DOCUMENT RESUME

ED 111 322 HE 006 850

AUTHOR Sprenkle, Virginia E.; Gillmore, Gerald M.
TITLE An Evaluation of Mastery Instruction Courses by Use of Student Ratings.
INSTITUTION Washington Univ., Seattle. Educational Assessment Center.
REPORT NO EAC-R-75-3
PUB DATE Jul 75
NOTE 33p.

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS Achievement Rating; *College Students; Educational Experiments; Educational Strategies; *Higher Education; Instructional Design; Learning Experience; Learning Theories; Lecture; *Rating Scales; *Student Evaluation; *Teaching Methods; Teaching Procedures
IDENTIFIERS *University of Washington

ABSTRACT

Seven mastery instruction courses offered in different quarters at the University of Washington are examined. The study method involved obtaining student ratings of these courses for comparison among themselves and with student ratings of traditional lecture method courses in the same subject field. For this purpose a Mastery Instruction Student Rating (MISR) Form was devised. It was found that: (1) mastery instruction courses received stable student ratings from different student samples, Fall and Winter Quarter; (2) a majority of items (18 out of 30) on the MISR Form discriminated among the mastery instruction courses at significant levels; (3) no systematic differences in mastery instruction and traditional lecture courses in the same subject field were observed; (4) mastery instruction courses received significantly more favorable ratings on 12 of 17 student rating items when their ratings were compared to the population of University of Washington Form D student ratings. (Author/LBH)

EAC REPORTS

EDUCATIONAL ASSESSMENT CENTER

University of Washington, 1400 Campus Parkway, PB–30, Seattle, Washington 98195 (206) 543–1170

Instructional Evaluation

Educational, Psychological, and Survey Research

Test Administration, Construction, Scoring, and Analysis

Educational Assessment Center

University of Washington

July 1975

76-3

An Evaluation of Mastery Instruction Courses by Use of Student Ratings[1, 2]

Virginia E. Sprenkle

Gerald M. Gillmore

This study was designed to investigate seven mastery instruction courses offered in different quarters at the University of Washington. This method involved obtaining student ratings of these courses for comparison among themselves and with student ratings of traditional lecture method courses in the same subject field. For this purpose a Mastery Instruction Student Rating (MISR) Form was devised.

The results of this study can be summarized as follows: (1) Mastery instruction courses received stable student ratings from different student samples Fall and Winter Quarter. (2) A majority of items (18 out of 30) on the MISR Form discriminated among the mastery instruction courses at significant levels of .05 and beyond. (3) No systematic differences in mastery instruction and traditional lecture courses in the same subject field were observed. (4) Mastery instruction courses received significantly more favorable ratings on twelve of seventeen student rating items when their ratings were compared to the population of University of Washington Form D student ratings.

Educational Assessment Center Project: 503.

An Evaluation of Mastery Instruction Courses by Use of Student Ratings[1, 2]

Virginia E. Sprenkle

Gerald M. Gillmore

The reasons for current dissatisfaction with the educational process at the college level are often justified: for example, classes are typically quite large, raising the student-faculty ratio to such levels that personal contact between students and instructors is seriously hampered. In fact, the manner in which instructional material is presented, student performance is defined and evaluated, the grades are distributed has changed but little since the founding of Harvard in 1639 (Johnston & Pennypacker, 1971). In recent years, however, college students increasingly are demanding an alternative to an allegedly dehumanizing process that results from the standard three lectures a week and the machine graded multiple-choice exam; alternatives are also sought regarding the familiar grading process described by the normal curve.

Meanwhile, attempts to improve college teaching techniques and student performance have recently appeared. Notable are the efforts made by F. S. Keller (1968) and replicated and extended by others. Keller has developed a teaching technique, called the Personalized System of Instruction (PSI), which is a particular form of mastery instruction.

Several features seem to distinguish mastery learning from conventional teaching procedures. Fundamentally, in most "traditional" learning situations, achievement is based on how much one learns in a given amount of time. In contrast mastery instruction has every student learn an equivalent amount of material but allows variation in time to learn the material. For clarification, the distinctive features of this instructional method are summarized by Keller (1968) as follows:

1) The go-at-your-own-pace feature, which permits a student to move through the course at a speed commensurate with his ability and other demands upon his time.

2) The unit-perfection requirement for advance, which lets the student go ahead to new material only after demonstrating mastery of that which preceded.

3) The use of lectures and demonstrations as vehicles of motivation, rather than sources of critical information.

4) The related stress upon the written word in teacher-student communication; and, finally;

5) The use of proctors, which permits repeated testing, immediate scoring, almost unavoidable tutoring, and a marked enhancement of the personal-social aspect of the educational process. (p. 83)

The proponents of this instructional method show much enthusiasm for its success in teaching (Block, 1971; Bloom, 1971; Ferster, 1968, Johnston & Pennypacker, 1971; Keller, 1968; McMichael & Corey, 1968), yet, to date, the amount of published data from the research on the use of mastery instruction in college courses is small. The purpose of this study was to evaluate mastery instruction as practiced in seven undergraduate courses at the University of Washington by use of student opinion. Subsequent reports (Anderson, Note 1) will deal with student achievement. A secondary purpose was to develop a rating form, patterned after the forms of the University of Washington Instructional Assessment System, which would discriminate among mastery instruction courses. The seven mastery instruction courses at the U of W are similar in concept to those reported by Keller (1968), Ferster (1968), and McMichael & Corey (1969). Details of the classroom format and instructional procedures of these courses are reported subsequently.

PREVIOUS EVALUATIONAL STUDIES

Some descriptions of research studies done in mastery learning courses may clarify the important effects of this new instructional approach. The results of over forty major studies, as summarized by Block (1971) indicate that mastery learning has markedly positive effects on student cognitive and affective development and their learning rate. In general, mastery strategies enable about three-fourths of students to

learn to the same performance standards as the top fourth of students learning under conventional, group-based instructional approached. Studies also indicate that students with average I.Q. scores seemed to learn as well with mastery instruction as students with above average I.Q. scores under a traditional approach (Block, 1971). Furthermore, where most students have achieved the prerequisite learning, mastery instruction is able to almost eliminate the effects of individual differences in achievement level (Block, 1971; Keller, 1968). However, most of these studies were carried out at the pre-college level.

On the college level, fewer studies using mastery strategies have appeared, probably due to the dearth of college courses taught by this instructional method. Keller (1968) found over a period of two years with 400 general psychology students that 65 per cent to 70 per cent of the students received A's or B's, and that each time the course was repeated using these learning principles it produced a larger percentage of A's or B's, and fewer D's and F's. Keller stressed the importance of the "personal-social interaction" component in mastery instruction as a key feature for the course's high rate of student achievement. Ferster's (1968) study with 79 introductory psychology students confirmed Keller's observations of the success of this instructional method with college students: 90 per cent received A's, 4 per cent B's, and 6 per cent C's. In addition, Ferster found that the students' test results and comments on the frequent unit exams often suggested defects in his learning materials which he promptly revised. Another advantage of this instructional strategy encouraged the instructor to seek out students who were having difficulty or lagging behind since a close tab was kept on all students' progress from the frequent testing.

More recent studies (McMichael & Corey, 1969; Johnston & Pennypacker, 1971; Green, 1969, cited in Block, 1971, pp. 121-122) followed the same strategy as Keller (1968) and Ferster (1968) but in addition, compared mastery instruction to traditional lecture courses. Green (1969) found that 150 undergraduates in freshman physics not only enjoyed the course but performed as well on a final examination as students learning under the "traditional lecture-discussion-demonstration approach." Green agreed with Keller's appraisal that the student proctors

added a valuable personal-social dimension to the course. Similarly, McMichael and Corey (1969) found that the approximately 200 students in the experimental, mastery learning group rated the course more favorably than did comparable students taught by conventional methods. These investigators found upon direct comparison of final examination scores between the groups taught the same subject material that the mastery instruction group received higher scores.

The student ratings of two classes based upon the Keller method (Johnston & Pennypacker, 1971) were also favorable. All students in both sections agreed that they would prefer to take other courses taught according to the mastery plan. Furthermore, they all expressed the belief that the course material had been better learned via this method as compared to the "typical" college approach. These students also seemed to appreciate being given, at the beginning of the course, a complete description of all criteria by which their efforts were being judged for grading purposes. As one function of mastery instruction advance criteria enable the student and professor to know, at any moment, where the student is with respect to the goals of the course. Johnston & Pennypacker (1971) conclude their report with the statement:

> The single most conspicuous effect of the instructional techniques described here is their overwhelming popularity among the students. There have been almost no data to the contrary, despite the increased time and effort currently required by these procedures and the high standards for successful performance which have been set compared to more traditional courses. Indications of such popularity come from every source--from questionnaires, through the managers (student proctors), from other professors, . . . and from the students themselves. (p. 244)

Studies reviewed thus far have not indicated the presence of any problems with the mastery instruction method. Yet, on the University of Washington campus, some professors and students have expressed misgivings about the method. One unusual feature and possible disadvantage

of mastery instruction is that the professor becomes more of a manager than a teacher. Students learn the subject matter from manuals and student proctors. Some students desire more contact with their professors rather than the proctors for obtaining such information and for answering their questions. Furthermore, the self-pacing element in mastery instruction allows students to delay beginning and completing the units. As a result, some students do not complete the course in one quarter. Other students have mentioned that courses taught by mastery instruction require more preparation hours and take time away from their other work.

HYPOTHESES

This study was designed to investigate mastery instruction at the University of Washington by means of testing the following hypotheses.

1. There will be no systematic difference in student ratings of the same course from one quarter to the next.

2. The Mastery Instruction Student Rating (MISR) Form will discriminate among mastery instruction courses.

3. Differences will be found between the student ratings of mastery instruction sections and traditional lecture sections teaching the same course.

4. Differences will be found between the student ratings of mastery instruction courses and the student ratings of all courses using the comparable, standard University of Washington student evaluation form.

METHOD

The Sample

The subjects for this study were college undergraduates, the exact number of which was dependent upon the number of student rating forms returned from the student survey. One hundred and eighty-three (183) subjects responded to the MISR Form in seven mastery instruction courses of the University of Washington; 255 students returned the University of Washington Student Rating Form D from four lecture courses which dealt with the same subject material as 5 of the mastery instruction courses, Table 1 shows both the academic area of each course and the number of student surveys received from each course at the end of Autumn and Winter Quarters.

From the total number of student ratings received an even split between underclassmen and upperclassmen is represented.

Table 1

Courses and Number of Subjects

Mastery instruction (Experimental group)			Traditional courses (Control group)		
Course	Quarter	Number of student ratings used	Course	Quarter	Number of student ratings used
Computer-A	Autumn	21	Computer	Autumn	20
Computer-W	Winter	17	Computer	Winter	22
Biology-A	Autumn	32	Biology	Autumn	89
Biology-W	Winter	47	(Only mastery instruction section taught Winter Quarter)		
Engineering-A	Autumn	36	(No comparable Engineering course available)		
Engineering-B	Winter	13			
Physics-W	Winter	17	Physics	Winter	124
TOTAL		183	TOTAL		255

The control group of four sections taught by the conventional lecture method covered the same subject matter as five of the mastery instruction courses. The selection of the control group was random with the provision that the instructors have used Form D for their course evaluations. The number of sections offered for each course was small, ranging from one to four sections, except for the computer course where there were 8 sections. Hence, there was often little choice in the selection of the comparison group. The section to be matched with the mastery instruction computer course was selected somewhat differently, however. The instructor of the control group had agreed to give the same final examination as used by the mastery instruction section, the results of which will be reported elsewhere (Anderson, in preparation). Since performance data were compared on these two sections, the student ratings were compared also.

The seven mastery instruction courses differed mainly in the amount of self-pacing allowed the students and the emphasis upon lectures. Otherwise, all the courses shared the following features of mastery learning based on Keller's (1968) plan:

a. criterion referenced tests
b. frequent testing
c. repeated opportunities to pass the unit tests
d. required mid-term and final examinations
e. few lectures
f. learning materials handed out to students and/or a manual which teaches most of the subject matter
g. student proctors who administer and correct the mastery unit tests
h. some self-pacing

Table 2 has been diagrammed as a composite of the available information concerning the mastery instruction courses in the study.

It is notable that six weeks into Fall Quarter the biology course instructor switched to mastery instruction, because the professor only had mastery instruction materials ready for the last half of the course. As a result, student ratings were given twice; at five weeks the U of W student rating for lecture courses was given and at ten weeks the MISR form for rating the five weeks of mastery instruction.

The Instrument

The Mastery Instruction Student Rating (MISR) Form was purposely designed in accordance with the present University of Washington student ratings format. (See Gillmore, Note 1, for specific information on the development of the U of W Instructional Assessment System.) Consequently, the MISR Form was designed in order to be optically scanned for data collection purposes and follow the three-section format of the five other U of W Student Rating Forms. The comparable, standard University of Washington evaluation form used in this study by the traditional lecture courses is Form D.

The items in section 1 are global in nature and appear on all the University of Washington Student Rating Forms, including the MISR Form,

Table 2

Description of Mastery Instruction Courses

A	A'	B	B'	C	D	E	
yes	yes	yes	yes	yes	yes	yes	criterion reference tests
must take a test every Tuesday		a test whenever students want; but all take mid-term at same time		a test whenever students want; must have finished unit			frequent testing
2	2	3	3	3	3	3	repeated chances to pass unit test
yes	yes	yes	yes	yes	yes	yes	mid-term and final exam required
occasionally, up to prof.		generally once a week		2-3 a q'ter	upon demand	none	a few lectures
yes	yes	yes	yes	yes	yes	yes	learning material or manual primary conveyor of subject matter
yes	yes	yes	yes	yes	yes	yes	student proctors giving test, correcting, and tutoring
A unit test every week, make-up test two days later		midterm & final given to all at same time, so students not allowed to plan entire schedule		allows the most self-pacing. A student can do as many units as he wishes			variable amount of self-pacing
depends on how many units pass and grades on mid-term and final examinations							determination of grade
yes	yes	yes	yes	yes	yes	yes	final exam
one	one	one	one	more than one		more than one	forms of final exam

A A computer programming course taught Autumn Quarter, 1974
A' The same computer programming course taught Winter Quarter, 1975, by the same professor as A
B An introductory biology course taught Autumn Quarter, 1974
B' The same introductory biology course taught Winter Quarter, 1975, by the same professor as B
C An electrical engineering course taught Autumn Quarter, 1974
D The same electrical engineering course taught Winter Quarter by a different professor from C
E A physics course taught Winter Quarter, 1975

except item 4 which has been changed to fit mastery instruction more closely (see Table 3). The purpose of these first 4 items is to provide a general indication of students' attitudes toward the course as a whole, the content of the course, and the instructor's overall contribution and effectiveness. The inclusion of these items on all forms allows comparisons to be made university-wide, college-wide and department-wide.

Likewise, the seven items in section 3 are included on all U of W Student Rating Forms, including the MISR Form (see Table 4). These items are intended to provide student information for course selection. Accordingly, the results are published at the end of each quarter (with the instructor's permission).

Section 2 on the U of W Student Rating Form contains diagnostic type items, tailored for different course types: large lecture, small lecture-discussion, seminar discussion, problem-solving, skill acquisition, and now, mastery instruction. The 11 items in this section are designed to help instructors discover weaknesses in their courses and teaching with the implicit purpose of facilitating course improvement. These 11 items on the MISR Form focus on the key characteristics of mastery instruction: the learning material/manual, the self-pacing, the availability of student proctors, and the frequent tests.

Table 3

Items Within Section 1

Items common to both forms:

1. The course as a whole was:
2. The course content was:
3. The instructor's contribution to the course was:

Items unique to Form D:

4. The instructor's effectiveness in teaching the subject matter was:

Items unique to MISR:

4. The Proctor's contribution to the course was:

Table 4

Items Within Section 3

(Common to Both Forms)

16. Use of class time was:
17. Instructor's interest in whether students learned was:
18. Amount you learned in the course was:
19. Relevance and usefulness of course content is:
20. Evaluative and grading techniques (tests, papers, projects, etc.) were:
21. Reasonableness of assigned work was:
22. Clarity of student responsibilities and requirements was:

The specific items for the rating form were generated by identifying and defining these important components of the mastery instruction courses at this university. Professors of these courses were given copies of the items and requested to comment on the items' content, scope, and appropriateness for assessing mastery instruction. Content validity was assumed for two reasons: first, by the plan and procedures used to construct the items, and secondly, since the potential users of the student rating information, the instructors, agreed on the appropriateness of the items.

Of the 11 items in section 2 on the MISR Form, three are identical to Form D items numbers 5, 14, and 15. Four other items (numbers 6, 7, 9, and 12) represent similar concepts but focus on the learning material or student proctors and instructors rather than on just the instructor as does Form D. Finally, four items (numbers 8, 10, 11, and 13) are unique to the MISR Form (see Table 5). The similar items on Form D and the MISR Form were intended for comparative purposes.

Section 4 deals exclusively with mastery instruction items. It consists of 8 items which focus on how the students feel about mastery instruction as a method in comparison to the traditional lecture courses they have had (see Table 6). Normally, this section is for optional items

Table 5

Items Within Section 2

Items common to both forms:

5. Course organization was:
14. Answers to student questions were:
15. Availability of extra help when needed was:

Items similar on both forms:

Form D

6. Sequential presentation of concepts was:
7. Explanations by instructor were:
9. Instructor's use of examples and illustrations was:
12. Instructor's enthusiasm was:

MISR

6. Sequential presentation of concepts by learning material was:
7. Clarity of explanations by learning materials was:
9. Examples and illustrations in the learning material were:
12. Instructor's and TAs' enthusiasm was:

Items unique to Form D:

8. Instructor's ability to present alternative explanations when needed was:
10. Quality of questions or problems raised by instructor was:
11. Contribution of assignments to understanding course content was:
13. Instructor's ability to deal with student difficulties was:

Items unique to MISR:

8. Helpfulness of discussion with TA after test was:
10. The high criterion to pass the test was:
11. The large number of tests given was:
13. Having the opportunity to repeat tests on a given unit was:

Table 6

Items Within Section 4

(MISR Only)

23. If given the opportunity, what are the chances you would take another course taught by this method of instruction?
24. Compared to the traditional lecture method, the personal contact between the instructor (and T.A.) and the students used in this course is:
25. The consistency between the learning materials and the tests was:
26. In general, compared to the traditional lecture method, this method of instruction is:
27. Did you find this instructional method to be a more efficient way of learning than the traditional lecture method?
28. Compared to other classes did you put more time and effort into this class?
29. Did you feel your learning significantly improved by studying between the first and later test on the same unit?
30. Would you have preferred more lectures in this course?

on the University of Washington Student Rating Forms. Thus, about half of the MISR Form is similar to Form D which is approximately the proportion of similarity and dissimilarity among all the U of W Student Rating Forms.

Six response categories were used for the first 26 items: Excellent, Very Good, Fair, Poor, Very Poor. The last four items (27-30) used the following five response categories: Very Much So, Probably So, Average, Probably Not, Not at All..

A pilot study was conducted Fall Quarter. No problems with the MISR Form were evident, although a few words were changed to make some statements more appropriate and comprehensible. For example, "this self-paced course" was replaced with "this instructional method" in the final form given at the end of Winter Quarter. This suggestion was made by a professor who wanted to use the MISR Form but felt that his course was not entirely self-paced. In addition, he had told his students they were not taking a genuine self-paced course even though he followed the basic Keller Plan. This obstacle was overcome by removing the restricting descriptors. The only additional change was to reduce ambiguity in question thirty, section 4; "Would you have preferred more informal lectures?" on the Fall Quarter Form was changed to "would you have preferred more lectures?" on the Winter Quarter Form.

MISR Form Administration

The Mastery Instruction Student Rating Form was administered to three courses at the end of Fall Quarter and to the same three courses, plus one, at the end of Winter Quarter. The responsibility for the administration of the evaluation forms was with the instructors, as is the usual custom. Generally the method is inefficient in mastery instruction courses since the students are never together as a full class due to the self-pacing feature. Furthermore, the instructors usually gave their student proctors the responsibility of giving students the student rating form. In the engineering courses and physics course the forms were left on a centrally located table, available for students to fill out on their own initiative. In the biology course, where the students met as a group once a week for a lecture or film the student rating return was more

successful. The computer programming course had one meeting of the class as a whole before the final examination, at which time the student rating forms were administered. The number of returned forms averaged 70 per cent over both quarters for the computer programming course. Completing a student rating form is not compulsory at the University of Washington, so if a student wishes, he may hand in an empty form.

Statistical Methods

The procedures for testing the research hypotheses of this study are as follows: First, the stability of the MISR instrument, research hypothesis number 1 was tested by means of a *t* test and by computation of omega squared (Hays, 1963). If the rating instrument is assessing characteristics of the instructor and learning material which are relatively unchanging over two quarters' time, then small *t* scores and omega squared coefficients should reflect such stability. Thus, Fall and Winter Quarter ratings of the same course were compared for six courses: the two mastery instruction computer programming courses taught by the same instructor with the same learning materials both quarters, the two mastery instruction biology courses taught by the same instructor with improved and more organized learning materials Winter Quarter, and the two mastery instruction engineering courses taught by different instructors using the same learning materials. The intraclass correlation formula (Ebel, 1951) was used to analyze the data pertaining to research hypotheses number 2, which concerned item discrimination.

The student ratings of the seven mastery instruction courses grouped into three, and the four control group courses grouped into three were compared pairwise by means of *t* tests in order to find what differences are evident from student opinion concerning the two teaching methods (research hypothesis number 3). In addition, the items on the MISR Form were compared with the items on Form D of all courses covering the same subject matter (research hypothesis number 4), also by means of *t* tests.

Limitations

There is one notable caution and limitation in this study: the courses and instructors could not be sampled randomly. As a result, there may be a selection bias for there was no way to guarantee random population of

instructors or students in the mastery courses. In order to obtain a large enough sample, all known professors teaching mastery instruction who offered to participate in the study were accepted. Furthermore, these instructors chose to use the mastery method. In the case of the students, selection bias may also have been a factor, e.g., the upperclassmen with previous experience at this University may have chosen specific courses because of the instructor's reputation while freshmen and sophomores may have chosen more randomly. Further, there was no way to make certain that comparable students enrolled for the same courses Fall and Winter Quarters. Even though the faculty understood the desirability of minimizing changes in course content and procedures during Fall and Winter Quarter, the instructors were still free to do otherwise. An additional difficulty was the problem of obtaining a large enough sample of student ratings from the mastery instruction courses since these courses rarely met as a whole class after the first meeting.

RESULTS

A. Item mean difference in 3 mastery instruction courses on Fall and Winter Quarter student ratings

Table 7 presents the *t* values computed between the item means on Fall and Winter Quarter student ratings using the MISR Form for 3 pairs of mastery instruction courses. It can be seen in this table that two pairs of courses, the computer programming sections taught by the same instructor both quarters and the engineering sections taught by two different instructors Fall and Winter Quarters, showed no systematic differences in their student ratings over the two quarters; that is, the Fall and Winter student samples of these two pairs of courses did not differ significantly in their opinion and assessment on any of the items. An omega squared was computed for each *t* value greater than 1.0 since a *t* score less than 1.0 is assumed to have an omega squared of zero. This statistical measure was used to assess the strength of the statistical association between the Fall and Winter Quarter student ratings. A small value indicated that very little of the variance of the dependent variable (student ratings) was attributable to the independent variable (Fall and Winter Quarters).

Table 7

Differences in Item Means from Fall to Winter Quarter Student Ratings for Each Pair of Mastery Instruction Courses

Items	Computer programming course		Biology courses		Engineering courses	
	t value	Omega squared	t value	Omega squared	t value	Omega squared
1	1.70	.05	1.21	.01	-.94	
2	.55		-.24		-.53	
3	-.13		-.84		-.20	
4	.33		1.48	.02	.17	
5	.33		.18		-.64	
6	1.61	.04	.28		-1.26	.01
7	1.95	.08	.48		-.23	
8	-.78		5.87*	.30	-.05	
9	2.06	.08	1.62	.02	-1.34	.02
10	.44		-.12		-.40	
11	-.55		2.33*	.05	-.18	
12	-.57		2.81*	.08	.51	
13	.45		3.30*	.11	-.88	
14	.54		1.36	.01	.57	
15	.38		4.20*	.17	1.13	.01
16	1.35	.02	2.29*	.05	-.28	
17	.49		.89		-.97	
18	.36		1.24	.01	-.57	
19	.93		1.44	.01	-.46	
20	1.43	.03	1.65	.02	.03	
21	.52		1.97	.04	-1.06	.00
22	-.57		2.34*	.06	-.52	
23	-.79		3.14*	.10	-----	instructor's own items
24	.16		2.72*	.08	-----	
25	.55		.03		-----	
26	-.26		3.77*	.14	-.06	
27	.40		-2.57*	.08	.38	
28	-1.61	.05	.57		-.80	
29	-.82		-1.35	.01	1.58	.03
30	-.91		2.90*	.09	-.46	

*Significant at the .05 level.

Small values were found on the items between quarters for the computer programming sections and the engineering sections. This provides related evidence that the two courses have very stable ratings, i.e., the change in the quarter accounted for little or none of the variance.

In contrast, data from the two biology sections do show differences in the student ratings between Fall and Winter Quarters; Fall Quarter ratings were significantly lower than Winter Quarter ratings for eleven items. This is probably due to the switch from lecture instruction to the mastery instruction method in the middle of Fall Quarter, since the instructor only had mastery instruction learning materials prepared for the last half of the quarter, and these materials were not previously used or tested. The instructor made improvements in his own organization and completed the preparation of the mastery instruction materials for Winter Quarter. The Fall Quarter student ratings may have been affected by the incompleteness of the learning materials used and by the instructor's initial attempt. Most of the biology student rating items that showed a significant t value pertained to the repeated testing feature of mastery instruction (items numbered 8, 11, 13) and those items that compared mastery instruction to the traditional lecture method (numbers 23, 24, 26, and 27). In sum, the Winter Quarter students rated the biology course significantly more favorably on eleven of the thirty items; the exception was item 27, "Did you find this instructional method to be a more efficient way of learning than the traditional lecture method?" which was rated significantly higher by the Fall Quarter students. This result is interesting in that one could speculate that the immediate contrast with lecture methods leads to a higher rating Fall Quarter.

The omega squared values for the two biology sections showed several items, especially numbers 8, 15, and 26, with large values indicating that the change in quarters accounted for 30 percent, 17 percent, and 14 percent of the variance on those items respectively. Consequently, the two biology sections were judged not to yield stable student ratings from Fall to Winter Quarters.

In sum, the data for the first research hypothesis showed that for two out of the three pairs of mastery instruction courses stable MISR

student ratings were obtained across Fall and Winter Quarters. Differences shown by the biology course data were explained by the incompleteness of the learning materials Fall Quarter.

B. Item discriminations for the MISR Form

Since the assessment of the "goodness" of a course is the primary consideration in student evaluation of courses, then it is necessary for the properties of a set of courses to be such that an item discriminates among the set, otherwise the item is worthless in assessing the "relative" goodness of each particular course. Eighteen of the thirty items on the MISR Form were found to discriminate significantly among the mastery instruction courses as shown in Table 8. This table presents the F ratios, the significance levels, and the item reliability coefficients[1] for each item on the MISR Form. The reliability of these items indicates the extent to which these courses and instructors would receive similar ratings from a different equivalent sample of students.

The seven mastery instruction courses were divided into four groups for the analysis of variance: one group was composed of the Fall and Winter sections of the computer programming course; another group consisted of the Fall and Winter sections of the engineering course; the Winter biology section only composed the third group, and the fourth group consisted of the one section of the physics course. Spuriously low item reliabilities were considered a threat if each of the seven courses had been treated as seven separate groups, since the same instructors had taught two of the three paired sections of mastery instruction courses. Further, the Fall section of the biology course was deleted from this analysis because it was perhaps not a "normal" mastery instruction course since the biology learning materials were not complete for the five weeks rated as were the materials for the other six mastery instruction courses. In addition, the t tests computing the item means between the two quarters,

---

[1] The formula used was $R = \frac{MS_B - MS_W}{MS_B}$ where MS_B is mean square between classes and MS_W is mean square within classes.

Table 8

Item Discriminations Using Ebel's Intra-item Reliability Formula

Item	Reliability	F Ratio	F Probability
1	.72	3.59	.01
2	.10	1.14	NS
3	.89	9.31	.001
4	.74	3.88	.01
5	.64	2.80	.05
6	.47	1.90	NS
7	.57	2.30	.05
8	.84	6.43	.001
9	.63	2.68	.05
10	.00	.98	NS
11	.71	3.50	.01
12	.92	12.93	.001
13	.00	.40	NS
14	.89	9.36	.001
15	.85	6.63	.001
16	.77	4.37	.01
17	.90	9.97	.001
18	.32	1.47	NS
19	.81	5.14	.001
20	.00	.89	NS
21	.87	7.80	.001
22	.00	.80	NS
23	.00	.02	NS
24	.00	.71	NS
25	.02	1.02	NS
26	.85	6.66	.001
27	.68	3.13	.01
28	.77	4.36	.01
29	.00	.26	NS
30	.49	1.97	NS

as mentioned above, had shown some significant differences due to the low ratings Fall Quarter. Therefore, only the Winter Quarter biology course data were used in order to avoid spuriously high item reliabilities.

The items that are the best discriminators received the largest variation among the student raters between courses relative to differences within course ratings. The nondiscriminating items are those with an F probability greater than the .05 significance level. Half of the eighteen discriminating items were significant at the .001 level. With one exception (item 26) these items (numbers 3, 8, 12, 14, 15, 17, 19, and 21) are all concerned with the instructor's and TA's input into the course. Item 26 concerns the comparison of the traditional lecture and mastery instruction methods. Despite the emphasis on learning materials in the mastery instruction method, the instructors and TAs of these courses still appear to be the most discriminating factor in distinguishing a "good" course from a "bad" one.

Other categories of items on the MISR Form were those that refer specifically to the learning materials, the mastery unit tests, the comparison between the traditional lecture method and the mastery instruction method of teaching, and the ten standard items on all U of W student rating forms.

Of the three items concerned specifically with the learning material (numbers 6, 7, and 9) two were found to discriminate (at the .05 level): item 7, "Clarity of explanations by the learning material was," and item 9, "Examples and illustrations in the learning material were." Item 6, which deals with the sequential presentation of concepts did not discriminate among the courses. It would now be considered a poor item and is a candidate for deletion from the MISR Form.

Only one item out of five that referred to mastery unit tests discriminated among the courses: number 11 (significant at the .01 level) which asked about the large number of tests. The other four nondiscriminating items in this category were numbers 10, 13, 25, and 29.

Six items on the MISR Form involved a comparison of traditional lecture and mastery instruction methods: numbers 23, 24, 26, 27, 28, and 30. Three of these items discriminated among the mastery instruction

courses, items 26, 27, and 28, (significant at the .001, .01, and .01 levels, respectively). Item content is as follows: item 26, general assessment of traditional lecture method versus the mastery instruction method; item 27, is the mastery instruction method a more efficient way of learning?; and item 28, does a mastery instruction course require more time and effort?

Four items, numbers 2, 18, 20, and 22, of the ten standard items on all U of W student evaluation forms did not discriminate among mastery instruction courses.

In sum, the data for the second research hypothesis show that the MISR Form was able to discriminate among these mastery instruction courses on the majority of the items.

C. Item mean differences between mastery and traditonal lecture sections of the same course

Table 9 shows t values for the comparative student ratings between sections of the same course taught by two different methods of instruction, mastery and traditional lecture. The table also states which quarter ratings were used for the computation of t values. These data indidate that neither method of instruction is rated consistently more favorably than the other. That is, computer programming sections did not differ significantly from each other on similar items. Yet, the mastery instruction biology section (Winter Quarter data), received significantly better ratings than the traditional lecture section on eleven items. The mastery instruction physics section received significantly poorer ratings on four items and was rated significantly more favorably than the lecture section on only one item.

Five items, (numbers 4, 8, 10, 11, and 13), were excluded for comparisons between mastery instruction and traditional lecture instruction since they were not comparable on the MISR Form and the U of W Form D used by the traditional secture sections in this study.

D. Item mean differences between mastery instruction student ratings and the composite ratings on the U of W Form D

The MISR Form is considered to be most similar to the standard U of W Form D and was even designed to include similar items. Approximately 7200

Table 9

t Values for Item Mean Differences between Mastery and Traditional Lecture Sections of the Same Course

Items	Computer programming courses[1]	Biology courses[2]	Physics course[3]
1	-1.16	3.30**	-.72
2	-1.04	2.76**	-.20
3	-.12	.98	-2.30*
4	--	--	--
5	-1.08	-.49	-1.40
6	-1.02	1.36	-.61
7	.58	.79	-2.09*
8	--	--	--
9	.40	.10	-2.44*
10	--	--	--
11	--	--	--
12	1.07	3.77*	-1.39
13	--	--	--
14	-.11	6.30**	3.12**
15	-1.03	7.78**	-.80
16	1.08	.67	-1.26
17	-1.14	5.29**	-1.29
18	.19	3.25**	.28
19	-1.24	2.98*	-1.54
20	.90	3.25**	-.35
21	-1.36	3.18**	-2.38*
22	-.28	4.51**	-1.26

*Significant at the .05 level.

**Significant at the .01 level.

[1]Fall and Winter Quarter data were summed for each section and then the *t* value was computed.

[2]Fall Quarter traditional lecture section data were compared to Winter Quarter mastery instruction section ratings.

[3]The ratings for both sections were Winter Quarter data.

Items 4, 8, 10, 11, and 13 were omitted since they were not comparable items on the two student rating forms.

student ratings on Form D obtained Fall Quarter were compared to the 183 mastery instruction student ratings. The null hypothesis was formed such that the 7200 students were considered the population of Form D users. Therefore, a t test was performed to determine if the 183 student ratings could be considered a sample from that population. Table 10 shows significant differences on twelve of the seventeen items compared. This indicates that mastery instruction student ratings are significantly different from the population item means. With one exception all significantly different items were in favor of mastery instruction courses.

Four items were significantly different at the .05 level, numbers 1, 14, 15, and 21. Item 1 is a global item, "The course as a whole was," which is considered a good indicator of the total effect of the course on the students. Items 14 and 15, "Answers to student questions were," and "Availability of extra help when needed was" are concepts emphasized by the mastery instruction method. Further, that mastery instruction students were more favorable about the "Reasonableness of assigned work," item 21 was revealing; the research literature on mastery instruction suggests that students often found they had to work harder in mastery courses than in the traditional lecture courses.

The eight other significantly different items, numbers 2, 3, 7, 9, 18, 19, 20, and 22, were significant at the .01 level. Of particular interest were three of these items which deal directly with major tenets of the mastery instruction method: item 3, "The instructor's contribution to the course"; item 20, "Evaluative and grading techniques (tests, papers, projects, etc.) were"; and item 22, "Clarity of student responsibilities and requirements was." Item 3 was the only negatively significant item. This is probably due to the de-emphasis in mastery instruction of the instructor's overt contribution to the course since he rarely lectures, although he is generally available to answer questions, and his main responsibility is organizing the course and learning materials. Three other items, numbers 2, 18, and 19, in this significant item group rated what students felt they had learned from the course. These items, which asked students about the course content, the amount learned, and the relevance and usefulness of the course content, were rated more favorably by the students in the mastery instruction courses.

Table 10

t Values for Item Mean Differences between the Mastery Instruction Student Ratings and the Standard U of W Form D Student Ratings

Items	t value	Items	t value
1	2.47*	12	.44
2	4.37**	13	--
3	-3.62**	14	2.49*
4	--	15	2.56*
5	.90	16	-.45
6	-.85	17	1.31
7	3.26**	18	5.49**
8	--	19	4.88**
9	2.89**	20	4.50**
10	--	21	2.36*
11	--	22	5.26**

*Significant at the .05 level.

**Significant at the .01 level.

Items 4, 8, 10, 11, and 13 were omitted since they were not comparable items on the two student rating forms.

In sum, the data indicate that mastery instruction courses as a group are rated significantly more favorably than the many courses using Form D. It is possible that the mastery instruction method addresses itself more specifically to some of the important classroom characteristics assessed by the student ratings.

DISCUSSION AND CONCLUSIONS

Proponents of the mastery instruction methods for college teaching have proclaimed that one of its most conspicuous effects is an overwhelming popularity among the students (Ferster, 1968; Keller, 1968; Johnston and Pennypacker, 1971; McMichael and Corey, 1969). These investigators reported that mastery instruction can produce markedly greater student interest in an attitude toward the subject learned than the more usual instructional methods. One purpose of the present study was to test several hypotheses related to the efficacy of the mastery instruction method over the traditional lecture method in the college classroom, this included an assessment of student opinion of the instructional method. A second purpose was the development of a reliable and discriminating student rating form for evaluating the mastery instruction courses.

The subjects used in this study formed two groups: the experimental group was composed of 183 undergraduates at the University of Washington in seven mastery instruction courses taught by five different instructors; the control group consisted of 255 undergraduates in four traditional lecture courses taught by three different instructors. All four traditional lecture courses and four of the mastery instruction courses were matched sections of the same course; that is, there were four courses from different academic areas that offered a section taught both by the traditional lecture method and the mastery instruction method.

Optimally, when making a comparison of one instructional method with another, random samples from populations of courses taught by each method of instruction should be used. Unfortunately, such was not possible at the University. Even though the mastery instruction courses were not a random selection, the matched courses taught by the traditional lecture method were randomly selected whenever possible. Further, the small number of mastery instruction courses used in this study as well as the

selection bias inherent in the instructor's choice to teach by this method seriously limits its generalizability. Nonetheless the implications are interesting and worthy of further research under more controlled conditions.

Instructional Methods

The first major finding was that the mastery instruction sections as a group were not consistently rated more favorably by the students than the traditional lecture sections of the same courses on similar items as suggested by previous research (Johnston & Pennypacker, 1971; McMichael & Corey, 1969). One pair of courses received similar student ratings from the two groups of students taught by different methods of instruction. The data from the second pair of courses showed significant differences on student rating items in favor of the mastery instruction section, while in the third pair of courses the traditional lecture section received more favorable ratings. These results suggest that neither instructional method is necessarily superior to the other as rated by students. Some mastery courses are better and some traditional course are better.

In contrast, the larger sample of data comparing mastery instruction student ratings to the population of University of Washington Form D student ratings did show significant differences between the two instructional methods. Mastery instruction courses received significantly more favorable ratings on the majority of items. These results suggest that the mastery method deals more effectively with many important classroom characteristics assessed by the student ratings. These results seem to conflict with those just mentioned above, yet the comparison group here is 421 courses taught by various and unknown methods. From these two conflicting findings there is the implication that if an instructor changes to mastery instruction he will not necessarily receive more favorable ratings, especially if his preparation is less than complete as apparently illustrated by the relatively low Fall Quarter biology ratings. Yet, since mastery instruction courses are rated more favorably than averaged ratings of 421 other "more traditional" problem-solving courses, the method appears as a viable and effective method of instruction.

Further, data suggest that the instructor's input may make a great contribution to favorable student ratings within the mastery method. The

student rating items that best discriminated among the mastery instruction courses were concerned with the instructor and TA's contribution to the course. On the other hand, the two engineering sections were taught by different instructors, using the same materials, and showed stability. This indicates that the course materials may interact with the instructor. But, the two instructors could also have been equivalent in teaching skills. This question requires further research.

The MISR Form

A second important result of this study was the production of a reliable student rating instrument, the Mastery Instruction Student Rating Form, which was also satisfactory to the instructors of the mastery instruction courses. These instructors had requested that a student rating instrument based on the specific characteristics of mastery learning be devised. Instructors now have available the MISR Form and increased confidence upon student judgments as worthy for assessing the "goodness" of mastery courses. For example, eighteen of the thirty MISR items yielded reliabilities from .57 to .92. In contrast, six items were found to be "unreliable," there are several possible reasons for the unreliability of six items on the MISR Form. First, the items may be invalid and unsuitable for discriminating among the courses evaluating their "goodness." Second, the population of these particular variables was small; a larger sample may find these unreliable items to be reliable. Third, it is possible these items asked the students to assess concepts that are relatively constant across mastery instruction courses.

Conclusions

Clearly, there are unanswered questions about mastery instruction from this study, but several implications are suggested. First, a comparison of two instructional methods on the college level may lead to no consistent, significant differences in student ratings. This is probably due to the college student's ability at this stage in his academic career to adjust to various instructional methods, and thus, the finding of no significant difference in student performance between the two methods in this study. Secondly, the serendipidous finding that the mastery instruction biology section, which was incomplete and unorganized Fall Quarter,

was rated significantly lower by students than the well-developed Winter section suggests that students are discriminating judges of courses and that the quality of an instructor's input is important. "Good" college instructors will usually keep their courses interesting and updated; they probably obtain favorable student ratings regardless of the instructional method they use as long as certain conditions are met: enthusiasm about the approach, organization, ability to arouse student interest, ability to give good explanations, and ability to stimulate the intellectual activity of the students. In short, such characteristics may transcent the specific instructional method used by the instructor.

Reference Notes

1. Anderson, E. The certainty of information in instructional decision making (EAC Report). Seattle: University of Washington, Educational Assessment Center, 1975. In preparation.

2. Gillmore, G. M. A brief description of the student rating forms of the University of Washington Instructional Assessment System (EAC Report 276). Seattle: University of Washington, Educational Assessment Center, 1974.

References

Block, J. H. (Ed.), Mastery Learning: theory and practice. New York: Holt, Rinehart and Winston, 1971.

Bloom, B. S. Individual differences in school achievement: a vanishing point? Education at Chicago, 1971, 1, 1, 4-14.

Ebel, R. L. Estimation of the reliability of ratings. Psychometrika, 1951, 16, 407-424.

Ferster, C. B. Individualized instruction in a large introductory psychology course. The Psychological Record, 1968, 18, 521-532.

Green, B. A. A self-paced course in freshman physics. In J. H. Block (Ed.), Mastery Learning: theory and practice. New York: Holt, Rinehart and Winston, 1971.

Johnston, J. M., & Pennypacker, H. S. A behavioral approach to college teaching. The American Psychologist, 1971, 26, 3, 219-244.

Keller, F. S. "Good-bye Teacher..." Journal of Applied Behavior Analysis, 1968, 1, 79-89.

McMichael, J. S., & Corey, J. R. Contingency management in an introductory psychology course produces better learning. Journal of Applied Behavior Analysis, 1969, 2, 79-83.

Footnotes

[1]This report is based on Master's thesis research by the senior author.

[2]The authors wish to sincerely thank the instructors who voluntarily participated in this research.