The Alberta Journal of Educational Research *Vol. XXIII, No. 4, December, 1977*

LAYMAN E. ALLEN
and
JOAN ROSS
University of Michigan

Improving Skill in Applying Mathematical Ideas: A Preliminary Report on the Instructional Gaming Program at Pelham Middle School in Detroit

Performance of 237 students in ten inner-city eighth-grade math classes was assessed under five different sets of instructional conditions to measure their ability to solve two types of problems: (1) those that involve only computation and (2) those that involve not only computation, but also recognition of the relevance of a particular mathematical idea that is indispensable to the solution of the problem. Results indicate that the combination of playing EQUATIONS (an instructional mathematics game) over a two-year period and then working intensively with IMP (Instructional Math Play) Kits for two weeks enables students to apply mathematical ideas (in the sense studied in this experiment) better (at the .0001 level of significance) than any of the other four sets of conditions: (1) playing EQUATIONS alone, (2) playing EQUATIONS and being taught explicitly by a teacher the ideas presented in the IMP Kits, (3) being taught the ideas in an ordinary math class without playing EQUATIONS, and (4) participating in an ordinary math class without explicit teaching of the ideas or playing EQUATIONS. (Mr. Allen is Professor of Law and Research Scientist at the Law School and the Mental Health Research Institute at the University of Michigan; Mrs. Ross is Research Associate at the Mental Health Research Institute.)

Educators working with the EQUATIONS game and associated materials quickly become aware that applying mathematical ideas is much more difficult than merely computing with those ideas. Of course, "applying" an idea is a somewhat vague notion; this preliminary report will be addressed to one aspect of application — namely, recognition that an idea is indispensably relevant to the solution of a problem. Consider the following pair of problems: the first is the C-type (computation), and the second is the R-type (relevance).

1. 6 - (1-3) = __?__.

	A	*B*	*C*	*Yes*	*No*
2.	- - 1 3 6	8	____	___	___

By writing an X in the Yes or No column, indicate whether or not all of the numbers and operations in Column A can be appropriately ordered and grouped (inserting parentheses wherever necessary) to form an expression equal to the number in Column B. If your answer is Yes, write that expression in Column C.

The indispensably relevant idea for solving each of the problems is the subtraction of negative numbers. In the C-type problem, the very statement of the problem clearly and explicitly indicates that subtracting a negative number is involved. That is neither so clearly nor so explicitly cued in the statement of the R-type problem. Those who understand how to subtract negative numbers can easily do the first problem correctly. But many of those who can solve a C-type problem involving subtraction of negative numbers fail to solve a corresponding R-type problem involving negative numbers. In general (in the groups we have studied), about two-thirds of those who solve C-type problems fail to solve a corresponding R-type problem that involves the same idea. The R-type problem is considerably harder than the C-type in this example because the student must recognize from less clear and less explicit cues that subtracting negative numbers is an indispensably relevant idea for solving the problem. "Understanding" an idea in the R-sense (being able to solve R-type problems) includes understanding it in the C-sense, but it also involves something more. R-sense understanding includes the capability of selecting from among a storehouse of ideas understood in the C-sense, those that are indispensably relevant for solving a particular problem. The question to which this study is addressed is whether skills in applying mathematical ideas can be improved by learning procedures which emphasize exposure to situations that are rich in opportunities for such application, at levels of complexity appropriate for each learner.

The EQUATIONS Game and the IMP (Instructional Math Play) Kits

The rules that define the EQUATIONS game establish a problem-generating and problem-solving interaction between small groups of students, an interaction that can easily be controlled to provide a highly individualized learning experience for each of the participants. It is a RAG (Resource Allocation Game) where the resources involved are mathematical ideas. (For details, see Allen, 1972.) The IMP Kits are 16-page pamphlet-simulations of a computer playing EQUATIONS where the computer is programmed to play like a good teacher, rather than like a good player. Each kit presents a lesson on one mathematical idea. For example, in the following situation.

RESOURCES:	+ - - 1 3 3 ( )	FORBIDDEN: 2
		PERMITTED:
________	= __5__	REQUIRED:
SOLUTION	GOAL	

on its turn to move, the computer might move the + from RESOURCES to FORBIDDEN, thus extinguishing the SOLUTION (3 + 3) - 1 and presenting the learner, in effect, with the question:

> Is it still possible (after the + is FORBIDDEN) to construct an expression equal to the GOAL of 5 from the remaining RESOURCES?

If the learner by challenging the computer's move, in effect, answers "NO," then the computer will direct the learner to a comment that teaches a lesson in the subtraction of negative numbers:

> Your challenge that all SOLUTIONS have been extinguished is incorrect. The SOLUTION 3 - (1-3) is still possible. Notice how a pair of minus signs can be used to give the effect of addition. Since 3-5 = -2 and 3 - -2 = 5, the GOAL can be achieved even though the + is FORBIDDEN. Go on to the next IMP Kit.

There are at present five versions of each of 21 ideas, or a total of 105 kits. The first 21 of these IMP Kits have been published and are available to those interested. (For details, see Allen & Ross, 1975.)

Method

Subjects

All of the students in each of ten of the fourteen eighth-grade classes in mathematics at Pelham Middle School participated in the study. The ten classes were chosen to include all four of the classes in which the EQUATIONS game had been used during the prior two years as part of the regular instructional program in mathematics and two other classes of each of the three participating teachers. Pretest and/or posttest data were collected on 237 of the students enrolled in these ten classes.

Experimental Treatments

The following five different sets of experimental conditions were represented in the ten classes:

- I an EQUATIONS class in which the IMP Kits were used in five class periods during the two-week experiment and the regular once-a-week EQUATIONS tournament was continued;
- E an EQUATIONS class in which the game was played for the five class periods without any explicit teaching of the 21 IMP Kit ideas:
- TE two EQUATIONS classes in which the game was played for five class periods and the teachers explicitly taught the 21 ideas presented in the IMP Kits;
- 0 three non-EQUATIONS classes in which the ordinary classroom procedure was continued without change with no special attention given to the 21 IMP Kit ideas; and
- T three non-EQUATIONS classes in which the teachers explicitly taught the 21 IMP Kit ideas for five class periods.

The set of conditions of greatest interest to the researchers was that of the I group in which students individually played through the IMP Kits, completing as many of the set of 105 as they could in the five periods. Before

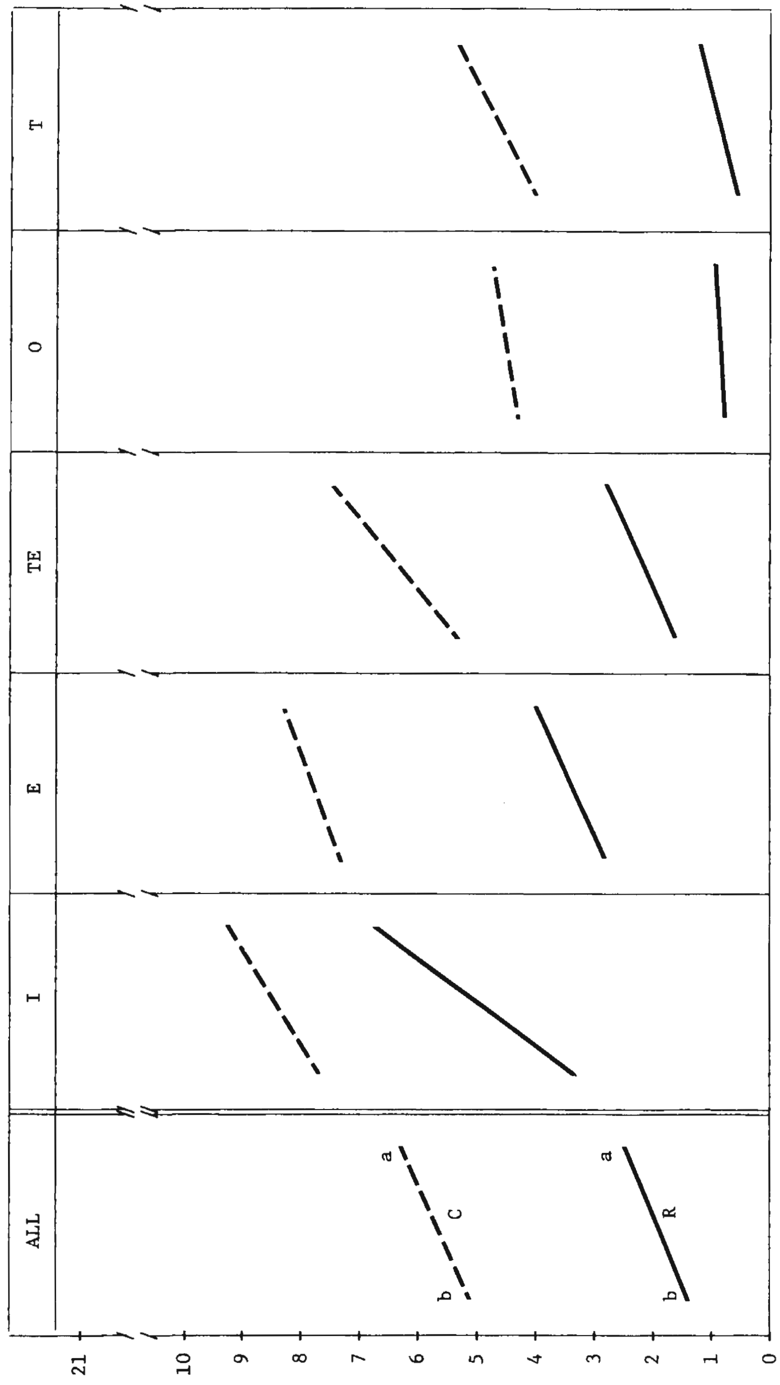

Figure 1. Mathematics pretest and posttest scores on C Tests and R Tests of five experimental groups of eighth-grade classes, Pelham Middle School, Detroit, 1974.

	ALL			I			E			TE			O			T		
	b	a	a − b	b	a	a − b	b	a	a − b	b	a	a − b	b	a	a − b	b	a	a − b
C																		
N	206	202	188	23	23	23	18	17	17	46	44	44	58	61	54	61	57	50
$\bar{X}$	5.10	6.32	1.26	7.70	9.26	1.57	7.39	8.35	0.82	5.33	7.48	2.07	4.31	4.74	0.44	4.02	5.32	1.42
Sx	2.50	3.23	2.36	2.29	2.75	2.39	2.70	2.87	2.40	2.33	3.69	2.43	1.81	2.20	2.28	2.05	2.61	2.11
Signif.			.0001			.005			ns			.0001			ns			.0001
R																		
N	207	202	190	23	23	23	17	17	16	46	44	44	60	60	56	61	58	51
$\bar{X}$	1.40	2.43	1.03	3.39	6.78	3.40	2.88	4.12	1.00	1.65	2.84	1.11	0.83	0.95	0.11	0.59	1.41	0.92
Sx	1.70	2.79	1.81	1.56	3.16	2.27	2.00	2.32	1.93	1.83	2.62	1.79	1.15	1.21	0.95	1.01	1.82	1.29
Signif.			.0001			.0001			ns			.0005			ns			.0001
C − R																		
N	202	198	181	23	23	23	17	17	16	46	44	44	58	59	53	58	55	43
$\bar{X}$	3.73	3.93	0.20	4.30	2.48	⁻1.83	4.59	4.24	⁻0.19	3.67	4.64	0.95	3.45	3.85	0.40	3.59	3.98	0.42
Sx	1.88	2.22	2.60	1.99	1.86	2.35	2.83	2.17	3.74	1.71	2.51	2.56	1.62	1.90	2.12	1.83	2.23	2.35
Signif.			ns			.005			ns			.05			ns			ns

KEY

Tests:

- – – – – – C test
- ——— R test
- a (after) posttest
- b (before) pretest
- C C test
- R R test

Treatments:

- I IMP Kits
- E EQUATIONS
- TE Taught 21 Ideas, EQUATIONS
- O Ordinary classroom activities
- T Taught 21 Ideas

this group started on the IMP Kits, one class period was devoted to teaching members of the class how to use the kits.

Dependent Variables

The effects of the various sets of experimental conditions were measured by two different forms of a pair of specially-constructed tests targeted at the 21 mathematical ideas presented in the IMP Kits. The first of the pair of tests is called a C test; it contains only C-type items. The second test is called an R test; it contains only R-type items. Two different forms of the C test were used (Form C and Form D), as well as two different forms of the R test (Form E and Form F). In each of the ten classes in which these tests were administered, the students were divided into eight groups — G1, G2, . . . G8. Each student received a C test and R test as pretests, and each received alternative forms of the two tests as posttests as follows:

	Order of Administration	*Groups*							
		G1	*G2*	*G3*	*G4*	*G5*	*G6*	*G7*	*G8*
Pretest	1	C	C	E	F	D	D	E	F
	2	E	F	C	C	E	F	D	D
Posttest	1	F	E	D	D	F	E	C	C
	2	D	D	F	E	C	C	F	E

Using Ca (*a*fter) to denote the score on the C posttest and Cb (*b*efore) to denote the C pretest score (and similarly for the R pretests and posttests), outcome measures of three dependent variables can be specified as follows:

1. Ca-Cb	Increase in performance on C test
2. Ra-Rb	Increase in performance on R test
3. (Ca-Ra) - (Cb-Rb)	Decrease in difference in performance on C test and R test

Results

The scores for each of the five experimental groups, summarized in Figure 1, were significantly higher (at the .0001 level) on the C test than on the R test both on pretests and on posttests. The mean pretest score for all students on the C test was 5.10, while for the R test it was 1.40 (maximum score = 21), a ratio of about 3.6 to 1. On the posttests the ratio decreased to 2.6 to 1 with mean scores of 6.32 and 2.43, respectively.

Three of the experimental groups had significant differences between pretest and posttest scores on the C test (measured by Ca-Cb). The IMP Kit group (I) had a mean pretest of 7.70 and 9.26 on the posttest, significantly higher at the .001 level, and the groups that were explicitly taught the 21 ideas (TE and T) went from 5.33 to 7.48 and from 4.02 to 5.32, respectively, significantly higher on the posttest at the .0001 level.

The same three experimental groups had significant differences between pretest and posttest scores on the R test (measured by Ra-Rb). The IMP Kit group and the T group were significantly higher on the posttest (.0001) with mean pretest to posttest scores of 3.39 to 6.78 and 0.59 to 1.41, respectively, whereas the TE group was significantly higher at the .0005 level with scores of 1.65 to 2.84.

Only two of the experimental groups showed significant changes in the gap between C-sense understanding and R-sense understanding from the pretests to the posttests [as measured by (Ca-Ra) - (Cb-Rb)]. The IMP Kit group achieved a 1.83 reduction in its CR Gap, significant at the .005 level, in moving from a pretest gap of 4.30 to a posttest gap of 2.48. On the other hand, the T group increased its CR Gap by 0.95, significant at the .05 level, with a pretest gap of 3.67 and a posttest gap of 4.64.

In comparing the test scores of the experimental groups with each other, only those pairs in the total collection that qualify by the highly conservative Scheffé procedure (see Winer, 1971) at the .05 level of significance are reported as being significantly different. The results of the between-group comparisons are summarized in Figure 2.

Tests: C

Pretest — Significance Lines and Levels

	T	O	TE	E	I
T		ns	.0005	.0001	.0001
O			ns	.0001	.0001
TE				.001	.0001
E					ns

Posttest — Significance Lines and Levels

	O	T	TE	E	I
O		ns	.0001	.0001	.0001
T			.0005	.0001	.0001
TE				ns	ns
E					ns

Pretest - Posttest — Significance Lines and Levels

	O	E	T	I	TE
O		ns	ns	ns	.005
E			ns	ns	ns
T				ns	ns
I					ns

Tests: R

Pretest — Significance Lines and Levels

	T	O	TE	E	I
T		ns	.0005	.0001	.0001
O			ns	.0001	.0001
TE				ns	.0001
E					ns

Posttest — Significance Lines and Levels

	O	T	TE	E	I
O		ns	.0001	.0001	.0001
T			.001	.0001	.0001
TE				ns	.0001
E					.0001

Pretest - Posttest — Significance Lines and Levels

	O	T	E	TE	I
O		ns	ns	.001	.0001
T			ns	ns	.0001
E				ns	.0001
TE					.0001

Tests: C - R

Pretest — Significance Lines and Levels

	E	I	TE	T	O
E		ns	ns	ns	ns
I			ns	ns	ns
TE				ns	ns
T					ns

Posttest — Significance Lines and Levels

	TE	E	T	O	I
TE		ns	ns	ns	.0001
E			ns	ns	ns
T				ns	ns
O					ns

Pretest - Posttest — Significance Lines and Levels

	TE	T	O	E	I
TE		ns	ns	ns	.0001
T			ns	ns	.0005
O				ns	.0005
E					ns

KEY

Treatments:

- I IMP Kits
- E EQUATIONS
- TE Taught 21 Ideas, EQUATIONS
- O Ordinary classroom activities
- T Taught 21 Ideas

Figure 2. Significance levels of pairs of sets of conditions that significantly differ from each other on test scores. (The only pairs of sets of conditions shown in this table as significantly different from each other are those in the collection of pairs that satisfy the highly conservative Scheffé procedure at the .05 level.)

On the pretest scores seven of the pairs of groups were significantly different on the C tests, six pairs were different on the R tests, but none were significantly different on the amount of the CR Gap. The significance lines written over the names of each of the experimental groups in Figure 2 should be interpreted as follows:

1. Groups whose names *do not* appear under a common line *do* differ significantly from each other;
2. Those whose names *do* appear under a common line *do not* differ significantly from each other.

Hence, on C pretest scores the IMP Kit group was significantly higher than the TE, O, and T groups (at the .0001 level); the E group was higher (.001) than the TE group and higher (.0001) than the O and T groups; and the TE group was higher (.0005) than the T group.

Pre C T O TE E I

On the R pretest scores the IMP Kit group was again significantly higher than the TE, O, and T groups (.0001); the E group, higher than the O and T groups (.0001); and the TE group, higher than the T group (.0005).

Pre R T O TE E I

The lack of any significant difference on the pretest scores between any of the pairs of experimental groups with respect to the amount of CR Gap is indicated by the appearance of all of the names of the groups under a common line.

Pre C-R E I TE T O

On the posttest scores one less pair of groups was different on the C test; one more pair, different on the R test; and there emerged at this time a pair different with respect to the CR Gap. On the C posttest scores both the IMP Kit group and the E group were significantly higher (.0001) than the T and O groups, and the TE group was higher (.0005) than the T group as well as higher (.0001) than the O group. On the R posttest scores, the IMP Kit group was significantly higher (.0001) than every one of the other groups; the E group, higher (.0001) than the T and O groups; and the TE group, higher (.001) than the T group and higher (.0001) than the O group. With respect to the amount of CR Gap on posttest scores, the one pair significantly different resulted primarily from the large pretest-posttest improvement in the R test score of the IMP Kit group; thus the CR Gap of the IMP Kit group turned out to be significantly smaller (.0001) than that of the TE group.

On the improvement indicated by the difference between posttest and pretest scores, the IMP Kit group clearly emerges as the group that achieved the greatest improvement. The only other group that was significantly higher than any of the other groups on any of the three improvement measures was the TE group. In improvement on C test scores only one pair of groups was significantly different: the improvement of the TE group was greater (.005) than that of the O group. In improvement on R test scores the IMP Kit group was significantly greater (.0001) than every one of the other groups, and the TE group was greater (.001) than the O group. With respect to what is probably the most important measure of all — the extent of the improvement in reducing the CR Gap — the IMP Kit group is the only group significantly better than any of the other groups. It deserves emphasis that the IMP Kit group turned out significantly better on this measure than every other group except the E group — and better than the E group, although not significantly so. The improvement of the IMP Kit group in reducing the CR Gap was greater (.0005) than that of the O and T groups and greater (.0001) than that of the TE group.

The significant differences on the C pretest and R pretest scores among the experimental groups deserve close scrutiny. Most (10 of the 13) of the differences are differences between EQUATIONS groups and nonEQUATIONS groups, and nearly half (6 of 13) are differences between the IMP Kit group and other groups. This raises the question as to whether the EQUATIONS groups generally and the IMP Kit group in particular were not simply more capable students at the beginning of the experiment. If so, perhaps it is not surprising that they improved more in learning to apply mathematical ideas during the two-week experiment. The next question is: Given that the students in the EQUATIONS groups were more capable at the end of their eighth-grade year when the experiment was conducted, were they also more capable two years earlier when they entered seventh grade?

School records indicate that the Stanford Arithmetic Test — Advanced (computation) was administered to all entering seventh-grade classes two years earlier and that 112 of the students in this study participated. From the scores recorded for this sample of the 237 students in the study for whom there is this indication of mathematical capability at the time of entry to the seventh grade, it appears that there was no significant difference between any of the pairs of the experimental groups at that time. In particular, there was no significant difference between the IMP Kit group and any of the other groups. Also, when data for the three EQUATIONS groups are combined and those for the two nonEQUATIONS groups are also combined, there is no significant difference in mean scores between the EQUATIONS groups and the nonEQUATIONS groups. The data are summarized in Table 1.

TABLE 1

MATHEMATICAL CAPABILITIES TWO YEARS EARLIER: SCORES ON STANFORD ARITHMETIC TEST—ADVANCED (COMPUTATION)[a]

Group	All	I	E	TE	O	T	EQ	nEQ
N	112	22	14	25	23	28	61	51
$\overline{X}$	46.54	48.95	52.07	44.36	44.96	43.50	47.79	44.16
Sx	11.48	9.59	11.06	10.93	10.05	13.70	10.78	12.20

[a] Administered in September, 1972, for 112 of the 237 students in this study enrolled in Pelham Middle School eighth-grade classes in May, 1974.

The EQUATIONS and nonEQUATIONS groups were quite different two school years later when this experiment was undertaken, as was the IMP Kit group compared to all other groups except the E group. On the C pretest the EQUATIONS groups had a mean score of 6.38, significantly higher at the .0001 level than the 4.16 mean score of the nonEQUATIONS groups. On the R pretest the EQUATIONS groups were also significantly higher (.0001); the mean scores were 2.36 to 0.71. With respect to the CR Gap the 0.50 difference between the means of the two groups was not significant. The data for the EQUATIONS groups compared to the nonEQUATIONS groups are summarized in Table 2. The data for the other comparisons are in Figure 2.

TABLE 2
COMPARISON OF EQUATIONS AND NONEQUATIONS GROUPS

		Pretest b		Posttest a		Posttest-Pretest a − b	
		EQ	nEQ	EQ	nEQ	EQ	nEQ
C Test							
	N	87	119	84	118	84	104
	$\overline{X}$	6.38	4.16	8.14	5.02	1.68	0.91
	Sx	2.62	1.94	3.35	2.41	2.44	2.24
	Signif.	.0001		.0001		.05	
R Test							
	N	86	121	84	118	83	107
	$\overline{X}$	2.36	0.71	4.18	1.18	1.72	0.50
	Sx	1.94	1.08	3.17	1.55	2.20	1.19
	Signif.	.0001		.0001		.0001	
C-R							
	N	86	116	84	114	83	98
	$\overline{X}$	4.02	3.52	3.96	3.91	−0.04	0.41
	Sx	2.06	1.72	2.44	2.06	2.99	2.22
	Signif.	ns (.0589)		ns		ns	

Discussion

This study provides strong support for the proposition that skills in applying mathematical ideas can be improved by learning procedures that are rich in opportunities for such application at appropriate levels of complexity for each student. Interpreted most favourably, the results show that the combination of playing EQUATIONS over a two-year period and then working intensively with the IMP Kits for two weeks enables students to apply mathematical ideas (in the sense studied in this experiment) better than any of the other four sets of conditions do: better than just playing EQUATIONS alone, better than playing EQUATIONS and being taught explicitly by the teacher the 21 ideas presented in the IMP Kits, better than being taught the 21 ideas in an ordinary traditional mathematics class, and better than being in an ordinary traditional class without any special teaching of the ideas — and, furthermore, better in each case by a highly conservative test at an extreme level of significance (.0001). It should be acknowledged immediately that there are some questions with respect to this most favourable interpretation which require further investigation. At the start of the experiment the IMP Kit group was clearly performing at a higher level of achievement than were the other groups. This superior performance seems linked to their two-year experience in playing EQUATIONS. The available evidence indicates that the EQUATIONS and nonEQUATIONS groups were not different upon their entry to the seventh grade, but that after two years of different experience with respect to whether or not they played EQUATIONS, the EQUATIONS group was significantly better in both computing with and applying these 21 mathematical ideas. The emphasis of the IMP Kit experience is clearly in the direction of improving skills in applying mathematical ideas, although it does improve both computing and applying. In terms of reducing the difference between understanding a mathematical idea in the sense of computing correctly with it and understanding it in the sense of being able to apply it in a context where it must be recognized to be relevant (the CR Gap), playing through the IMP Kits clearly is more effective than any of the other methods tried in this experiment (except possibly playing EQUATIONS alone, where the effect is in the right direction but is not significant).

For classrooms in which learning to apply mathematical ideas is still a problem, the implications of the findings of this study are obvious: learning environments so structured are effective — and should be used.

In an earlier study (Allen & Main, 1976), the objectives of designing structured learning environments of the type studied here were described as the enhancement of both the affective and cognitive dimensions. It was shown there that attitude as measured by absenteeism was profoundly affected through the use of such structured learning environments, and the prediction was ventured that in intelligent hands the achievement of knowledge should be, too. In this inner-city school in Detroit the learning structure we are designing has been in such hands. We still need to learn more, but the efforts there have advanced our understanding one solid step along the way.

We gratefully acknowledge the indispensable relevance of the cooperation of four dedicated educators at Pelham Middle School whose enthusiastic support made this study possible: Lewis Jeffries, Principal; Gloria Jackson, Mathematics Department Chairperson; and William Beeman and Harold Hauer, Mathematics Teachers.

References

Allen, L. E. *EQUATIONS: The game of creative mathematics.* New Haven: Autotelic Instructional Materials Publishers, 1972. (Originally published, 1964.)

Allen, L. E. RAG-PELT: Resource Allocation Games—Planned Environments for Learning and Thinking. *Simulation and Games,* 1972, *3,* 407-438.

Allen, L. E., & Main, D. The effect of instructional gaming upon absenteeism: The first step. *Journal for Research in Mathematics Education,* 1976, *7,* 113-128.

Allen, L. E., & Ross, J. Instructional gaming as a means to achieve skill in selecting ideas relevant for solving a problem. *International Journal of Mathematics Education and Science Technology,* 1975, *6,* 475-488. The 21 published IMP Kits are available from the Autotelic Instructional Material Publishers, 1490 South Boulevard, Ann Arbor, MI. 48104 ($15).

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw Hill, 1971.