Psychology of Learning: improving pupil performance

2

THE ART OF TEACHING AND THE SCIENCE OF LEARNING

B.F. Skinner took up the ideas of behaviourism in the 1930's and after extensive experimental research, much of it based on the performance of laboratory rats and pigeons, extended his conclusions to the higher reaches of the phylogenetic chain, to human behaviour.

Skinner's success is attributable to his isolation of a few highly repeatable phenomena in conditioning - many of which Pavlov and Thorndike had studied and named before him (Hilgard and Bower, 1975) - which were then used as a basis for analyzing more complex forms of behaviour. Following on from Watson's psychology, Skinner continued to reject mentalistic or cognitive explanations of behaviour and was interested only in the observable and repeatable.

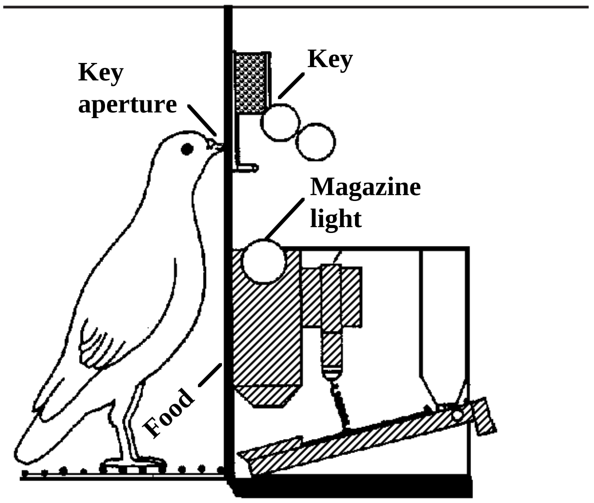
Operant Conditioning

His early work made extensive use of a piece of apparatus which is now commonly termed the 'Skinner box', shown in Figure 2-1. In such a box, the caged animal (usually a rat or pigeon) presses a lever or pecks at a key and receives food pellets as a reward. It is from this experimental base that Skinner developed the linear teaching machine. A close look at the procedures involved in this 'operant conditioning' is required to appreciate the evolution of such machines, and subsequent criticisms of their use.

The type of conditioning studied by Skinner is similar to the trial and error learning discussed by Thorndike (1898). This differs from Pavlov's procedure in that the confined animal emits the response (presses the lever) which is, according to Skinner, under 'voluntary' control. Whereas classical or Pavlovian conditioning is 'involuntary' and the response is under the control of the autonomic system; it consists of elicited behaviour, such as salivary secretion or pupillary constriction to light.

At the start of the operant conditioning procedure for a rat an unconditioned, hungry animal is placed in the Skinner box. Initially, the rat is encouraged to associate the sound of the food magazine with the delivery of food pellets in the food tray. Pellets are discharged from the magazine periodically by the experimenter to encourage

Figure 2-1. The Skinner Box



this. During the preliminary training the lever is disconnected from the magazine and at this stage the delivery of food is not contingent upon any behaviour exhibited by the rat.

When the rat has been adapted to the environment and moves to the food tray each time it hears the magazine in operation, operant conditioning can begin. The lever is connected to the food magazine so that each time the lever is depressed by the rat a pellet of food is dispensed at the food tray. The rat can then be left to explore the environment. It will show a rather low rate of lever pressing as it explores, but not a zero rate. It will press the lever a few times during an hour at a rather irregular rate (Skinner, 1938) and in doing so will receive the pellet of food, which will act as a reinforcer of the behaviour which immediately preceded it (the pressing of the lever), making it more likely to be repeated. Eventually, the relationship between lever pressing and dispensing of food is made and the rat presses the lever at an increasing rate, as described by Skinner:

> On the first day of conditioning a first response was made five minutes after re-

> lease. The reinforcement had no observable effect upon the behaviour. A second response was made 51.5 minutes later, also without effect. A third was made 47.5 minutes later followed by a fourth 25 minutes after that. The fourth response was followed by an appreciable increase in rate showing a swift acceleration to a maximum. The intervals elapsing before the fifth, sixth, and following responses were 43, 21, 28, 10, 10 and 15 seconds respectively. (Skinner, 1938, p.67-68)

Shaping Behaviour

Not all behaviours can be considered to be discrete units and for some behaviours the experimenter may have to wait hours, days or even weeks before they emerge. The original probability of the response in its final form is very low; in some cases it may even be zero. Fortunately, there is a procedure for producing the response more quickly, the experimenter takes control of the food magazine by disconnecting the lever from the food magazine and rewards successive approximations to the desired response.

> To get the pigeon to peck the spot as quickly as possible we proceed as follows: We first give the bird food when it turns slightly in the direction of the spot from any part of the cage. This increases the frequency of such behaviour. We then withhold reinforcement until a slight movement is made towards the spot.... We continue by reinforcing positions successively closer to the spot, then by reinforcing only when the head is moved slightly forward, and finally only when the beak actually makes contact with the spot. We may reach this final response in a remarkably short time. (Skinner, 1953, p.92)

This procedure is termed 'shaping' the behaviour and it enables the experimenter to bring a rare response to a very high probability in a short time. Having established the response the pigeon will continue pecking the key and receiving food until sated. The animal is usually reduced to 80% of its normal body weight in order to provide the motivation necessary to maintain the lever pressing or key pecking.

Extinction of a Response

If the lever is disconnected from the food magazine the animal's response rate will decline, a property noted by Pavlov, which he termed 'extinction' of the response. This may be used to produce a differential response, as is the case in discrimination training. For example, a pigeon can be trained to peck at a key only when it is illuminated by light of a particular colour (green), but not when it is another colour (red). Initially , it is trained to peck at the unilluminated key until a stable rate is observed. The training then proceeds with a food pellet being delivered for each peck only when it is green. When it is red the animal receives no reinforcement. This soon produces a differential rate of pecking: a high rate for green and an increasingly lower rate to red as the response extinguishes. The special stimulus which must be present in order for a reinforcement to occur (green key) is called a Discriminated Stimulus or S^D.

Schedules of Reinforcement

When each response receives a reinforcement the procedure is termed 'continuous reinforcement'. Of course, under such a schedule of reinforcement the animal soon becomes sated and refrains from responding at a high rate until it is again hungry. Skinner discovered that a high rate of responding could be maintained under a variety of reinforcement schedules, without the necessity to reward each and every response, a technique that 'gets more responses out of an organism in return for a given number of reinforcements' (Skinner, 1953, p.99).

He distinguished between schedules which are controlled by the system outside the organism and those controlled by the animal's behaviour. An example of the first is a schedule based on the passage of time, eg. the animal receives a reinforcement every 5 minutes. When a reinforcement is given every fiftieth response, it is an example of the second type of control. Both types of these 'intermittent schedules' can be either fixed or variable. Thus, a reinforcement in a fixed interval schedule is given after the set time has elapsed, eg. every five minutes; with a variable-interval schedule the reinforcement is given on **average** after a set time, eg. on average every five minutes, with the intervening interval varying between a few seconds or as long as ten minutes. This type of schedule produces behaviour which is extraordinarily persistent and very difficult to extinguish. A fixed ratio schedule also produces a very high rate of response, but the most powerful of all the schedules is the variable-ratio, where the reinforcement occurs after a given number of responses on **average**, the pigeon or rat becoming a victim of an unpredictable contingency of reinforcement.

Cultural Engineering

The results of Skinner's early work were published in his 'Behavior of Organisms' (1938), but his next major work is a somewhat startling contrast, until it is realized that Skinner's first career was journalism. 'Walden Two' (1948), a novel, 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 (see Chapter 4), 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).

In addition to the question of comparative effectiveness researchers were also investigating the nature of the components of the teaching programme. There are six components of an efficient linear programme (1) an ordered sequence of items, through which the student works, (2) at his own pace, in (3) short steps (4) making few errors in his (5) written, constructed responses, to which he receives (6) immediate knowledge of results. Each has received attention in experimental evaluations of the effectiveness of programmed instruction.

A Logical Sequence?

Five interesting experiments compared logically sequenced programmes with a presentation of the items in random order and, surprisingly, three of the experiments showed no difference, one showed no difference on a retention test and the remaining comparison showed a difference in favour of the logical structure. The results were obtained with short programmes and it was concluded that longer programmes would require a more ordered sequence. Gagné and his collaborators (Gagné, 1962a; Gagné and Paradise, 1961; Gagné et al, 1962) analyzed a number of tasks (solving algebraic equations) and found evidence for hierarchies of subordinate learning sets, thus providing a basis for logical structure within programmes. A subordinate learning set being defined as what the subject needed to know in order to be able to perform a new task. However, Mager (1961a) complicated matters by demonstrating that when students in an electronics course were given a free hand to learn by questioning an instructor, they took a path which bore little resemblance to the logical sequence in which the course was taught. The instructors tended to move from parts to wholes, whereas the students typically moved from smaller wholes to larger wholes. A seemingly straightforward issue was not resolved by these competing results. Mager captured their essence when he claimed that:

> Where it is necessary to teach one thing before another, do so. But be careful! There isn't as much reason for this kind of sequencing as instructors like to think. (Mager, 1961b)

Later, McKeachie (1974a) also commented on the ability of learners to withstand quite considerable distortions of the original programme sequences, suggesting that the motivational effect of 'surprisingness' is sometimes neglected and that 'randomness may not make sense conceptually but it may be more fun than a sequence so logical and with steps so small that one knows exactly what is coming next.'

A Self-paced System?

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.

Short Steps?

Skinner proposed that programmes should have short steps, which would lead to few errors in the learning sequence and, initially, this was controlled by the physical size of the frame window in a Skinner machine. Much of the early evidence supported short step size. Evans, Glaser and Homme (1960) tested programmes teaching conversion to unfamiliar number bases, with 30, 40, 51 and 68 steps and found the 51 and 68 step versions to be superior to the others on both immediate and delayed retention. Coulson and Silberman (1960) obtained similar results for programmes teaching psychology items. But, Smith and Moore (1961) found no difference when teaching spelling by means of programmes of 1, 128, 830 and 546 steps.

One of the most fascinating series of early experiments employed training films, such as the assembly of an automobile ignition distributor (Maccoby and Sheffield, 1958; Margolius and Sheffield, 1961; and Weiss, Maccoby and Sheffield, 1961). The length of film viewed before permitting practice was varied in each experiment and it was found that more learning came from gradually increasing step size rather than from maintaining short or long steps. When the students were given control of the film sequence, they also gradually increased the length. Students only permitted to practice short steps showed a gradually deteriorating performance, adding to a body of evidence suggesting student impatience and boredom with long programmes having only short steps.

Although the balance of the early studies was in favour of short steps, Hartley suggested that the following issues were also 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)

An Active Response?

One of the major differences between operant and classical Pavlovian conditioning was claimed by Skinner to be the active, 'voluntary' response of the animal to a given stimulus. This aspect of the early animal research is included in programmed instruction by the demand for an overt, observable, constructed response, usually a written response to a stimulus item in the programme. Unfortunately, this characteristic of the linear programme was by no means unequivocally supported by the early research. The great majority of the studies showed no statistically significant differences between the amount learned from overt and covert responses. Schramm identified 16 research investigations demonstrating this result. Since there was no difference, and since covert or 'thinking' the response took less time, the covert mode could be considered more efficient. A minority of cases demonstrated some difference in favour of the overt, constructed response, mainly for complex subject matter (Cummings and Goldstein, 1962).

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. Schramm concluded that the results 'are going to cause researchers in this field to do a great deal of thinking about the principles behind programmed instruction' (p. 10). Of course, when the step size is such that error levels are reduced to a minimum, overt responding may become superfluous.

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.

Immediate Knowledge of Results?

The majority of the early studies did support another of the central tenets of Skinnerian programming, the provision of immediate knowledge of results (Schramm, 1964, p.10). Meyer (1961) found significantly more learning from immediate knowledge when compared with a group that waited for the next meeting to receive results. However, not all the research demonstrated advantages for immediate knowledge of results. Glaser and Taber (1961) did not find any difference when the knowledge of results was varied from 100 percent continuous knowledge of results to 50 or 25 percent knowledge, or when it was received in a variable ratio. They concluded that if the probability of error is low, as in typical linear programmes, knowledge of results is less important than when the probability of error is high.

Later studies confirmed the complexity of the issue. Grundin (1969) found that overt responding interacted with frequency of feedback, making the knowledge of results superfluous, or even detrimental. 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)

Intrinsic Programming

One 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. The student turns to the page number associated with his answer, which may contain the next unit of information, if the answer is correct, or the correctional information followed by the original question, if the answer is wrong.

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).

The term 'intrinsic' programming was used to indicate that the necessary programme of alternatives was built into the material itself such that no external programming device was required, in contrast to the early experiments which were using computers as 'extrinsic' programming devices.

Schramm found few research evaluations of the intrinsic methods, but those which were available showed no differences between the two methods. Hartley (1974) agreed with Schramm's overall 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.

Few Errors?

The larger frames of information and the method of responding in Crowder's method naturally allows for more errors in the programme, but in this case this is seen to be a positive virtue. Leith (1968) demonstrated that for 10 and 16-year-olds there was no relationship between success on test performance and numbers of errors made. This was confirmed by Elley (1966). A further study by Leith showed that making and overcoming errors may sometimes help learning. Children were taught to calculate in bases other than ten. One group learned four different bases, another learned two (with twice as much practice on each) and a third group learned only one base, but with four times the practice. Each group read the same number of frames. The group with two bases made most errors when they changed to the second base, but were best on a transfer test to an entirely new base.

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).

The Art of Scientific Application

The results of these early research studies do seem to cast doubt on most of the characteristics of linear programmes, which were derived from Skinner's psychological theories. This may well be because the translation of psychological principles into educational practice has not always been as precise or accurate as Skinner claimed (Bugelski, 1971). For example, linear programmes embody the Skinnerian principle of shaping behaviour, the student being led step by step to the final goal. But in normal laboratory usage shaping refers to a sequential act. The pigeon is trained to turn clockwise in a circle and then to peck at a key and then climb on a perch. When training is completed the acts are carried out in succession by the animal, but this is not so for the student who ends the programme by carrying out a completely new act, one which he may never have carried out before. The student does not go through the entire sequence of steps in the programme when performing the last operations.

Thorndike had used the words 'right' and 'wrong' as rewards and punishments, with the consequence that knowledge of results was adopted as a perfectly satisfactory secondary reinforcer in teaching machines. 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.

Later Research: Meta-analysis

Three decades and more further on there is little discussion concerning the merits of one form or another of programmed instruction. 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.

There are, however, new techniques for evaluating the results of the many early and varied studies of programmed instruction, which attempt to provide estimates of the magnitude of the benefits to be gained from such educational innovations and interventions. Kulik, Cohen and Ebeling (1980) have used a statistical procedure called 'meta-analysis' (Glass, 1976) for the statistical analysis of a large collection of results from individual studies for the purpose of integrating findings. In a traditional review an innovatory method may be shown to be better than a more conventional approach in most cases, but such a review does not say whether the new method wins 'by a nose or a walkaway' (Glass, 1976). Kulik, Cohen and Ebeling 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. An effect size of 1.0 would indicate that the difference achieved by the new method is equal to an increase of one standard deviation of the control group's score, shifting the position of an average pupil in the control group from the 50th percentile point to the top 20% of the class. An effect size of 0.8 and above demonstrates a large effect of the experimental treatment, which certainly has educational significance; an ES of 0.2 and less is small and educationally insignificant; an ES of 0.5 is a medium effect and the experimental treatment warrants further investigation and consideration.

Kulik, Cohen and Ebeling (1980) 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 analyzed 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 analyzed 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 thirty 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)

REFERENCES

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

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.

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

Coulson, J.E. and Silberman, H.F. (1960) 'Effects of three variables in a teaching machine', *Journal of Educational Psychology*, 51, 135-143.

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.

Cummings, A. and Goldstein, L.S. (1962/1964) 'The Effect of Overt and Covert Responding on Two Kinds of Learning Tasks.' 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.

Elley, W.B. (1966) 'The role of errors in learning with feedback', *British Journal of Psychology*, 36, 296-300.

Evans, J.L., Glaser, R., and Homme, L.E. (1960) 'An investigation of variation in the properties of verbal learn ing sequences of the teaching machine type.' In A.A. Lumsdaine and R. Glaser (Eds) *Teaching Machines and Programmed Learning*. Washington, D.C.: Department of Audio-Visual Instruction, National Education Association.

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. and Paradise, N.E. (1961) 'Abilities and learning sets in knowledge acquisition', *Psychological Monographs*, 75(15), No.518.

Gagné, R.M., Mayor, J.R., Garstens, H.L. and Paradise, N.E. (1962) 'Factors in acquiring knowledge of a mathematical task', *Psychological Monographs*, 76(7), No.526.

Glaser, R. and Taber, J.I. (1961/64) Investigations of the Characteristics of Programmed Learning Sequences. 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.

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

Grundin, H.U. (1969) 'Response mode and information about correct answers in programmed instruction: a discussion of experimental evidence and educational decisions'. In A.P. Mann and C.K. Brunstrom (Eds) *Aspects of Education al Technology, III*. London: Pitman.

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

Hilgard, E.R. and Bower, G.H. (1975) *Theories of Learning.* Englewood-Cliffs, N.J.: Prentice-Hall.

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

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.

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

Maccoby, N. and Sheffield, F.D. (1958) 'Theory and experi mental research on the teaching of complex sequential procedures by alternative demonstration and practice'. In G. Finch and F. Cameron (Eds) *Symposium on Air Force Human Engineering, Personnel and Training Research*. Washington, D.C.: National Academy of Sciences, National Research Council.

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.

Mager, R.F. (1961a) 'On the sequencing of instructional content', *Psychological Reports*, 405-413.

Mager, R.F. (1961b) 'On the sequencing of instructional content'. In I.K. Davies and J. Hartley (Eds) *Contributions to an Educational Technology*. London: Butterworths.

Margolius, G.J. and Sheffield, F.D. (1961) 'Optimum methods of combining practice with filmed demonstration in teaching complex response sequences: serial learning of a mechanical assembly task'. In A.A. Lumsdaine (Ed) *Student Response in Programmed Instruction. A Symposium on Experimental Studies of Cue and Response Factors in Group and Individual Learning from Instructional Media*. Washington, D.C.: National Academy of Sciences, National Research Council.

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.

Meyer, S.R. (1961) 'Report on the initial test of a junior high school vocabulary program'. In A.A. Lumsdaine (Ed) *Student Response in Programmed Instruction. A Symposium on Experimental Studies of Cue and Response Factors in Group and Individual Learning from Instructional Media*. Washington, D.C.: National Academy of Sciences, National Research Council.

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.

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

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, 86-97.

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

Smith, W. and Moore, J.W. (1961/1964) 'Size-of-step and Achievement in Programmed Spelling.' Lewisburg, Pa.: Bucknell University. 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.

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.

Thorndike, E.L. (1898) 'Animal Intelligence', *Psychological Review*, Monograph Supplement, 2, No.8.

Weiss, W., Maccoby, N. and Sheffield, F.D. (1961) 'Combining practice with demonstration in teaching complex sequences: serial learning of a geometrical construction task'. In A.A. Lumsdaine (Ed) *Student Response in Programmed Instruction. A Symposium on Experimental Studies of Cue and Response Factors in Group and Individual Learning from Instructional Media*. Washington, D.C.: National Academy of Sciences, National Research Council.