Behavior Research Methods & Instrumentation
1980, Vol. 12 (2), 256-258

SESSION XIII
MICROPROCESSOR USERS' GROUP

MISHA PAVEL, *New York University, Presider*

An "intelligent" program to teach logical thinking skills

LARRY E. WOOD
Brigham Young University, Provo, Utah 84601

An "intelligent" program was developed on a microcomputer to help students improve their logical reasoning skills. The program is based on the popular two-person game of "Mastermind" and provides feedback for less than optimal performance. In addition to playing the game with a student, the program checks to see if he or she is using all available information. If not, the program indicates the information that is being overlooked. High performance at successive levels of difficulty produces reliable improvement on equivalent forms of standardized tests of logical thinking skills.

This paper describes an interactive program written in BASIC on an 8080-based microcomputer to improve logical reasoning skills in college students. The basis of the program is the popular game of "Mastermind," which requires a considerable amount of logical analysis, both inductive and deductive. These logical skills are fundamental to most problem solving endeavors. Because it has been shown that games and simulations are valuable educational tools (Allen, Allen, & Ross, 1970; Fletcher, 1971), it seemed potentially productive to combine the educational value of games with their obvious motivational value to produce an interactive exercise to teach logical thinking skills. For other successful examples of computerized educational games, see Brown and Burton (1978) and Burton and Brown (1979).

The game of Mastermind was chosen because combinations of inductive and deductive reasoning skills are required to play the game effectively and efficiently. Also, the difficulty level can be varied through a rather broad range as one becomes more proficient at playing. In its popular form, Mastermind is a two-person game where one person is designated the "codemaker" and the other is the "codebreaker." The codemaker's task is to create a "code" consisting of a set of two to six pegs, each of which may be one of five different colors. These are kept hidden from the codebreaker, whose task is to "break" the code by "guessing" a set of pegs. The codemaker then provides feedback indicating whether any of the guessed pegs are the correct color, in the correct position, or both. The codebreaker then uses this information for the next guess. The object of the game is to minimize the number of trials required to break the code.

The program reported in this paper is a modification of the game of Mastermind; it not only plays the game with a person by acting as the codemaker in the usual way but also analyzes the guess made by the person and provides feedback when the guess isn't consistent with information already provided on previous trials.

The major purpose of the project was to see if the corrective feedback during the game would serve to improve a person's logical skills in the process of teaching him or her to play the modified game more effectively. The effectiveness of the program was evaluated using the Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1964), a standardized test of logical thinking skills. Alternate forms of the test were administered as pre- and posttest measures of logical thinking skills to a group of students who worked with the program and a group of students who did not.

METHOD

Hardware

The hardware configuration includes a Billings microcomputer, which consists of an 8080A microprocessor, 32K of dynamic RAM, a real-time clock, dual CalComp floppy disks, a Soroc 120 CRT terminal, and a Centronics 701 printer.

Software

The programs were written in Microsoft BASIC. Two programs were used in the experiment. INFER is simply a computerized version of the game of Mastermind that was used to familiarize the participants with the game. It generates a number two to eight digits long in which each of the digits can have the value 1-9. It then accepts guesses from the participant and provides feedback on the guesses in the form of a blank if there are no correct digits at all, a slash (/) for each digit correct but in the wrong position, and an asterisk (*) for each correct digit

  0005-7878/80/02056-03$00.55/0

in the correct position. The participant continues to input guesses based on the feedback provided by the computer until he or she finds the correct number. A sample is illustrated in Figure 1.

The second program used in training logical skills is a modification of the first and is called INFERX. In addition to generating numbers for the participants to guess, it also generates two to six additional numbers that are simulated trials (called statements by the program) and provides appropriate feedback. The participant is then instructed to utilize the information provided in the simulated guesses to finish the game with a minimum number of trials. Therefore, the program not only tests each subsequent guess against the original number but also compares each guess with each of the simulated guesses to determine if the participant has failed to consider some of the information that has been provided. Whenever the participant fails to consider some of the information, an error message is given containing the relevant statement number(s). Participants in the study were required to play three games in succession without making any such errors before they could move on to the next level of difficulty. For each game the program recorded the number of "correct" guesses, the number of "incorrect" guesses, and the time to play the game. This information was also displayed to the participants. A sample game is provided in Figure 2.

Procedure

The Watson-Glaser Critical Thinking Appraisal was administered to 133 volunteers from introductory psychology courses. On the basis of their performance, 14 pairs of participants were chosen whose scores were the same. One participant from each pair was randomly assigned to interact with the computer program; one was assigned to a control group and took the alternate form of the test. The participants who interacted with the computer were required to play two or three games of INFER to become familiar with the rules. Play began with two-digit numbers. Following this, the participants began interacting with the INFERX program, starting with two-digit numbers, where each digit could be 1-5, and two statements (simulated trials). As the criterion of three successive games without an error was reached, the experimenter increased the number of statements to six in increments of one. At that point, the length of the numbers was increased by one digit, and the number of statements was decreased to three. The difficulty of the games was increased in this manner until the participant reached the criterion with four-digit numbers and six statements. The time required to reach this level of proficiency varied from 3 to 6 h and was accomplished in sessions lasting approximately 1 h.

In addition to the feedback provided by the computer, the experimenter, who was very proficient at playing the game, provided general strategic suggestions to the participants whenever progress seemed to be unreasonably slow. The suggestions consisted of reminders about the meaning of the feedback, of the need to go back over all the statements each time, and of ways in which information from two or more statements could be integrated to make more effective guesses. In addition, the participants were encouraged to "think aloud" so that the experimenter could assess the appropriateness of their logical thinking.

Okay, I have a three-digit number in mind where each digit can be 1-9, you can begin guessing.

```
Guess 1? 123 Nothing
Guess 2? 456 *
Guess 3? 789/
Guess 4? 546/
Guess 5? 447 Nothing
Guess 6? 859**
Guess 7? 858 **
Guess 8? 855 You got it!!!!!
```

Figure 1. A sample game from the program INFER.

I will think of a three-digit number from the digits 1-9.

```
Statement 1      727 Nothing
Statement 2      885 *
Statement 3      387 *
Statement 4      263 /
Your Guess Is ?  345 /
Your Guess Is ?  785 * -- Contradicts Statement 1
Your Guess Is ?  684 //*
Your Guess Is ?  486 Good !!!!! You have the correct number
```

You took 115 sec for that game, and you made three correct guesses and one incorrect guess.

Figure 2. A sample game from the program INFERX.

RESULTS AND DISCUSSION

Watson-Glaser Scores

The difference scores on the pre- and posttest measures of the Watson-Glaser showed mean increases for the experimental and control groups of 3.1 and –.5, respectively. This change was statistically reliable for the experimental group [$t(13) = 3.2$, $p < .01$, one-tailed test], but not for the control group [$t(13) = -.33$, $p > .05$]. While this may appear to be a rather modest change for the experimental group, it does translate into a mean change of 10 percentile points based on the norms for liberal arts college freshman on the Watson-Glaser. The subjective reports of the participants were unanimously positive in their feelings of having improved their logical skills, and all were certain that they had performed much better on the posttest than they had on the pretest. This was not true, however, because the raw change scores ranged from –2 to 10. It was expected that the program would benefit mostly those who were the least proficient in logical thinking as indicated by pretest scores. There was a trend in this direction, as indicated by a negative correlation between pretest scores and gain scores, but it was not reliable ($r = -.21$, $p > .05$).

Errors

The program was designed to encourage participants to be more careful in considering all available information before responding. Having to play three errorless games in succession, in a sense, forced subjects to do so. In this regard, it is interesting to consider the changes in the patterns of errors that occurred between the pretest and the posttest for the Watson-Glaser. In general, the errors can be classified into two categories: those resulting from a failure to utilize all the available information to best advantage and those resulting from assuming more information than was actually given.

The participants were divided into two groups: the seven who improved the most and the seven who improved the least. For those who improved the most, there was approximately an equal reduction in both types of errors. For those who improved the least, there was an obvious change in the kinds of errors made, but only a slight change in the number of errors. On the posttest relative to the pretest, subjects made fewer errors in assuming more information than they had been given but more errors in failing to utilize all the information that had been provided. This result seems paradoxical in light of the fact that the program was designed to force participants to make the best use of the information given. Instead, it seemed to have the effect of simply causing subjects to be more conservative about making inductive inferences and resulted in their trading one type of error for another.

Game Performance

Those participants who gained the most from the program also were relatively more efficient at doing so, as indicated by a reliable negative correlation between the Watson-Glaser change scores and the number of games required to reach final criterion ($r = -.53$, $p < .05$, two-tailed test). There was no relationship between initial scores on the Watson-Glaser and the number of games required to reach final criterion ($r = .08$).

CONCLUSIONS

The initial results of the project appear promising. It seems reasonable to assume that logical skills are not particularly transient. Hence, the fact that there was a reliable improvement as a result of a few hours interaction with one particular program is basis for encouragement. Obviously, the program needs to be refined to meet the needs of those who failed to improve overall but became more conservative in making inferences.

Additional work is in progress to make the program more intelligent so that it will play the game from the point of view of the participant. The intent is to provide feedback from an integration of information across all previous statements rather than each individual statement. Of course, this represents a program of another magnitude of complexity, and it simply may not be feasible given the limitations of the speed and memory capacity of the microcomputer. The current version of INFERX takes about two-thirds of the available memory of the system. However, since the Microsoft BASIC has chaining capabilities, memory probably will not be the constraining factor. The limiting factor will most likely be the length of time it takes the program to evaluate the responses of the participant. While computer games have powerful motivational effects, their effectiveness is quickly lost if students are required to wait much longer than 10-20 sec for the computer's response.

REFERENCES

Allen, L. E., Allen, R. W., & Ross, J. The virtues of nonsimulation games. *Simulation and Games*, 1970, **1**, 319-326.

Brown, J. S., & Burton, R. R. Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 1978, **2**, 155-192.

Burton, R. R., & Brown, J. S. An investigation of computer coaching for informal learning activities. *International Journal of Man-Machine Studies*, 1979, **11**, 5-24.

Fletcher, J. L. The effectiveness of simulation games as learning environments. *Simulation and Games*, 1971, **2**, 425-454.

Watson, G., & Glaser, E. M. *Watson-Glaser Critical Thinking Appraisal.* New York: Harcourt, Brace & World, 1964.