

## DOCUMENT RESUME

ED 076 027

EM 011 023

AUTHOR Hammer, Mark; Henderson, C. Obert  
 TITLE Improving Large Enrollment Undergraduate Instruction With Computer Generated, Repeatable Tests.  
 INSTITUTION Washington State Univ., Pullman. Bureau of Economic and Business Research.  
 PUB DATE 72  
 NOTE 21p.; Reprint from Proceedings of 1972 Conference on Computers in Undergraduate Curricula, Atlanta, June 12-14, 1972, pp. 209-216

EDRS PRICE MF-\$0.65 HC-\$3.29  
 DESCRIPTORS Administrator Education; College Instruction; \*Computer Assisted Instruction; Enrollment Trends; Feedback; Flexible Progression; \*Higher Education; Individualized Instruction; \*Large Group Instruction; Objective Tests; \*Test Construction; \*Testing; Testing Problems; Test Reliability  
 IDENTIFIERS CGRT; \*Computer Generated Repeatable Testing

## ABSTRACT

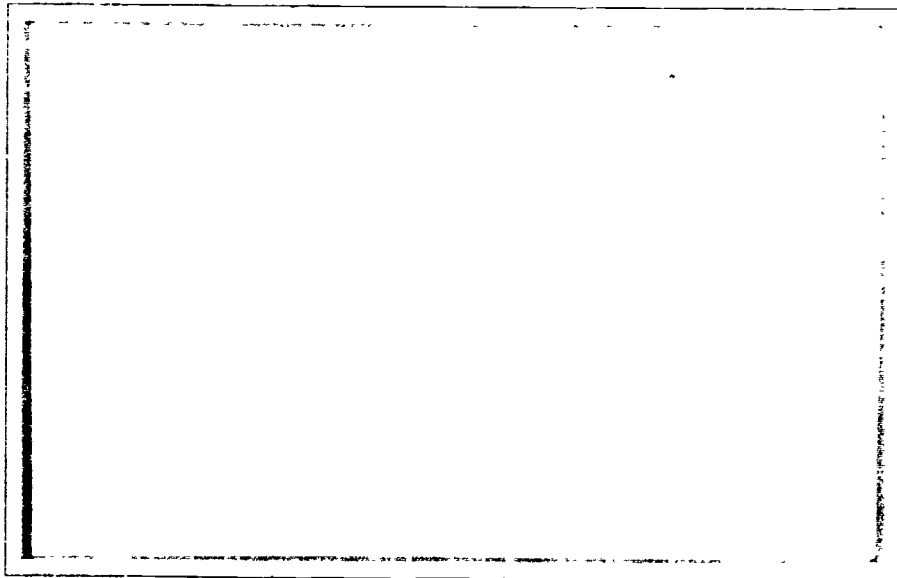
In addition to the usual drawbacks of large enrollment college classes such as impersonal atmosphere, discouragement of questions, and insensitivity to individual differences, the testing methods that tend to be associated with large classes can be detrimental to the learning process. Objective tests decrease the level of intellectual mastery required from recall to mere recognition, tend to be used as evaluative devices rather than as learning devices, provide slow feedback, and encourage a loafing-cramming approach to course subject matter. Donald Jensen's computer generated, repeatable testing (CGRT) attempts to overcome these difficulties by providing frequent tests with immediate feedback, flexible scheduling, test forms, and a method of coding fill-in responses. An attempt was made to implement a CGRT type system for an introductory Personnel Administration course. Student attitudes towards the course and their performance were both very good, although there did seem to be some problems with unreliability of the tests. Some additional implications of CGRT and possibilities for the future are also discussed. (RH)

FORM 55 0

PRINTED IN U.S.A.

ED 076027

**Bureau of Economic and Business Research**



**College of Economics and Business**

**WASHINGTON STATE UNIVERSITY**

**Pullman, Washington 99163**



EMO 11 023

FILMED FROM BEST AVAILABLE COPY

ED 076021

U S DEPARTMENT OF HEALTH  
EDUCATION & WELFARE  
OFFICE OF EDUCATION  
THIS DOCUMENT HAS BEEN REPRO  
DUCED EXACTLY AS RECEIVED FROM  
THE PERSON OR ORGANIZATION ORIG  
INATING IT. POINTS OF VIEW OR OPIN  
IONS STATED DO NOT NECESSARILY  
REPRESENT OFFICIAL OFFICE OF EDU  
CATION POSITION OR POLICY

IMPROVING LARGE ENROLLMENT UNDERGRADUATE INSTRUCTION

WITH COMPUTER GENERATED, REPEATABLE TESTS

by

Mark Hammer and C. Robert Henderson

Department of Business Administration  
Washington State University  
Pullman, Washington 99163  
Phone 509-335-3507

Proceedings of the 1972 Conference on Computers in  
Undergraduate Curricula, Atlanta, June 12-14, 1972;  
Southern Regional Education Board, pp. 209-216.

IMPROVING LARGE ENROLLMENT UNDERGRADUATE INSTRUCTION  
WITH COMPUTER GENERATED, REPEATABLE TESTS

I. The Problem

Large enrollment classes are increasingly characteristic of undergraduate education, most especially for introductory, freshman-sophomore level courses. This trend toward larger lecture courses has been accelerated by recent budgetary squeezes and the resulting pressure for improved academic productivity.

When compared with small classes, large enrollment classes have several serious disadvantages. For one, they are impersonal: Socratic dialogue is impossible, classroom questions are disruptive, and personal acquaintance with instructors is discouraged. For another, they are insensitive to individual differences: large lectures must be aimed at the "average" student, with detrimental consequences for both fast and slow learners.

Perhaps the most serious of the large-enrollment disadvantages are those surrounding the examination procedures which are forced upon instructors by sheer class size. For example, essay examinations are all but precluded by the impossibility of the grading task they impose. The typical substitution of "objective" (true-false, multiple-choice) tests for essay tests tends to reduce the intellectual rigor of the course by changing the required level of learning from mastery (recall level) to familiarity (recognition level).

In addition to changing the level of learning required, the imperatives of large enrollment instruction effectively force a change in the educational role of the test itself. Tests in small classes may be utilized primarily as learning devices which provide both student and instructor with diagnostic information on the student's level of understanding. Tests in large classes, however, are harder to utilize as learning devices. The essay exams, frequent quizzes, in-class recitation, and rapid feedback which are possible in smaller classes are effectively precluded for use in large lecture sections; large-class tests are much more likely to be infrequent (two or three major tests per semester), to cover correspondingly larger blocks of subject matter, and to have longer feedback periods (if the tests are returned at all; finals frequently are not). The cumulative result is that large-class examinations are used for evaluation rather than for diagnosis, and the potential value of the test as a learning device is forfeited. The common practice of posting test grades while not returning tests themselves confirms the exclusively evaluative role of the examination process.

Finally, large-enrollment tests are likely to be aversive (anxiety-arousing: dissatisfying) to students. Several factors are responsible for this aversiveness. First, the study habits of students are commonly observed to follow a "loaf-cram" pattern, with crams coming just before tests. Second, when exams are infrequent, the subject matter to be learned during one cram is greater. Third, the "perform now or never" nature of the test situation, coupled with intense emphasis on grades, creates a high-tension situation for the student. Neither the loaf-cram study schedule nor the pre-exam anxiety are conducive to effective learning.

## II. Computer Generated, Repeatable Testing:

### A Promising Development

The limitations of large-enrollment instruction have been systematically assessed by psychologist Donald Jensen, who has proposed and evaluated a variety of potential solutions (Jensen, 1966, 1968, 1969; Jensen and Prosser, 1969). The most promising of Jensen's approaches to date is computer generated, repeatable testing (CGRT).

CGRT encompasses several important changes from typical large-class testing procedures (Prosser & Jensen, 1971). First, tests are given more frequently, typically biweekly. Second, students are allowed to schedule tests at their own convenience, within broad limits. This is made possible by the provision of multiple test forms. Third, immediate feedback is provided on test performance; students are given the correct answers to test questions as soon as they have completed a test. Fourth, students can repeat tests until they earn a grade which satisfies their aspirations. Finally, testing for mastery (recall) is possible through the use of a procedure for coding responses to fill-in questions (Prosser & Jensen, 1971, p. 297).

The procedure used in CGRT to accomplish these changes is to prepare a large number of test questions for each subject matter segment of the course and to read them into a computer. The computer is programmed to generate independent test forms, each of which contains a stratified random sample of questions from the bank in computer storage. Thus, literally hundreds of tests can be generated with no two being the same. Having pre-printed a supply of tests on the computer (in batch mode), a testing room is scheduled to be available for convenient hours during the exam week. Students may come in when they feel most ready, take an exam, get immediate feedback, and

return to do additional studying if their first score is not satisfying. Prosser and Jensen (1971, p. 301) have reported that CGRT has been successfully implemented in several institutions in a variety of subject areas including psychology, economics, accounting, chemistry, speech therapy, and English. Among the benefits said to be associated with these implementations are higher student achievement, lowered anxiety and antagonism surrounding examinations, and better attitudes generally toward both subject matter and instructors.

### III. Implementing CGRT: Our Experience

The theory behind CGRT made sense to us, and we had heard favorable reports on the effects of repeatable testing from Jensen and others. We decided that it was worth a trial run and agreed to attempt it. Since both of us anticipated teaching one section of an introductory Personnel Administration course, we agreed to cooperate in developing CGRT for both sections. These decisions were made in the early summer of 1971, and we aimed for Fall semester 1971 implementation.

#### Creating the Test Bank

The first obstacle to be contended with was the required bank of test questions. Prosser and Jensen (1971) reported that the number of test questions available for any one test should exceed the number of questions on that test by six to ten times to assure adequate variation among the test forms. More recently Jensen has said that a 10 to 1 ratio is a desirable minimum (personal communication). Prosser and Jensen also noted (correctly!) that the preparation of this number of test questions is a formidable task.

Since we did not have enough time to create a complete test bank before the beginning of the fall semester, we adopted a text which had a fairly large number of accompanying objective test questions. Some of these test questions were contained in an instructor's manual and some were in a student workbook which was available to accompany the text. We adopted the workbook and included the questions from it in the question bank, thereby providing students with pre-exposure to a number of questions over the text as well as with motivation to utilize their workbooks as study aids. The task of supplementing the questions accompanying the text and of preparing questions over class lectures was divided among ourselves and a teaching assistant.

#### Obtaining Computer Programs

The second obstacle to be overcome in order to implement CGRT was obtaining the computer capability needed. We initially anticipated using the system developed at Indiana University by Prosser and Jensen (1971), but two problems developed. First, a telephone conversation with Jensen convinced us that it would probably take as much programming time to convert the Prosser-Jensen system to our computer (IBM 360-67) as to develop our own from scratch. Second, we had wanted to improve on the Prosser-Jensen system in several respects, the most important one being the capacity to stratify the test bank by test item type. Without such a stratification the proportion of question types on any given test could vary randomly: the number of true-false items on a given 20-question test might vary, for example, from 7 on one test to 14 on another. In the interest of achieving uniform difficulty among test forms, we felt that each form should have the same proportion of question types.



We finally decided to create our own CGRT system. Being short on both time and money, we decided to program only the test generation capability, and to postpone the mark-sense scoring and computer tallying capabilities which are part of the Prosser-Jensen system. After specifying the capacities of the program we wanted, we located a computer programmer who agreed to write the programs for \$300.00. To our programmer's credit and to our delight, the resulting programs have functioned flawlessly throughout their first semester of operation. A sample test is shown in Figure 1, which provides an idea of the format of the tests generated by these programs.

#### Developing Policies and Procedures

For testing purposes the 14-week semester was divided into seven two-week units, and a test scheduled for each unit. Students were allowed to take a maximum of three (later changed to four) tests during a six-day period from Wednesday of the second week of the unit through the following Monday. This testing interval covered the period from the last lecture of a unit until the first lecture of the next unit (students had two lectures and one small discussion group weekly).

A testing room was manned by an instructor or an assistant for six different scheduled periods, including one period Saturday evening and another Sunday evening. Testing room procedure called for a student to sign for a test in a log book and to indicate there his discussion section and the form number of the test he received. Upon completing the test, the student would cut the "Responses" column from the test questions (Figure 1) with a pair of scissors provided, and hand it to the instructor on duty.

Figure 1

NAME SAMPLE PERSONNEL TEST FORM NO. 10	QUESTIONS	RESPONSES	ANSWERS	NAME SAMPLE PERSONNEL TEST FORM NO. 10
	TRUE OR FALSE			
	1. A JOB EVALUATION PROGRAM CAN CONTRIBUTE TO THE IMPROVEMENT OF EMPLOYEE SAFETY WITHIN A COMPANY.	1. ITEM 6 1 1 003 ANSWER: T	1. ITEM 6 1 1 003 T C-S PAGE 546.	
	2. THE CONSTITUTIONALITY OF THE FAIR LABOR STANDARDS ACT WAS UPHELD BY THE UNITED STATES SUPREME COURT IN THE MT. CLEMENS POTTERY CASE.	2. ITEM 6 2 1 009 ANSWER: F	2. ITEM 6 2 1 009 F C-S PAGE 589.	
	3. IF MOST GROUP INCENTIVE SYSTEMS THE INCENTIVE PAYMENTS ARE NOT SOLELY BASED UPON UNITS OF PRODUCTION.	3. ITEM 6 3 1 009 ANSWER: T	3. ITEM 6 3 1 009 T C-S PAGE 644.	
	4. THE PRODUCTION LIMITS THAT A GROUP ESTABLISHES FOR ITS EMPLOYEES IS KNOWN AS A "BUDGET."	4. ITEM 6 4 1 002 ANSWER: T	4. ITEM 6 4 1 002 T C-S PAGE 601.	
	MULTIPLE CHOICE			
	5. THE PROFILE METHOD OF JOB EVALUATION: A. INVOLVES THE USE OF THE MANAGEMENT GRID B. IS USED TO EVALUATE CLERICAL JOBS C. CONTAINS CERTAIN FEATURES OF THE FACTOR COMPARISON METHOD D. PERMITS THE EVALUATION OF BOTH THE EMPLOYEE AND HIS JOB	5. ITEM 6 1 2 003 ANSWER: C	5. ITEM 6 1 2 003 C C-S PAGE 563.	
	6. EFFECTIVE FEBRUARY 1, 1968, THE MINIMUM WAGE FOR THOSE WORKING IN EMPLOYMENT COVERED PREVIOUSLY BY THE FAIR LABOR STANDARDS ACT BECAME: A. \$1.20 B. \$1.40 C. \$1.90 D. \$1.80	6. ITEM 6 2 2 004 ANSWER: C	6. ITEM 6 2 2 004 C C-S PAGE 587.	
	FILL-IN			
	7. THE _____ SYSTEM OF JOB EVALUATION PERMITS JOBS TO BE CLASSIFIED AND GROUPED ACCORDING TO A SERIES OF PREDETERMINED WAGE CLASSES OR GRADES. THE FEDERAL CIVIL SERVICE SYSTEM IS AN EXAMPLE OF THIS TYPE.	7. ITEM 6 1 3 005 ANSWER: O	7. ITEM 6 1 3 005 O JOB GRADE PAGE 548.	
	8. WHEN A SALESMAN IS ADVANCED MONEY WHICH MUST BE REPaid OUT OF SURVEY COMMISSIONS, HE IS SAID TO BE ON A STRAIGHT COMMISSION AUGMENTED WITH A _____.	8. ITEM 6 1 3 008 ANSWER: W	8. ITEM 6 1 3 008 W DRAWING ACCT. P 610.	
	9. THE "WALDEN" AND THE "ELEVATION" ARE TWO CHARACTERISTICS OF A _____.	9. ITEM 6 2 3 005 ANSWER: V	9. ITEM 6 2 3 005 V WAGE CURVE LEC 11-12.	
	10. UNDER THE _____ SYSTEM OF INCENTIVES, THE AMOUNT OF WAGES IS DIRECTLY PROPORTIONAL TO OUTPUT.	10. ITEM 6 3 3 002 ANSWER: C	10. ITEM 6 3 3 002 C PIECEWORK PAGE 604.	

END OF TEST FORM NO. 10 SAMPLE PERSONNEL TEST

The instructor would take the "Answers" column, which had been previously cut off, line up the correct answers with the student's responses, and grade the student's test. This grade was then marked on both the "Responses" column, which was kept for recording, and on the "Answers" column, which was returned to the student.

Having agreed that an arbitrary, pre-established criterion schedule for grading was preferable, the use of grade "curving" we adopted a fairly exacting standard, viz., 95%+ = A, 90%+ = B, 85%+ = C, 80%+ = D, and below 80% = F. We assured ourselves that students could be expected to attain levels higher than are typically demanded because: a) some of the test questions used were taken from their workbook, giving them pre-exposure to some items, b) up to half of the test questions were of the True-False type, and c) any chance variation in test difficulty worked in the students' favor since only the highest test score was counted. Even with these considerations the grading standards seemed to us plenty rigorous, but we reasoned that we could be lenient in final grading if they turned out to be too demanding.

#### IV. Results

##### Student Attitudes

Twice during the semester feedback was solicited from students on several aspects of CGRT. The first set of student ratings was obtained in the fifth week of the semester, which was just after the second CGRT unit test; the second set was gathered in the thirteenth week, after the sixth test. Both sets asked for open-ended comments on several specific goals and mechanics of the CGRT technique, as well as an overall evaluation of CGRT.

The open-ended responses were favorable overall, with two exceptions. Specifically, the answering students practically all had favorable responses to inquiries on fairness in evaluation and grading, repeatability, frequency, student-scheduling, and availability of immediate feedback on performance. There was also substantial agreement on two criticisms of our CGRT program: test unreliability and excessively high grading standards. Both of these criticisms will be discussed below.

For the overall evaluation, students were asked on both occasions to rate CGRT "in comparison with other testing procedures you have seen" on a 7-point scale from "much worse" to "much better." The student responses are summarized in Table 1. On the average, students rated CGRT "slightly better" on both occasions (mean scores were 5.0 and 4.8 respectively). However, Table 1 shows that the distribution of ratings shifted from the first to the second evaluation; while the modal response decreased from

Table 1  
Student Ratings of CGRT vs. Conventional Tests

Percent (Number Responding)

CGRT Rating	Much Worse	Considerably Worse	Slightly Worse	Average	Slightly Better	Considerably Better	Much Better
After 2nd CGRT Test (N=71)	7.0 (5)	5.6 (5)	7.0 (5)	7.0 (5)	36.6 (26)	11.3 (8)	
After 6th CGRT Test (N=64)	1.6 (1)	3.1 (2)	7.8 (5)	9.4 (6)	31.2 (20)	28.2 (18)	18.8 (12)

"considerably better" to "slightly better," the number of "much worse" and "considerably worse" ratings decreased and that of "much better" ratings increased. Additionally, students were asked on the second evaluation occasion to indicate whether they would choose a class with a) CGRT or b) Conventional testing, if all other things were equal. Answering students, 49 (or 77%) chose CGRT.

Test Performance

Student performance on tests has exceeded our expectations. Table 2 shows the grade distributions for each of six tests that have been administered to date, as well as for the six-test average grades. There seems to be a general trend toward higher grades, and after six tests the distribution of average grades is skewed upward with a distinct mode at the "B" grade level.

Table 2  
CGRT Grade Distributions for Six Biweekly Tests  
Number of Students (N=81)

Grade	Test						Six-Test Average
	1	2	3	4	5	6	
A	25	30	38	46	36	36	13
B	29	26	26	26	26	22	40
C	13	12	12	6	11	16	20
D	9	9	3	2	6	4	5
F	5	4	2	1	2	3	3

Our initial expectation was that our achievement standards might have been too high. Our early doubts were amplified by student responses on the first questionnaire; many students complained that our standards were too high and unrealistic. However, after six tests, almost two-thirds of the students have averages of "B" or better. It appears to us that the distribution of final grades will be higher than the distributions either of us has seen recently in this course.

However, our standards may be too high. It is quite clear to us that the higher grades reflect a considerably higher level of effort on the students' part. We asked students on the first questionnaire how much time they were spending on this course, and how this time compared with that spent on other courses. Of the 60 responding, 2 claimed they were spending less time compared with 54 who reported they were spending more time. Whether or not it is legitimate to utilize techniques which effectively extort a disproportionate amount of the student's study time is a question with which we have only skirmished, but which appears likely to be controversial.

We had expected some expression of resentment on the questionnaires over the increased study time which students were devoting to the course. To our surprise, the students generally expressed gratitude for being allowed the opportunity to improve their scores by repeating tests. Given the overall favorability of sentiments expressed and the pattern of test performance observed, it seems clear to us that our students are both learning more and liking it better!

In addition to improvements in performance and attitude, several other phenomena associated with CGRT deserve comment. First, it is quite evident

to us that CGRT has eliminated a great deal of the aversiveness normally associated with the testing experience. Students come to the testing room relaxed and, occasionally, in a playful mood. They frequently ask questions both before and after taking their test, and the most common response to having their tests graded is to grab their text and check on incorrect responses. In short, tests are really functioning as learning devices which stimulate further study.

A second phenomenon associated with CGRT concerns student attitudes toward the instructors, who are increasingly being viewed in a coaching role, rather than in an adversary role. Having shared the task of test item construction, and having settled on a fixed and exacting set of standards for grades, we are more prone to honestly encourage each student to do his best. When a student does well, we are elated along with him. When one does poorly, his disappointment is also ours. The students seem to sense that we are really on their side and appear more prone to relate to us as helpers.

A final phenomenon associated with CGRT is that students are becoming aware that they have direct and immediate control over their own grades. When this realization is coupled with the opportunity to repeat tests until a satisfactory grade is earned, the effect is that the student's ability to rationalize a poor test score is eliminated. We emphasize this point because we think that it may be one of the most important observations to be made in connection with our experience.

To illustrate: we suspect that a substantial number of college students having actual grade point averages of "C" or lower really prefer to think of themselves as "A" or "B" students. Professors who have observed closely the typical post-examination behavior of students will agree, however, that inferior performance doesn't necessarily threaten one's self-image. Why?

Because there are so many good, plausible explanations for poor performance: "Misleading test question," "incompetent instructor," "lousy text," "testing room too hot," "headache (didn't sleep all last night)," "my great aunt died," "my girl left me and I'm all messed up." These familiar rationalizations (and countless others) are all invoked by students to effectively convince themselves and others that they are really better students than the record indicates.

None of this nonsense is effective under CGRT, and we think that this may explain much of the increasing scarcity of C's, D's, and F's in our grade distributions. Interestingly enough, a number of the best students have shown signs of the same effect. Some seem quite incapable of settling for anything less than a perfect score. For example, students who have earned an A- (19 or 20 correct) frequently return a second and a third time in attempts to make the perfect score.

#### Test Reliability

It was mentioned earlier that test reliability was the subject of considerable student criticism. It seemed that students all too often received lower test scores in spite of greater preparation. Our perusal of the patterns of test scores confirmed that there was at least some problem, since there were occasional instances in which a student would get, for example, a B on the first test followed by a F on the second. We were therefore led to investigate the reliability problem further.

We did a check on the test-retest reliability of one of the CGRT unit tests using four different groups of students. For a class of 112 freshman and sophomore Introduction of Business students, the reliability



coefficient was .25. Reliability for a group of 23 advanced personnel students was only slightly better at .38. The highest reliability was obtained with a group of 14 MBA students, where the figure was .61. Finally, students in the present CGRT course were given two tests during one class period (both for credit), and the resulting reliability figure for these 76 students was .46.

These reliability figures were disappointingly low. The students were all too right--apparently the process of randomly selecting test questions from an item bank results in a wider variation in overall test difficulty level than we had anticipated. As a result of this information we have been thinking about ways to improve test reliability. The most promising approach now seems to us to be that of stratifying the question bank by concept, rather than by textbook chapter or by time period (e.g., Week 8). This procedure would have the effect of reducing the variance in test difficulty attributable to variance in topical coverage. We are beginning to think more in terms of clusters of fairly equivalent questions being associated with each key objective or concept to be covered. Of course, a second sure-fire way to improve reliability is to increase the test length; so far our tests have had 20 questions each. Whether or not the increased reliability of a 30-question test would offset the disadvantages of the longer test is not yet clear.

#### V. Possible Implications of CGRT

The following speculations are offered to suggest the range of potential impact possible if CGRT proves successful.

1. Many large-enrollment, introductory courses have multiple sections and multiple instructors, and it is no secret, among students at least, that substantial differences exist among sections which are attributable to different instructors. It seems to us that there are too many instances of multiple-section introductory courses where substantial differences in course content exist. Where a certain course is a prerequisite to others, or is required for a major, substantial differences among sections of multi-section courses cause untold problems for instructors of advanced courses and student advisors. Clearly, standardization of courses at the introductory level is needed.

The possibility of cooperation among instructors for the purpose of developing a test item pool for a course suggests cooperation in defining the goals for the course. It seems plausible, if not likely, that instructors should be able to reconcile whatever differences exist among themselves and agree on specific course goals and the associated test pool questions and criteria for satisfactory performance.

One interesting question suggested by the above is, "What would happen if a department were to require that instructors assigned to a certain multi-section introductory course participate in establishing a mutually acceptable set of course objectives, a test item pool, and the level of satisfactory performance?" Surely some groups of instructors could do this with little inconvenience; almost as surely some could not. However, it may be that those instances where irreconcilable differences exist are precisely those where departmental-level intervention is appropriately exercised to eliminate minority individuals or factions from teaching

the introductory course. This may sound severe, but it boils down to the reasonable proposition that introductory courses should concern themselves with consensus-level subject matter.

This should not be taken to imply that the course in question should be highly structured in either content or method: one group of instructors might, for example, decide that their "consensus topics" should constitute 25% of the course requirements, and the remaining 75% would be open to the individual instructor's preference. Furthermore, the methods used by the instructor to cover the consensus topics would be quite open.

2. If the development of consensus-level test items pools is a practicable possibility, and these were to become available for major undergraduate courses, a number of interesting advantages might be realized. For example, take a transfer student who has taken an introductory math course at another institution: is he satisfactorily prepared to begin work in advanced courses? The availability of consensus test pool would make it possible to give the student a subject matter mastery test which would pinpoint any areas of weakness.

Such tests might be useful in determining whether students should be given credit for various combinations of prior work. The effect of such a practice might well be to shift the criteria for acceptability from such arbitrary consideration as, "Was his institution accredited." or "What text did he use." or "Where and when did he take the course?" to "Does he now understand the critical concepts."

3. Another major advantage of the existence of CGRT tests would be that superior students could be invited and challenged to proceed at their own pace and to demonstrate their competence as soon as they are ready.

4. A further implication of the widespread availability of CGRT test item banks is that independent and off-campus study could be greatly facilitated. If course objectives and requirements were specified and made available along with sample CGRT tests, all eligible applicants could be invited to demonstrate their competence on any available CGRT test, and to claim credit and advanced standing for doing so.

Incidentally, CGRT tests would seem to be ideally suited to correspondence study. For one thing, numerous sample tests could be provided to the correspondent-student. For a second, the immediate feedback on test performance possible with CGRT would be a dramatic improvement over the long-delayed feedback typical of correspondence course tests. Finally, the use of the same CGRT exams being used in parallel courses on campus would insure the comparability of the two courses in subject matter coverage.

5. CGRT appears to be highly compatible with several concepts associated with the audio-tutorial approach to learning (Postlethwait, Novak & Murray, 1969). Student scheduling, repeatability, and prompt feedback from frequent quizzes are features of both. The concept of providing mini-courses and requiring learning for mastery (Bloom, 1968) suggests that CGRT test pools could be geared to mini-courses and the criterion level stated. Furthermore, specifying objectives in behavioral terms (Mager, 1962) is a step which should naturally precede preparation of the specific test bank items which operationalize those objectives.

#### VI. Further Development Anticipated

CGRT has been surprisingly successful, and we have plans for expanded implementation and for more systematic experimental evaluation. Marl Hammer has recently received a \$12,000 grant to develop a more sophisticated CGRT

has recently received a \$12,000 grant to develop a more sophisticated CGRT computer system and to give CGRT a more thorough evaluation compared with conventional testing techniques. As one result of this project, computer programs and documentation should be available by September 1972.

#### REFERENCES

- Bloom, B. S., Learning for mastery, UCLA CSEIP Evaluation Comment, 1968, Vol. L, No. 2.
- Campbell, Donald T., and Julian C. Stanley, Experimental and quasi-experimental design for research in teaching. In N. C. Gage (ed.), Handbook of Research in Teaching. Chicago: Rand McNally, 1963.
- Hammer, Mark A., C. Obert Henderson, and LeRoy Johnson, A report on four techniques for improving large enrollment instruction. Unpublished summary of report given at Northwest Universities Business Administration Conference, Portland, October 30, 1971.
- Jensen, D. D., Testing and large class instruction. Growth and Change at Indiana University, 1966, Vol. III, Section XLI.
- Jensen, D. D., A system for improved large enrollment instruction. Unpublished manuscript, Indiana University, 1968.
- Jensen, D. D., An efficient, effective, and humane system of large class instruction. Unpublished manuscript, Indiana University, 1969.
- Jensen, D. D., & F. Prosser, Computer-generated, repeatable examinations and large class instruction. Paper presented at Midwest Psychological Association Meeting, Chicago, 1969.
- Mager, Robert F., Preparing Instructional Objectives. Palo Alto, Calif.; Fearon, 1962.
- Postlethwait, S. N., J. Novak, and H. T. Murray, Jr., The Audio-Tutorial Approach to Learning, 2nd ed., Minneapolis: Burgess Publishing Co., 1969.
- Prosser, Franklin, and D. D. Jensen, Computer generated repeatable tests. AFIPS [American Federation of Information Processing Societies] Conference Proceedings, 1971, 38, pp. 295-301.