# Chapter 5

# Mastery Learning Methods

Visual-based and audio-visual methods have not always demonstrated significant gains in student performance. Simple presentations that activate the visual memory system, such as line drawings that teacher may use on a chalk board, are sufficient to capitalise on the visual storage system. Thus, there is usually no difference between this form of traditional presentation and newer visual-based systems.

Mastery methods, however, do seem to have consistently superior results when compared with traditional methods. Mastery methods can be traced back to the early part of this century, but it is Skinner's work that provides the initial impetus for those methods that are currently in use. The results of Skinner's early work on operant conditioning were published in his 'Behavior of Organisms' (1938), but his next major work, a somewhat startling contrast, was 'Walden Two' (1948), a novel, which was written as a programme that outlined the utopia attainable by adopting a behaviouristic, experimental approach toward social problems. The scientists in 'Walden Two' dispensed with traditional 'trial-and-error' freedom because there were efficient ways of attaining desired goals, based mainly on positive reinforcement. Cultural engineering was not a pejorative term in this brave new society, although the book was frequently read as a dystopia. A more technical treatment of the 'Walden Two' ideas is to be found in 'Science and Human Behaviour' (1953), dedicated to his colleague Fred Keller, who later developed the Personalized

System of Instruction, based on Skinnerian principles. In the section dealing with education, he refers to 'knowledge' as 'the entity which is traditionally said to be maximized by education' (p.408). He acknowledges that it is a complex term, but continues with the assertion that most knowledge acquired in education is verbal. He claims that the acquisition of such knowledge does not mean that education is merely rote learning, because the student also comes to understand the facts. For example, in the study of history this is because:

> The individual agrees with a statement about a historical event in the sense that he shows high probability of making the statement himself. The growing understanding with which he reads and re-reads a passage describing a period in history may also be identified with the growing probability that he will emit verbal responses similar to those which comprise the passage. (Skinner, 1953, p.409)

## THE APPLIANCE OF SCIENCE TO TEACHING

With such a view, linking verbal behaviour to knowledge as the entity to be maximized in education, the stage is set for one of the most important announcements in the history of educational technology. 'The Science of Learning and the Art of Teaching' (1954) was first read at a conference at the University of Pittsburgh, in March, during which there was a demonstration of an experimental teaching machine. It was published later in the year in the influential 'Harvard Educational Review.'

Skinner claimed that it was a great shock to turn from the exciting prospect of an advancing science of learning to that branch of technology which is most directly concerned with the learning process, education. In the first place he questioned the reinforcements used, though he did acknowledge that they had changed since the early years of the century and were not all based on major aversive methods of control, now they were based on 'minor' aversive events such as the teacher's displeasure, low marks etc. He then asked how the contingencies of reinforcement were arranged, for example, when is an arithmetical operation reinforced as 'right'? The conclusion was that many minutes and in many cases many hours or even days may intervene between a child's response and the teacher's reinforcement, even though it could be demonstrated that, unless explicit mediating behaviour has been set up, the lapse of only a few seconds between response and reinforcement destroys most of the effect. He calculated that during the first four years of education 50,000 reinforcements would be necessary to attain efficient mathematical behaviour,

but that in a traditional class situation it would be possible for the teacher to provide a pupil with only a few thousand. In order to provide sufficient reinforcements the only possible conclusion was that human learning would require instrumental aid, and the simple fact was that 'as a mere reinforcing mechanism, the teacher is out of date' (Skinner, 1954, p.94).

With the introduction of instrumental aid, complex patterns of behaviour could receive gradual elaboration, with the whole process of becoming competent in any field being divided into a very large number of very small steps. The maintenance of the behaviour in strength at each stage could be accomplished by applying the techniques of scheduling which had been developed in the studies of other animals, but more probably would be most effectively arranged through the design of the material to be learned.

> By making each successive step as small as possible, the frequency of reinforcement can be raised to a maximum, while the aversive consequences of being wrong are reduced to a minimum. (Skinner, 1954, p.94)

Skinner demonstrated two prototype machines at Pittsburgh. The first taught arithmetic, with the material to be learned appearing in a square window. Missing numbers were made to appear by moving sliders with numbers printed on. When the problem was completed the child turned a knob on the front of the machine. The machine 'senses the composed answer, and if it is correct, the knob turns freely and a new frame of material moves into place.' It would not turn if the answer was incorrect and the child could then re-set the sliders until a correct answer was obtained. Errors were recorded automatically. The second machine was similar, but with more sliders and was designed to teach both arithmetic and spelling.

Skinner foresaw objections to the use of such devices in the classroom, including the challenge that children were being treated as mere animals and that an essentially human intellectual achievement was being analysed in unduly mechanistic terms. He countered this by suggesting that the behaviours elicited by the instrumental aids were the 'very behaviours which are taken to be the evidences of such mental states or processes' and that the 'behaviour in terms of which human thinking must eventually be defined is worth treating in its own right as a substantial goal of education' (1954, p.96).

With regard to possible costs, he argued that a country which annually produces millions of fridges, automatic dishwashers and washing machines

can certainly afford the equipment necessary to promote high standards of education.

Echoing Thorndike he stated that the teacher has a more important function than to say 'right' or 'wrong' and that 'it is beneath the dignity of any intelligent person' to mark a set of papers in arithmetic. The more important work, the teacher's relations with her pupils, cannot be duplicated by a machine, and instrumental help would improve these relations by freeing the teacher. He concluded:

> There is a simple job to be done. The task can be stated in concrete terms. The necessary techniques are known. Nothing stands in the way but cultural inertia. But what is more characteristic of the modern temper than an unwillingness to accept the traditional as inevitable? We are on the threshold of an exciting and revolutionary period, in which the scientific study of man will be put to work in man's best interests. Education must play its part. (Skinner, 1954, p.97.)

With the new technology, which was to apply the newly emerging science of learning to the age-old art of teaching, we can see how Skinner drew parallels between his work with animals and the action of an ideal teaching environment. As with the animals in the Skinner box the human learner must make a response which is observable and can be described in operational terms. The learner's behaviour is shaped by the programme, which consists of discriminated stimuli (programme frames), with the correct responses being immediately reinforced (knowledge of results), on a continuous schedule. Incorrect responses are extinguished, because they receive no reinforcement.

Within a few years there was such interest in programmed learning that Schramm (1964) in his review of the research concluded that 'no method of instruction has ever come into use surrounded by so much research activity.' By 1957 Skinner was demonstrating a machine in which a frame of material appeared in a window near the centre of the machine and the student wrote his answer in a window to one side of the frame. By moving a lever the response made by the student moved under a transparent cover (it could be seen, but not altered, making the device cheat-proof) and at the same time the correct response was uncovered in the upper corner of the frame. If the student recognized that his response was correct he moved another lever which punched a hole next to the correct response and shifted the programme on to the next frame. Such machines were used to teach a course in human behaviour to

Skinner's Harvard University students, with nearly 200 students completing the course.

Skinner compared the machine to a private tutor because there is constant interchange between programme and student, unlike lectures and textbooks. The programme 'insists' that a given point be thoroughly understood and presents just sufficient material for which the student is ready, thus ensuring that the student comes up with the correct answer, which receives positive reinforcement and holds the student's interest, according to Skinner.

It was eventually realized that the essential feature of the new method of teaching was the 'programme' rather than the machine and this led to the development of the programmed linear text in which 'frames' were printed on one side of the page and answers on the reverse, a form used in Holland and Skinner's influential text 'The Analysis of Behavior: a program for self-instruction' (1961).

## THE RESEARCH: DOES IT WORK?

The major research question in the decade that followed Skinner's announcement of the forthcoming revolution in education was 'Do students learn from programmed instruction?' Claims were made that student performance could be increased, giving 90% of students scores of 90% on evaluation tests, and the commercial stakes were high, particularly for companies with educational interests. Schramm (1964) attempted to answer this question and located 190 reports of original research on programmed instruction, which included some experiments with programmed films and television. Most of the research (40%) was conducted with college students as subjects, with smaller numbers of secondary (20%) and adult/military (20%) subjects and a minority of primary/pre-school (10%) subjects. From this mixture of research Schramm concluded that many different kinds of students - college, high school, secondary, primary, pre-school, adult, professional, skilled labour, clerical employees, military, deaf, retarded, imprisoned - do indeed learn from linear programmes either on machines or as texts. Using programmes such students learned mathematics and science at different levels, foreign languages, spelling, electronics, computer science, psychology, statistics, business skills, reading, and many other subjects. For each of the groups of students and the different kinds of subject matter the early experimental evidence demonstrated that 'a considerable amount of learning can be derived from programmes' (Schramm, 1964, p.4) either by comparing pre- and post-tests or the time and trials to reach a given criterion of performance.

When the question is changed to 'How well do students learn from programmes as compared to how well they learn from other kinds of instruction?' Schramm could not answer quite so confidently. Of the 36 studies which compared programmed instruction with traditional teaching, 18 showed no significant statistical difference when the two groups were measured on the same criterion test. But 17 showed a statistically significant superiority in favour of the programmes, with only one showing a difference in favour of classroom students. Schramm cautiously concluded that 'the results should not discourage us about the amount of learning derived from programmes' (p. 5).

It was assumed that a major advantage of the new technology would be that the student could proceed at her own pace. Surprisingly, the early research failed to confirm this. Seven of the studies found no difference between external and individual pacing, with students taught by machine, text or television, compared with two studies in favour of individualised pacing. Carpenter and Greenhill (1963) varied the pace from 20 percent below to 10 percent above the average of class self-pacing without decreasing the overall amount of learning. Hartley (1974), in a later review of the research from 1954-1974, also found a wealth of non-significant differences in test performance between individual and group learning in either paced or unpaced situations, and concluded that the research indicated three main points:

(i) Self-pacing produces considerable administrative difficulties, because even like-ability groups spread out enormously in terms of time taken to complete a programme.
(ii) Group-pacing methods are often technically complex to set up: the gains (if any) may not be worth the technical effort involved.
(iii) Some learners have difficulty in judging what is an appropriate pace at which to go.

(Hartley, 1974, p.281)

The last point has implications for programme writers because, although the learner can control the speed at which he works through the programme, the overall pace is largely determined by the writer.

Hartley also suggested that the following issues were important:

(i) the ability of the learners involved (university students found small steps irritating)
(ii) the age of the learners (small steps were more appropriate for small children)
(iii) the pre-knowledge of the learners (small steps were more suitable than large steps for learners with little relevant background)
(iv) the confidence of the learners (small steps were more suitable for learners who were afraid of the task in hand eg. statistics for female arts undergraduates)
(v) the kind of subject matter (small steps might be useful for statistics, or for subjects with their own built-in logic, but not for more over-inclusive subjects such as, for example, literary appreciation)
(vi) the language employed (complex language can still be confusing even in small steps).

(Hartley, 1974, p. 282)

The great majority of the studies also showed no statistically significant differences between the amount learned from overt and covert responses. Since there was no difference, and since covert or 'thinking' the response took less time, the covert mode could be considered more efficient.

Pressey (1963) offered a direct challenge to the Skinnerian programme ideal when he tested the first unit of the Holland-Skinner psychology programme (54 frames) against the same material re-written in good prose paragraphs. There was no difference in the final performance, but the prose version took less time for the students.

Leith (1968) demonstrated that the need for overt responding depended on the type of task and the background of the learner. Conceptual tasks, for example, are clearly different from response-learning tasks. English spelling was found to be better accomplished by writing the responses, while co-ordinate geometry or the structure of genetic materials were learned more successfully when covert responses were required. Even recall, retention and transfer of complex circuit diagrams were carried out just as efficiently by thinking the responses as by drawing them if the students had a background of 'O' level Physics. Those students with less than 'O' level knowledge of Physics did poorly when thinking the responses but did as well if allowed to make overt responses.

Hartley (1974) concurred, adding that overt responses may be superior when the learners were young children, when the material was difficult and the programmes lengthy and when novel or specific terminology was taught.

Later studies confirmed the complexity of the issue. When learners were required to work out each stage of a correction for themselves, receiving feedback stage by stage, Tait, Hartley and Anderson (1973) found that they performed better than learners receiving more global feedback. McKeachie (1974b) commented that

> Programmed instruction proponents were understandably aghast to find that immediate knowledge of the correct response (expected to be a reinforcer) did not facilitate learning in programmed instruction. (p.8)

A major change in programming styles, which contested the mode of responding, size of step and type of feedback, was introduced by Crowder in the late 1950's. This intrinsic or branching style of programming presents larger amounts of information, usually a paragraph, for the student to study. There is an immediate test, usually a multiple-choice question, and the test result is used to determine the content of the next frame in the teaching sequence. If the answer is correct, she is automatically given the next unit of information and the next question. But if the answer is incorrect the preceding unit of information is reviewed, the nature of the error is explained and she is retested (Crowder, 1960). There is a separate set of correctional materials for each wrong answer that is included in the multiple-choice test.

This 'intrinsic' programming technique was devised to operate with a sophisticated device for handling programmed materials on microfilm, but could equally be applied to what became known as 'scrambled' texts, in which each answer has a different page number associated with it.

Crowder claimed that human learning takes place in a variety of ways and that these ways vary with the abilities and knowledge of the students, the nature of the subject matter etc. The intrinsic programming method would provide the necessary feedback control in this complex series of interactions. This is different to simply providing knowledge of results to the student because the test result is used to control the behaviour of the teaching machine, the primary purpose being to determine whether the communication was successful, in order that corrective steps may be taken if it has failed (Crowder, 1960, p.288).

Hartley (1974) agreed with Schramm's earlier conclusion concerning the lack of differences between linear and branching programmes, but observed that some studies demonstrated a saving in time taken to complete the programme, in favour of branching programmes, with older and intelligent learners. This strikes at the heart of Skinnerian psychology, because Skinner had argued most strongly for maximizing the positive consequences of learning and a minimizing of the negative aspects.

Having determined that research had indicated that each of the original characteristics was not a sine qua non of programmed learning, Hartley concluded that 'Skinner's techniques have not been shown to be universally valid. They have not, however, been shown to be valueless' (1974, p.286).

To Bugelski (1971) the use of knowledge of results as a reinforcer makes it impossible for anyone to determine whether the teaching machine really represents an application of a psychological principle. Indeed, it must be remembered that the animals in Skinner's experiments were reduced to 80% of their normal weight; they were in a deprived state before engaging in the required behaviour and once satisfied stopped performing.

There is also evidence that different groups of students react in different ways to the type of reinforcement or feedback. McKeachie (1974b) reports a number of studies showing that informing a child of the correctness of his response increases achievement for middle class children, while other children may learn more effectively when given praise or tangible rewards. Means and Means (1971) found that below average students did well on a mid-term exam when told that they had previously performed well on an aptitude test; whereas above average students did better when told that they had done poorly on the aptitude test.

Research peaked in 1967 and has been declining steadily since that date, although there continues to be interest in individualised modes of learning and the old techniques of programming have been given a new lease of life in some applications of computer-based learning.

Kulik, Cohen and Ebeling (1980) calculated the size of the effect of the programmed instruction intervention, using as the index Glass's Effect Size (ES), defined as the difference between the treatment (programmed instruction) and control (traditional classroom) mean scores, divided by the standard deviation of the control group.

They found 56 studies comparing programmed instruction with conventional methods in higher education. A majority of these favoured programmed instruction (40), the remaining 16 favouring conventional methods. Only 25

studies reported statistically significant differences between methods, with 21 favouring programmed instruction. At first sight this seems to produce a very favourable result for programmed instruction. However, when Glass's Effect Size is calculated it is only 0.24. In other words, the effect of programmed instruction in a typical class is to raise student achievement by about a quarter of a standard deviation unit. This is a small effect in educational terms, according to Cohen (1977), but it does mean that in a programmed class 60% of pupils attain at least the average score of the conventional class, compared with 50% of conventional pupils. Medium to large effects were observed in a third of the studies.

When examination scores were analysed the average score of the conventional class was 64.8%; the average of the programmed class was 67.1%. The two groups differed by 2.3% points on average, based on scores taken from 56 studies.

The same technique was applied by Kulik, Schwalb and Kulik (1982), but with results from secondary school rather than higher education. A total of 47 studies were analysed, with 23 of the studies favouring programmed instruction and 24 favouring conventional classes. However, only 19 studies reported statistically different results, with a majority (12) favouring programmed instruction. The average value of the Effect Size was 0.08, the typical programmed instruction student gaining an advantage of less than one tenth of a standard deviation unit. Such an effect is trivial. It implies that 53% of programmed students perform at least at the average level of the conventional group, which has 50% of students at this level. The ES did vary from strongly negative to strongly positive and when study features were analysed size of effect was found to be significantly related to the subject matter taught, with social sciences showing a particularly high effect size of 0.57. For mathematics it was surprisingly small and negative (-0.01) and it was small for the science studies. The conclusion was that:

> .... in general programmed instruction did not improve the effectiveness of secondary school teaching. In the typical study, programmed instruction failed to raise student achievement on final examinations. It did not make students feel more positively about the subject matter they were studying or about the quality of teaching at their schools. Nor did it reduce the role that aptitude plays in determining how much students learn in secondary school classes. (Kulik, Schwalb and Kulik, 1982, p.137)

Finally, the results of studies investigating computer-based learning also contribute to the overall picture of small or trivial effects. Kulik, Bangert and Williams (1983) examined the use of computers in several educational modes: managing, tutoring, simulation, programming and 'drill and practice'. The results for 'drill and practice' are relevant given Skinner's early claim for automation of this aspect of teaching. However, they do not show a great advantage for the computer in this role, with a small average ES of 0.27, when computer and conventional classes were compared.

The results of forty years of experimental evaluations demonstrate that the expectations set by Skinner for his technological revolution have not been fulfilled. Programmed instruction was supposed to make learning more efficient and enjoyable. The research, however, shows that student reactions to programmed instruction were not discernibly different from reactions of students taught conventionally and that differences in attainment were trivial or small, in educational terms, in most cases. Clearly the job of improving education was more difficult than Skinner had predicted from his animal learning theory.

It may well be that learning theories are irrelevant to the solution of such problems, as Gagné (1962b) and McKeachie (1974b) have suggested. After all, the history of automated instruction goes back at least as far as 1860, before Thorndike's work was published, when Halcyon Skinner developed and patented a device for teaching spelling; and Pressey began building machines, which performed all the essential tasks carried out by later programmed devices, as early as 1915, without a supporting learning theory. Perhaps Leith (1968) is correct in suggesting that 'we have been bamboozled in our interpretation of programmed learning by the colourful exploits of a few flamboyant cheer-leaders.' (p.1)

## MASTERY LEARNING

Although early attempts at developing a scientific approach to learning were not always as successful as their proponents anticipated, they have led to an approach to learning that does produce consistent benefits in terms of achievement and attitudes to learning, benefits that are educationally significant.

Mastery learning is embodied in Bloom's (1968) 'Learning For Mastery' and Keller's (1968) 'Personalized System of Instruction.' Bloom's approach was developed to exploit mastery learning in the schoolroom, whereas Keller developed his system for higher education. However, both have been applied in many different contexts and have been found to be very powerful methods for increasing student performance in a wide range of activities.

## KELLER'S PERSONALISED SYSTEM OF INSTRUCTION

Keller describes the first steps towards his system as having taken place during an evening brain-storming session involving Keller and three other psychologist colleagues, who had been invited to set up a new department of Psychology in the University of Brazilia, in 1963. They were given absolute freedom to follow their own dictates and fancies in carrying out the task.

Keller's group was overtly 'operant' in its orientation, or as Keller put it 'tarred with the brush of reinforcement theory.' They were excited by Skinner's Natural Science Course at Harvard University, and all were convinced that traditional teaching methods were hopelessly out-of-date, and were impressed by the teaching machine and programmed-instruction movement.

They were not the only ones disillusioned with traditional methods of instruction. In the late 1950s and early 1960s educational researchers raised some unsettling questions about the effectiveness of most college teaching. In a number of major studies, investigators found that they could reduce by more than two-thirds the amount of time students spent in the classroom without affecting end-of-course student achievement (Kulik and Jaksa, 1977). Dubin & Taveggia (1968) went even further and demonstrated that such typical classroom activities as lectures and discussions might be eliminated completely without altering student achievement. Pooling data from 91 studies conducted between 1924 and 1965 they showed that students learned as much from independent study as they did from more conventional approaches to teaching. In fact, one of the chapters in Dubin and Taveggia's 'The Teaching-Learning Paradox' is titled "THERE ARE NO DIFFERENCES", in very large capital letters.

On the basis of these research results many innovators believed that students could learn as much through the newer approaches as they had learned from the more traditional approaches. And some even believed that they might actually improve learning.

According to the behaviouristic approach of Keller and his colleagues students are more likely to perform well if they find satisfaction in their studies

or in behaviouristic terms 'The consequences of learning are important instructional contingencies.'

This means that positive consequences (instructor praise, good grades, feelings of achievement) are considered to be much more effective facilitators of learning than negative consequences (boredom, failure, or other forms of punishment).

External rewards are important, but it was thought that for adult learners they were not nearly as important as the general internal rewards (feelings of achievement, satisfaction or accomplishment).

Keller's 'Personalized System of Instruction' (PSI), which is also known as the 'Keller Plan', is based on these principles. It specifies objectives and provides reinforcement for their successful achievement, as well as giving the student more options and opportunities for personal interaction than traditional instructional systems.

The system was first publicly announced in 1968 in Keller's article 'Goodbye, teacher', which appeared in the new Journal of Applied Behavior Analysis. It begins with the chant:

Goodbye scholars, goodbye school
Goodbye teacher, darned old fool

and goes on to describe Keller's reasons for being disillusioned with the traditional classroom and formal education. Much of the Personalized System of Instruction can be seen as a development of Keller's observations of military training — perhaps one reason for the emphasis on instruction, rather than 'education' or 'teaching'. There are five main elements in PSI as described in the 1968 paper:

1. The unit-perfection requirement for advance, which lets the student go ahead to new material only after demonstrating mastery of the preceding unit.
2. 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.
3. The stress upon the written word in teacher-student communications.
4. 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.

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

**MASTERY**

Keller indicates that the mastery requirement means an essentially perfect performance. It is acceptable to interpret this as each student answering nine out of ten questions correctly, provided the single error is corrected. Any further compromise with mastery is unwise because:

> the requirement of eighty-five percent will automatically raise questions as to why eighty-three percent isn't 'close enough.' There will be requests for 'partial credit.' An adversary system with 'bargaining' about grades quickly returns to the classroom. The mutual respect and high morale upon which the course depends deteriorates. The course policy statement should specify that the criterion for a pass is perfection.
> (Keller and Sherman, 1974, p.37)

**SELF-PACING**

One of the essential characteristics of mastery learning is the time spent in study and this leads to the emphasis, in the Keller Plan, placed on a completely student-paced programme of study.

It is important to stress that the system places no restraints on the student's study time - thus a student may elect to take a unit quiz at her convenience or when she is most ready to demonstrate mastery, rather than at a time for all students which is dictated by the instructor.

The self-pacing allows for individual differences in rate of study and is essential if the level of achievement is held constant. This is because in essence it is not possible to hold both level of individual achievement and rate of completion constant. Traditional lecture methods, for example, hold rate of completion constant and, because of this, level of achievement varies widely. However, in PSI courses tests can be taken at any time and can be repeated as required, and the result is that while the average student evidences a steady progress, at any one time individual students will be working on different units and taking different tests. Here we have mastery level of achievement held constant and so the rate of completion varies.

## WRITTEN MATERIALS

Keller in 'Goodbye, Teacher' summarizes his use of written materials by stating that PSI places stress upon the written word in teacher-student communication. Although programmed texts are used, Keller emphasizes that the steps in a PSI programme are not frames in a set; they are more inclusive and can be better described as reading assignments or laboratory exercises. Also, he states in a later description of PSI that television, computers and teaching machines may all be used in the system but should not be equated with the system.

PSI courses are based upon a standard textbook, journal articles, other readings and are supplemented by a study guide. This guide breaks the course into units. Each guide consists of an introduction, statement of objectives, procedure, study questions, and supplementary materials as dictated by the specific course.

Keller and Sherman (1974) indicate that a well-written procedure section tells the students what to do, how to self-test comprehension, how to decide whether to proceed or review and how to decide when they have finished. The study guide may include any further information that helps the students to complete their work successfully, the main criterion for a study guide being that it works.

## STUDENT PROCTORS

These are usually students who have successfully completed the course. Their job is to score quizzes, discuss answers, tutor students who are experiencing difficulties, and provide feedback to the instructor on the course in general.

According to Keller, the proctor is a vital component of the system, being a primary source of external reward for the mastery of each unit; they constitute the personal contact between student and system and provide immediate feedback to the students. The proctor allows interaction between the student and the system and a degree of flexibility that can only come from human interaction: credit can be given to a student if he appears to understand the item tested, even though the answer to the item was incorrect due to some minor error. Also the proctor can discuss correct responses in order to determine whether the answer is guessed or arrived at for invalid reasons.

The proctor is helped in his work by 4 factors:

1. his experience with a similar course in which he was recently successful;
2. a weekly proctor's meeting in which every question on every test

may be discussed;
3. a handy list of acceptable answers to the questions, to jog his memory if needed;
4. usually, relevant knowledge from advanced courses within the same subject-matter field.

In addition to the proctors there is the classroom assistant whose tasks include distribution of assignments, study questions, checking of supplies, maintenance of progress charts etc. Later, he may take the place of the instructor at various meetings and in general give the instructor time to alter the course as and when necessary.

The instructor's role is that of a manager of learning. He selects and analyses the material to be mastered, he decides how to present it, and constructs the various questions based upon it. Keller emphasizes that the textbooks must be carefully read before selection to determine their suitability for students who will be questioned upon it in detail - errors, confusions and contradictions that might go unnoticed in a conventional system will stand out as 'unsightly blemishes' in PSI.

The instructor is also responsible for dividing material into suitable units, devising suitable study questions to help students and test questions for the mastery test. He will receive a great deal of information on everything he has devised for the course and as a result he will be responsible for the continual modification of the system, as he aims for perfection.

**LECTURES**

Lectures are used for enrichment rather than to transmit information, which in PSI is done through the use of study guides and the text. Lectures are given to demonstrate the excitement of the discipline; and may be used to show films or for other special events such as visiting speakers. They serve a purely supplemental function. As Keller and Sherman point out:

> A few lectures are useful. The instructor can serve as a model, impart style, even provide inspiration. In fact we may expect a very polished and challenging presentation now that such events are infrequent. Such lectures must be fine indeed; otherwise they will not attract an audience. Since the lecture is given for motivational reasons rather than for the transmittal of essential information, PSI lectures are 'extras' and not required. (Keller and Sherman, 1954, pp.41-42)

## PERSONALIZED SYSTEM OF INSTRUCTION: THE RESEARCH

In the 'Review of Research in Education' (1975) McKeachie & Kulik summarized the results of the available evaluations of PSI:

1. Keller plan is an attractive teaching method to most students. In published reports students rate Keller plan more favourably than teaching by lecture.
2. Self-pacing and interaction with tutors seem to be features most favoured by students. There are reports of higher-than-average student drop-out rates for Keller courses.
3. Content learning, as measured by final examinations, is adequate: in published studies final examination performance always equals and usually exceeds performance in lecture sections.
4. Students almost invariably report that they learn more from PSI and report expending more time and effort.

In addition to these generally favourable reports evidence was also accumulating which showed the relative importance of the various components of PSI in contributing to the system's overall successful performance.

## STUDENT RATINGS OF THE COURSE

Gallup (1970) found favourable comments on a psychology PSI course from 98% of students and Born & Herbert (1971) found that only 7 of 145 students would not recommend the course to other students. When looking at Library Science, Knightly & Sayre (1972) found that the self-paced mode of instruction was considered better than lecture courses by 100% of students; the course being rated as 'above average' or 'one of the best' by 93%.

Perhaps it should be noted here that in the early days many academics tended to describe PSI as a de-humanizing, programmed approach to instruction, and stress its links with Skinner's behaviourist orientation. Students who studied such courses, however, liked its freedom, self-determination and personal interaction! Kulik et al (1990) show that there is a medium effect size of 0.4 for measures of course satisfaction.

## CONTENT LEARNING

Kulik, Kulik and Smith (1976) presented details of 39 studies which satisfied their basic requirements of sound experimental design. In 38 of the 39 studies final exam performance was better in the PSI course. In 34 of the comparisons the difference was considered to be statistically reliable. The

authors state that the differences were large enough to have practical significance: they show that for a typical result a student, who is average on a standard test, scoring 50% in a typical lectures-based course, would score 75% if she takes the PSI courses, indicating an Effect Size of 0.66. An early meta-analysis (Kulik, Kulik, and Cohen, 1979) found 61 studies which compared PSI and conventional teaching. PSI final examinations were 8 percentage points higher, giving an average effect size of 0.49. This was confirmed in the later meta-analysis (Kulik et al, 1990), although Slavin (1987) has criticised some aspects of the analysis.

But what about the long-term impact PSI has on students? Again, Kulik, Kulik and Cohen (1979) indicated that PSI comes out with better results. In fact the differences were greater after the retention period (which varied from 3 weeks to 15 months), suggesting that final exams, if anything, underestimate the magnitude of PSI's effect.

Another way in which the effects of PSI can be investigated is the transfer study, where students study, for example, MATHS1 by either PSI or conventional methods. They then study MATHS 2 together in a conventional class. How does PSI affect performance in MATHS 2 ? The transfer studies all indicate that those students who take the PSI option for the initial course outperform conventionally taught students on the second course, even though the second course is a conventionally taught course.

## COMPONENT ANALYSIS

There can be no doubt that PSI is effective, the question is, why is PSI so effective? Various studies have attempted to throw some light on this question, and have looked at the various components of PSI.

Feedback seems to be important and there are several studies in which groups receiving delayed feedback were compared with groups receiving immediate feedback. The achievement of the delayed feedback groups was significantly lower than groups receiving immediate feedback.

Relaxation of the mastery criterion leads to a decline in the effectiveness of PSI, as was anticipated by Keller. Semb in 1974 showed that a high mastery criterion (100%) produced better final test performance than a low mastery criterion (60%) and this was confirmed in a later study (Phillips and Semb, 1975) and this was confirmed in later meta-analytic studies (Kulik et al, 1986), in which mastery level was positively related to the effect size. With mastery set at 91-100% the average effect size was large (0.82), but fell to medium below this criterion level (0.4).

Surprisingly, the proctors don't seem to be an essential feature of PSI and in a Keller PSI course with a section which included self-grading (the proctors only acting as administrators) both groups performed equally well. Written feedback when compared with feedback provided by the proctor and showed no significant differences.

Research studies have also failed to show that self-pacing is important. Several studies show no difference on final examination or student evaluation for teacher-paced and self-paced groups (Beneke and Taylor, 1975; Calhoun, 1976; Lewis, 1972).

Although Keller recommends lectures in PSI courses as motivational agents, some instructors have built in conventional lectures in addition to the individualised work, but they did not improve the performance of PSI students. Minke and Carlson (1973) found that adding lectures did not change the percentage of students passing, the withdrawal rate or the overall evaluation of students passing the course.

## BLOOM'S LEARNING FOR MASTERY

Bloom (1968) indicated that a learning strategy for mastery may be derived from the work of Carroll (1963), with support from the ideas of Morrison (1926), Skinner (1954) and Bruner (1966). Block (1971) traces the influences back to Washburne's (1922) Winnetka Plan and Morrison's approach at the University of Chicago's Laboratory School of 1926. In both systems mastery was defined in terms of particular educational objectives and mastery of each unit was required of students before they proceeded to the next unit. An ungraded diagnostic-progress test was administered at the completion of each unit to provide feedback on the adequacy of the student's learning. The test either indicated mastery, usually set at a level of 80-90%, or it highlighted the material the student still had to master. On the basis of the results of the diagnostic test each student's original instruction could be supplemented with appropriate remedial materials so that he could complete his unit, by obtaining a score reaching the mastery criterion. Although popular well into the 1930s these ideas eventually disappeared due primarily to a lack of the technology to sustain them (Block, 1971). They didn't emerge again until the late 1950s and 1960s, as a corollary of programmed instruction. The basic idea underlying programmed instruction being that the learning of any behaviour, no matter how complex, rested upon the learning of a sequence of less complex component behaviours. Theoretically, by breaking down complex behaviour into a chain of component parts and by ensuring mastery of every link in the chain, it would be possible for any student to master even the most complex skills.

The research in programmed instruction showed that it worked very well for some students, especially those requiring small steps, drill and frequent reinforcement, but it was not so effective for all students, and it did not provide a useful mastery model in itself.

Bloom (1968) based his 'Learning For Mastery' (LFM) on the model proposed by J. Carroll (1963). This requires 'that the task can be unequivocally described and that means can be found for making valid judgement as to when the learner has accomplished the learning task - that is, has achieved the learning goal which has been set for him.'

Carroll's model advocates a 'criterion referenced' assessment procedure in which the student's performance is judged according to how well he has done by comparison with some predetermined criterion. This contrasts with the traditional 'norm referenced' assessment in which a given student is judged against the performance of his colleagues.

Criterion referenced tests provide a standard against which all student performances can be measured, and can ensure a minimum standard of competence. The driving test is a good example of this kind of assessment: all learners take the test and if they don't reach the criterion level they fail but may repeat the test until they reach the set level of competence. Norm referencing of a group at a driving school may result in a good distribution of grades from fail to A-grade, but this would not necessarily mean that even the A-grade students were able to reach the criterion level. It would mean that they could out-perform their colleagues, but the overall standard could still be below that required for safe driving on the highway.

Bloom (1968) transformed Carroll's conceptual model into an effective working model for mastery. He argued that if aptitudes were predictive of the **RATE** at which (and not the **LEVEL TO** which) a student could learn a given task, then it should be possible to fix the degree of learning expected of students at some level of mastery and to systematically manipulate the variables in Carroll's model so that all or almost all students attained it. He also reasoned that if aptitude for a subject was normally distributed in the student population and they were provided with uniform instruction in terms of quality and learning time, then achievement on completion of instruction would be normally distributed.

However, if the learner were to receive optimal quality of instruction and learning time then a majority of students could expect to attain mastery and there would be little or no relationship between aptitude and achievement.

## LEARNING FOR MASTERY STRATEGIES

Bloom (1968) acknowledged the fact that there were many different strategies for mastery learning.

He suggested that one ideal strategy would be to have a good tutor for each individual student, but acknowledged that this is not possible on a large scale on economic grounds. Nevertheless, the tutor-student relationship is a useful model to consider when working out the details of a more economically viable strategy.

Other strategies suggested include allowing students to go at their own pace, guiding students as to the most suitable courses, and providing different tracks or streams.

Bloom and his colleagues, following in Morrison's footsteps at the University of Chicago, experimented with various mastery strategies using diagnostic procedures and alternative methods and materials as supplements to regular classroom instruction. With such an approach the aim was to bring all students up to mastery standard within the regular term or period of calendar time in which the course was usually taught.

To develop mastery learning it is necessary, according to Bloom, to define what mastery is and whether or not a student has achieved it. Specification of objectives and content of instruction are, therefore, necessary preconditions. The translation of these into evaluation procedures is also necessary. Implicit in defining outcomes and evaluation procedures is the distinction between the teaching-learning process and the evaluation process. The teaching-learning is intended to prepare a student for achievement of a specific level of competence in an area of learning. The summative evaluation at the end of a course is intended to determine the extent to which this has been achieved. This is not a competitive approach to education because Bloom sees much of learning and development being destroyed by primary emphasis on competition.

The LFM approach relies on setting standards of mastery and excellence, followed by a strategy to bring as many students as possible to this standard. One method employed was to base the standards for the new approach on those achieved by the top group prior to the introduction of the mastery approach. In this way students were informed of the performance required but were not in competition for grades, they were to be judged on the basis of levels of mastery actually obtained by students in the previous year. This enabled a more co-operative approach, students helping each other without the fear that special advantage was being given.

Bloom's mastery approach was based on materials and methods used in previous years, within a similar timetable, in the belief that requiring extensive training of teachers would limit the acceptance of the new approach.

The operating procedures were used as supplements to the regular teaching. They were devised to ensure mastery of each learning unit in such a way as to reduce time required by altering quality of instruction and ability to understand the instruction. Two main operating procedures were adopted.

The course was broken down into small units of learning, such as a chapter in a textbook, a time unit of the course or a well defined portion of a course, usually a week or two of learning activity.

Following the breaking down of subject matter, brief diagnostic progress tests were devised, which were then used to determine the extent of student mastery of a given unit. Such tests also indicated what the student still had to accomplish when he failed to reach the mastery criterion. These diagnostic tests were referred to as Formative Evaluation.

Frequent formative evaluation paces the students and helps motivate them, according to mastery theory, ensuring that each unit of learning is mastered before subsequent learning tasks are started. Early parts of the course may require more frequent formative testing than later sections, but there will be some variation in testing in various parts of a course.

Bloom believed that, for students who thoroughly mastered the unit, formative evaluation would reinforce the learning and reduce course anxiety about achievement.

Students lacking mastery of a particular unit had the specific deficits identified and particular instructional materials or processes were recommended to correct the learning. A very specific prescription was seen as being essential for students failing a formative test.

Thus, the main purpose of the formative evaluation was to determine the degree of mastery of a given learning task and to identify those aspects of the task which were not mastered.

The term 'summative evaluation' was given to the much more general assessment of the degree to which the 'larger outcomes have been attained over the entire course.' Such summative evaluation could be used to grade pupils and to report to parents or administrators. It was felt that several skills or concepts which combine to make a broader ability should be presented before a summative examination is administered. Summative tests should not be thought of only as end of course or final examinations, they could equally be used every four or six weeks to assess mid-term attainment of the various abilities taught in a course.

Bloom, Hastings and Madaus (1971) suggest that it is the level of generalization which differentiates summative from formative evaluation, and they indicate that summative assessment is linked to Tyler's more general descriptions of desired behaviours, whereas formative evaluation is more closely allied to Gagné's detailed prerequisite capabilities for each large aim.

Bloom found that students do attempt to work on their difficulties if they are given specific suggestions of what to do. The best procedure identified was to have small groups of students (two or three) meet regularly for up to an hour each week to review the results of formative tests and to help each other overcome difficulties. Tutorial assistance was found not to be as popular as peer group meetings for secondary and higher education students. Other types of learning resource prescriptions, recommended by Bloom, were:

1. Rereading pages of the original instructional materials.
2. Reading alternative materials.
3. Use of specific pages of workbooks or programmed texts.
4. Use of selected a-v aids.

Ryan and Schmidt (1979) identified the most successful corrective strategies as being those which include objectives plus a problem testing the objectives of the previous lesson, discussion of the problem, specific prescriptions for using the text, class notes and handouts, and alternative resources, such as texts, workbooks, games and kits. When correctives consisted of objectives or problems only their effectiveness was considerably diminished.

## AFFECTIVE AND COGNITIVE CONSEQUENCES OF LFM

Bloom suggested that if a system of formative and summative evaluation informs the student of his mastery of the subject, he will come to believe in his own mastery and competence and this will change his view of himself and the world.

He will certainly begin to respond more positively to the subject he has mastered. Mastery will provide the necessary reassurance and reinforcement to help the student to view himself as adequate thus contributing to positive mental health:

> Finally, modern society requires continual learning throughout life. If the schools do not promote adequate learning and reassurance of progress, the student must come to reject learning - both in the school and later in life. Mastery learning can give zest to

> school learning and can develop a life-long interest in learning. It is this continual learning which should be the major goal of the educational system. (Bloom, 1968/81, p.174)

Bloom's initial description of the cognitive outcomes described results for an experiment conducted from 1965 to 1967. Using the standards of grading adopted in 1965, prior to the introduction of a mastery approach, the aim was to use a variety of techniques during ensuing years to turn all students into 'A' students (ie. all students in the years 1966 and 1967 would obtain the same grades, using the same criteria, as were obtained by the top students of 1965). The results indicated that for 1967 80% of students obtained the level of mastery previously associated with 'A' grade students and by 1967 this had risen to 90% of students. The techniques certainly appeared to work in these initial studies and led to further investigations.

## LEARNING FOR MASTERY: THE RESEARCH

Bloom (1976) summarized the early research on the mastery approach in 'Human Characteristics and School Learning.' These were 'a series of small experimental and school-based studies' in which students from the University of Chicago contrasted mastery with traditional methods.

Cognitive entry behaviours were measured at the beginning of each learning task, using a formative diagnostic test, and students were tested again at the end of the teaching session. Those students in the mastery groups received additional tuition, as indicated by the formative diagnostic test, if they failed to reach the set mastery criterion. They were then tested again after these 'corrective procedures' using a parallel form of formative test.

If Bloom's theory is correct we would expect a student's performance on one learning task to be highly correlated with performance on the next task, in the traditional group. The best students in this group would perform well on the first task and this would enable them to perform at an equally high level on the next task.

This would not be the case in the mastery group because those students who were failing to reach mastery on the initial task would receive appropriate materials and time to enable them to reach a position previously associated with the more able students. Having gained this position they would be just as likely to perform at a higher level in the next task. Thus, their performance on the original task would not be predictive of their performance on the second task. That at least is the theory, and it also predicts that there would be high correlations between one task score and the next task score for traditional

courses, but much smaller (or even zero) correlations between one task and the next for mastery groups.

Bloom (1976) quotes studies covering subjects such as second language acquisition, matrix algebra, probability and an 'imaginary' science. The correlations between achievement on the first learning task in the series and the final (summative) achievement are:

| | **Mastery** | **Non-mastery** |
|---|---|---|
| Anderson (1973) | 0.31 | 0.68 |
| Arlin (1973) | 0.19 | 0.49 |
| Block (1970) | 0.44 | 0.78 |
| Levin (1975) | 0.59 | 0.72 |
| Ozcelik (1974) | 0.36 | 0.48 |
| Median | 0.36 | 0.68 |

These results demonstrate an overall agreement with Bloom's prediction from his theory, although the mastery correlations do not reach the zero figure that Bloom had hoped for. There is a definite trend for a much reduced correlation between initial performance and final performance under mastery conditions. The larger correlation of 0.68 for non-mastery students indicates the extent to which performance on the initial learning task will influence the final achievement when there are no correctives for defective learning.

Block quotes the work of Kim (1971) and his colleagues (Lee et al., 1971) as showing the power of LFM. In these studies, using a group-based, mastery learning strategy, with pupil-teacher ratios of 70:1, Kim and his colleagues converted much of the Korean elementary and middle-school system into a mastery model. The pilot study taught 272 seventh grade students an eight-session unit on geometry. Kim found that 75% of the mastery group scored at the mastery criteria of 80% correct, compared with only 40% of the traditionally taught students reaching this level.

On the basis of this successful experiment Kim expanded his next study to 5,800 students, who studied maths and English over an eight week period. In English, 75% of mastery students reached the 80% criterion, compared with 28% of the traditional group; in maths, 67% of mastery students reached the criterion, against 39% of the traditional group.

The third experiment was much larger in scope. Rural and urban pupils learned maths, English, physics and biology over a full school year. There were 25,887 students in the experiment. The results were quite impressive, with

2,000-3,000 more students gaining mastery in maths, physics and biology when compared with traditional methods.

Similar results were obtained by Lee et al. (1971), for arithmetic and science when taught to 12,504 pupils in 5th and 6th grades, with 27-32% more pupils reaching the 80% criterion under mastery conditions.

These differences in performance demonstrate the scale of achievement possible for students who are not reaching their maximum potential under normal instructional circumstances. Block (1974) applauds the magnitude of the successful application of mastery learning:

> Here a difference of a few percentage points, let alone differences of 20 to 30 points, between the percentage of students reaching a high performance level under mastery conditions and the percentage reaching the same level under non-mastery learning conditions translates into differences of thousands of students who are achieving levels in their learning they might never ordinarily have reached. (Block, 1974, p.34)

Further research on mastery learning in the classroom was summarized by Block and Burns (1977) in 'Review of Research in Education' (vol. 4), edited by L.S. Shulman. In this later paper the authors reported a meta-analysis on the research, giving an average effect size for the LFM-taught students of 0.83, which would move the average student in the non-mastery group from the 50th position in class to the top 20% of the class, a large educational effect. Keller studies were also included in this review.

## LFM AND PSI: A REVIEW OF THE RESEARCH

The research on mastery learning as found in the approaches of Bloom and Keller was summarized in a comprehensive review by Block and Burns (1977). In this paper the authors reported a meta-analysis on the research, giving the effect size for LFM achievement comparisons and a combined retention effect size for PSI and LFM.

A total of 97 comparisons were identified, involving different types and numbers of students. Mastery students scored higher than non-mastery students 89% of the time, and statistically significantly higher 61% of the time. For

retention (combining PSI and LFM) the mastery students were almost always superior to the non-mastery groups, significantly so 63% of the time. The average effect size for the LFM-taught students was 0.83. A similar large effect size was obtained from studies of LFM and PSI which included a retention test, with an average retention effect size of 0.67.

In addition, Block and Burns looked at the variability in learning. According to Bloom's theory the mastery approach will lead to greater uniformity in learning when compared to traditional methods, which should show increasing variability. In 80 studies comparing variance in achievement test scores, mastery students showed less variation in their scores 74% of the time, and less variability in retention scores 85% of the time. The reduction in variance was approximately 50% for both achievement and retention measures.

## KINDS OF LEARNING

Block and Burns also looked at learning in qualitative terms. They found some evidence to suggest that mastery strategies may be more likely to elicit higher order than lower order behaviours. Several researches found that mastery students were superior on essay-type questions that included comprehension, application, analysis, synthesis and evaluation skills, and that mastery students were more able to deal with these complex high order skills than the lower order multiple-choice knowledge questions.

There was some evidence, however, that for retention periods of 6 or 18 months it was the lower order behaviours associated with knowledge, rather than the complex application or analysis-synthesis, that were retained at a higher level in the mastery group.

The mastery approaches typically elicited more favourable affective responses from students, in particular there was a positive impact on students' interest in and attitudes to the subject matter, and also on self-concept, academic self-confidence, attitudes to co-operative learning and towards instruction. Although there is some evidence to suggest that mastery set at the 100% level adversely affects such attitudes and that an optimum figure of 90% maintains both the positive affective and achievement responses.

## TIME CONSEQUENCES

A group of studies investigated the amount of time spent in learning in mastery and non-mastery groups. Uniformly, they found that mastery strategies had a homogenizing effect on differences in study time. One study found that differences were reduced from 7 to 1, to 4 to 1 by mastery methods, and that

there was progressively less variability in the mastery study time. In the final unit of eight there was 90% less variability.

Block and Burns comment that this reduction in individual differences in study time is purchased at the price of additional study time for the slower students. Over the short term there is an increase in the average study time required for mastery of 70-80%, compared with non-mastery. In the long term this may be reduced to 20-50%.

The extra time is available from two sources. The first is the teacher who is encouraged to prepare and organize the instructional plans, procedures and materials outside of class time and prior to the instruction. This allows more time for interaction with the students. The student can also be encouraged to exert more effort, especially in overcoming procrastination, and various incentive schemes have been used, such as frequent testing, varying the size of learning units, teacher/peer pacing of instruction. The emphasis is clearly on techniques that encourage teachers and students to spend more time learning as opposed to non-learning activities.

The work of Block and Burns, and others, has been challenged by Slavin (1987), who has suggested that those educationally significant results have only been found in studies which have used teacher-prepared tests, rather than standardised assessment tests. This is a controversial area and Kulik at al (1990) retaliated with a further meta-analysis based on 108 studies (two thirds of which considered PSI, the rest LFM). The average effect size was 0.52, when examination performance was measured. This put the average student at the 70th percentile. The effects were stronger for less able students, but not significantly so (ES for low aptitude student 0.6; high aptitude 0.4). The weak effect of mastery on standardised tests was confirmed, but only 4 studies were considered in the analysis. There seems to be less variation in student performance in mastery programmes (70% of that found with traditional methods). Students are also very positive about mastery programmes, with an effect size of 0.6; and attitudes towards the subjects generally improved (effect size: 0.4).

Kulik et al (1990) conclude that mastery programmes do have positive effects on student achievement, with LFM showing a slightly larger effect size (LFM 0.59; PSI 0.48). Those studies which had the highest effect sizes tend to come from the social sciences, use locally developed tests, are teacher-paced, and have more feedback for mastery students, who are required to achieve high standards of performance (100%). With these features the average effect size is 0.9. If standardised tests are used the effect size is still remarkably high (0.8). The disagreement between Slavin and Kulik at al. continues! (Slavin, 1990)

## ACHIEVEMENTS OF THE BEHAVIOURAL SYSTEMS

In many ways the work of Bloom and Keller in their mastery learning systems represents the summit of achievement for the neo-behaviourist 'systems' approach to instruction. Both were built on firm, though differing, psychological principles and of all the applications of the behavioural science approach to the technology of education, LFM and PSI represent the only methods which consistently produce significant educational results. Clark (1983) attributes this to the emphasis placed on method, defined by Glaser (1976) as 'the conditions which can be implemented to foster acquisition of competence,' rather than on the media involved in the delivery of instruction, as represented by the physical science approach.

It is fitting that Benjamin S. Bloom, whose work has been central to the development of instructional science for a generation, should have the last words:

> If you can be moved to try these ideas with a few teachers in your school for even as short a period as three months, you can determine the validity and limits of these ideas where they belong - in your classrooms and with your teachers and students.

He continues:

> However, neither further opportunity for education nor increased financial support for education will do much to improve the education of each of our students. The answer does not lie in additional funds, new fads, or major and sweeping changes in the organization of our educational system. As I see it, the solution lies in our views about students and their learning. These have grown out of our practices and they will not be changed until we alter these practices. When the changed practices succeed in promoting more effective learning, both teachers and students will change their views. (Bloom, 1978, p.563)

And he went even further when considering Slavin's criticisms:

> It is a crime against mankind to deprive children of successful learning when it is possible for virtually all to learn to a high level. (Bloom, 1987, p.508)

**REFERENCES**

Beneke, W.M. and Taylor, D.C. (1975) Effects of Unit Deadlines on Performances in a PSI Course. Paper presented at the annual meeting of the American Psychological Association, Chicago.

Block, J.H. (1971) Mastery Learning. London: Holt Rinehart & Winston.

Block, J.H. (1974) Schools, Society and Mastery Learning. London: Holt, Rinehart and Winston.

Block, J.H. and Burns, R.B. (1977) 'Mastery learning'. In L.S. Shulman (Ed) Review of Research in Education. Volume 4. Itasca, Ill.: F.E. Peacock, 3-49.

Bloom, B.S. (1964) Stability and Change in Human Characteristics . New York: John Wiley.

Bloom, B.S. (1968/81) 'Learning For Mastery', (UCLA-CSEIP) The Evaluation Comment, 1(2). In B.S. Bloom (Ed) All Our Children Learning. London: McGraw-Hill.

Bloom, B.S. (1976) Human Characteristics and School Learning. New York: McGraw-Hill.

Bloom, B.S. (1978) 'New views of the learner - implications for instruction and curriculum', Educational Leadership, 35, 562-76.

Bloom, B.S. (1987) 'A response to Slavin's mastery learning reconsidered', Review of Educational Research, 57(4), 507-508.

Bloom, B.S., Hastings, J.T. and Madaus, G.F. (1971) Handbook on Formative and Summative Evaluation. London: McGraw-Hill.

Born, D.G. and Herbert, E. (1971) 'A further study of personalized instruction in large university classes', Journal of Experimental Education, 40, 6-11.

Bruner, J.S. (1966) Toward a Theory of Instruction. Cambridge, Mass.: Harvard University Press.

Bugelski, B.R. (1971) The Psychology of Learning Applied to Teaching. Indianapolis: Bobbs-Merrill.

Calhoun, J.F. (1976) 'The combination of elements in the personalized system of instruction', Teaching of Psychology, 3, 73-76.

Carpenter, C.R. and Greenhill, L.P. (1963) Comparative Research on Methods and Media for Presenting Programmed Courses in Mathematics and English. University Park, Pa.: Division of Instructional Services, Pennsylvania State University.

Carroll, J. (1963) 'A model of school learning', Teachers College Record, 64, 723-733 .

Clark, R.E. (1983) 'Reconsidering the research on learning from media', Review of Educational Research, 53, 4, 445- 459.

Cohen, J. (1977) Statistical Power Analysis for the Behavioural Sciences. (Revised edition) New York: Academic Press.

Crowder, N.A. (1960) 'Automatic tutoring by intrinsic programming'. In A.A. Lumsdaine and R. Glaser (Eds) Teaching Machines and Programmed Learning. Washington, D.C.: Department of Audio-Visual Instruction, National Education Association.

Dubin, R. and Taveggia, T.C. (1968) The Teaching Learning Paradox. University of Oregon Press: Eugene.

Gagné, R.M. (1962a) 'The acquisition of knowledge', Psychological Review, 69(4), 355-365.

Gagné, R.M. (1962b) 'Military training and principles of learning', American Psychologist, 17, 83-91.

Gagné, R.M. (1965) The Conditions of Learning (1st Edition). London: Holt, Rinehart and Winston.

Gallup, H.F. (1970) Individual Instruction in an Introductory Psychology Course. Paper presented at the meeting of the American Psychological Association, Washington, D.C. (Sept.)

Glaser, R. (1976) 'Components of a psychology of instruction: towards a science of design', Review of Education al Research, 46(1), 1-24.

Glass, G.V. (1976) 'Primary, secondary, and meta-analysis of research', Educational Researcher, 5, 3-8.

Hartley, J. (1974) 'Programmed instruction 1954-1974: a review', Programmed Learning and Educational Technology, 6(11), 278-291.

Holland, J.G. and Skinner, B.F. (1961) The Analysis of Behaviour. New York: McGraw-Hill

Keller, F.S. (1968) 'Goodbye, Teacher', Journal of Applied Behaviour Analysis, 1, 79-87.

Keller, F.S. and Sherman, J.G. (1974) The Keller Plan Handbook. Menlo Park-California: Wa. Benjamin.

Kim, H. (1971) 'Mastery learning in the Korean middle schools', UNESCO Regional Office for Education in Asia, 6(1).

Knightly, J.J. and Sayre, J.L. (1972) 'Self-paced instruction for library science students', Journal of Education for Librarianship, 12, 193-197.

Kulik, C.C., Kulik, J.A. and Bangert-Drowns, R.L. (1986) Effects of Testing for Mastery on Student Learning.. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 1986.

Kulik, C.C., Kulik, J.A. and Bangert-Drowns, R.L. (1990) 'Effectiveness of mastery learning programmes - a meta-analysis', Review of Educational Research, 60(2), 265-299.

Kulik, C.C., Schwalb, B.J. and Kulik, J.A. (1982) 'Programmed instruction in secondary education: a meta-analysis of evaluation findings', Journal of Educational Research, 75(3), 133-138.

Kulik, J.A., Bangert, R.L. and Williams, G.W. (1983) 'Effects of computer-based teaching on secondary school students', Journal of Educational Psychology, 75(1), 19-26.

Kulik, J.A., Cohen, P.A. and Ebeling, B.J. (1980) 'Effectiveness of programmed instruction in higher education: a meta-analysis of findings', Educational Evaluation and Policy Analysis, 6(2), 51-64.

Kulik, J.A. and Jaksa, P. (1977) 'PSI and other educational technologies in college teaching', Educational Technology, 17, Sept. 12-19.

Kulik, J.A. and Kulik, C.C. (1989) 'Meta analysis in education,', International Journal of Educational Research, 13, pp.221-340.

Kulik, J.A., Kulik, C.C. and Bangert-Drowns, R.L. (1990) 'Is there better evidence on mastery learning - a response', Review of Educational Research, 60(2), 303-307.

Kulik, J.A., Kulik, C.C. and Cohen, P.A. (1979) 'A meta-analysis of outcome studies of Keller's personalized system of instruction,' American Psychologist, 34(4), 307-318.

Kulik, J.A., Kulik, C.C. and Smith, B. (1976) 'Research on the personalized system of instruction', Programmed Learning and Educational Technology, 13, 21-30.

Lee, Y.D. et al. (1971) Interaction Improvement Studies on the Mastery Learning Project - Final Report on Mastery Learning Programme. Educational Research Centre, Seoul National University. (Nov.)

Leith, G.O.M. (1968) Second Thoughts on Programmed Learning. Occasional paper No.1. London: National Council For Educational Technology.

Lewis, J.D. (1972) Manipulation of Two Variables in a Contingency-managed Statistics Class. Paper presented at the West Coast PSI Conference. California State University.

McKeachie, W.J. (1974a) 'Instructional psychology', Annual Review of Psychology, 25, 161-193.

McKeachie, W.J. (1974b) 'The decline and fall of the laws of learning', Educational Researcher, 3, 7-11.

McKeachie, W.J. and Kulik, J.A. (1975) Effective college teaching. In F.N. Kerlinger (Ed) Review of Research in Education Vol.3. Itasca, Illinois: F.E. Peacock.

Means, R.S. and Means, G.H. (1971) 'Achievement as a function of the presence of prior information concerning aptitude', Journal of Educational Psychology, 62, 185- 187.

Minke, K.A. and Carlson, J.G. (1973) A Comparison of Two Lecture Systems in Unit Mastery Instruction. Paper presented at the annual meeting of the American Psychological Association, Montreal. (Aug.)

Morrison, H.C. (1926) The Practice of Teaching in a Secondary School. University of Chicago Press: Chicago.

Phillips, T.W. and Semb, G. (1975) A Comparison of Pre-quiz and Post-quiz Lectures in a Contingency-managed University Course. Paper presented at the annual meeting of the American Association, Chicago. (Sept.)

Pressey, S.L. (1963/64) 'A Puncture of the Huge "Programming" Boom?' In W. Schramm (Ed) The Research on Programmed Learning: an annotated bibliography. (0E- 31034) Washington, D.C.: U.S. Department of Health, Education and Welfare.

Ryan, D.W. and Schmidt, M. (1979) Mastery Learning: Theory, Research, and Implementation. Toronto, Ontario: The Ontario Institute for Studies in Education.

Schramm, W. (1964) The Research on Programmed Learning: an annotated bibliography. (0E-31034) Washington, D.C.: U.S. Department of Health, Education and Welfare.

Semb, G. (1974) 'Personalized instruction: the effects of mastery criteria and assignment length on college student test performance', Journal of Applied Behaviour Analysis, 7, 61-69.

Skinner, B.F. (1938) The Behaviour of Organisms. N.Y.: Appleton-Century-Croft.

Skinner, B.F. (1948) Walden Two. N.Y.: Macmillan.

Skinner, B.F. (1953) Science and Human Behaviour. N.Y.: Macmillan.

Skinner, B.F. (1954) 'The science of learning and the art of teaching', Harvard Educational Review, 24(2), 86-97.

Skinner, B.F. (1968) The Technology of Teaching. N.Y.: Appleton-Century-Croft.

Slavin, R.E. (1987) 'Mastery learning reconsidered', Review of Educational Research, 57(2), 175-213.

Slavin, R.E. (1990) 'Mastery learning re-reconsidered', Review of Educational Research, 60(2), 300-302.

Tait, K., Hartley, J.R. and Anderson, R.C. (1973) 'Feedback procedures in computer assisted arithmetic instruction', British Journal of Educational Psychology, 43, 161-171.

Tyler, R.W. (1932) 'The construction of examinations in botany and zoology', Service Studies in Higher Education. Ohio State University, Bureau of Educational Research Monographs, 15, 49-50.

Washburne, C.W. (1922) 'Educational measurements as a key to individualizing instruction and promotions', Journal of Educational Research, 5, 195-206.