The Effects of Varying Schedules of Reinforcement on Human Task Performance

ROBERT D. Pritchard, DALE W. LEONARD, CLARENCE W. VON BERGEN, JR., AND RAYMOND J. KIRK

Purdue University

The present research utilized four schedules of financial reinforcement (Hourly, Fixed Ratio, Variable Ratio, and Variable Ratio-Variable Amount) in an organizational simulation setting. Subjects were hired for what they perceived to be a real job of 4 weeks duration, which required them to learn self-paced material about electronics. Each subject worked for 1 week under each of the four schedules of reinforcement. Performance was lowest under the Hourly schedule, the FR and VR and VR-VA schedules produced higher performance. Attitudes were best under FR. The results were discussed in terms of their practical implications as well as their implications for theory, particularly expectancy-valence theory.

Since the statement of Thorndike's (1911) Law of Effect, psychologists have been arguing that increases in performance can be obtained if valued rewards are made contingent on that performance. This point of view is most clearly articulated in the industrial/organizational area by the current emphasis on expectancy-valence models of motivation (e.g., Vroom, 1964; Georgopoulos, Mahoney, & Jones, 1957; Porter & Lawler, 1968; Lawler, 1971; Mitchell & Biglan, 1971; Dachler & Mobley, 1973; Campbell & Pritchard, 1975). Although these expectancy-valence models stress the establishment of performance-reward contingencies (instrumentalities), they clearly imply that performance will be maximal the closer the performance-reward contingency approaches a perfect relationship.

This point of view is in direct contrast to the experimental psychology literature which suggests that performance will be higher under conditions where the performance-reward relationship is not perfect, i.e., partial reinforcement. This operant conditioning literature (e.g., Skinner, 1938;
Reese, 1966; Reynolds, 1968; Ferster & Skinner, 1957) deals with a variety of schedules of reinforcement, the most common of which are Fixed Interval (FI), Variable Interval (VI), Fixed Ratio (FR), and Variable Ratio (VR).

Another type of schedule which is not so well researched in the animal literature is Variable Ratio—Variable Amount. This is a schedule which not only varies the number of responses needed to obtain the amount of reinforcement (VR) but also varies the amount of reinforcement given (VA). The classical example of this type of schedule is a gambling situation, especially a slot machine. Other examples would include some managerial compensation systems, and financial returns for entrepreneurs. The expectation is that such a schedule would result in the highest performance of all. While most of the research on the effects of schedules has been done with animals, there has been considerable supportive research on the effects of FR schedules on humans in the form of industrial piece-rate payment systems (e.g., Mariott, 1957; Lawler, 1971), training (e.g., Schramm, Note 4), vigilance (e.g., Holland, 1958), educational settings (e.g., Lipe & Jung, 1971), and clinical settings (e.g., Krasner, 1971). However, much less attention has been given to the more complex schedules using human subjects.

The few studies that have explored the more complex schedules suggest that humans react to these schedules in a manner similar to animals (Bijou, 1957a and b; Bijou, 1958; Orlando and Bijou, 1960; Fattu, Mach, and Auble, 1955; Long, Hammack, May, & Campbell, 1958; Brackbill, 1958; Verplank, 1956).

As several authors have noted (Nord, 1969; Campbell, 1971; Jablonsky & DeVries, 1972), the operant conditioning literature has direct implications for industrial/organizational psychology, yet attempts to utilize the schedules of reinforcement are rare indeed. One exception is a study by Yukl, Wexley, and Seymore (1972) which examined the effectiveness of pay incentives under variable and continuous reinforcement schedules. This research was conducted under conditions more similar to an organizational environment. In a simulated job situation, subjects worked for 1 hr/day for a period of 2 weeks. Subjects were paid $1.50/hr for the first week. At the beginning of the second week subjects were randomly assigned to one of three incentive conditions: a 25¢ incentive with a continuous reinforcement schedule, a 25¢ incentive with a 50% variable ratio schedule, and a 50¢ incentive with a 50% variable ratio schedule. It was found that pay incentives were more effective in motivating increased production when employed with a variable ratio schedule than when used with a continuous reinforcement schedule, and that production gains were significantly higher for the group given the larger magnitude of reinforcement. The findings are consistent with the operant conditioning literature
in demonstrating the effects of size and schedule of reinforcements (Bandura, 1969; Cohen, 1969).

Taken as a whole, the literature suggests that the less conventional schedules may have even greater effects on performance than the traditional FR, piece-rate type schedule. Probably the most surprising thing is that these schedules have received so little attention.

The present study is an attempt to deal with some of these schedules with complex human task behavior, simulating an organizational setting. Four of the schedules are explored. The first is an hourly pay system which closely approximates the FI schedule. Also explored are FR, VR, and VRVA. The FI, FR, and VR schedules were selected because of their frequency of use in industrial settings and/or operant conditioning research. The VR-VA schedule was selected because even though there is little literature on this schedule, one would expect it to be quite powerful, and it characterizes the reinforcement schedules in a number of work settings.

METHOD

Overview

The basic procedure of this study was to hire subjects for what they felt was a real job of 4 weeks duration. The job dealt with learning self-paced material on electricity, electronics, and transistors. The first group of subjects worked for 4 weeks on a straight hourly pay system, and their data were used to refine the task material. The second group of subjects also worked for 4 weeks, 1 week each under: (a) hourly, (b) fixed ratio, (c) variable ratio, and (d) variable ratio-variable amount payment systems. Performance and attitude data were collected throughout the duration of the experiment.

Task

The basic task material consisted of three programmed texts: Basic Electricity, Basic Electronics, and Basic Transistors (New York Institute of Technology, 1963, 1964, and 1964, respectively). These books had been developed for technical training and assumed no electrical or electronics knowledge. The three basic books formed a series, and were geared to a population with at least some high school.

Due to the nature of the design, it was necessary to divide the three volumes of task material into work units. Each work unit was to represent 12 hr of work for the average subject. Since the pay of subjects on the three pay schedules other than hourly was to be determined by how many work units they completed, these units had to be constructed so that they would take nearly equal time to complete.

The first step in making these divisions was to divide the task material into 1/2-hr units by estimating the time to complete each section. The
second step was to actually have a group of subjects go through the task material under conditions similar to those to be used in the actual experiment, and use the data produced by these subjects to divide the task material into the 1/2-hr units.

To this end, eight subjects were hired and worked on the task material for 4 weeks (5 hr/day, 5 days/week). They were paid $2/hr and worked in the same room and used the same apparatus as did the subjects in the actual experiment. Also, they were of the same approximate ability level as the subjects in the experiment. Data were collected on the time taken to complete each unit of the task. The mean time per unit was calculated, and these data were used to revise the divisions of the task material for the actual experiment so that the average subject would complete each task unit in 30 min.

In addition to the actual study material, a short (three to eight item) multiple choice test was prepared for each unit. The items for these tests were taken from the published texts and, where needed, additional items were written by graduate students in electrical engineering. Thus, a subject studied the given unit of task material and, when finished, took the test associated with that material. Once he passed the test, he would proceed with the next unit of task material.

Subjects

The first step in the selection of subjects was to place advertisements in the local newspaper and to post flyers in various markets, shopping centers, and around high schools. These ads announced a 4-week, part-time summer job for males between 17 and 19 years of age which would pay “around $2 per hour, depending on what you do.” The ad gave a telephone number to call for more information. Applicants who called were not given detailed information about the job, but they were told it “involved going over reading and study materials and required no special skills.” It was felt that actually explaining that the job dealt with electricity and electronics materials might discourage some subjects who felt they had little background for such a job.

Callers who were interested were told to go to the company’s office (in downtown Lafayette, Indiana) to complete an application. Candidates were scheduled so that between two and ten applied at one time.

Applicants in these sessions were told that the company did contract work and they now had a contract to evaluate the effectiveness of certain types of programmed instruction. They were told that they would be going through task material in electricity, electronics, and transistors. They were also told that no special background was required to do the job. Finally, they were told that the pay “depended on what they did,” but
should average $2/hr. Since some of the applicants would be working entirely under the $2/hr condition, it was felt that a description of the other payment systems should not be made at this time.

After answering any questions, they completed an application blank, the Otis–Lennon Mental Ability Test, Advanced Level Form J (Otis & Lennon, 1967), an arithmetic test, and an electricity–electronics test. The arithmetic test was developed for this project, and consisted of arithmetic operations necessary for completing the task material. It covered such areas as multiplication and division of whole numbers, fractions, and decimals; scientific notation; and solving simple equations. It was felt that subjects who had no arithmetic ability should be rejected, since they would be unable to get through the task material. In fact only 2 or 3 applicants had to be rejected for this reason.

An electricity–electronics test was also developed for this research and consisted of items that most people would not know unless they had considerable knowledge of electricity and electronics. Sample items were “Describe Coulomb’s law of electric charges”; “Define diode . . . transducer . . . rectifier.” It was felt that a clear test of the effects of the schedules on performance could only be made if all subjects had essentially no knowledge of electricity and electronics. Consequently, only subjects who scored very low on this test were selected.

After completing these instruments, subjects were thanked and told they would be contacted. In all, 57 people completed the application procedures.

After scoring the various tests, the final selection of subjects was made. The characteristics of the 24 subjects who were ultimately selected (8 for the initial four-week pilot and 16 for the actual experiment) are described in Table 1.

| TABLE 1 |
|---|---|---|---|---|
| | 8 Pilot subjects | | 16 Experimental subjects | |
| | $\bar{X}$ | $SD$ | Range | $\bar{X}$ | $SD$ | Range |
| IQ | 111.30 | 11.00 | 96–128 | 117.60 | 13.30 | 98–146 |
| Age in months | 208.00 | 8.35 | 196–225 | 213.10 | 9.53 | 280–236 |
| Years of education | 11.40 | 0.92 | 10–13 | 11.40 | 0.56 | 11–13 |
| Arithmetic$^a$ | 7.50 | 1.85 | 5–11 | 4.63 | 3.31 | 0–10 |
| Electricity– Electronics$^b$ | 1.00 | 2.45 | 0–7 | 1.25 | 3.05 | 0–12 |

$^a$ Score is number of errors.
$^b$ Score is number correct.
Operation of Schedules

The testing procedure required each subject to read and study a specified segment of task material which, based on the pilot data, should take, on the average, 1/2 hr to complete. When the subject felt he knew the material, he took a test over that material, entering his answers in a computer-controlled console. The test would be scored, and the subject would be informed, through his console, of the number of items correct, whether he passed the test, and how much reinforcement (pay) he was to receive, if any.

The entire procedure of entering the answers and receiving feedback on the test took less than 30 sec. If no reinforcement was given, the test-taking procedure was ended. If reinforcement was to be given, the counter on the computer controlled console immediately started to click, accumulating to the actual amount earned.

It was pointed out to the subjects that each of the four tables (with two consoles on each) worked on a different payoff system. They were told that the first table (hourly) paid a straight $2.00/hr. The second table (FR) paid off $3.00 for every three tests passed. The third table (VR) also paid $3.00 when it did pay off, but the computer determined when it actually did pay off; and it could vary from paying off three times in a row ($9.00 for three tests) or it might go six or seven tests without paying off. The fourth table (VR - VA) also paid off intermittently, but the amount of payoff could vary from $.50 to $10.00. It was explained that if they worked at an "average pace," all the schedules would result in earnings of about $2.00/hr.

Data input, test scoring, delivery of reinforcement, and data output were all controlled on-line by a small computer interfaced to each of eight response-input consoles. This procedure is described in detail in Pritchard, Leonard, VonBergen, and Kirk (Note 3). The basic mode of operation was for a subject to approach the instructor when he felt he had mastered one unit of task material and was given the test for that unit. Once he had answered the test by hand he entered his answers in the computer-controlled console. The test was scored by the computer and the number of correct answers was indicated on the console, as well as whether he had passed the pre-established criterion (75% correct) for passing that test. The subject knows the criterion in advance. If he failed the test he returned it to the instructor and was given the same task material for further study. He could not retake a test over that material for at least 5 min. Once he felt he was ready, he once again approached the instructor and was given an alternate form of the test over that material.

Once the subject had passed the test, if any reinforcement was to be given, a counter on the console began to operate. Each pulse of the counter equalled $.10, so if a $3.00 reinforcement was to be given, the
counter pulsed until 30 units showed on the counter.

The actual reinforcement for each schedule was predetermined in the computer program. For the hourly schedule, the internal clock of the computer merely kept track as to whether or not an hour had elapsed and then delivered a sequence of 20 pulses to the two hourly counters, corresponding to the 20 dimes needed for a $2/hr pay rate. In the Fixed Ratio schedule the computer kept track and reinforced each third set of test data for a given subject which met the pass-fail criterion. The computer then pulsed the reinforcement counter of that particular response console 30 times corresponding to a $3 reinforcement. (Recall that each test covered 12 hr of material. Thus, three tests were 90 min, at $2.00/hr this results in $3.00). The Variable Ratio schedule was somewhat similar to the FR schedule except that the $3 payment was made after a variable number of passed tests. Whenever a test which passed criterion was inputed on one of the VR boxes, the computer referred to an internal memory table, 99 locations in length. Thirty-three of these locations had been selected randomly to contain a reinforcement code. The computer merely stepped along this table once per each passed test and delivered the $3 reinforcement whenever one of the 33 pay locations was encountered. Thus, on the average, every third test was reinforced, analogous to the FR schedule.

Table 2 describes the frequency distribution of reinforced tests. With the schedule employed, as indicated in the table, the longest series of nonreinforced tests encountered by a given subject was seven, with occasional reinforcements occurring consecutively. When the computer reached the end of the 99 location table, it recycled again to the first table entry and cycled through the table again. With a total of 99 locations and

<table>
<thead>
<tr>
<th>Number of intervening nonreinforcements</th>
<th>Frequency VR</th>
<th>Frequency VR–VA</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

1 Zero intervening nonreinforcements indicates that two reinforcements occurred in succession. These include one instance in each schedule in which a sequence of three consecutive reinforcements occurred.
no marker to identify the beginning of the series, it was virtually impossible for any subject to anticipate the number of passed tests required for reinforcement.

A second randomly generated table, similar to that used for the VR schedule, was employed for the Variable Ratio–Variable Amount schedule. Again, the reinforcement locations were randomly selected, with each location containing the amount of reinforcement to be delivered, if any. The computer stepped along a 99 location table and delivered reinforcement if it encountered a nonzero amount in a particular location following a passed test. Table 2 also presents the reinforcement sequences used for this schedule. The amounts of reinforcement ranged from $0.50 to $10.00 with an average $3.00. There were four reinforcements each of $0.50, $1.00, $1.50, $2.00, $2.50, $3.00, and $3.50; and two of $5.00, $7.00, and $10.00, for a total of $100.00 delivered per 100 tests passed.

Subjects were actually paid at the end of each week. Thus, all the money they had earned on a given schedule was paid at the end of that schedule.¹

Thus, although the schedules varied considerably, if the performance of all subjects was equal to the performance in the pilot, all subjects would earn $2.00/hr. As performance increased in the FR, VA, and VR–VA schedules, pay would correspondingly increase.

Procedures and Experimental Design

As described previously, two experimental sessions were run—the first to refine the task materials, and the second to actually conduct the experiment. The orientation for the subjects in the pilot session was identical to that for the actual experiment described below, except that they were told that the pay would be $2.00/hr for the entire time they would work.

The 16 individuals selected for the actual experiment reported for an orientation the Saturday before their first work week. Eight subjects reported for the morning shift and eight for the afternoon shift. In these orientation sessions the apparatus was explained and the test-taking procedures were demonstrated and clarified.

The experimental room contained four large (3 × 8 ft) tables with two

¹ One might argue that from a theoretical point of view, subjects should be actually paid at the time the counter indicated a reinforcement. However, the concern here is to be able to generalize to an actual job. Immediate payment was considered impractical in an actual job and thus the present weekly system was employed. Furthermore, one could argue that if a subject knew he had earned $3.00, this was the equivalent of actually receiving the money. Whether actually receiving the money at the time of “reinforcement” would make a difference in the results is an empirical question.
consoles on each table and a chair in front of each console. It was explained that the consoles were the mechanism by which they took tests, found out how well they did on the tests, and which determined their pay. These consoles were hooked up to a computer in another room and the wires and actual computer were shown to all subjects. This "computer room" also included a teletype, numerous rolls of paper tape, tools, and miscellaneous gear that added realism to the setting. In fact, all the equipment was indeed being used, none of it was "planted."

It was stressed that they would work the first week on one of the tables (under one of the schedules) but that they would shift to a new table on the Monday of each week. Thus, each person knew from the start that he would work for 1 week on each schedule.

The subjects were told that the researchers were also interested in the reactions of the employees during the four weeks. To obtain these reactions they would be given a short questionnaire on Monday and Friday of each week. Before leaving the Saturday orientation, subjects were administered a short battery of personality measures.

On the following Monday, subjects reported for their respective shifts (7:30 AM to 12:30 PM or 1:30 PM to 6:30 PM) and the apparatus, pay system, and procedures were briefly reviewed. They were given the first unit of task material and started working.

Test length varied from three to eight items, and in order to pass the test the subject had to reach a criterion of 75% correct. Once a test was taken, subjects were required to return to the supervisor and obtain the task material appropriate to the next test. In the event of test failure, the subject was given the same task material and was required to restudy the material. When he was ready, an alternate test over that material was taken.

The behavior of the supervisor encompassed two major activities: distribution of the test and task material and maintenance of order. Thus, rather than performing instructional activities the supervisor's behavior could more appropriately be characterized as that of a monitor or proctor. The supervisor advised the subjects that there were no scheduled breaks but that breaks were self-determined and could be taken by simply leaving the work area so as not to disturb their fellow employees who were still taking tests. An unlimited supply of coffee and soft drinks were provided for the subjects on a self-service basis.

Each Friday subjects were administered a comprehensive test over the material studied during that week. Because different subjects progressed at different rates, the length of the comprehensive test varied from subject to subject.

The experiment was a Latin Square repeated measures design with four treatments. The four treatments were the four schedules: hourly, FR, VR,
and VR-VA. Each subject worked 1 week under each pay schedule, and the order was determined by a randomly selected balanced Latin Square.

Dependent Variables

Performance dependent variables were those related to test-taking behavior. They included number of tests taken, number of tests passed, percent correct on all tests taken, percent correct on tests passed, time between each test taken, time between each passed test, and earnings. Scores on the comprehensive tests taken at the end of each week comprised the final performance-dependent variable.

In addition to those performance-related dependent variables, attitude and personality data were also collected. One group of measures was given at the orientation the same day the subjects learned to use the apparatus. These included eight items from the Locus of Control Scale (Rotter, 1966), seven items from the Protestant Ethnic Scale (Mirels & Garrett, 1971), one item assessing the degree to which subjects felt they should exert a great deal of effort on this type of job, and two items dealing with the subjects’ need for money.

A second group of measures was given at the end of the work day on Monday and Friday of each of the 4 work weeks. The first instrument was a measure of job satisfaction. This consisted of the 16 most relevant items from the short form of the Minnesota Satisfaction Questionnaire (Weiss, Dawis, England, & Lofquist, Note 5). Subjects also rated the degree to which the pay system made the job more interesting, the extent to which the pay system made the subject want to work harder, the degree of control they had over the job, and the extent to which the job was seen as manipulative. Self-ratings of effort were measured by two 9-point Likert Scales. The first item asked how hard they were working, with answers of “As hard as I possibly can,” through “About average,” to “I am taking it easy.” The second item asked them to rate the percent (10 to 90%) of their total possible effort that they were putting in on the job. Perceptions of equity were also measured by two 9-point Likert items. The stem of the first item was “Compared to what the other guys are getting paid for what they do”: Answers ranged from “I’m getting far less money that I should,” through “I’m getting paid about right,” to “I’m getting paid far more than I should.” The second item asked subjects to compare how much they do on the job with how much they should get out of it. Responses ranged from “Highly overpaid,” through “Paid about right,” to “Highly underpaid.” Finally, subjects were asked on Monday to estimate the amount of money they thought they would make, and to rate the degree of attractiveness of that amount of money. On Fridays they were asked to rate the attractiveness of the amount of money they had actually earned. On each Monday and Friday subjects were also asked a series of
questions about their expectancies and instrumentalities. These are discussed below.

In addition to these questionnaires, subjects were asked to sign up for a 1/2 hr time period on the Saturday after the last working day. They were to be paid for this time, and it was explained that they would get their last paycheck and go through an exit interview at this time. Subjects appeared in groups of two and were given a semistructured interview dealing with various aspects of the job.

RESULTS

The data analyses to be presented fall into four major categories: checks on the manipulations, performance results, attitude—personality results, and a description of the results of the exit interviews.

Checks on the Manipulations

Since the purpose of the various schedules of reinforcement was to make pay more or less contingent upon performance, it is important to ascertain whether the subjects actually perceived that the various schedules did indeed result in differential behavior—reward contingencies. In line with classical expectancy-valence models of motivation (e.g., Campbell & Pritchard, 1975; Dachler & Mobley, 1973; Mitchell & Biglan, 1971; Porter & Lawler, 1968) three sets of perceptions were measured. Specifically, measures were made of perceptions of effort-performance expectancy [e.g., “On this job the more effort I put in (the harder I work) the more _material_ I can get thru in a day”], performance—reward instrumentality (e.g., “The more _material_ I can get thru in a day the more _money_ I make”), and effort—reward expectancy [e.g., “The more _effort_ I put in (the harder I work) the more _money_ I make”]. These were taken on Monday and Friday of each week. No differences were predicted for effort—performance expectancy since the degree of relationship between effort and performance should not be effected by the actual reward system. However, performance—reward instrumentalities and effort—reward expectancies should be highest for the FR, VR, and VR—VA schedules since the greater the effort and/ or performance, the greater the reward. The hourly schedule should be lowest since effort and/or performance was not related to pay. Within the FR, VR, and VR—VA schedules, FR should show the highest instrumentality and expectancy since pay was clearly and consistently tied to behavior. The VR schedule should be lower since frequency of pay did not directly follow from performance, and FR—VA should be lowest of the three since both amount and frequency of pay did not directly follow from performance.

The data indicate that the effort—performance expectancies, as expected were not different from schedule to schedule on Mondays. However, by Friday, expectancies for VR and VR—VA were lower. The per-
formance—reward instrumentalities and effort—reward expectancies show exactly the pattern anticipated, and the overall $F$s are highly significant. Figure 1 presents these means graphically. In each case planned comparisons showed the hourly schedule to be significantly lower than any of the other schedules and the FR schedules higher than the VR—VA. Although in the predicted direction, VR was not significantly lower than FR.

These data clearly indicate that subjects did perceive that pay was contingent upon performance in the FR, VR, and VR—VA schedules, and that these contingencies were highest for FR and VR, followed by VR—VA. Thus, the manipulations were successful.

Performance Results

From the performance data collected, seven dependent variables were analyzed. These were: (1) the total number of tests passed on each schedule, (2) the total number of tests taken on each schedule, (3) the mean percent correct on passed tests, (4) the mean percent correct on all tests taken (whether passed or failed), (5) the mean inter-passed test time (IPTT), i.e., the mean time between passing two successive tests, (6) the mean time between taking two tests regardless of whether the subject passed the tests or not, and 7) the total earnings on each reinforcement schedule.

In addition to the seven performance variables above, two scores on the weekly criterion tests were analyzed. The two scores were the percent correct and the percent correct multiplied by the number of half-hour task units covered by the criterion test. Since the amount of material covered on a criterion test varied considerably, the latter adjusted test score was used to weight the criterion test performance by the amount of material covered by the test.

The performance data are summarized in Table 3. The single variable of greatest importance is number of tests passed. The means for each schedule are presented in Fig. 2. This figure clearly shows that the hourly schedule resulted in the fewest tests passed, FR and VR next and about

---

2 If a subject failed the test the first time, an equivalent form test was administered. A total of three tests had been developed for each task unit. Thus, a subject could not employ a strategy of asking for the test, looking at the items, and then studying the material to find the answers.

3 In one case, missing data were present due to the fact that one subject quit after the first week. A replacement subject was hired and worked the remaining 3 weeks. Thus, no data were available for the new subject for one condition (VR—VA). Consequently, his VR—VA performance data were estimated based on the relative differences between VR—VA and the other three conditions for the remaining 15 subjects. (See Pritchard et al., Note 3 for a more complete discussion.)
equal to each other, and VR–VA showed the most tests passed. The overall F is highly significant ($p = .0005$). Planned comparisons indicated that hourly was significantly lower than the other schedules ($p < .01$), but VR–VA was not significantly higher than FR or VR.

Number of tests taken (Table 3) shows essentially the same pattern of results as number of tests passed. As expected, interest time and inter-passed test time show the same pattern. That is, the hourly schedule produced significantly longer times than the other schedules ($p < .01$) but FR, VR, and VR–VA were not significantly different from one another. Actual earnings also show the same pattern.

Data on the percent correct for the task material tests (Table 3) indicate

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hourly</th>
<th>FR</th>
<th>VR</th>
<th>VR–VA</th>
<th>$p$</th>
<th>MS error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of tests passed</td>
<td>41.8</td>
<td>59.3</td>
<td>59.2</td>
<td>64.8</td>
<td>.0005</td>
<td>209.74</td>
</tr>
<tr>
<td>Number of tests taken</td>
<td>81.6</td>
<td>106.6</td>
<td>112.4</td>
<td>116.1</td>
<td>.0002</td>
<td>436.5</td>
</tr>
<tr>
<td>Percent correct on tests passed</td>
<td>86.1</td>
<td>85.6</td>
<td>85.4</td>
<td>86.3</td>
<td>.5560</td>
<td>4.14</td>
</tr>
<tr>
<td>Percent correct on all tests</td>
<td>66.4</td>
<td>66.7</td>
<td>67.0</td>
<td>67.9</td>
<td>.9000</td>
<td>37.54</td>
</tr>
<tr>
<td>Time between passed tests (hr)</td>
<td>.73</td>
<td>.45</td>
<td>.44</td>
<td>.42</td>
<td>.0002</td>
<td>.037</td>
</tr>
<tr>
<td>Time between all tests taken</td>
<td>.36</td>
<td>.24</td>
<td>.24</td>
<td>.23</td>
<td>.0013</td>
<td>.010</td>
</tr>
<tr>
<td>Earnings (dollars/week)</td>
<td>46.25</td>
<td>58.50</td>
<td>58.19</td>
<td>62.81</td>
<td>.0320</td>
<td>254.73</td>
</tr>
<tr>
<td>Proportion correct on</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>comprehensive</td>
<td>.734</td>
<td>.666</td>
<td>.602</td>
<td>.609</td>
<td>.0030</td>
<td>.011</td>
</tr>
<tr>
<td>Weighted comprehensive score</td>
<td>31.62</td>
<td>40.59</td>
<td>37.23</td>
<td>40.29</td>
<td>.05</td>
<td>102.45</td>
</tr>
</tbody>
</table>
that the increased performance under the FR, VR, and VR-VA schedules did not reduce subjects' scores on the tests. There were no significant differences ($F < 1.0$) in percent correct for either the tests taken or test passed.

Data on the weekly comprehensive tests are also presented in Table 3. The results indicate that there were differences ($p < .01$) in retention when simple percent correct was used as the dependent variable, with the hourly schedule showing the highest comprehension. However, when scores on the comprehensive tests were weighted by the amount of material covered, the hourly condition showed the lowest mean, and the other three schedules were about equal. One explanation for this pattern of findings on the comprehensive tests is that a quantity-quality tradeoff was operating such that the hourly subjects displayed higher quality with lower quantity while, for the other schedules, the reverse was true. This argument is weakened, however, by the percent correct on the task material tests. Since the subjects in the hourly schedules did not get better scores on these tests than did the subjects in the other schedules, they were apparently not doing better quality work. Another explanation is that since the subjects in the hourly schedule passed fewer tests, they had less material to remember, and thus did better on the comprehensives. However, one point is clear from the weighted comprehensive data. For a given unit of time (1 week) the subjects in FR, VR, and VR - VA did, in fact, learn more material than did the subjects in the hourly condition.

**Attitude Results**

As was discussed above, a job attitude questionnaire was given at the end of the work day on Monday and Friday of each week. The first measure in this questionnaire consisted of a job satisfaction questionnaire. Total satisfaction (the sum of the 16 items) is presented in Fig. 3.
The pattern of means shows that on Mondays the hourly and VR – VA resulted in the highest overall satisfaction, but by Friday, the FR was highest. The differences within each day were not significant, but the change from Monday to Friday, by schedule, was significant ($p < .04$). The satisfaction item dealing with pay was of particular interest. Significant differences in pay satisfaction occurred on Monday, with FR resulting in the highest satisfaction (4.4), followed by VR (4.0), VR – VA (4.9), and hourly (3.6). Post hoc comparisons (Newman–Keuls) indicated that the hourly mean was significantly lower than the means for the other three schedules, but FR, VR, and VR – VA did not differ from each other. The pattern is the same on Friday, but the overall $F$ is not significant. The hourly schedule resulted in the lowest pay satisfaction on both Monday and Friday.

The next attitude measure to be discussed deals with subjects self-ratings of effort. Two items assessed effort and these items were summed. Means of this composite are presented in Table 4. Post hoc comparison indicated that perceived effort was significantly less under the hourly schedule than under any of the other schedules. There was a trend on Monday for subjects to feel greatest amount of effort under the FR schedule; while, by Friday, VR – VA was seen as resulting in the highest level of effort.

Subjects’ perception of the quality of their pay was also assessed. Means for the sum of the two questionnaire items are also presented in Table 4. The results indicate that subjects perceived the various schedules as being equally equitable.

Subjects felt the different schedules resulted in differential job interest (Table 4). They felt the hourly schedule resulted in the least job interest by

![Fig. 3. Total satisfaction by schedules and time.](image)
TABLE 4
Job Attitude Items by Schedule and Time

<table>
<thead>
<tr>
<th>Response</th>
<th>Schedule</th>
<th>Day</th>
<th>Hourly</th>
<th>FR</th>
<th>VR</th>
<th>VR–VA</th>
<th>P</th>
<th>MSW</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-9</td>
<td>Effect of payment method on job interest</td>
<td>M</td>
<td>5.1</td>
<td>6.9</td>
<td>7.1</td>
<td>7.3</td>
<td>.002</td>
<td>2.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>4.3</td>
<td>7.0</td>
<td>6.5</td>
<td>6.7</td>
<td>.0002</td>
<td>2.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F-M</td>
<td>-.81</td>
<td>.13</td>
<td>-.63</td>
<td>-.63</td>
<td>.18</td>
<td>1.63</td>
</tr>
<tr>
<td>1-9</td>
<td>Effect of payment method on effort</td>
<td>M</td>
<td>4.7</td>
<td>6.8</td>
<td>6.9</td>
<td>7.0</td>
<td>.002</td>
<td>3.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>3.5</td>
<td>7.2</td>
<td>6.9</td>
<td>6.3</td>
<td>.0000</td>
<td>3.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F-M</td>
<td>-.19</td>
<td>.38</td>
<td>.06</td>
<td>-.75</td>
<td>.09</td>
<td>3.58</td>
</tr>
<tr>
<td>2-18</td>
<td>Self-perceived effort</td>
<td>M</td>
<td>12.3</td>
<td>15.2</td>
<td>14.3</td>
<td>14.7</td>
<td>.0006</td>
<td>3.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>11.3</td>
<td>14.2</td>
<td>14.1</td>
<td>14.8</td>
<td>.0006</td>
<td>5.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F-M</td>
<td>-.10</td>
<td>.10</td>
<td>-.19</td>
<td>.06</td>
<td>.27</td>
<td>3.59</td>
</tr>
<tr>
<td>2-18</td>
<td>Perception of equity</td>
<td>M</td>
<td>9.6</td>
<td>10.2</td>
<td>11.0</td>
<td>9.6</td>
<td>.10</td>
<td>3.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>9.7</td>
<td>10.3</td>
<td>9.8</td>
<td>9.1</td>
<td>.51</td>
<td>4.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F-M</td>
<td>-.13</td>
<td>.06</td>
<td>-.13</td>
<td>-.56</td>
<td>.14</td>
<td>3.47</td>
</tr>
<tr>
<td>2-18</td>
<td>Personal control</td>
<td>M</td>
<td>12.1</td>
<td>13.4</td>
<td>13.6</td>
<td>12.9</td>
<td>.14</td>
<td>3.77</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>13.4</td>
<td>13.8</td>
<td>13.5</td>
<td>13.3</td>
<td>.61</td>
<td>1.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F-M</td>
<td>-.15</td>
<td>.04</td>
<td>-.06</td>
<td>-.38</td>
<td>.36</td>
<td>4.84</td>
</tr>
<tr>
<td>1-9</td>
<td>Pay expectation</td>
<td>M</td>
<td>50.75</td>
<td>67.88</td>
<td>62.60</td>
<td>62.63</td>
<td>.03</td>
<td>248.6</td>
</tr>
<tr>
<td>1-9</td>
<td>Valence of pay$^a$</td>
<td>M</td>
<td>5.3</td>
<td>6.6</td>
<td>6.2</td>
<td>5.7</td>
<td>.07</td>
<td>2.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>5.4</td>
<td>6.1</td>
<td>5.9</td>
<td>5.9</td>
<td>.72</td>
<td>2.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F-M</td>
<td>-.19</td>
<td>-.50</td>
<td>-.25</td>
<td>.19</td>
<td>.52</td>
<td>2.39</td>
</tr>
<tr>
<td>1-9</td>
<td>Feelings of manipulation</td>
<td>M</td>
<td>5.7</td>
<td>5.8</td>
<td>5.1</td>
<td>5.4</td>
<td>.60</td>
<td>2.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>5.6</td>
<td>6.0</td>
<td>5.9</td>
<td>5.4</td>
<td>.19</td>
<td>.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F-M</td>
<td>-.13</td>
<td>-.13</td>
<td>.44</td>
<td>-.50</td>
<td>.26</td>
<td>1.76</td>
</tr>
</tbody>
</table>

$^a$ The rating on Monday was valence of expected pay throughout the week, the rating on Friday was valence of actual pay earned.

far, and significantly so on Friday.

The perceived effect of schedule on effort (Table 4) followed a similar pattern. Hourly was significantly lower than the other schedules on both Monday and Friday.

The two items dealing with feelings of control over the work setting showed no schedule effects.

Pay expectations (Monday) are also reported in Table 4. The pay expectations in the hourly condition were significantly lower than for the other schedules. The valence of pay data followed essentially the same pattern as the pay expectation data.

Finally, subjects did not feel that any of the pay schedules manipulated them (Table 4) any more than any other schedule.

Analyses dealing with the biographical and personality data are presented in Table 5. This table presents correlations between the various
biographical and personality variables and performance as measured by the number of tests passed under each schedule. None of the correlations with biographical data are significant. Locus of control and perceptions of expected effort did not show significant correlations with performance. The need for money scale unexpectedly showed negative correlation with performance, and for the VR schedule, significantly negative. Correlations with the Protestant Ethic Scale showed a significantly negative correlation in the hourly schedule, but small positive correlations under VR and VR-VA.

*Exit Interviews*

Subjects were asked to come to an interview at the end of the project, and 9 subjects showed up for these exit interviews. The interviews were scheduled for 1/2 hr time blocks, and two persons were interviewed at a time.

The most informative questioning centered around the preference of the various subjects for the various pay schedules. Each subject was asked to rank order the four schedules from highest (1) to lowest (4) in terms of preference. Subjects preferred the FR schedule by giving it an average ranking of 1.4. VR and VR-VA each averaged 2.3, hourly averaged 3.9. The single subject who gave FR a ranking of only 3 did so because he expressed a great deal of enthusiasm for gambling and, thus, rated the two

---

**Table 5**

Correlations of Biographical Data and Personal Reactions Questionnaire with Number of Tests Passed, by Schedule

<table>
<thead>
<tr>
<th>Biographical data</th>
<th>Hourly</th>
<th>FR</th>
<th>VR</th>
<th>VR-VA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of education</td>
<td>.44</td>
<td>.21</td>
<td>-.30</td>
<td>-.34</td>
</tr>
<tr>
<td>Arithmetic score&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-.05</td>
<td>-.31</td>
<td>-.29</td>
<td>-.45</td>
</tr>
<tr>
<td>Electricity score&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-.13</td>
<td>.36</td>
<td>.20</td>
<td>.04</td>
</tr>
<tr>
<td>Intelligence</td>
<td>.25</td>
<td>.40</td>
<td>.19</td>
<td>.37</td>
</tr>
<tr>
<td>Age</td>
<td>.14</td>
<td>.05</td>
<td>-.26</td>
<td>-.38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Personality data</th>
<th>Hourly</th>
<th>FR</th>
<th>VR</th>
<th>VR-VA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal locus of control&lt;sup&gt;c&lt;/sup&gt;</td>
<td>.00</td>
<td>-.08</td>
<td>.29</td>
<td>-.21</td>
</tr>
<tr>
<td>Need for money</td>
<td>-.21</td>
<td>-.15</td>
<td>-.57*</td>
<td>-.32</td>
</tr>
<tr>
<td>Perceptions of expected effort</td>
<td>.13</td>
<td>-.00</td>
<td>-.28</td>
<td>.23</td>
</tr>
<tr>
<td>Protestant ethic</td>
<td>-.61**</td>
<td>-.20</td>
<td>.24</td>
<td>.18</td>
</tr>
</tbody>
</table>

<sup>a</sup> Arithmetic score = number of errors.

<sup>b</sup> Electricity score = number correct.

<sup>c</sup> High score = high internal locus of control.

<sup>*</sup> $r_{.05} = .48$.

<sup>**</sup> $r_{.01} = .61$. 
“gambling” schedules (VR—VA and VR) first and second. In his opinion, however, the FR schedule ranked much closer to his first two choices than did the hourly schedule, which ranked a distant last. The other two instances in which FR ranked other than first seemed to be related to the fact that these subjects happened to make inordinately large amounts of money on the schedule preferred first.

When asked what they liked about the FR schedule, a typical comment was that they felt they were in control. For example, one subject expressed his feelings this way: “I felt that with this schedule (FR) I was in control of what I did. If I came in in the morning with a headache or feeling tired, I knew I could go ahead and make my $10 (his daily minimum goal) and then see what happened after that. On the gambling schedule (VR—VA) you might work all morning and make nothing.”

The two “gambling” schedules, VR—VA and VR, ranked equal in preference and below FR, even though many subjects earned the most money under these schedules. Except for those one or two subjects who liked to gamble, many subjects expressed a fear of having a long “dry spell” if forced to work on one of these schedules for an extended time period, such as several months.

The hourly schedule was by far the least preferred of the four schedules. However, some subjects felt that this schedule seemed worse since they could compare it so readily with the other schedules. In fact, performance under the hourly schedule was lower during the last 2 weeks of the study than under the first 2 weeks. It appeared that over time, subjects clearly realized that they did not have to work as hard under hourly. This “contrast effect” would suggest that it might prove disadvantageous to have both an hourly type schedule and an incentive schedule operating in the same work situation.

Evaluations of the hourly schedule usually focused upon the low motivation it inspired; e.g., “It made you so you didn’t want to work,” or, “I knew I was going to get paid whether or not I did anything, so I just felt like goofing off.” Subjects admitted that they learned 10 longest breaks and covered the least material on this schedule. A few subjects felt that they may have “learned more” when working under the hourly schedule, but further questioning indicated that this impression was related to the fact that the subjects covered far less material during the hourly schedule and were, therefore, able to achieve higher scores on the weekly comprehension tests. (When corrections for amount of material covered were made, the hourly schedule was actually the poorest.)

**DISCUSSION**

Taken as a whole, the data clearly indicate that the various schedules of reinforcement had differential effects on performance. Of the four
schedules, the hourly schedule resulted in much lower performance than the other three schedules. In fact, the mean performance (number of tests passed) of the other three schedules was 46% higher than performance under the hourly schedule. The FR and VR schedules were approximately equal in their effect upon performance, with the VR-VA schedule producing higher, but not significantly higher level of performance.

Equally important are the findings regarding quality of performance; i.e., percent correct on tests. The data indicate that even though the FR, VR, and VR-VA schedules resulted in more tests being passed than in the hourly schedule, the percentage of correct answers on the tests was just as high. In addition, the data from the weekly comprehensive tests also show that more learning was taking place under the FR, VR, and VR-VA schedules. That is, although the raw scores show that the hourly condition had the highest raw percent correct on the comprehensives, when the amount of material covered is taken into consideration, and three partial schedules were superior to the hourly condition.

The attitude data also show some interesting results. There was a trend for satisfaction to be highest for VR-VA on Monday, but highest for FR by Friday. Also, satisfaction was never highest under the hourly condition and, by the end of the week, satisfaction was lowest under hourly. Satisfaction with pay did show significant differences on Monday, with the hourly schedule resulting in much lower pay satisfaction than the other schedules. A similar, but nonsignificant, pattern emerged on Fridays. Also, the hourly schedule was lowest in job interest, and significantly so by the end of the week.

One finding of considerable significance is the change in attitude from Monday to Friday. A fairly consistent pattern was for the VR-VA schedule to be preferred at the beginning of the week but for FR to be preferred by the end of the week. These findings suggest that subjects reacted to the schedules differently after experience. Clearly, the evaluation of such schedules must occur only after subjects have experience working under the schedule.

One problem anticipated with the use of partial schedule of reinforce-
ment, especially VR and VR-VA, was that feelings of inequity or mani-
pulation might occur on the part of subjects under these schedules. The
data indicate that no difference in feelings of equity or manipulation oc-
curred across the various schedules. Furthermore, responses to these
tems were near the neutral point for all schedules.

Results from the biographical and personality analyses indicated some 
significant correlations with performance. Unfortunately, no systematic 
patterns emerged from these analyses and due to the small sample size, 
examination of these correlations through subgroup analyses is impos-
ible. However, one possible implication emerges from these analyses. The
magnitude and direction of the correlations for different schedules was quite variable. If future research shows these differences to be reliable, it would suggest that interactions exist between individual differences and the effects of schedules. Specifically, people with certain characteristics would perform well under one schedule, but not so well under another. Future research should carefully explore interactions with individual differences.

Our results have certain implications related to theoretical considerations involving expectancy–valence theory. Expectancy–valence models suggest that performance will be highest when effort–reward expectancies are highest. In this study, the FR schedule should result in higher effort–reward expectancies than the VR–VA schedule, since a given amount of effort is more closely followed by a given amount of reward (pay) in the FR schedule as compared to the VR–VA. In fact, the checks on the manipulations and the exit interviews both support the fact that subjects perceived this to be the case. However, contrary to the prediction of expectancy–valence theory, performance was actually higher in the VR–VA condition than in the FR condition.

One could argue that since the performance differences were not significant, there is really no issue. However, the actual difference was quite large (9%), and, more importantly, the effort–reward expectancy data would predict the opposite performance pattern. That is, effort–reward expectancies (and performance–reward instrumentalities) were lower in VR–VA than in FR, yet, actual performance showed the reverse pattern.

One way expectancy theory could explain such findings is through the concept of valence. It is possible that subjects on a “gambling” schedule such as VR–VA simply expect that they will be lucky and make more money than average, and this greater amount of money increases the valence of the pay outcome, thereby increasing the force toward high effort. However, our data (Table 4) indicate that subjects under the VR–VA schedule actually expected to earn less money than under the FR schedule. Thus, this explanation is not supported.

A second explanation is that the money actually earned is somehow more valent under VR–VA than under FR because of, for example, the added pleasure of “winning” a large payoff of low probability. This would imply that the valence of money earned under the VR–VA would be higher than the valence under FR. Once again, our data do not support this line of reasoning. The valence of money earned under the two conditions (Table 4) was equal, even though subjects actually earned more money under the VR–VA.

The only way that expectancy theory seems to be able to handle these results is to somehow postulate that additional outcomes are tied to performance under the VR–VA as compared to the FR. It could be argued,
for example, that outcomes such as the anticipated pleasure of hitting a large jackpot, irrespective of the actual money involved, could be operating. Such a point of view remains to be supported.

It is interesting to note that while the incentive motivation techniques employed in this study resulted in substantial gains in performance, a field experiment similar to the study reported here, conducted by some of the present researchers (Pritchard, DeLeo, & Von Bergen, Note 2) was not so effective. This field experiment was conducted at an Air Force technical training base and employed essentially a FR reinforcement schedule. Subjects in this field experiment were of the same age, roughly the same ability, the task material was very similar, and it was also a technical training setting. Three incentive systems were employed with each of two technical training courses. The results indicated that the incentive system was substantially effective in only one course, under one of the three incentive systems.

The results of this field experiment are in marked contrast to those found in the present research. This is extremely important from an applications point of view since it brings into question the generalizability of our findings to a field setting. To explain the conflicting findings, one might argue that the incentives used in the present research were more powerful than those used in the field experiment. This argument may have some validity for two of the incentive conditions used in the field experiment, but the third system employed substantial financial incentives (up to $40/week) as well as a variety of nonfinancial rewards. Yet, in this high incentive condition, on performance effects were observed for one of the courses, and only one of the two major dependent variables showed increased performance in the other course. Thus, an explanation based on the strength of the incentives should probably be ruled out.

However, there are two arguments which could explain the difference in the findings of the two studies. The first deals with the amount of reward offered by the incentive system relative to the total rewards available in the situation. In the present research, the majority of the extrinsic rewards available to the subjects were controlled by the incentive system (i.e., the schedules). In the field experiment this was not nearly so true. Within the technical school environment, the instructors had substantial reward power in the way they treated the students and the students tended to perceive, correctly or incorrectly, that the technical instructors had some control over their future careers. Outside of the technical school environment, the military instructors also had substantial reward power.

To the extent that these sources of reinforcement generated contingencies between rewards and behavior incompatible with high performance, the effects of the incentive system would be weakened. The implication of this line of reasoning is that in a field setting, the majority of the trainees'
rewards, both within and outside the technical school environment, should come from the incentive system. To the extent that meaningful rewards are controlled by sources outside the system, the power of the system will probably be weakened.

The second explanation for the disparity in the findings of the field experiment and the present research deals with effort-performance expectancies. In the field experiment, the students in the course showing no performance effects were near the upper limit of performance (mean test scores approached 90%) and thus perceived little relationship between increased effort and increased performance. In the present study this was not the case since increased effort would result in increased performance. The issue revolves around the nature of the tests used in the training situation. Tests of low difficulty could be passed by nearly everyone with low effort, and increased effort would not substantially change performance. Tests of very high difficulty could result in a similar situation where a student comes to realize that increased effort will not result in passing the tests any better or faster, but that he must simply "plough through" the material. If he increases his pace, he finds he will fail the test. This line of reasoning suggests that effort-performance expectancy would be highest, and thus incentive effects strongest, in a situation where tasks or tests are of moderate difficulty.

Thus, it seems likely that if a field situation were generated where most rewards were controlled by the system and tests were of moderate difficulty for the trainees involved, results similar to those obtained by the present research would be more likely to be found.

The results of the present research can also be compared to those found in the study discussed above by Yukl et al. (1972), and those in a more recent study by Berger, Cummings, and Heneman (Note 1). Both used a similar methodology, as well as the same schedules, 50c VR 2 (VR with 50% responses reinforced with 50c), a 25c VR 2, and a 25c continuous reinforcement (i.e., FR 1). They found a similar pattern of findings, i.e., 50c VR 2 > 25c VR 2 = 25c CRF. The critical finding is that the variable ratio schedule showed stronger performance effects than the continuous schedule, even though the earnings were equal with equal performance. In contrast, our results show no difference in performance between FR 3 (analogous to their continuous reinforcement) and VR 3. Clearly, there were many differences between these two studies and that presented here. The task was different, as were the subjects, the method of reinforcement, the number of hours worked, frequency of reinforcement, and the use of a base rate of pay. However, the simplest explanation deals with the fact that our FR 3 is not the same as their CRF schedule. If one predicts that a FR 3 would produce higher performance than a CRF (e.g., Morse, 1966), the apparent conflict disappears.
Another, less parsimonious explanation deals with the exact type of schedule used. In our study, the amount of reinforcement per reinforced trial was constant (i.e., $3.00), and the frequency of reinforcement was different. In both of their studies, both the frequency and amount of reinforcement varied. This would suggest that the performance effects obtained in their results were due to the amount of reinforcement per reinforced trial. That is, it was not the variable nature of the reinforcement which influenced performance, but the magnitude of the anticipated reinforcement.

Turning to the more applied side of our results, it is hard to argue with the point that in many situations the institution of a VR-VA reward system would be impractical. However, there are a number of situations where such a system would be feasible. Clearly, a training program, particularly a computer-assisted training program, could utilize such a system. Token economies for the mentally ill or educational settings could also apply such techniques. Finally, a VR-VA schedule could be used to supplement other financial compensation in the form of bonuses.

The exit interviews provided some valuable insights about how such systems might be designed. Subjects reported that one of the things they liked about the system was the impersonal nature of the performance feedback. They felt that having the computer score their tests and report to them the number correct removed the stigma of an instructor informing them that they failed. To the extent such a procedure removed some of the anxiety from failing, for example, in low ability subjects, this could have positive effects.

The nature of the feedback was also significant on other grounds. Subjects reported that the immediate nature of the feedback was a positive feature. More importantly, however, was the manner in which both the performance and reinforcement feedback was given. The system was intentionally designed so that this feedback information accumulated. That is, the light indicating the number of correct test items would start to flash, one flash per correct answer. Thus, the subject did not know how many he had answered correctly until the light stopped flashing. An analogous situation existed for the reinforcement counters. The counter would start to count, with an audible click, in units of 10¢. In the VR-VA schedule, subjects did not know how high the counter was going to go until it stopped. Observation of the subjects, as well as the exit interviews indicated that subjects attended to both the flashes for number correct and the reinforcement counters very carefully. In fact, when the counters of the subjects at the VR-VA table started counting, every subject in the room was carefully attending to how much money was being made. Thus, the cumulative nature of the performance and reward feedback generated a certain amount of suspense which, according to the subjects, seemed to
break the monotony. Probably the most important issue raised by the present research is whether a VR–VA schedule will produce better performance than a classical FR schedule. Our results show a large absolute difference, but the lack of a statistically significant difference argues against the reliability of that finding. It remains for future research to clarify the effects of a VR–VA schedule.

REFERENCES


Bijou, S. W. Patterns of reinforcement and resistance to extinction in young children. *Child Development*. 1957, 78, 47–54. (a)


REFERENCES


Received: March 20, 1975.