The purposes of this paper are: (1) to investigate the effects of SQUARE ONE TV on children's problem-solving behavior and their attitudes toward mathematics and (2) to indicate some connections between the result and the National Council of Teachers of Mathematics (NCTM)'s Curriculum and Evaluation Standards for School Mathematics. This paper describes SQUARE ONE TV and compares its goals with those of the Standards. Half of the 48 fifth graders received treatment of 30 half-hour unaided viewing sessions in a group setting for six weeks. The Problem-Solving Activities (PSAs) coding systems are described. From pretest to posttest, children in the viewing group made significantly greater both P-score (problem solving) and M-score (mathematical completions and sophistication) gains on the PSAs than the nonviewers did. Possible reasons for the effects are considered. Goals of the SQUARE ONE TV are listed in the appendix. (YP)
A Study of the Effects of SQUARE ONE TV on Children's Problem Solving and Some Connections with NCTM's Standards

Edward T. Esty
Eve R. Hall
Shalom M. Fisch

Children's Television Workshop

A Paper Presented at the RAC/SIGRME Research Presession before the National Council of Teachers of Mathematics' Annual Meeting at Salt Lake City, April 17, 1990

Authors' address:
Children's Television Workshop
1 Lincoln Plaza
New York, NY 10023
Introduction

The purpose of this paper is twofold: To present information about a study entitled "Children's Problem-Solving Behavior and Their Attitudes toward Mathematics: An Evaluation of the Effects of SQUARE ONE TV," and, concurrently, to indicate some connections between that study and the document from the National Council of Teachers of Mathematics (NCTM) entitled Curriculum and Evaluation Standards for School Mathematics (NCTM, 1989). This dual purpose will necessitate some alternating between reporting of the study and some higher-level commentary from the point of view of the Standards.

The basic outline of the paper is this: We start with a very brief description of SQUARE ONE TV, and compare its goals with those of the Standards. Next we describe the study and its background, together with some results, again with commentary from the perspective of the Standards. Finally, we make some overall connections among SQUARE ONE TV, its evaluation, and the Standards, against a backdrop of current school mathematics.

A Brief Description of SQUARE ONE TV

SQUARE ONE TV is a television series about mathematics, produced by Children's Television Workshop (CTW). It is aimed at an audience of 8- to 12-year-old viewers, primarily watching at home (although some stations carry the program during school hours). The program is 30 minutes long, and is generally broadcast Mondays through Fridays on public television stations. The first two production seasons resulted in a total of 115 programs. Season III, which premiered in January, 1990, consists of 40 new programs. Season IV production is now under way.

SQUARE ONE TV has three goals that have guided its production:

Goal I is to promote positive attitudes toward, and enthusiasm for, mathematics, by showing that mathematics is a powerful and widely applicable tool; is aesthetically pleasing; and can be understood, used, and even invented, by non-specialists;

1The production of SQUARE ONE TV and the research summarized here have been supported by the National Science Foundation, the Corporation for Public Broadcasting, the Carnegie Corporation, and the U.S. Education Department. First season production was also supported by the Andrew W. Mellon Foundation and by IBM.
Goal II is to encourage the use and application of problem-solving processes (which are classified as problem formulation, problem treatment, problem-solving heuristics and problem follow-up); and

Goal III is to present sound mathematical content in an interesting, accessible and meaningful manner. (This content is categorized as numbers and counting; arithmetic of rational numbers; measurement; numerical functions and relations; combinatorics; statistics and probability; and geometry).

Each of these goals is further refined into a range of subgoals; the complete breakdown is fully explicated in Appendix I.

A Comparison with the Standards

We pause at this point to consider how the three goals of SQUARE ONE TV are related to the "five general goals for all students" set forth in the Standards (NCTM, 1989, p. 5). The production of SQUARE ONE TV started well before even the first version of the Standards was published, so the series was not designed with the specific goals of the Standards in mind. Nonetheless, the designers of the series and the show's Advisory Board have been active in mathematics education reform for many years, so it is reasonable to expect there to be some relation. The following table illustrates the correspondence:

<table>
<thead>
<tr>
<th>SQUARE ONE TV</th>
<th>Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Promote positive attitudes, enthusiasm</td>
<td>1. Learn to value mathematics</td>
</tr>
<tr>
<td>II. Encourage use of problem solving</td>
<td>2. Become confident in ability to do mathematics</td>
</tr>
<tr>
<td>III. Present sound mathematical content</td>
<td>3. Become mathematical problem solvers</td>
</tr>
<tr>
<td></td>
<td>5. Learn to reason mathematically</td>
</tr>
</tbody>
</table>

(The connection between III. and 5. on the last line of Table 1 is less direct than the others, but it should be noted here that
the wide range of mathematical content actually presented on SQUARE ONE TV corresponds closely to the NCTM content-specific standards; that is, Standards 5 through 13 for Grades 5-8 are consonant with the subgoals of Goal III. The only place the correspondence breaks down is with Standard 9, Algebra, which is largely absent in the series.)

Given that SQUARE ONE TV is a television series, it is ironic that the one "general goal" of the Standards that is not explicitly present in the SQUARE ONE TV statement of goals is "(4) that they [students] learn to communicate mathematically." (NCTM, 1989, p. 5.) There are many instances in the program where good mathematical communication between characters is modeled, and of course the series itself is an attempt to communicate mathematically with its audience, but the promotion of mathematical communication among its audience is not a goal of the program.

Each segment of every program in Seasons I, II and III has been carefully analyzed to determine which subgoals it incorporates; as a result one has detailed knowledge of how the goals are reflected in the programs. (This information appears in Schneider, Miller, McNeal & Esty, 1990.) A natural question that arises is the degree to which regular viewers of SQUARE ONE TV are affected by the material in the series that is directed toward the goals. We pursue this after a brief description of the role of research at Children's Television Workshop.

A Background for the Study: Research at CTW

Programs are created at the workshop via the "CTW model," in which three distinct groups -- production, content and research -- are brought together to work in concert under the leadership of an executive producer whose experience is grounded in television production. As a result, research has always played an important role at CTW.

The research used in the production of SQUARE ONE TV is both formative and summative. Other researchers have used these terms in a variety of ways, but for our purposes, "formative research" is a continuing line of research conducted before and during the production of material for the series. The aim of formative research is to insure that the preferences and needs of SQUARE ONE TV's target audience are represented in the production of the series; such research provides the production and content staffs with feedback on the comprehensibility and appeal of proposed material.

By contrast, "summative research" assesses the impact of SQUARE ONE TV after production has been completed; its aim is to examine whether the series has been successful in meeting its
goals (although, naturally, this sort of research also holds
formative implications for future production). One previous
summative study is described here.

"The Comprehension and Problem-Solving Study" (Peel, Rock-
well, Esty, & Gonzer, 1987), examined children's comprehension of
a sample of ten segments taken from the first season of the
series; the segments contained a wide variety of mathematical
content, but all concerned problem solving. Comprehension was
assessed on three levels: recall of the problems and solutions
shown, understanding of the mathematical principles underlying
but not fully explicated in the segments, and extension of
concepts shown in the segments to related problems.

The study found the segments' mathematical content to be
accessible to children across the series' target age range: third
through sixth graders were able to isolate and recall both the
problems and solutions which had been shown. Also, the content
appeared to be age-appropriate. Well over half of the third
graders gave satisfactory answers to questions which assessed
their understanding of the segments' underlying mathematical
content, and performance increased to over 80% for sixth graders;
thus, it seemed that the segments are neither so difficult that
no one could understand them nor so easy that third graders
performed as well as sixth graders. Moreover, many children were
able to extend the mathematical principles used in the segments
to solve related problems that had not been shown. Finally, the
study demonstrated that viewers perceived the segments' charac-
ters to be pleased with their own competence in solving problems.

Purposes of the Present Study

The present study builds upon this previous research in
several ways. While the study described above examined problem
solving in the context of problems presented in SQUARE ONE TV,
the current study presents children with new problems to see how
the series might affect problem solving in a more general sense.
And while children in "The Comprehension and Problem-Solving
Study" were asked to assess the feelings of the characters shown
in the segments, the present study examines the changes in the
children's own attitudes that might result from viewing SQUARE
ONE TV. In this way, the current study provides a direct,
experimental test of Goals I and II; that is, it attempts to
describe the changes in children's attitudes toward mathematics
and their use of problem-solving techniques that might arise as a
result of sustained viewing of SQUARE ONE TV.

The attitude component. In assessing Goal I ("To promote
positive attitudes toward, and enthusiasm for, mathematics"), we
developed an Attitude Interview. Here we have conceived of
"attitude" as pertaining to issues of motivation, confidence,
enjoyment, perceptions of usefulness and importance, and children's conceptions of what mathematics is, i.e. their "construct" of mathematics.

The aims described under the Goal I subgoals directed our creation of the specific interview questions used. Goal IA ("Mathematics is a powerful... tool, useful to solve problems... and to increase efficiency") was related to interview questions that assessed children's construct of mathematics and the degree to which children think of mathematics as useful and important. Goal IB ("Mathematics is beautiful and aesthetically pleasing") was related to questions about the children's enjoyment of mathematics. And Goal IC ("Mathematics can be used, understood, and even invented by non-specialists") was related to questions assessing children's confidence and motivation in using mathematics. The questions were open-ended and asked with respect to three domains: the problem-solving activities that children engaged in as part of the study; problem solving in general; and mathematics in and out of school.

Because the Attitude Interview consisted of open-ended questions aimed at developing an elaborate picture of individual children's beliefs and feelings regarding mathematics, interviewers were trained to ask follow-up probe questions to draw out and reveal the full complexity of the issues children raised. Detailed coding schemes were developed through an analysis of the children's responses. This analysis is still underway; reports on results will be available later in the year.

The problem-solving component. With regard to Goal II, the study examines the impact of SQUARE ONE TV on children's problem-solving actions (particularly problem treatment and follow-up) and the extent to which they use a variety of heuristics (e.g., constructing tables or graphs, looking for patterns, or working backwards) in problem solving. Further, the study assesses the impact of the series on the mathematical completeness and sophistication of children's solutions to nonroutine problems. The remainder of this paper will concentrate on the problem-solving component of the study, since it is virtually complete.

An Overview of the Study: Methodology

Subjects. The subjects for the study were fifth graders in four public elementary schools in a mid-sized southwestern city. (This site was chosen because it is one of the few cities in the country in which SQUARE ONE TV had not been part of the regular public television broadcast schedule prior to completion of data collection. Also, none of the participating schools had shown SQUARE ONE TV as part of classroom instruction.) All the schools used the same standard mathematics textbook
and curriculum. Moreover, the four schools were matched as pairs on the basis of standardized achievement test scores, racial/ethnic composition, and student socioeconomic status (SES). One pair of schools served mostly lower-SES children, and the other two schools were largely middle SES. One school in each pair was randomly designated as an experimental (viewing) school, and the other was designated as a control (nonviewing) school.

A total of 48 children, 12 from each school, participated in the part of the experiment described here. They were drawn from all of the regular fifth-grade classrooms in the four schools. Children within matching schools were matched as pairs on gender, race/ethnicity, achievement test scores, and eligibility for free lunch (used as a further indicator of SES).

Treatment. All fifth graders in the two experimental schools were exposed to programs from Seasons I and II of SQUARE ONE TV. They watched one program each weekday for six weeks, a total of 30 half-hour programs. The viewing took place during school hours, but not during regularly scheduled mathematics classes. The teachers in the viewing schools did not alter their usual mathematics instruction in any way. They did not use SQUARE ONE TV as part of their teaching, they did not comment on it, and they did not make any connection between the program and mathematics. Thus, the experimental exposure to the series consisted of sustained unaided viewing in a group setting.

The two control schools did not see SQUARE ONE TV at all; their schedule did not change from what it usually was.

Instruments. A variety of instruments were used, aimed at assessing problem-solving performance and attitudes. The problem-solving instruments, called Problem-Solving Activities (PSAs), were a range of mathematically rich, nonroutine, situations. Each PSA allowed children to demonstrate the problem-solving actions and heuristics of Goal II and to reach solutions through a variety of approaches. The PSAs all involved a manipulative component, and they were substantively different from problems traditionally encountered in elementary school mathematics.

The PSAs comprised three levels of complexity. Level C problems (the most complex), asked children to determine what is wrong with a complicated mathematical game and to fix it. (One of these is described in much greater detail later.) Level B PSAs (moderately complex), involved sorting party guests or price tags into piles that meet several conditions. Level A PSAs (the least complex) were combinatorics problems involving circus performers or stripes on a shirt.

At both the pretest and the posttest, three PSAs were administered to individual children over two 55-minute sessions on two
successive days. (The Attitude Interview, described earlier, was also conducted during the second day.) The procedure for administering each PSA was this: The researcher described a problem situation to the child, using a written script. The child was given time to work on the problem alone. Following this activity, the researcher used a series of standard probe questions to get at what the child was thinking during the work session. Special emphasis was placed upon having the child describe and assess the choices he or she made during the problem-solving process.

To prevent experimenter bias, the interviewers were not informed of the viewing/nonviewing status of the children. As an additional safeguard, interviewers had no contact with any classroom teachers. Further, special care was taken to insure that the children made no connection between the interviewing and SQUARE ONE TV.

The schedule used to administer the PSAs is shown in Table 2, below.

<table>
<thead>
<tr>
<th>Schedule of Instrument Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Groups</strong></td>
</tr>
<tr>
<td><strong>Experimental</strong> (Viewing)</td>
</tr>
<tr>
<td><strong>Control</strong> (Nonviewing)</td>
</tr>
<tr>
<td><strong>PRETEST</strong></td>
</tr>
<tr>
<td>Day 1: Two PSAs</td>
</tr>
<tr>
<td>Day 2: One PSA</td>
</tr>
<tr>
<td>Day 1: Two PSAs</td>
</tr>
<tr>
<td>Day 2: One PSA</td>
</tr>
<tr>
<td><strong>TREATMENT</strong></td>
</tr>
<tr>
<td>View 30 programs of SQUARE ONE TV</td>
</tr>
<tr>
<td>No change from normal schedule</td>
</tr>
<tr>
<td><strong>POSTTEST</strong></td>
</tr>
<tr>
<td>Day 1: Two PSAs</td>
</tr>
<tr>
<td>Day 2: One PSA</td>
</tr>
<tr>
<td>Day 1: Two PSAs</td>
</tr>
<tr>
<td>Day 2: One PSA</td>
</tr>
</tbody>
</table>

There were two versions of each level of PSA -- A and A', B and B', C and C'. Eight months of pilot testing determined that it is possible to use two variants of each PSA while maintaining the same level of difficulty from the pretest to the posttest. In fact, t-tests on pretest data from the main study revealed no significant differences in children's performance within each pair of problems, indicating that difficulty did not vary within pairs of problems. (Henceforth the pair A and A', for example, will be referred to as A*. )
One set of PSAs (either C, B, A or C', B', A') was administered to each child at the pretest, and the other at the posttest. Within each set, the most complex problem (C or C') was used first and the least complex last. Half the children at each school used one set for the pretest and the other set for the posttest; the order was reversed for the other children.

A Description of One of the PSAs

The following is a description of one variant of the most complex of the Problem-Solving Activities, PSA C'2.

The child is told about a person named Dr. Game, who owns a game factory. Dr. Game was recently dismayed to find that his factory had been broken into, and that some of his games had been changed in some way. The child has been hired by Dr. Game to find out what is wrong with one of these games.

The experimenter shows the child the equipment for the game, which consists of:

Two spinners, divided like this:

A coin, with "+" on one side and "X" on the other;

A number board with two elasticized loops -- one orange and the other green -- arranged so that they surround two sets of numbers, like this:

Two stand-up, cut-out players, one of whom wears a sign around its neck saying "Orange" and the other with a sign saying "Green";

Nine plastic chips.

The experimenter explains the rules of the game to the child. To play the game a spinner-person (not identified further) spins both spinners, getting two numbers, and flips the coin, getting addition or multiplication. Then he does the addition

This PSA is loosely based on a lesson from the Comprehensive School Mathematics Program (CSMP, 1979, IV, pp. 49-60).
or multiplication and finds the answer on the board. If the answer is inside the green loop then the Green player gets one chip; if the answer is inside the orange loop then the Orange player gets one chip. Whoever has more chips at the end of nine spins wins the game.

After the child is again told that there is something wrong with the game and that the task is to find what is wrong, the experimenter leaves the child to work alone. A kit of materials (paper, pencils, pens, a calculator, a ruler, a protractor, and some circular stickers) is available for the child to use if he or she wants to.

[What is wrong with the game is that it is unfair to Green. The probability of awarding each chip to Orange is 3/4, and the probability of Orange's winning more chips than Green by the end of the game is greater than 0.95.]

When the child has told the experimenter what he or she thinks is wrong with the game, the experimenter asks several standardized questions that encourage the child to describe his or her actions, thoughts and strategies. Then the next task is posed: to fix the game.

[The game can be fixed (or at least made fairer than it is) in a variety of ways: by moving the orange and green loops appropriately; by changing some or all of the numbers on the spinners; by changing the operations on the coin; by awarding more than one chip to Green if the answer is in the green loop; or by some combination of these.]

Again the child is left alone to work on this. The experimenter returns to the table when summoned or if the child seems no longer to be working productively. As before, the experimenter uses a set of carefully structured probe questions to get at what the child was doing and thinking during the period he or she was working on the problem.

The Design of the Study vis-a-vis the Standards

In designing this evaluation of SQUARE ONE TV we have taken very seriously the first of the standards dealing with evaluation, namely Alignment. Alignment, according to the Standards (p. 193), has three aspects: Any evaluation should be aligned with:

1. **The goals, objectives and mathematical content of the program.** Here, two of SQUARE ONE TV's goals -- I and II -- are the specific foci of the two principal parts of the study; with regard to Goal III (on mathematical content), the content (e.g. probability, combinatorics) that is represented in the problems that the subjects worked on was part of the content that appears in the series.

2. **The relative emphases given to various topics and processes and their relationships.** Here, we consider the main "topic" of the series to be problem solving, which is a
principal focus of both the show and the evaluation. As will become clear in the next section, the correspondence between problem solving as categorized (a) in the series and (b) in the evaluation is very close.

(3) instructional approaches and activities, including the use of calculators, computers and manipulatives. Here, we note first that the PSAs of the evaluation were embedded in the sort of story context that is typical of SQUARE ONE TV; in fact, any one of them could have been the basis of a studio sketch for the program. Calculators, computers and manipulatives are used in the series whenever they are natural and helpful, and, correspondingly, a calculator is provided as just one of several tools in a kit that subjects can use or not use as they wish. Further, a variety of manipulative devices are used by the writers to illustrate concepts through the television medium, just as they might be used by teachers in classrooms.

Coding Systems

There were two coding systems used in the problem-solving part of the study to quantify the children's performance on these PSAs. One, called the P-score, ("P" for "problem solving") analyzed problem-solving processes employed. The second, called the M-score, ("M" for "mathematical") measured the mathematical completeness and sophistication of the subjects' solutions. In both cases, the coding systems used the child's verbal reports and overt behaviors as sources of evidence, rather than coders' inferences. We will discuss the P-score coding system first, and provide two examples of its application.

P-scores. This system was directly tied to the statement of SQUARE ONE TV's Goal II: The behaviors of interest were the

---

3Note that if one considers "topics" in the series from the point of view of Goal III (taking it just as a count of instances of the various Goal III subgoals, as in the analysis in Schneider, et al. (1990)), then the evaluation is not aligned -- e.g. the heaviest Goal III subgoal in the series is IIIB (arithmetic of rational numbers), but that is not the focus of any of the PSAs.

4There are no segments in SQUARE ONE TV that deal with calculator or computer use per se.

5Aside from the alignment issue, the very heavy manipulative emphasis in the presentation of the PSAs is partly to make the subjects' actions when working alone more visible to the videotaping.
problem-solving actions and heuristics described in the subgoals of Goal II, slightly modified as shown in Table 3. The chain of reasoning that led us to this system of coding is essentially this: SQUARE ONE TV segments portray characters modeling mathematical problem solving in a variety of situations. These segments are coded on the basis of the Goal II subgoals that they depict. Children who watch the show repeatedly may, as a result, be more likely to use behaviors that would be categorized under these subgoals. To see if in fact this is true, it is reasonable to use this same system of subgoals to categorize the children's behavior.

Note that one would not expect children necessarily to mimic the behavior shown by SQUARE ONE TV characters exactly. The actors are carefully scripted, and what they do is often designed to illustrate problem-solving strategies as clearly as possible. Further, two of the PSAs (B* and C*) are very different from any particular segment that was included in the six weeks of shows, and the third one is only partly similar to one of the segments. Thus, the particular behaviors that children use on the PSAs may not be exactly the same as ones that they have seen during the treatment. Nonetheless, we can use the same subgoal system to categorize the children's behavior even if that behavior is not precisely what one would see modeled on the show itself. (It should be noted, too, that we found that all the behaviors that children consistently demonstrated were categorized somewhere in the coding system.)

So, with the minor modifications portrayed in Table 3, we simply took the full statement of Goal II and used it to code the children's problem solving as if their performances were segments on the show. On the basis of our experience with pilot testing, this seemed to be a reasonable approach, since the behaviors that were elicited in the pilot test phase were ones that could be categorized according to Goal II subgoals.

For each subject on each PSA, one or more coders would look at the entire videotape and verbatim transcript of the child's behavior, including the discussion between the interviewer and the child. The coders were blind as to whether the child was a viewer or nonviewer, just as the interviewers had been. The coder would award points (a maximum of two) for each instance of the child's use of a problem-solving action or heuristic. This

---

6Note that because of the slightly different introductions to the interviews, coders were not blind to whether the interview was from the pretest or the posttest. This is of little importance, however, because the viewer/nonviewer status was unknown. Also, one child mentioned SQUARE ONE TV in a PSA A* interview, but this happened to be the last interview coded by that coder.
was uniformly done in accordance with procedures laid out in a
detailed P-Score Coding Guide.

Table 3
A Comparison of the Series' and the Study's Interpretation of Goal II

<table>
<thead>
<tr>
<th>GOAL CLASSIFICATIONS FOR</th>
<th>THE SERIES</th>
<th>THE STUDY</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIA. PROBLEM FORMULATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recog., state a prob.</td>
<td>IIA1</td>
<td>Not used at all,</td>
</tr>
<tr>
<td>Assess value of solv.</td>
<td>IIA2</td>
<td>since interviewer</td>
</tr>
<tr>
<td>Assess possib. of solv.</td>
<td>IIA3</td>
<td>posed the problems</td>
</tr>
<tr>
<td>IIB. PROBLEM TREATMENT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall info</td>
<td>IIB1</td>
<td>IIB1</td>
</tr>
<tr>
<td>Estimate, approx.</td>
<td>IIB2</td>
<td>IIB2</td>
</tr>
<tr>
<td>Gather data</td>
<td>IIB3</td>
<td>IIB3</td>
</tr>
<tr>
<td>Calc. or manipulate</td>
<td>IIB4</td>
<td>IIB4x</td>
</tr>
<tr>
<td>Consider probabilities</td>
<td>IIB5</td>
<td>IIB5</td>
</tr>
<tr>
<td>Trial &amp; err; guess &amp; chk</td>
<td>IIB6</td>
<td>IIB6</td>
</tr>
<tr>
<td>IIC. HEURISTICS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale model</td>
<td>IIC1a</td>
<td>IIC1x</td>
</tr>
<tr>
<td>Diagram</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table, chart</td>
<td>IIC1b</td>
<td>IIC1x</td>
</tr>
<tr>
<td>Graph</td>
<td>IIC1c</td>
<td>IIC1x</td>
</tr>
<tr>
<td>Use objects; act out</td>
<td>IIC1d</td>
<td>IIC1x</td>
</tr>
<tr>
<td>Transform problem</td>
<td>IIC2</td>
<td>IIC2</td>
</tr>
<tr>
<td>Look for patterns</td>
<td>IIC3a</td>
<td>IIC3x</td>
</tr>
<tr>
<td>Look for missing info</td>
<td>IIC3b</td>
<td>IIC3y</td>
</tr>
<tr>
<td>Pertinent vs extraneous</td>
<td>IIC3c</td>
<td></td>
</tr>
<tr>
<td>Change point of view</td>
<td>IIC4a</td>
<td>IIC4</td>
</tr>
<tr>
<td>Generate new hypoths.</td>
<td>IIC4b</td>
<td>IIC4</td>
</tr>
<tr>
<td>IID. PROBLEM FOLLOW-UP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reasonableness</td>
<td>IID1</td>
<td>IID1</td>
</tr>
<tr>
<td>Alternative solns.</td>
<td>IID2</td>
<td>IID2</td>
</tr>
<tr>
<td>Alt. ways to solve</td>
<td>IID3</td>
<td>IID3</td>
</tr>
<tr>
<td>Extend to rel. probes</td>
<td>IID4</td>
<td>IID4</td>
</tr>
</tbody>
</table>
Two Illustrations of the P-Score Coding

We illustrate how the P-scores are assigned by considering highly abbreviated excerpts from the interviews of two children working on PSA C', the spinner game. These are transcripts of a tape that was prepared in the summer of 1989 to indicate how the interviews were conducted. These particular children were chosen almost at random from ones who had been coded by the time the tape was produced. There is no intention here of making any implications about pretest vs posttest, experimental vs control, or any gender, ethnicity, or SES comparisons. Also, their names have been changed.

The two interviews illustrate somewhat different approaches. Paul starts by noting "numbers repeat"; he then finds that some of the numbers in the orange loop are not obtainable, and concludes (incorrectly) that the game is unfair to Orange. Jodie notes a more general pattern connected with the size of the numbers, and attempts to fix the game by moving the loops. She does not check her solution in any detail, though, and it turns out to be no better than the original partition of the numbers.

Excerpts from an Interview with Paul

R: . . . [DESCRIBES THE PROBLEM] . . . but there's still something wrong with the game, okay? Now, I'm going to go over there while you're working on this. Why don't you take a little while to think about this. Don't rush. Take as much time as you like. And if you need something or you need to ask me any questions, please let me know. Then, when you think you've found out what's wrong with the game, call me back, and I'll come back and then we can discuss what you've done and what you were thinking, okay?

S: Okay.

[RESEARCHER LEAVES]

S: [EXAMINES SPINNERS AND POINTERS. PICKS UP PLAYERS. LOOKS AT NUMBER BOARD. COUNTS CHIPS.]

S CONTINUES TO EXAMINE THE GAME PIECES FOR 3.5 MINUTES

. . .

S: [LOOKS UP, SMILES] I think I figured it out.

R: You figured it out? Great. [RESEARCHER RETURNS]
What do you think is wrong with the game?

S: It repeats numbers?
R: It repeats numbers? Okay, how do you know that?

S: Because on the spinners is four and four, five and five, and three and three, and six and two--three times.

R: Hm. Okay. Well, I wasn't here while you were doing this, and I want to know what you were doing and what you were thinking. So what did you do first, and what were you thinking?

S: First, I was looking at this. [POINTS TO NUMBER BOARD]

R: Uh huh.

IIC3 (2 pts) One feature ruled out as extraneous. Use of "probably" does not qualify as IIB5, though

S: And I thought probably--there was a number repeated but there wasn't.

R: Hm. O.

S: Then I looked at the chips.

R: Hm.

S: And then there's nothing wrong with the chips.

R: Uh huh.

S: And I looked at the spinners, and then just figured it out!

R: ... I'd like you to try something else. You've told me what you think is wrong with the game and now Dr. Game wants to hire you to fix the game. So, um again, I'm going over there while you're working on this, and ..." [R LEAVES.]

IIB4y (2 pts) Manipulates spinners so that each one has 1 through 6.

S: [PUTS STICKERS ON SPINNERS, WRITES ON THE STICKERS. THIS CONTINUES FOR 1:30 MINUTES.]

[THE INTERVIEW CONTINUES, WITH S EXPLAINING THAT WITH THE ORIGINAL SPINNERS, "YOU'RE NEVER GONNA GET 1, 2, OR 3 ON THE NUMBER BOARD"].

S: ... two and two.
Is this IIB5, using probability? Can't be sure at this point.

IIB5 (2 pts) Use of probability confirmed.

---

R: Hm hm.

S: It's gonna be four. And if you spin again, you're never gonna be able to get one, two, or three [POINTS TO NUMBERS ON NUMBER BOARD]

R: Hm hm. Okay. And how does that fix it?

S: Um, it gives both sides a chance.

R: It gives both--what do you mean by that?

S: Um, like if they spin one number--might not be left out?

R: Hm hm. One number might not be left out? What do you mean? ...

S: ... if you're spinning it and then the number that it lands on--it might not get one of those numbers, right? [POINTS TO NUMBER BOARD]

R: Hm hm. Okay.

S: So, um, that leaves the green side a better chance to win.

R: Hm hm. It leaves the green side a better chance to win? How--how so?

S: Because if it [Orange] has six numbers off, um, and it [Green] has more--how is that side going to be able to win?

R: Uh huh. Uh huh. Okay. And so--okay. Okay. It has more numbers off? Can you show me or explain that to me?

---

IIB4x (2 pts) Calculates

S: [PUTS POINTERS SO THEY POINT TOWARDS ONES] Like one times one is one. [POINTS TO ONE ON BOARD]

R: Hm hm.

S: And before it didn't have one and one, so you couldn't get times one.

R: Hm hm.

S: One and two. [PUTS ONE POINTER TOWARDS TWO] ...
Excerpts from an Interview with Jodie

R: [RESEARCHER INTRODUCES PSA C', THE SPINNER GAME] Then when you think you've found out what's wrong with the game, let me know, and I'll come back and we'll talk about it.

S: [NODS TO SELF. SMILES SLIGHTLY TO THE CAMERA. SPINS POINTERS. DROPS COIN. AWARDS CHIP TO APPROPRIATE PLAYER.] Times.

[CONTINUES TO PLAY GAME FOR 3:10 MINUTES.]

[S CALLS RESEARCHER BACK.]

S: I figured out a catch to it. The orange guy's, though, is always gonna win because he has the lower numbers. [POINTS TO NUMBER BOARD] And every time you have a plus, it's always going to go in the lower numbers. [POINTS TO NUMBER BOARD AGAIN]

R: Okay. Say more about that.

S: Okay. If you have plus, 'cause, see, if you plus six [POINTS TO SIXES ON SPINNERS], it's only gonna be twelve. And the highest number is fifteen [POINTS TO FIFTEEN ON NUMBER BOARD], so if you always get plus, the orange guy's always gonna win. And you'd have to get a real high times number for the green guy to win.

R: Hm Hm. Okay. And so, one more time: so, what's wrong with the game is . . . ?

S: That -- that -- that the orange numbers are always gonna be-- that the orange guy is almost always gonna win. He has a better chance to win.

R: Okay. And how do you know that?

S: I figure that out because I kept-- every time I got times, it would go higher, and sometimes it'd go lower. But every time I got plus, it'd always go to the . . . .
IIB4y (2 pts)
Describes manipulation of the equipment

-- you took it around that direction. And then the orange one was to go this way. See? You'd have the lower numbers and the higher numbers [in each half of the board].

R: Why don't you show me. Show me what you mean.

S: Okay. [PICKS UP BOARD] Like, here's your orange one. And you put the orange one around these right here [RUNS FINGER ALONG SUGGESTED BOUNDARY], and back around here. And then you took your green one, and you--and you put it like this. Around here. See? How you have just this square on each of the same amount, and the you have the higher and the lower number.

R: Okay, so it would be like--

S: Down the middle.

R: This half [DIVIDES IN HALF WITH FOREARM TO DEMONSTRATE] --one, two, three, seven, eight, nine
would be orange?

S: Yeah.

R: And this side four, five, six, ten, eleven, twelve, would be green?

S: Yes.

R: Okay. And how would that fix the game?

S: Because you have the same amount of lower numbers and higher numbers for each color. And then that means if you get plus, you can get on this side, too. You don't always have it on the top.

R: Okay, if you have plus, you can get 'em--

S: If you have plus [PUTS HAND ON NUMBER BOARD], you can always get it. You have lower numbers that you could put it on. [MOTIONS TOWARDS BOARD]

R: Was there anything as you were thinking that you--you considered and then you thought, "No, that's not very helpful."

S: I thought about the numbers again. [POINTS TO SPINNERS]

R: What did you think about the numbers?

S: I thought that um maybe if you--if you um--made them higher numbers, and you made--you some real low like this [POINTS TO 2 ON SPINNER], and then you made some a little bit higher [POINTS TO 6 ON SPINNER], so they also help in this. [POINTS TO BOARD] Like, instead, like a six maybe, like a nine, or a ten or something.

R: Okay, and how would that help?

S: It's because when you add 'em [TURNS COIN OVER], like if you had a ten [POINTS TO ONE SPINNER], and you have a ten on this side [POINTS TO OTHER SPINNER], you have a twenty, and it also be able to go to the high side [POINTS TO NUMBER BOARD], the green side.

R: That's if--if you would--if you would leave it just like this. [POINTS TO NUMBER BOARD] If you'd leave the loop just like that.
M-scores. The M-score is a measure of the child's mathematical success with a PSA. It is derived from two sources: (1) a mathematical analysis of that the particular PSA involves, and (2) the range of mathematical ideas that subjects in the pilot testing phase expressed. Coding is again based on examination of the videotape and transcript, with guidance provided by a detailed M-score Coding Guide.

The M-score is unlike the P-score in several respects. First, since the M-score is designed to reflect how far the child got with the PSA, it takes into account only the child's most advanced, final thinking on the PSA. Unlike the P-score, it mirrors the ultimate destination of the problem solving, rather than the actions and heuristics used along the way. Second, the M-score scheme does not apply generally to all the PSAs because they differ from each other mathematically. Thus there is one set of scores for PSA A*, another for PSA B*, and a third one for PSA C*. Third, the M-score is not open-ended: a top score is awarded for a full and complete solution to the PSA, and these maximum scores are different for PSAs A*, B*, and C*.

The M-score coding scheme for each PSA is organized in a hierarchical fashion, with various numbers of points assigned to components that could be part of a child's solution to the PSA. For example, PSA C' involves nine different levels (six for determining what is wrong with the game, and three for fixing it). The "What's Wrong?" part involves a score for (1) what the subject thinks is wrong; (2) the reasons he or she gives for that being wrong; (3) a justification for asserting that the probability that the Orange player wins a chip is greater than one half; (4) the level of completeness and systematicity of the analysis of the Orange's advantage; (5) the equiprobability of the spinner numbers and operators; and (6) the relation between winning a single chip and winning the whole game. The second task, fixing the game, involved scores for: (7) changes to make the game fairer; (8) justifications for those changes; and (9) verific-

7There is an element of arbitrariness in setting these maximum scores, since for some people a "complete" solution may require, for example, generalizations of a solution to additional cases. The solutions that warranted maximum scores for all the PSAs, however, were set at a level that matched or exceeded every child's sophistication.
tion that the changes did in fact make the game fairer.

Within each of these levels, there may be several statements or methods that would earn various numbers of points. In some cases, different statements within a level are awarded the same score. (Of course many children do not progress much beyond level (3) in the "What's Wrong?" part of the PSA, nor beyond level (7) in the "Fix It" part.) As an example, there were essentially four different ways to fix the game mentioned earlier; each of those ways, perhaps in combination, could be pursued to make the game fairer, and each might result in the same total M-score.

A relation between P-scores and M-scores. Note that the P-scores and the M-scores are conceptually independent in the sense that a child's use of a large number of problem-solving actions or heuristics would not necessarily lead to a sophisticated or complete solution; and, conversely, a sophisticated and complete solution might be obtained despite a child's use of a very limited problem-solving repertoire. Generally speaking, though, one would expect that a greater use of problem-solving actions and heuristics would lead to more complete and sophisticated solutions. After all, this is what leads us to use the adjective "problem-solving" in the first place.

An Overview of Some Results

The principal results of the study can be summarized as follows:

- From pretest to posttest, children in the viewing group made significantly greater P-score gains on each of the three PSAs than the nonviewers did. (Two-way ANOVAs showed interactions of pre/post with viewer/nonviewer to be significant at the p < .001 level for PSAs A*, B*, and C*.) The viewers' pretest to posttest gains were significant (p < .001 for PSA A* and C*; p < .01 for PSA B*); the nonviewers did not make significant gains. Further, in each PSA there was a significant P-score difference at the posttest between the viewers and nonviewers (p < .001 in each case).

Figure 1 shows the combined mean P-scores of the two groups at the pretest and posttest, with an interval of one standard

---

8The M-scores for Paul and Jodie were 8 and 13, respectively; the maximum possible score (which no child attained) was 26.

9The pairwise correlations among P-scores for PSAs A*, B* and C* were all positive and significant at the p < .01 level, or even more significant. The combining of P-scores was done via principal components analysis.
deviation above and below each mean.

It is clear from Figure 1 that there is substantial overlap between the viewers' and nonviewers' P-scores at the pretest. At the posttest, however, the viewers' P-scores increased significantly, while the nonviewers' did not. At the posttest, then, there was much less overlap between the two groups.

**Figure 1**
Mean P-scores (all PSAs combined) for viewers and nonviewers on pre- and posttest, with 1 SD above and below the means

---

From pretest to posttest, children in the viewing group made significantly greater M-score gains than nonviewers on two of the three PSAs. (Two-way ANOVAs showed interactions of pre/post with viewer/nonviewer on PSA A* (p < .01) and PSA C* (p < .001).) From pretest to posttest, the viewers' M-scores increased significantly on PSAs A* and C* (p < .001). Further, the

---

10 The decline in the nonviewers' mean combined P-score is marginally significant (p < .10). The nonviewers' P-scores declined significantly in PSAs A* and B* (p < .01), but not in C*. 

21
difference between the two groups at the posttest was significant in PSA C* ($p < .001$) and marginally significant in PSA A* ($p < .10$).\textsuperscript{11}

Figure 2 shows the mean total\textsuperscript{12} M-scores of the two groups at the pretest and the posttest, with an interval of one standard deviation above and below each mean.

\textbf{Figure 2}

Mean M-scores (all PSAs combined) for viewers and nonviewers on pre- and posttest, with 1 SD above and below the mean

---

\textsuperscript{11}M-score changes in PSA B* were not significant for either group. Something akin to a ceiling effect appeared to be operating in the sophistication and completeness of children's solutions to this problem at both pretest and posttest. Thus there was little change from pretest to posttest.

\textsuperscript{12}For summary purposes only, the M-scores from the three PSAs were combined simply by adding them. The correlations among the M-scores for PSAs A*, B* and C* were not all significant, so any combination of M-scores across the three PSAs should be interpreted with caution.
The same pattern observed for P-scores is apparent here: At the pretest there is substantial overlap between the two groups. However, at the posttest the viewing group's M-scores were significantly higher, resulting in much less overlap. The nonviewers's M-scores did not change significantly from pretest to posttest on any of the PSAs.

Even though the P- and M-scores are conceptually independent, in this sample they were significantly correlated ($r = .52; p < .001$); higher P-scores tended to be associated with higher M-scores. For reasons detailed in a full report on the study (Hall, Esty, Fisch, et al., in preparation) we hypothesize that there is a causal relationship between the P-scores and M-scores: an increase in P-score (a greater tendency to use problem-solving actions and heuristics) leads to an increase in M-score (mathematical sophistication and completeness of solution).

There were no significant gender differences in children's M-scores at either the pretest or the posttest. Further, the changes in children's M-score performance from pretest to posttest did not interact significantly with their gender.

Similarly, gender did not have a significant main effect on children's P-scores. Both boys and girls who watched SQUARE ONE TV improved significantly ($p < .01$) from pretest to posttest, and there was no difference between boys and girls in the viewing group at either the pretest or the posttest. Thus, it appears that SQUARE ONE TV had a similar effect on the boys and girls in the viewing group.

Middle-SES children received higher P-scores than low-SES children did ($p < .01$), and higher M-scores on two of the three PSAs ($p < .01$ for PSA A*; $p < .05$ for PSA C*). But, as in the case of gender, the changes in children's P-scores and M-scores did not interact significantly with SES, indicating that SQUARE ONE TV exerted a similar effect on the low- and middle-SES children in this sample.

In this study minority (i.e. African-American and Latino) children were largely of lower SES, and nonminority (i.e. Anglo) children were of middle SES. Thus a pattern similar to the one found for SES emerged when the data were analyzed by ethnicity. That is, nonminority children received higher P-scores than minority children ($p < .05$), and marginally higher M-scores in PSA C* ($p < .10$), but there was no significant interaction between SES and P-scores or M-scores. This indicates that SQUARE ONE TV affected minority and nonminority children similarly.

There was, however, a marginally significant ($p < .10$) three-way interaction among gender, condition and pretest/posttest; this was attributable to a drop ($p < .05$) from pretest to posttest in the nonviewing girls' P-scores.
Ten months before the study started, school district personnel administered an annual standardized mathematics achievement test to all fifth graders in the district. The children's scores on this achievement test were not significantly correlated with their P-scores or M-scores on any of the PSAs. (The correlations between P-scores and standardized test scores range from -.18 to .11; the correlations between M-scores and standardized test scores range from -.07 to .02.) Further, there were no significant correlations between scores on standardized mathematics tests and pre- to posttest changes in P-scores or M-scores. (Of course this research says nothing about how viewing SQUARE ONE TV might affect standardized test scores!)

A set of more detailed analyses was carried out to explain the sources of the viewers' significantly increased P-scores. Some of the results, briefly, are these:

- A large percentage (an average of 42%) of the problem-solving actions and heuristics that the viewers used in the posttest were new, i.e. actions that they had not used in the pretest. This proportion for viewers was significantly larger than the average of 25% observed for nonviewers (p < .005 for PSAs A* and C*; p < .10 for PSA B*).

- For each of the PSAs we tallied the number of problem-solving actions and heuristics for which there was an increased use from pretest to posttest. Averaged over the three PSAs, viewers increased in their use of 11.7 of the 17 actions and heuristics, while nonviewers increased in only 4.0 of the 17.

- A more fine-grained study was undertaken of the relationship between the representation of specific Goal II subgoals in the treatment and viewers' subsequent use of particular problem-solving actions or heuristics. The situation here is complex because children's use of specific problem-solving actions or heuristics is a function of at least three factors: (1) the influence of SQUARE ONE TV, (2) what the children would bring to the problem normally, without any influence from SQUARE ONE TV, and (3) the kinds of behavior that would be appropriate to use on the particular problem. As a result, generalizations are difficult to make in this area. However, in many cases viewers (more than nonviewers) used particular problem-solving techniques that were especially appropriate or powerful in their solutions of certain PSAs.

It is interesting to speculate, more generally, on what it is about SQUARE ONE TV that produces such an effect on children's problem solving. We know that the subjects were exposed to a considerable amount of problem-solving material over the course of the six weeks of treatment. The 30 programs they watched included a total of 201 segments, of which 116 explicitly involve
some kind of problem solving. These include more than 500 instances of one of the subgoals from problem treatment, heuristics, or problem follow-up, an average of about 17 per day. All of this is viewed in a zany, humorous, context that we know is highly appealing to many children of this age. Are children who watch this material learning new problem-solving techniques that were not in their repertoires previously? Or are they recogniz- ing for the first time that techniques they already knew about are in fact applicable? Or are they becoming more motivated to use techniques that they already recognized as being useful? Analysis of the attitude data that is now under way, in conjunc- tion with the problem-solving data, will help to sort out the mechanisms that underlie the effects of the program on problem solving.

Some Summary Comments

We have discussed SQUARE ONE TV, the Problem-Solving Activi- ties, and NCTM's Standards, against a background of current school mathematics. A diagram of links among them might look like this:

![Diagram](image)

**Figure 3**

Link #1, between the Standards and SQUARE ONE TV, was discussed earlier when we showed the consonance between the goals of the television program and those espoused in the Standards.

There are two links between SQUARE ONE TV and the PSAs. Link #2 in Figure 2 is the one of alignment, as described in the Standards. The other, #3, is an arrow from SQUARE ONE TV to the PSAs; this stands for the statistically significant effects that
sustained viewing of the series has on problem-solving performance.

Link #4 is intended to signify a bond between the PSAs and the Standards in the sense that the PSAs appear to be good examples of the kinds of assessment instruments that the Standards advocates, particularly for instructional feedback and program evaluation. (See especially Table 3.1 on pp. 200-1 of the Standards, which lists characteristics of types of assessments. Many of these characteristics are descriptive of the PSAs.)

The other three links are drawn in a different font, to indicate that the relationship is one of dissonance, not consonance. Link #5 reminds one that SQUARE ONE TV presents mathematical content and an emphasis on problem solving that is not found in the typical elementary mathematics program.

Link #6 recalls that activities similar to the PSAs are not part of the ordinary elementary school mathematics curriculum either. Rarely are children given opportunities to work on open-ended, nonroutine, problems that allow many approaches and have more than one solution.

Finally, link #7 reminds us of the disparity between school mathematics as currently practiced and the vision for improvement that the Standards provides.

References:


GOAL I. To promote positive attitudes toward, and enthusiasm for, mathematics by showing:

A. Mathematics is a powerful and widely applicable tool useful to solve problems, to illustrate concepts, and to increase efficiency.

B. Mathematics is beautiful and aesthetically pleasing.

C. Mathematics can be understood, used, and even invented, by non-specialists.

GOAL II. To encourage the use and application of problem-solving processes by modeling:

A. Problem Formulation
   1. Recognize and state a problem.
   2. Assess the value of solving a problem.
   3. Assess the possibility of solving a problem.

B. Problem Treatment
   1. Recall information.
   2. Estimate or approximate.
   3. Measure, gather data or check resources.
   4. Calculate or manipulate (mentally or physically).
   5. Consider probabilities.
   6. Use trial-and-error or guess-and-check.

C. Problem-Solving Heuristics
   1. Represent problem: scale model, drawing, map, picture, diagram, gadget, table, chart, graph; use object, act out.
   2. Transform problem: reword, clarify, simplify; find subgoals, subproblems, work backwards.
   3. Look for: patterns; missing information; distinctions in kind of information (permanent or extraneous).
   4. Reapproach problem: change point of view; reevaluate assumptions; generate new hypotheses.

D. Problem Follow-up
   1. Discuss reasonableness of results and precision of results.
   2. Look for alternative solutions.
   3. Look for alternative ways to solve.
   4. Look for, or extend to, related problems.

GOAL III. To present sound mathematical content in an interesting, accessible, and meaningful manner by exploring:

A. Numbers and Counting
   1. Whole numbers.
   2. Numeration: role and meaning of digits in whole numbers (place value); Roman numerals; palindromes; other bases.
   3. Rational numbers: interpretations of fractions as numbers, ratios, parts of a whole or of a set.
   4. Decimal notation: role and meaning of digits in decimal numeration.
   5. Percents: uses; link to decimals and fractions.
   6. Negative numbers: uses; relation to subtraction.

B. Arithmetic of Rational Numbers
   1. Basic operations: addition, subtraction, division, multiplication, exponents; when and how to use operations.
   2. Structure: primes, factors, and multiples.
   3. Number theory: modular arithmetic (including parity); Diophantine equations; Fibonacci sequence; Pascal's triangle.
   4. Approximation: rounding; bounds; approximate calculation; interpolation and extrapolation; estimation.
   5. Ratios: use of ratios, rates, and proportions; relation to division; golden section.
C. Measurement

1. Units: systems (English, metric, non-standard); importance of standard units.
2. Spatial: length, area, volume, perimeter, and surface area.
3. Approximate nature: exact versus approximate, i.e., counting versus measuring; calculation with approximations; margin of error; propagation of error; estimation.
4. Additivity.

D. Numerical Functions and Relations

1. Relations: order, inequalities, subset relations, additivity, infinite sets.
2. Functions: linear, quadratic, exponential; rules, patterns.
3. Equations: solution techniques (e.g., manipulation, guess-and-test); missing addend and factor; relation to construction of numbers.
4. Formulas: interpretation and evaluation; algebra as generalized arithmetic.

E. Combinatorics and Counting Techniques

1. Multiplication principle and decomposition.
2. Pigeonhole principle.
3. Systematic enumeration of cases.

F. Statistics and Probability

1. Basic quantification: counting; representation by rational numbers.
2. Derived measures: average, median, range.
5. Data processing: collection and analysis.

G. Geometry

1. Dimensionality: one, two, three, and four dimensions.
2. Rigid transformations: transformations in two and three dimensions; rotations, reflections, and translations; symmetry.
3. Tessellations: covering the plane and bounded regions; kaleidoscopes; role of symmetry; other surfaces.
5. Perspective: rudiments of drawing in perspective; representation of three-dimensional objects in two dimensions.
6. Geometrical objects: recognition; relations among; constructions; patterns.
7. Topological mappings and properties: invariants.