skills.

^{xi} Learning communities have generated a great deal of interest; see especially Matthews (1994a and 1994b).

^{xii} Learning communities have also been used for ESL classes — for example, by pairing a computer class with an ESL class concentrating on computer-related literature and vocabulary — but I will not discuss these here.

^{xiii} Such study groups are reminiscent of those that are at the heart of Uri Treisman's approach to teaching math, though they are more informal.

^{xiv} Gudan, Clack, Tang, and Dixon (1991); Tokina (1993); Tinto, Goodsell-Love, and Russo (1994); MacGregor (1991).

^{xv} Boylan, Bliss, and Bonham (1997) found that only 14 percent of community colleges had any systematic evaluation. A recent SHEEO survey found very few states able to comment on the impact of remedial policies on student success, and most studies seem to be somewhere in the planning stages (Russell, 1998, p. 26 and Appendix G). Most studies that purport to describe effective programs rely on nominations of programs by various observers, not on outcome measures; see, for example, the programs profiled in McCabe and Day (1998) or in Roueche and Roueche (1993).

^{xvi} This is true if none of the students who failed to survive to the post-test gained at least a year. Since some of these students may have benefited from their period in adult ed, one might say that the results support gains among 40 percent to 64.7 percent of students — too broad a range to have much confidence about the outcomes. Other results indicate that students with less than 21 contact hours gained an average of .76 years, while those with more than 120 hours gained .96 years — a difference that, whether it is statistically significant or not, strikes me as being trivial in practical terms. See *Student Outcomes Research Project* (1996), Tables 2 and 4. This report, to its great credit, spends a great deal of time clarifying the limitations of pre-test/post-test comparisons and uses various other measures of success including interviews with teachers and students.

^{xvii} The TALS, developed by the Educational Testing Service, includes document, quantitative, and prose literacy components; only the first two were used.

^{xviii} The gain from 454 to 488 in San Diego is equivalent of a gain from 227 to 244 on each component. ETS divides the TALS scores into five levels of proficiency: scores below 225 are considered the most basic; then levels 2, 3, and 4 are distinguished by 50-point increases. An

average increase of 17 points is therefore only one-third of the variation within these levels, and is unlikely to move an individual from one level to another.

^{xxix} This finding is based on an article sent to me for refereeing, and therefore anonymous.

^{xxx} Similar results are available for the CUNY system. For example, they indicate that 36.5 percent of Associate degree students who passed all the basic skills tests they needed had graduated eight years later, compared to 14.9 percent of those who did not pass all of them and 33.9 percent of those who took no remediation (CUNY Office of Institutional Research 1998, Table 12).

^{xxxi} See also Rosenbaum (1998), who clarifies the much slower progress through postsecondary education for students who have done poorly in high school.

^{xxxii} The results in Table 2 are also available for CLAS test scores. The CLAS test is a "rising junior" exam required of all students in Florida before they start their junior year; while it is only a test score, it is one of great importance to students who want to transfer to four-year colleges.

^{xxxiii} I have summarized these debates and the evidence in Grubb (1996), Ch. 2. See also Lavin and Hyllegard (1996) and Rouse (1995). The empirical evidence on balance is against the hypothesis of cooling out, since most community college students would not otherwise have gone to postsecondary education at all. In addition, the critics of community colleges tend to rely on ancient "evidence" about the role of counselors in Clark (1960), although approaches to counseling have changed dramatically since the 1950s. If there is any truth to the charge of "cooling out," my argument is that it occurs by accepting "non-traditional" students and then teaching them in traditional ways; see Grubb and Associates (1999), especially Ch. 10.

^{xxxiv} Oral communication, Susan Brown, Council for Latino Issues Forum, San Francisco, December 1997.

^{xxxv} There is, to be sure, a large advice literature about how best to teach adults, but this is based largely on experience rather than empirical evidence of any sort. Similarly, the synthetic lists of recommendations about good teaching, like the widely cited "seven principles for good practice" (Chickering and Gamson, 1991), are based on a mixture of evidence from the K-12 literature, experience, and student ratings.

^{xxxvi} For a more extended argument about the creation of such communities of practice through peer observation, see Grubb and Associates (1999).

^{xxxvii} This is taken from interviews with community college students, in Grubb (1996), Ch. 2.

^{xxxviii} The dominant approach to dropouts in higher education has followed Tinto's (1987) model, which assumes that the extent of academic and social integration into a college explains dropping out. But this model is much too restrictive for community colleges since it fails to include the many external factors — fiscal reasons, complex lives, issues of identity and commitment — that affect community college students. Therefore quantitative analysis should look for causes beyond Tinto's model. Some of these, like the reasons for experimenting, are difficult to quantify and are probably best examined through interviews.

^{xxxix} See the battle between Kane and Rouse (1995a, 1995b) and me (Grubb, 1993, 1995) about whether program completion is necessary for economic benefits to materialize. Other work with the SIPP data (Grubb, 1997) and a survey of the available literature (Grubb, 1998) indicates that the benefits of taking courses without completing credentials is on the average quite low and quite variable.

^{xxx} For an example of a large study without any comparison group, the National Study of Developmental Education surveyed results for developmental students only in a variety of postsecondary institutions. The findings — e.g., that 24 percent of developmental students in community colleges persisted until graduation — are impossible to interpret without knowing more about persistence of other students at the same institutions. See Boylan and Bonham (1992); other results from this study came out in subsequent issues of *Research in Developmental Education*.

^{xxxii} Community college students often take courses according to the time of day they are taught, to fit into their complex schedules. One possible design, therefore, would be to randomly assign different pedagogical approaches to different times of day. Then students could either choose a course or be assigned randomly according to the time of day they prefer; any particular class would have a combination of self-selected and randomly assigned students, and these two groups could be compared to see if the students who select a particular course are different on any dimensions from randomly assigned students.

^{xxxiii} In some states, placement in remedial courses is mandatory if students score below certain levels on diagnostic tests; in other states remediation is voluntary.

^{xxxiv} For a similar recommendation in the context of job training programs see Friedlander, Greenberg, and Robins (1997).

^{xxxv} See, for example, the work of Knapp & Turnbull (1990) and Knapp & Associates (1995). They defined "the conventional wisdom" and "alternatives to conventional practice" almost precisely as I have defined skills and drills versus meaning-centered approaches. They then compared the effects of classes with different numbers of practices drawn from the list of "alternatives."

^{xxxvi} One instructor we observed had devised a way of reinforcing material in four different ways, and was highly conscious of using different materials — written materials, oral instruction, films, computer-based materials, etc. — to fit different "learning styles." While the skills presented in her class were quite basic, students seemed more engaged than in most remedial classes.

^{xxxvii} Where individuals are assigned to remediation based on a basic skills tests with continuous results, information is available on the subsequent education experiences of those just above and just below the cut-off score for assignment to remediation. Alternatively, analysis of completion as a function of scores at entry plus remediation could reveal whether remediation is effective for different groups of entering students. Most community colleges have these data in their files, though they are often scattered in different data systems and research of this kind seems always to be a low institutional priority.

^{xxxviii} From those below standard in one subject, 7 percent who did not complete the required course graduated, while 28 percent who passed the required course graduated. If we assume that remedial courses ought to bring students to the level of those who did not need remediation, then 45 percent should have graduated; therefore $(28-7)/(45-7)=55$ percent benefited. Other assumptions of what remediation can hope to achieve obviously generate different conclusions.

REFERENCES

- Boylan, H., Bliss, L., & Bonham, B. (1997). Program components and their relationship to student success. *Journal of Developmental Education, 20*, 2-9.
- Boylan, H., & Bonham, B. (1992). The impact of developmental education programs. *Review of Research in Developmental Education, 9*.
- Cameron, S., & Heckman, J. (1993, January). The Non-equivalence of High School Equivalents. *Journal of Labor Economics, 11*.
- Chickering, A.W., & Gamson, Z.F. (1991). *Applying the seven principles for good practice in undergraduate education*. San Francisco, CA: Jossey-Bass Publishers.
- Clark, B. (1960). *The open door college: A case study*. New York: McGraw-Hill.
- Cross, K.P. & Angelo, T. (1993). *Classroom assessment techniques: A handbook for college teachers, (2nd ed.)*. Ann Arbor: National Center for the Improvement of Postsecondary Teaching and Learning.
- CUNY Office of Institute Research and Analysis (1998, February). *Basic skills and ESL at the City University of New York: An overview*. New York: CUNY.
- Eaton, Judith S. (1994). *Strengthening collegiate education in community colleges*. San Francisco: Jossey-Bass.
- Edelsky, C. (1990, November). Whose agenda is this anyway ? A response to McKenna. *Educational Researcher, 9*,7-11.
- Friedlander, D., Greenberg, D., & Robins, P. (1997). Evaluating government training programs for the economically disadvantaged. *Journal of Economic Literature, 35*,1809-1855.
- Goto, S. (1995, October). *The evolving paradigms of developmental/remedial instruction in the community college*. Berkeley: Graduate School of Education, University of California.
- Goto, S. (1998, October). *The threshold: Basic writers and the open door college*. Unpublished doctoral dissertation. Berkeley School of Education, University of California.
- Gowan, S. (1992). The politics of workplace literacy: A case study. New York: *Teachers College Press*.

-
- Gowen, S., & Bartlett, C. (1997). Friends in the kitchen: Lessons from survivors. In G. Hull (Ed.), *Changing Work, Changing Workers: Critical Perspectives on Language, Literacy, and Skills*. (pp. 36-51) Albany: State University of New York Press.
- Grubb, W. N. (1993). The varied economic returns to postsecondary education: New evidence from the class of 1972. *Journal of Human Resources*, 28, 365-382.
- Grubb, W.N. (1995). Response to comment. *Journal of Human Resources*, 30, 222-228.
- Grubb, W. N. (1996). *Working in the middle: Strengthening education and training for the mid-skilled labor force*. San Francisco: Jossey-Bass.
- Grubb, W.N. (1997). The returns to education and training in the sub-baccalaureate labor market, 1984-1990. *Economics of Education Review*, 16,231-246.
- Grubb., W.N. (1998, August). *Learning and earning in the middle: The economic benefits of sub-baccalaureate education*. Occasional Paper, Community College Research Center, Teachers College, Columbia University.
- Grubb, W.N., & Associates (1999). *Honored but invisible: An inside look at teaching in community colleges*. New York and London: Routledge.
- Grubb, W.N., & Kalman, J. (1994). Relearning to earn: The role of remediation in vocational education and job training. *American Journal of Education*, 103, 54-93.
- Gudan, S., Clack, D., Tang, K., & Dixon, S. (1991). *Paired classes for success*. Livonia, MI: Schoolcraft College.
- Hull, G. (undated). *Alternatives to remedial writing: Lessons from theory, from history, and a case in point*. Paper prepared for the Conference on Replacing Remediation in Higher Education, National Center for Postsecondary Improvement, Stanford University. Berkeley: School of Education, University of California.
- Kane, T., & Rouse, C. (1995a). Comment on W. Norton Grubb, The varied economic returns to postsecondary education: New evidence from the class of 1972. *Journal of Human Resources*, 30, 205-221.
- Kane, T., & Rouse, C. (1995b). Labor market returns to two- and four-year colleges. *American Economic Review*, 85, 600-614.
- Knapp, M. S., & Turnbull, B. J. (1990, January). *Better schooling for the children of poverty: Alternatives to conventional wisdom. Volume I: Summary*. Washington, DC: U.S. Department of Education, 1990. (ERIC Document Reproduction Service No. ED 314 548)

-
- Knapp, M., & Associates. (1995). *Teaching for meaning in high-poverty classrooms*. New York: Teachers College Press.
- Knapp, M., Shields, P., & Turnbull, B. (1993). *Academic challenge for the children of poverty. Volume 1: Findings and conclusions*. Washington, DC: U.S. Department of Education.
- Labaree, D. (1997, Spring). Private goods, public goods: The American struggle over educational goals. *American Educational Research Journal*, 34, 39-81.
- Learning Assessment and Retention Consortium (LARC). (1989a, August). *Student outcomes study: Fall 1988 follow-up study of students enrolled in remedial writing courses in fall 1986 and remedial reading courses in fall 1987*. Santa Ana, CA: Rancho Santiago College.
- Learning Assessment and Retention Consortium (LARC). (1989b, October). *Student outcomes study: Mathematics, year 3*. Santa Ana, CA: Rancho Santiago College.
- Lavin, D., & Hyllegard, D. (1996). *Changing the odds: Open admissions and the life chances of the disadvantaged*. New Haven: Yale University Press.
- Losak, J., & Morris, C. (1985). *Comparing treatment effects for students who successfully complete college preparatory work* (Research Report No. 85-45). Miami, FL: Miami-Dade Community College, Office of Institutional Research.
- MacGregor, J. (1991, Fall). What differences do learning communities make? *Washington Center News*, 6, 4-9.
- Martinson, K., & Friedlander, D. (1994, January). *GAIN: Basic education in a welfare-to-work program*. New York: Manpower Demonstration Research Corporation.
- Matthews, R.S. (1994a, Spring). *Notes from the field: Reflections on collaborative learning at LaGuardia*. Long Island City, NY: Office of the Associate Dean for Academic Affairs, LaGuardia Community College.
- Matthews, R.S. (1994b). Enriching teaching and learning through learning communities. In T. O'Banion & Associates (Eds.), *Teaching and learning in the community college*. (pp. 179-200). Washington DC: American Association of Community Colleges.
- McCabe, R., & Day, P. (1998, June). *Developmental education: A twenty-first century social and economic imperative*. Mission Viejo, CA: League for Innovation in the Community College and The College Board.

-
- McGrath, D., & Spear, M.B. (1991). *The academic crisis of the community college*. New York: State University of New York Press.
- McKenna, M.C., Miller, J.W., & Robinson, R.D. (1990). Whole language: A research agenda for the nineties. *Educational Researcher*, 19, 3-6.
- Morris, C. (1994, November). *Success of students who needed and completed college preparatory instruction*. (Research Report No. 94-19R). Miami: Institutional Research, Miami-Dade Community College.
- Murnane, R., Willett, J., & Boudett, K.P. (1995). Do high school dropouts benefit from obtaining a GED? *Educational Evaluation and Policy Analysis*, 17, 133-148.
- Quinn, L., & Haberman, M. (1986, Fall). Are GED certificate holders ready for postsecondary education? *Metropolitan Education*, 2, 72-82.
- Richardson, R.C., Jr., Fisk, E.C., & Okun, M.A. (1983). *Literacy in the open-access college*. San Francisco: Jossey-Bass.
- Romberg, T., & Carpenter, T. (1986). Research on teaching and learning mathematics: Two disciplines of scientific inquiry. In M.C. Wittrock, (Ed.), *Handbook of Research on Teaching*, (Third edition, pp.850-873). New York: McMillan.
- Rosenbaum, J. (1998, October). *Unrealistic plans and misdirected efforts: Are community colleges getting the right message to high school students?* Occasional Paper. New York: Community College Research Center, Teachers College, Columbia University.
- Roueche, J. E., & Roueche, S. D. (1993). *Between a rock and a hard place: The at-risk student in the open-door college*. Washington, DC: American Association of Community Colleges.
- Rouse, C. (1995). Democratization or diversion: The effect of community colleges on educational attainment. *Journal of Business and Economic Statistics*, 13, 217-224.
- Russell, A. (1998, January). *Statewide college admissions, student preparation, and remediation policies and programs*. Denver: State Higher Education Executive Officers.
- Schultz, K. (1997). Discourses of workplace education: A challenge to the orthodoxy. In G. Hull, (Ed.), *Changing Work, Changing Workers: Critical Perspectives on Language, Literacy, and Skills*. Albany: State University of New York Press.
- Shaughnessy, M. (1977). *Errors and expectations : A guide for the teacher of basic writing*. New York: Oxford University Press.

-
- Sticht, T. (1979). Developing literacy and learning strategies in organizational settings. In H. O'Neill, (Ed.), *Cognitive and Affective Learning Strategies*. New York: Academic Press.
- Sticht, T., Armstrong, W. B., Caylor, J. S., & Hickey, D. T. (1987). *Cast-off youth: Policy and training methods from the military experience*. New York, NY: Praeger Publishers.
- Student outcomes research project, final report*. (1996, December). Division of Adult and Continuing Education, City University of New York.
- Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition*. Chicago: University of Chicago Press.
- Tinto, V. & Goodsell-Love, (1995). *A longitudinal study of learning communities at LaGuardia Community College*. Washington, DC: National Center on Postsecondary Teaching, Learning, and Assessment, Office of Educational Research and Improvement, U.S. Department of Education. (ERIC Document Reproduction Service No. ED 380 178)
- Tinto, V., Goodsell-Love, A., & Russo, P. (1994). *Building learning communities for new college students: A summary of research findings of the Collaborative Learning Project*. Washington D.C.: National Center on Postsecondary Teaching, Learning, and Assessment, Office of Educational Research and Improvement, U.S. Department of Education.
- Tinto, V., Russo, P., & Kadel, S. (1994). Constructing educational communities: Increasing retention in challenging circumstances. *AACC Journal*, 64, 26-29.
- Tokina, K. (1993). Long-term and recent student outcomes of freshman interest groups. *Journal of the Freshman Year Experience*, 5, 7-28.
- Tokina, K., & Campbell, F. (1992). Freshman interest groups at the University of Washington: Effects on retention and scholarship. *Journal of the Freshman Year Experience*, 4, 7-22
- Traub, J. (1994, Sept. 19). Class struggle. *New Yorker*, 70, 76-90.
- Weisberg, A. (1988). *Computers, basic skills, and job training programs: Advice for policymakers and practitioners*. New York, NY: Manpower Demonstration Research Corporation.
- Weinstein, C, Dierking, D., Husman, J., Roska, L., & Powdrill, L. (1998). The impact of a course on strategic learning on the long-term retention of college students. In J. Higbee and P. Dwinnel (Eds.), *Developmental Education: Preparing Successful College Students*, pp. 85-96. Columbia, SC: National Research Center for the First-Year Experience and Students in Transition.

Worthen, H. (1997). *Signs and wonders: The negotiation of literacy in community college classrooms*. Unpublished doctoral dissertation, School of Education, University of California, Berkeley.