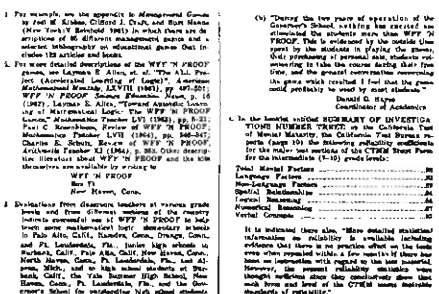
thorne effect", are not just those features exactly what we want to build into the learning situation arranged to facilitate optimum learning?

The fact that the class also took place in the summer with volunteer students could have at least three effects unrelated to WFF 'N PROOF: the teacher may have responded more enthusiastically to them (and they to him, having chosen to study over the summer), creating a motivational situation similar to that generated by the novelty of using the game. (This, however, may be a better explanation of the learners' motivation to understand and appreciate the particular subject matter than of the change in IQ scores.) It is also possible that students who volunteer to learn beyond normal requirements—regardless of their initial abilities—may have more potential to learn and improve their intellectual skills regardless of the type of training. And finally, the approaches of the teachers in the experimental and control groups may have had differential effects beyond those created by their differing teaching methods. We cannot be sure about the effects of any of these factors (and possibly others, too), but one finding in the study would suggest that they are subsidiary, at least, to the more specific aspects of the game. Since these factors would affect virtually any intellectual task, we would expect that if they were predominant, language IQ as measured by the CTMM would have increased equally with nonlanguage. Such is not the case, however; the junior high experimental group increased only 3.7, the control, only 1.1, and neither one increased significantly. Thus it would appear that we must look to other aspects of the game situation for a more adequate answer.

The properties of WFF 'N PROOF which would have produced the changes observed are somewhat less variable, but still far from unambiguous. The competition of being in a game may have been the major variable, and thus the increase in non-language IQ score may have been mainly a motivational measure reflecting the increased desire to succeed in this area. But, it is difficult to imagine that just *any* game would have had the same properties. Indeed, the students probably engaged in other games during both the junior high experimental and control group classes. And the closer those other games are to being "educational", the closer they would be to representing the variable we believe is isolated by the present study.

Just what is this variable? Our guess is that it is a combination of elements which include the achievement motive induced by the game (which could be both a motivational factor in taking tests—including the CTMM—as well as a motivational factor in learning the problem-solving skill taught by WFF 'N PROOF), the repeated practice at a symbol-handling task which the game provides, and the increased interest in the learning situation by both student and teacher. The motivational properties are perhaps the most dramatic to watch: discipline problems are virtually eliminated, students ask to play the game at recess and outside of class hours, etc. In this connection, it is relevant that the drop-out rate for all classes at the 1963 Burbank Summer School was more than 14 percent, while the drop-out rate among the students enrolled in the WFF 'N PROOF classes was exactly zero. These motivational properties carry over to the practice, for the student is provided relatively long hours of practice at solving logic problems in a way that is fun, rather than boring or tiring. That an increase in general problem-solving skills should thus result is not surprising, for the moves required by the game have virtually the same content as all logic exercises, and are similar in content to many mathematical exercises. Thus the game provides a way of making relevant and interesting in a social situation the problem-solving behavior required—we could say it provides a series of reinforcements for correct responses in problem-solving. This behavior has not only an apparent, but we think also a real, connection with the problem-solving tasks required in tests of problem-solving skills.

FOOTNOTES



THE JOURNAL OF EDUCATIONAL RESEARCH
(Volume 60, Number 1, September 1966)

Programed Games and the Learning of Problem-Solving Skills: The WFF 'N PROOF Example

LAYMAN E. ALLEN
Yale Law School

ROBERT W. ALLEN
Jordan Jr. High School, Burbank, California

JAMES C. MILLER
Yale University

THE OLD SAW is that learning algebra somehow improves one's ability to reason effectively. A related, but somewhat different and perhaps more interesting question is whether learning some mathematical logic improves one's reasoning—say, as measured by one of the standard IQ tests. On the other hand, would experience in playing games that provide practice of intricate problem-solving skills result in increased IQ scores? Finally, would the experience of learning some mathematical logic by means of intricate problem-solving games produce significant increases in the IQ scores of the players? The present study sheds some light on the third question and indicates the importance of further investigation of the first two.

Many games have been developed in recent years to enhance motivation and otherwise facilitate learning in a variety of fields.[1] One such game, (or rather, programed series of games) called WFF 'N PROOF, has been designed by the senior author to provide practice in abstract thinking, to teach some mathematical logic, and to help develop more favorable attitudes toward symbol-handling activities in general.[2] While evidence in the classroom to date suggests that this kit of games can be an effective aid in teaching logic, many of the broader implications of its use have as yet not been fully examined.[3] Will experience in playing such games enhance general problem-solving skills, modify attitudes about learning, or encourage other significant changes in the learning situation? The present study is an inquiry directed to the first of these broader implications: whether learning a game like WFF 'N PROOF aids in the development of more general problem-solving skills.

The subjects in this study were 57 junior high and high school students in the public school system in Burbank, California. The junior high school students were pre- and post-tested approximately six weeks apart by the California Test of Mental Maturity, Junior High Level, 1957 S-Form and the others by the Advanced Level, 1957 S-Form. The experimental group consisted of 35 students enrolled in summer school at John Burroughs High School in 1963. Members of the experimental group played WFF 'N PROOF for 45 minutes to an hour each day, five days a week for the 29 days of the summer session. In each two-hour class, the remaining time was devoted to reading and discussing the rules of the game and the concepts underlying them, as well as taking periodic tests to determine how effectively players were learning. The number of tests taken varied from individual to individual; some took five, others as many as 20. Students were instructed to take tests only when they felt they could construct any proof of the type taught by that phase of the game, and were allowed to proceed further in the game only when they in fact received 100 percent correct on the test for that phase. This method of testing and progressing insures (1) an understanding by each learner in some depth of each concept that has been introduced before new concepts are undertaken, and (2) a high degree of success for each learner on the tests since he is tested only when fully prepared.

In playing WFF 'N PROOF, which consists of a dice-like set of cubes imprinted with logical symbols which two, three, four or five persons roll in a competitive situation to determine who can most astutely construct a logical system and a proof of a theorem in that system, students were grouped on the basis of prior performance in playing the games. Thus the groups playing the game were hierarchically ranked and played with students of their own level unless a lower player "challenged" a player who ranked higher in the heirarchy, in which case, if he won, he was allowed to displace that higher player. The games formed, therefore, a kind of continuing "tournament" in which no one was eliminated, but only promoted or demoted. A series of graded problems were used to determine grades for the course: success with the simplest one assured a C; the next hardest, a B; and the hardest, an A.

The control group, which consisted of 22 junior high school students enrolled in regular 1963 fall classes, simply took the CTMM twice at six-week intervals with regular classwork intervening.

While an attempt was made to obtain a sample similar to that of the experimental group, certain differences resulted: there were more boys (7 to 16) in the junior high experimental group than in the control (10 to 12); the junior high experimental group's average initial IQ score (by CTMM) was also ten points higher than the average IQ score of the control group (114.5 vs. 104.5). These differences are taken into account in presenting the results here. The control group also contained no high school students; therefore, in all comparisons made below, only the results of the junior high sample of the overall experimental group ($N = 23$) are included. The teacher of the experimental group during the six-week summer session was also different from those of the control group for the six weeks in the fall.

The dependent variable for all of the comparisons was change in non-language IQ score from pre- to post-testing. This choice was in part directed by practical considerations (the CTMM is readily available and widely used in California) and partly by the notion that general problem-solving skills are most validly measured by the non-language section of an intelligence test. While there may be some doubt as to whether such tests in fact measure problem-solving skills, our examination of the validating work done with the CTMM leads us to conclude that it was both reliable and valid enough to present research results in terms of it with confidence, once the limitations of such an approach are admitted as a reminder.[4]

The overall comparison of mean change scores in the junior high experimental and control groups were found to be highly significant. The mean change in non-language IQ score for the junior high experimental group was +17.3, while the mean change for the control group was +9.2. By a t test this difference between the two groups could have occurred by chance about twice out of a hundred times ($p = .02$).[5] What this indicates at the simplest level is that while some positive change occurred in the control group—perhaps from a practice effect of taking the test twice (although the California Test Bureau indicates that there is no practice effect on this test[6]), or possibly from other curricular or environmental learning over the elapsed interval between tests—the change in the junior high experimental group was significant above and beyond whatever "normal" changes one could expect.

Further interpretation of these results, however, must take into account both initial IQ score differences, and the difference in sex distribution in the two groups. A linear positive relationship between initial IQ score and change would, in fact, indicate that the difference between the junior high experimental and control groups might merely be a matter of group selection; a larger change by boys than girls would limit the generality of the difference to males, since the larger change group was predominantly male. While the results bear out the latter possibility, they do not bear out the former. There is no positive correlation between initial position and change in either the junior high experimental ($r = -.33$) or control ($r = -.35$) groups; in fact, there is a low negative correlation in both of them. There is a suggestion of a curvilinear relationship in both of these groups, with extreme quartiles changing more than middle, but by t tests neither is significant (junior high experimental, $p = .08$; control, $p = .15$). Furthermore, when initial IQ score is covaried with change, to bring initial positions between the two groups equal and thus measure *corrected* change, the result is even more significant ($p < .01$). Thus initial IQ score difference is not an important variable facilitating the overall increase in IQ score that resulted.

An analysis of differences between boys and girls, however, indicates that virtually all of the main experimental differences were contributed by boys. Mean change for boys in the junior high experimental group was +18.7, while change for boys in the control group was only +3.9 (the difference yields a significant t well beyond .01). Mean change for girls in the junior high experimental group was +14.3, while change for girls in the control group was +13.6 (the difference is not significant). The puzzling feature of this result is not that the boys produce most of the change in the junior high experimental group (in fact they do not), but rather, that by simply retaking the test, the girls change almost as much as the boys change from exposure to the logic games. One possible explanation is that the girls benefited more from remembering the previous test, and received little additional benefit from playing the logic games, whereas the boys remembered little of the pre-test but benefited greatly from the motivational or learning properties of the competitive games. Some might argue that this would make some intuitive sense from what is known about male and female roles in learning: women frequently do better at rote memory tasks, while men are conditioned to compete in all phases of their work and play and are thus more likely to be affected by a competitive learning situation. But whatever explanation is adopted (and quite frankly, we are puzzled by these results), it is a fact that virtually all of the overall difference between junior high experimental and control groups is contributed by the boys in the present study.

Table 1.—Comparison of Changes after a Six-Week Interval in Language IQ Scores and Non-Language Scores on the California Test of Mental Maturity, Junior High Level, 1957 S-Form

Control Group (22 junior high students enrolled in regular classes at Jordan Junior High School in the 1963 fall term)	Language IQ Scores			Non-Language IQ Scores		
	Pre.	Post	Change	Pre.	Post	Change
GIRLS						
G 1	99	117	+18	90	118	+28
G 2	106	119	+13	91	95	+ 4
G 3	99	106	+ 7	95	91	- 4
G 4	93	92	- 1	98	100	+ 2
G 5	116	121	+ 5	71	98	+27
G 6	97	97	0	98	109	+11
G 7	113	93	-20	105	107	+ 2
G 8	100	100	0	68	106	+38
G 9	88	88	0	77	94	+17
G10	143	136	- 7	116	130	+14
G11	125	126	+ 1	113	126	+13
G12	118	120	+ 2	126	137	+11
Girls' Averages	108	110	+ 1.5	96	109	+13.6
BOYS						
B 1	121	101	-20	101	100	- 1
B 2	143	138	- 5	116	134	+18
B 3	93	115	+22	81	89	+ 8
B 4	91	83	- 8	113	122	+ 9
B 5	93	103	+10	87	76	-11
B 6	147	142	- 5	151	158	+ 7
B 7	109	104	- 5	85	94	+ 9
B 8	101	101	0	109	105	- 4
B 9	100	99	- 1	107	117	+10
B10	85	104	+19	119	113	- 6
Boys' Averages	108	109	+ 0.7	107	111	+ 3.9
Boys' and Girls' Averages	108	109	+ 1.1	101	110	+ 9.2

Experimental Group (23 junior high students enrolled in the WFF 'N PROOF class at Jordan Junior High School in the 1963 summer term)	Language IQ Scores			Non-Language IQ Scores		
	Pre.	Post	Change	Pre.	Post	Change
GIRLS						
G 1	101	120	+19	108	114	+ 6
G 2	130	116	-14	127	161	+34
G 3	128	138	+10	129	138	+ 9
G 4	88	94	+ 6	97	128	+31
G 5	105	109	+ 4	121	122	+ 1
G 6	123	104	-19	101	100	- 1
G 7	117	118	+ 1	120	140	+20
Girl's Averages	113	114	+ 1.0	115	129	+14.3
BOYS						
B 1	89	89	0	94	91	- 3
B 2	102	109	+ 7	87	109	+22
B 3	137	134	- 3	125	136	+11
B 4	134	122	-12	125	144	+19
B 5	122	124	+ 2	122	116	- 6
B 6	116	128	+12	121	133	+12
B 7	131	133	+ 2	121	148	+27
B 8	102	121	+19	85	123	+38
B 9	114	118	+ 4	108	128	+20
B10	132	134	+ 2	132	164	+32
B11	108	102	- 6	112	138	+26
B12	125	132	+ 7	123	153	+30
B13	123	125	+ 2	110	126	+16
B14	111	133	+22	99	127	+28
B15	104	111	+ 7	70	105	+35
B16	144	156	+12	144	136	- 8
Boys' Averages	118	123	+ 4.8	111	130	18.7
Boys' and Girls' Averages	117	120	+ 3.7	112	130	17.3

In speculating about the results of the study, it is essential to consider the complex conditions of the classroom in which WFF 'N PROOF was taught. There are numerous aspects of the game itself, of the classroom situation, and of the administration of the tests which could have been responsible for the observed differences. With respect to the classroom situation it should be noted that the game was played under conditions quite unlike those of an ordinary class; the teacher involved knew that he was utilizing a new method, and in the course of the class he presumably became more committed to the use of the game. The observed differences, then, could have been partly the result of his communication of his special interest and enthusiasm to the students—a kind of Hawthorne effect—having to do with the playing of the game only insofar as it affected the teacher. However, whatever mysteries might be lurking beneath the tag "Haw-