The Thinking Aloud Technique and Problem Solving Ability

E. G. FLAHERTY
Nasson College

ABSTRACT

The effects of overt verbalization and practice on problem solving ability were examined. The 100 secondary school students who served as Ss were divided into four groups: (1) those who received practice word problems and solved problems while thinking aloud, (2) those who did not practice but solved problems while thinking aloud, (3) those who practiced but did not verbalize, and (4) those who received no practice and did not verbalize. Analysis of variance revealed that neither overt verbalization nor practice significantly influenced problem solving scores. However, Ss who were required to think aloud made significantly more computational errors than those who worked without verbalizing.

ALTHOUGH THE DEVELOPMENT of problem solving ability is a field which has evoked considerable interest among mathematics educators, productive research has failed to keep pace with the expressed concern (8, 13). The researchers who have tried to make some progress in the area of human problem solving have employed a number of different techniques in their work. Some have employed problems specifically designed to check for flexibility or rigidity in problem solving (5), while others have looked for evidence of an analytical or global reaction to a problem (14).

In many cases the solution to a problem reveals little about the way in which the result was reached. To gain information about an individual's problem solving processes, the researcher must find a method of getting the S to reveal the steps he is following so that an observable sequence of processes will be available for analysis. The technique of requiring an S to think aloud and then analyzing the resulting protocol has been growing in favor in recent years. However, some disagreement persists regarding the efficacy of the thinking aloud technique in giving an adequate account of a person's thought processes during problem solving. The question arises as to whether an S reacts differently to a problem when he is asked to verbalize as he works than he would if overt verbalization were not required (2, 3, 6). One of the purposes of the present study is to consider this question.

In their investigations Paige and Simon (9) used the thinking aloud technique to follow the methods Ss employed in attacking problems. These researchers reported that they gained insight into the S's problem solving ability by requiring the Ss to think aloud and then analyzing the resulting protocol. Since problems describing physically impossible situations had been found helpful by previous researchers in identifying problem solvers as verbal or physical, two problems of this type were included on the test.

In their investigations Paige and Simon (9) used the thinking aloud technique to follow the methods Ss employed in attacking problems. These researchers reported that they gained insight into the S's problem solving ability by requiring the Ss to think aloud and then analyzing the resulting protocol. Since problems describing physically impossible situations had been found helpful by previous researchers in identifying problem solvers as verbal or physical, two problems of this type were included on the test.

Problem

The study was designed: (1) to develop a coding system to analyze thought processes of students as they attempt to solve problems, (2) to compare the problem solving behavior of Ss who are required to think aloud with those who are not required to verbalize overtly; (3) to consider the effect of practice in thinking aloud in solving problems.

Method

Subjects. The Ss consisted of 100 senior high school students who had agreed to take part in the study. All Ss were just completing their second year of algebra at the time of the testing. The students were identified by their teachers as being of above-average ability on the basis of their general classroom performance.

Materials. The problem solving test consisted of 6 word problems of the type usually found in second year algebra books (a copy of the test is included in Flaherty) (4). For Ss in the practice groups two additional problems were presented prior to the actual administration of the test.

Since problems describing physically impossible situations had been found helpful by previous researchers in identifying problem solvers as verbal or physical, two problems of this type were included on the test.

Procedure. A 2 (practice versus no practice) x 2 (overt verbalization versus no verbalization) factorial design was used.
of four groups: (1) those who received practice word problems and solved problems while thinking aloud, (2) those who did not practice but solved problems while thinking aloud, (3) those who practiced but did not verbalize, (4) those who received no practice and did not verbalize.

Half of the Ss were required to think aloud while solving the problems; the verbal responses of these Ss were taped. The Ss in the overt verbalization group met with the E individually. The E explained to the S that the purpose of the experiment was to learn how people think when they solve problems. The S was asked to think aloud as he worked each problem. He was encouraged to vocalize all his thoughts concerning the problem even if they seemed trivial and unimportant. If he stayed silent for a minute or so, the E prodded him—asking him what he was thinking or doing—thus stimulating him to think aloud.

The Ss in the non-verbalization group were not required to meet with the investigator individually. The written solutions to the problems which the students submitted were analyzed to check on the procedures these Ss employed.

As was mentioned earlier, one of the purposes of this study was to prepare a coding procedure to analyze the processes which students use as they attack problems. On the basis of a pilot study conducted with six secondary school students, the E devised and revised a list of categories which were representative of a S's reactions as he solved the word problems. The final modified list of coding system variables is given in Table 1.

**Scoring procedure.** Each of the six test problems was worth two points. Success was determined on the basis of the processes which the S employed. Thus an S was not penalized for errors which were purely computational. An S received full credit for the correct process; arriving at the correct answer was not necessary. For example, if an S decided to set up an equation to solve a problem, wrote the correct equation, but made an error in solving the equation, he received the full two points. If an S planned successive approximations but made an arithmetical mistake in the course of carrying through his plan, he was not penalized. Of course, if an S was unable to say or write enough to indicate he was using a particular process or if he made a structural error which indicated a failure to understand the principle required to solve the problem, then he received no points for the problem.

In scoring the two problems containing contradictory information, the E used the following procedure: two points indicated that the S had noted both the grammatical and physical implications of the problem and thus had recognized that a contradiction existed; one point signified the S had substituted a physically possible problem for the given problem or the S had given a direct phrase-by-phrase grammatical translation into equation form.

**Results**

The maximum possible score on the problem solving test was 12 points. In this study the scores ranged from 2 to 12, with a mean of 8.18 and a median of 8. Analysis of variance showed that the requirement to think aloud did not significantly affect problem solving scores (F < 1.00). Practice also failed to reach statistical significance (F < 1.00). Similarly, Verbalization x Practice interaction was not significant (F < 1.00).

Scores on the problem solving test give some indication of the S's problem solving behavior. However, a more accurate description of a S's performance may be obtained by considering his problem solving approaches as measured by the coding system variables listed in Table 1. Categories 1, 2, 4, 5, 6, 12, and 13 could not be coded accurately for the Ss who did not think aloud and were, therefore, eliminated from consideration. Thus, to determine whether differences existed between verbalization and non-verbalization Ss with respect to the categories on which both could be coded, the means and standard deviations for both groups were found for each of ten categories. When t-tests were performed, it was determined that the only significant difference between the two groups was in the area of computational errors (t = 3.29, p < .01). Overt verbalization Ss made more errors of this type.

Only seventeen of the 100 Ss recognized the physical impossibilities posed by two of the test problems. In addition to these seventeen Ss, twenty-six Ss recognized the contradiction in one of the problems but failed to realize that a second of the problems also described a physically impossible situation. The Ss who responded to the physical cues in the contradictory problems were classified as physical problem solvers, while those who relied on the grammatical aspects of both problems were classified as verbal problem solvers.

**Table 1: Coding-System Variables**

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Misreads problem</td>
</tr>
<tr>
<td>2.</td>
<td>Rewords problem</td>
</tr>
<tr>
<td>3.</td>
<td>Draws diagram</td>
</tr>
<tr>
<td>4.</td>
<td>Indicates familiarity with type of problem</td>
</tr>
<tr>
<td>5.</td>
<td>Notes need for auxiliary information</td>
</tr>
<tr>
<td>6.</td>
<td>Lacks a systematic approach</td>
</tr>
<tr>
<td>7.</td>
<td>Recalls definition or auxiliary information</td>
</tr>
<tr>
<td>8.</td>
<td>Fails to use correct auxiliary cues</td>
</tr>
<tr>
<td>9.</td>
<td>Unsuccessful, adopts new approach</td>
</tr>
<tr>
<td>10.</td>
<td>Fails to retain model of solution</td>
</tr>
<tr>
<td>11.</td>
<td>Makes computational errors</td>
</tr>
<tr>
<td>12.</td>
<td>Indicates concern about method</td>
</tr>
<tr>
<td>13.</td>
<td>Signifies inability to solve problem</td>
</tr>
<tr>
<td>14.</td>
<td>Uses equations</td>
</tr>
<tr>
<td>15.</td>
<td>Uses deduction and arithmetic</td>
</tr>
<tr>
<td>16.</td>
<td>Stops without solution</td>
</tr>
<tr>
<td>17.</td>
<td>Makes structural errors</td>
</tr>
</tbody>
</table>
Thus thirty Ss were classed as verbal and thirty-seven were identified as physical. The remaining Ss either responded verbally to one problem and physically to the other or showed no reaction to one or both of the problems.

Scores for the sixty-seven Ss who had been classified as physical or verbal were computed based on the four test problems which had solutions. Although the mean for those who concentrated on physical cues was slightly higher than for those who relied on grammatical aspects, the difference between means was not statistically significant.

Previous research indicates that Ss who use auxiliary cues and internal representations of the physical aspects of a problem might be less impaired by instructions to think aloud than Ss who concentrate on verbal cues. This possibility was examined. Among Ss who were required to think aloud, fifteen were classified as verbal problem solvers and twenty-two as physical problem solvers. The means of the scores for these two groups were computed based on the four test problems which had solutions. The difference between means of the two groups was found to be statistically significant ($t = 3.06, p < .01$). Thus, the requirement to think aloud appeared to exert an adverse effect on the problem-solving behavior of the verbal problem solvers.

Discussion

Some psychologists have expressed reservations about the thinking aloud technique and its possible effect on the results of their research. However, they have found the method yielded quite a good amount of information while being fairly simple to employ. They are, therefore, reluctant to discard its use although very little experimentation has been done to test the efficacy of the technique specifically. The few studies which have examined the possibility of the technique affecting results have presented inconclusive reports (6, 11).

In general, the results of the present study appear to support de Groot's (3) confidence in the use of the thinking aloud technique. This psychologist, who has been interested in the analysis of thinking processes for many years, asserts: "If we want to simulate at all, if we want to try out program models of cognitive processes, we need good, old-fashioned introspection along with a modern methodology of protocol analysis."

There is, however, one respect in which this study sounds a note of caution. It was possible to divide some of the Ss into verbal and physical problem solvers according to the definitions proposed by Paige and Simon (9). Previous work (1, 7) suggests that if two similar tasks are performed simultaneously, the interference will be greater than if two dissimilar tasks are performed simultaneously. Since thinking aloud is closely related to the habitual mode of thought of the verbal problem solver, his problem-solving ability may be impaired by instructions to verbalize while solving problems. On the other hand, physical problem solvers usually construct some type of internal representation of the physical situation and, therefore, may be less likely to be influenced by instructions to think aloud. The result of the present study indicated that the requirement to think aloud may have somewhat impaired the problem-solving performance of the verbal solvers. Although the methods used to study this question were admittedly crude and any conclusions drawn can be only very tentative, the results suggest that the differences between verbal and physical problem solvers are worth exploring in greater detail using more problems and a larger number of Ss.

The finding that practice failed to influence the results of the present study may have been somewhat unexpected. Previous researchers have tended to accept the idea that practice would improve results. One possible explanation for the lack of a significant relationship between practice and problem-solving score in this study may be that the number of practice problems was too small to have a facilitating influence on performance. Also the practice problems may have been so similar to those which students had encountered previously in their algebra classes that the practice period had little import on their ability to solve the test problems.

At present, understanding of internal problem-solving processes is certainly not well developed. But the fact that research is continually progressing in the area of cognitive development suggests that a promising and stimulating future lies ahead.

REFERENCES