exploration, experimentation, observation, and other problem-solving heuristics.

3) The presence and importance of female characters was increased. The Season I evaluation revealed that boys in both the viewing and non-viewing conditions outperformed girls on the comprehension tasks on two of the four shows. Thus, to help support girls reasoning and interest, the visibility of the female characters was increased in Season II. The shows highlighted the friendship and problem-solving collaboration between Cro and Pakka, a young female mammoth, and made Dr. C, the 20th century female scientist, a more active role model. The episodes also featured the female characters more often in problem-solving roles and introduced a new female human character, Sooli.

These three changes were designed to further facilitate children's comprehension of the processes and concepts of science and technology, and to further support their interest in science and technology. By providing more positive female role models, the new shows also aimed at enhancing girls' interest.

SEASON II EVALUATION

The findings and recommendations from the first evaluation raised several interesting questions regarding the impact of *CRO* on children's understanding of and interest in science and technology. The revisions to the show for Season II were designed to enhance children's comprehension and to further engage children, especially girls, in science and technology. As a result, CTW commissioned and collaborated in the design of a Season II evaluation to further investigate the impact of *CRO* on the comprehension of and interest in *CRO* related topics. Whereas Season I's evaluation focused primarily on the effects of *CRO* on children's comprehension, Season II's evaluation primarily focuses on the impact of *CRO* on children's interest in science and technology. Due to the link between interest and knowledge (Fink, 1991; Renninger, 1992), there is also a secondary focus on children's understanding of the scientific concepts introduced in some of the shows.

In keeping with the goals of the series, and building off the findings from the first season evaluation, the study had three major goals:

- To assess the series' impact on children's interest in the science and technology topics presented in Season II's shows.
- To examine the series' impact on children's broader interest in science and technology topics and activities that were not presented in Season II's shows .
- To assess the series' impact on children's understanding of the science and technology principles presented in each show.

METHOD

DESIGN OF STUDY

The study employed an Experimental versus Control group pre/post test design. Children assigned to the Experimental group viewed Season II episodes of *CRO* and children assigned to the Control group viewed *Where on Earth Is Carmen Sandiego.* This show was chosen as the Control show because it matched *CRO* on three critical dimensions: 1) It has the same target age group; 2) It is fully animated; and 3) It has educational goals. The critical difference between *CRO* and *Where on Earth Is Carmen Sandiego* is subject content. Whereas *CRO* focuses on science and technology, *Where on Earth Is Carmen Sandiego* focuses on geography. Furthermore, the popularity of *Where on Earth Is Carmen Sandiego*, which at the time of this study was among the top 20 Saturday morning children's programs (Nielsen Media Research, 1994), provides a strong comparison for the appeal of *CRO*.

Children assigned to the Experimental condition (henceforth referred to as the *CRO*-group) watched all eight Season II episodes of *CRO* and children assigned to the Control condition watched eight episodes of *Where on Earth Is Carmen Sandiego* (henceforth referred to as the *CSD*-group). The effect of condition was assessed by comparing: 1) show appeal ratings, 2) pretest/posttest ratings on interest scales, 3) activity preferences, 4) show recall, and 5) comprehension scores. (All measures are described in detail in the Material section.)

PARTICIPANTS

One hundred children (51 Boys and 49 Girls), ranging from 5 to 10 years (mean age = 7.8 years) participated in the study. Children were recruited from three after-school programs in Pittsburgh, PA. Within each site, children were assigned to either the *CRO*-group or the *CSD*-group. Ninety percent of the children were African American, 9% were

Caucasian, and 1% were Asian. Most children were from lower SES (Socio-Economic Status) families as indicated by participation in the state lunch assistance program: 72% received full assistance and 12% received partial assistance. Ethnic and gender composition was similar across sites.

Children were divided into three age groups: Young -- children in kindergarten through grade 2 (3 five-year-olds were in each condition, one kindergartner per condition); Middle -- children in third grade; and Old -- children in fourth grade or higher. For children who had either skipped a grade(s) or been held back (approximately 10% of the sample), we used a combination of their age and grade to determine their age-group assignment (e.g., a 9-year-old in grade 2 would be assigned to the Middle group). Table 1 shows the mean ages and number of children (shown in parentheses) in each condition by age group and gender. No significant differences or interactions were found for age by condition or gender, $p > .1$.

AGE GROUP	**CONDITION**			
	Experimental (*CRO*-Group)		**Control (*CSD*-Group)**	
	Boys (n)	**Girls (n)**	**Boys (n)**	**Girls (n)**
Young	6.5 (11)	6.5 (10)	6.5 (10)	6.1 (9)
Middle	8.4 (8)	8.0 (10)	8.6 (8)	8.0 (8)
Old	10 (6)	9.3 (6)	9.6 (8)	9.2 (6)
Number of Children	**(25)**	**(26)**	**(26)**	**(23)**

Table 1. Mean Ages of Children by Condition & Gender

SITES

Pittsburgh was chosen as the research site because *CRO* is not currently broadcast in that

area and so all children would have similar and minimal prior exposure to the show. To maintain a reasonably high level of validity, we chose to conduct the evaluation in after-school programs rather than during class time. In these after-school programs children could interact freely, choose among a variety of activities and games, and occasionally view entertainment videos. Thus the evaluation study was placed in a play-like context that was more comparable to the home viewing context than an in-class context, where the show might have high novelty and high appeal in comparison to the regularly scheduled school activities.

STIMULI

Children viewed either eight *CRO* episodes taken from Season II (see Table 2) or eight episodes of the animated version of *Where On Earth Is Carmen Sandiego* (Fox Network), which focuses on geography along with other topics from the arts and sciences (see Table 3).

Episode Name	Device & Science/Technology Content
Here's Lookin' At You, Cro!	Mirrors & Periscopes: Light & Reflection
Just a Stone's Throw Away	Catapults: Collection & Storage of Energy
It's Snow Problem	Windmills: Wind Power, Mechanical Advantage
Turn Up The Heat	Hot Shower: Heat Transfer, Conductors & Insulators
Pakka's Cool Invention	Cooling Fan: Wheels & Belts, Waterwheels
No Time For Steamer	Timekeepers: Measurement of Time
The Legend of The Big Thing	Traps & Triggers: Mechanics of Snares & Traps
Laugh, Mammoth, Laugh	Flotation Device: Flotation & Buoyancy

Table 2. Viewing Order of the eight *CRO* Episodes with Content Descriptions

Episode Name	Theme & Locations
Music To My Ears	Music: Sydney, Australia; Vienna, Austria; Bonn, Germany
A Higher Calling	UFO's: Stonehenge, UK; Palenque, Mexico; Easter Isl., Chile
Stolen Smile	Art: Amsterdam, Holland; Giza, Egypt; The Louvre, Paris
By A Whisker	Cats: London, UK; Siberia, Russia; Gobi Desert, Mongolia
Dinosaur Delirium	Dinosaurs & DNA: Agra, India; Mount Everest, Nepal
Rules of the Game	Chess: St. Petersburg, Russia; Bavaria, Germany; Xiían, China
Chapter and Verse	Children's Literature: Copenhagen, Denmark, London, UK
The Good Old Bad Days	Low-Tech Challenge[1]: Tokyo, Japan; San Francisco, CA

Table 3. Viewing Order of the eight *CSD* Episodes with Content Descriptions

DESIGN OF MEASURES & MATERIALS

Theories suggest that interest in an activity results from the combination of both attraction to the activity and sufficient knowledge to support engagement with the activity (Fink, 1991; Renninger, 1992). In other words, although children may be attracted to certain activities about which they know nothing, a certain level of background knowledge is needed for an enduring interest in the activity to develop. Thus *CRO*, by introducing children to concepts in science and technology, and presenting them in appealing contexts that support their value (i.e., solving everyday problems and overcoming obstacles), is designed to enhance children's

[1] The first seven episodes of the *Carmen Sandiego* link the geography content with a secondary educational topic from the arts and sciences. The eighth episode is unusual in that the storyline is built around a challenge to commit a crime without using any high-tech gadgets that are often Carmen's trademark.

interest in these topics. This theoretical framework motivated the design of a battery of measures to examine the impact of *CRO* on children's interest and knowledge, and the relationship between them.

Baseline Measures

Science Achievement Assessment.

Children were given a 40- item multiple choice test to measure their familiarity with and understanding of various science and technology principles (see Appendix A). This test was adapted from the Michigan Educational Assessment Program (Michigan State Board of Education, 1994), which was designed to assess starting fifth grade students' acquisition of the conceptual material covered in the first through fourth grade science curriculum. Thus the test contained items that ranged in difficulty and familiarity for children in our target age group.

Hands-on Activities and Reading Interest Assessment

This measure was used prior to treatment to establish children's initial attraction to items that were to be used in our behavioral measures (to be described below) and to ensure that children in both viewing conditions had similar levels of interest in the various activities -- both technology-related and unrelated -- prior to treatment. Children were given a booklet containing a list of book titles with a picture of each book cover, pictures of objects made from LEGO DACTA building kits, pictures of pattern puzzles, and pictures from the boxes of computer software games. Children were read a brief description of each item and were asked to indicate their interest in each hands-on activity and book using a four-point scale (1 = Definitely Not to 4 = Definitely Yes) (see Appendix B).

Interest Measures

There has been a long and considerable history of research on children's interest towards science. However, there have been three major criticisms of this work: 1) a lack of consistency

in the use of the concept of "interest," 2) a lack of consistency in the concept of "science," and 3) inadequate assessment instruments. Researchers have often used the term "interest" when in fact they were measuring different affective factors such as attitude toward science or experience with science activities. These measures may tap into components of interest (e.g., attitude shares some aspects of appeal) but may not reflect both the affective and cognitive components. Furthermore, some have used other terms when in fact they are measuring interest (Schiefele, Krapp, & Winteler, 1992). The inconsistent use of these terms makes it difficult to analyze the relationship between interest and achievement.

The second criticism concerns the lack of clarity and consistency of what is meant by "science." Researchers have used a wide range of constructs, often mixing them within a single study. For example, questions about the processes or procedures of scientific reasoning, the content or mode of instruction in science class, particular extra-curricular activities, or science as a profession or human service or societal benefit, have all been used to assess science interest (Kyrnowsky, 1987). Additional problems with these measures concern the explicit use of the term "science"; it is likely that subjects who complete interest scales interpret questions using constructs of science that are different from those intended by researchers. This is consistent with past research that has shown that children -- and even college students -- do not always hold clear and accurate conceptions of what "science" is (e.g., Crelinsten, de Boerr, & Aikenhead, 1991; Fisch, Yotive, McCann, Garner, Chen, & Ozaeta, in preparation; Fleming, 1988).

Since children in our target age group can often have a limited or biased interpretation of what "science" means -- viewing science as what they do in their classroom or as matching stereotypes (often negative) from the media -- we decided not to use the word "science" in the show or in the interest measures. Instead we operationalized interest as a preference for <u>specific</u> activities and/or topics, so that children would have a clear understanding of what we were asking them about. Furthermore, using specific activities or topics is consistent with the

conceptualization of interest as a content-specific motivational characteristic (cf., Dewey, 1913; Schiefele, Krapp, Prenzel, Heiland, & Kasten, 1983). For example, children may express an interest in earthquakes, which is content-specific, but show little interest in earth science, which is more content-general and susceptible to biased interpretations of "science."

The final concern has been the quality of the measurement instruments themselves. Most assessments of interest have relied on paper-and-pencil measures, such as self-reports, rating scales, and sentence completion tasks. Paper-and-pencil assessments provide fairly limited data as compared to in-depth interviews and observations and in most studies there has not been any indication of how or if these kinds of assessments correlate with behavior and other more in-depth measures. Furthermore, according to some theoretical views, individuals can be reflectively unaware of their interest, and thus used alone measures that ask individuals to explicitly identify their interest can be unreliable (Renninger, 1992).

To provide a richer, more reliable measure of interest, and to investigate the relationship between interest and behavior, we employed a multi-method approach, using a combination of paper-and-pencil, interview, and observational measures. Using multiple methods not only provides checks and balances for each method, it also provides a richer picture of how *CRO* may impact on children because each method has its own unique strengths and weaknesses. Paper-and-pencil measures provide straightforward, quantitative comparisons which in-depth interviews can then support and enrich with the child's detailed explanations and examples. Observational measures can provide additional converging evidence on children's interest, and can also be used to measure interest that may be below explicit awareness. Because impact on behavior is more difficult to attain, behavioral observations provide a stronger test of the impact of *CRO* on children's interest than the self-report measures. However, all these approaches taken together provide a broader, richer, and more reliable picture of how *CRO* affects children.

The multi-method approach was particularly useful in examining the impact of children's

interest in the eight topics covered in the Season II shows. The topics and activities that appeared in the show often paralleled ones that children rated in the paper-and-pencil measures and that they had the opportunity to play with or read about during the activity periods. For example, children's interest in catapults -- a topic introduced in one of the Season II shows -- was assessed using all the different methods: interest in making and learning about catapults was assessed in a paper-and-pencil measure, children were given the opportunity to build, read about, and play with a computer game about catapults during activity periods, and children were asked if they engaged in any activities outside of the activity period relating to catapults.

In the Season I evaluation, interest measures were only administered after the treatment. The present study utilizes a pretest-posttest design, comparing each child's ratings before viewing the shows to their ratings after viewing the shows. This design provides us with an additional level of control, in that pretest measures of the attraction, or value of each activity, can then be compared to children's actual behavior and posttest ratings, and used to infer the impact of the show on children's interest.

Paper-and-Pencil Assessments

The paper-and-pencil instruments developed for the evaluation avoid relying on the word "science" by assessing children's preferences for specific activities or topics related to the episodes that children viewed, and compare them to children's preferences for equally well-specified activities or topics that are unrelated to events that occurred in the episodes. Some of these measures focus on children's interest in learning about certain topics whereas others focus on children's interest in doing specific activities related to those topics. These two categories of behavior, learning about vs. doing, were included in the assessments because research has suggested that children rate items higher when they involve physical activity (e.g., building a catapult) as opposed to when they involve cognitive activity (e.g., reading about how catapults work) and because girls and boys sometimes demonstrate their interest in

different ways (Renninger, 1992).

There were four paper-and-pencil measures designed to assess children's interest in science and technology topics and/or activities: Show Appeal Ratings, TV Show Interest, Learning and Doing Interest, and Club Interest. The latter three measures focused on children's interest in three types of topics or activities: science and technology topics or activities introduced in the Season II *CRO* episodes, science and technology topics or activities not presented in the Season II shows, and non-science and technology topics or activities.[2] These three categories of items allowed us to assess whether *CRO* only impacted on children's interest in and attraction to those topics and activities that were introduced in the episodes or whether *CRO* also affected children's interest in science and technology topics or activities not introduced in Season II shows. Each measure examined a different way in which a child might express an attraction to or interest in science and technology, such as reading about a science topic, watching a science television show, doing hands-on activities, or joining clubs. None of the paper-and-pencil items measured children's interest in the topics or activities introduced in *Where on Earth Is Carmen Sandiego*; however, in one measure children were asked to rate their interest in viewing this and several other shows.

The interest measures used a Likert-type response scale with each point on the scale represented by a verbal descriptor; three of the measures (Show Appeal Ratings, TV Show Interest, and Club Interest) also included a picture of a face that reflected the attitude associated

---

[2] In designating an item as either technology or non-technology, we made a distinction between items that treat technology as a goal in itself (e.g., learning about how catapults work) and items that treat technology as a tool in service of another goal (e.g., using a camera to take pictures). Topics or activities that fit the former were considered "technology" and those that fit the latter were considered "non-technology". Without this distinction, virtually any modern topic or activity could be considered technology-related in some way.

with each descriptor. The Show Appeal Ratings scale was presented after each of the eight episodes. The other questionnaires -- Club Interest, Learning & Doing Interest, and TV Interest -- were administered prior to and after the treatment. (See Appendix C.)

Show Appeal Ratings

After viewing each episode of *CRO* or *CSD*, children were given a one page rating sheet and were asked to rate the show on a five-point scale, where 1 = Terrible and 5 = Great. Points on the scale were represented by a verbal label and a face that reflected the label. For example, a rating of 5 was associated with a smiling face and the label "great." (See Appendix C.1.)

TV Show Interest

The TV Show Interest measure provided a concrete set of items to assess the impact of *CRO* on attraction to and interest in science. Children were read a brief description of a television program and were asked to indicate how often they had seen the show (1 = Never to 4 = More than 20 times) and their interest in seeing the show (1 = Definitely Not to 4 = Definitely Yes). There were three science-related shows (*CRO*, *Bill Nye the Science Guy,* and *Beakman's World)* and 10 control shows (*Where on Earth Is Carmen Sandiego*, *The Little Mermaid*, *The X-Men*, *Mighty Morphin' Power Rangers*, *CBS Storybreak*, *Brains and Brawn*, *Bobby's World*, *Conan & the Young Warriors*, *Dog City*, and *California Dreams)* (Appendix C.2.)

Learning & Doing Interest

Each page of the questionnaire consisted of two pictures. The question associated with the top picture assessed children's interest in learning about a topic. These picture showed a child

reading a book. A "thought bubble" over the child's head illustrated what the child was learning by showing that child engaged in an activity. A caption under the picture described what the child was learning about; for example, "This child is learning about catapults." The questions associated with the bottom picture on each page assessed children's interest in doing an activity. These pictures were identical to the pictures displayed in the "thought bubbles." However, the captions below these pictures described what the child was doing; for example, "This child is making a catapult." Each page focused on an activity that was linked to a different topic. Twelve topics and related activities were included in the questionnaire: eight *CRO*- related Technology (Mirrors & Periscopes, Catapults, Windmills, Insulation, Wheels & Belts, Timekeepers, Traps & Triggers, Making Things Float); two Non-*CRO* Technology, i.e., technology content that did not appear in Season II episodes of *CRO* (Flying Machines and Levers); and two Non-Technology (Coins and Photography). For each picture, children were asked to indicate their interest in learning about the topic or doing the activity (e.g., "I think learning about catapults would be...," or "I think making a catapult would be ...") using a five-point scale ranging from 1 = Very Boring, to 5 = Very Interesting (Appendix C.3.)

Club Interest

This assessment was also used in the Season I evaluation. Children were told to pretend that their after-school program was going to start a bunch of clubs and were asked to indicate if they would want to join the clubs. There were two *CRO* -related clubs (Inventors' and Scientists' clubs) and seven control clubs (Garden, Writers', Homework, Recycling, Computer, Photography, and Music). The four-point rating scale used to indicate their interest ranged from 1 = Definitely Not to 4 = Definitely Yes. After children had ranked their interest in joining each of the clubs, they were asked to indicate which of the nine clubs they would most want to join and which club they would least like to join. (See Appendix C.4.)

Interview Assessments

The first part of the interview was designed to examine children's interest in the topics presented in the *CRO* or *Where on Earth Is Carmen Sandiego* episodes. (The second part of the interview examined comprehension, and is discussed in the Comprehension Measures section below.) Each interest interview focused on the two previous shows that children had viewed. Children were asked about their attitudes towards the shows, what they liked and disliked about the shows, whether they wanted to find out more about the topics in the shows, and whether they engaged in any show-related activities after viewing. (See Appendix D.)

Observational Assessment

Children participated in four one-hour activity periods during which they could choose among three types of activities: reading (books or pamphlets), hands-on activities (making objects with LEGO DACTA construction sets, or solving pattern puzzles) or playing computer games. Each type of activity consisted of items that were either *CRO*-related (e.g., a DACTA kit for building a windmill, a book on light and energy), Non-*CRO* Technology (e.g., a computer game about building machines), or Non-Technology (e.g., a cookbook, a computer game on ocean life). Children's choices of activities, as well as their behavior while engaging in these activities, were used to assess their interest. (See Appendix E.)

Comprehension Measures

Children's understanding of the science and technology principles introduced in the *CRO* episodes was assessed in the second part of the in-depth interviews. The comprehension section consisted of three parts: A Free-Recall Task, a Sorting Task, and an Explanation Task. The Free-Recall task assessed what children remembered from all eight episodes of either *CRO* or *Where on Earth Is Carmen Sandiego.* The Sorting and Explanation Tasks asked children in both groups to reason about the devices and science principles that were introduced in four of

the *CRO* episodes: 1) *Just a Stone's Throw Away*, 2) *Turn Up The Heat*, 3) *No Time For Steamer*, and 4) *Laugh, Mammoth, Laugh.*

Free-Recall Task

The Free-Recall Task was designed to evaluate what children focused on in each of the episodes they viewed. The accuracy and content of children's recall are indicators of comprehension and can also be viewed as an indication of their attention to or interest in the show. Children in the *CRO*-group were asked what they remembered from the shows and their recall was coded according to whether they discussed non-technology content (e.g., characters, animation quality, music, story structure) or technology-related content (e.g., how the catapult worked, or how the characters made the shower hot). For example, the following is considered a non-technology recall statement from *No Time for Steamer* because it focuses on the storyline: "All the mammoths were going to a party and Steamer got left behind." In contrast, a technology content statement from the same episode focuses on the device and the goal: "They wanted to make a machine that would count time so that everyone would baby-sit for the same time, so they poured sand in this cone and when all the sand ran out it was time for the next person to baby-sit." Children in the *CSD*-group were also asked to recall the shows they had viewed. Their recall was coded according to whether they discussed content that was related to the geography or arts and sciences content (e.g., Van Gogh's artistic style and the museum in Amsterdam, or the Opera house in Sydney, Australia) or whether it was unrelated to the geography or arts and sciences topics (e.g., characters, animation, music, story structure).

Sorting Task

Sorting tasks have been used extensively to infer underlying knowledge representations (cf., Chi, 1989, 1985; Chi, Feltovich, & Glaser, 1981). One advantage of this kind of task is

that it does not rely on verbal ability. This makes it especially attractive for work with children because their ability to verbalize their knowledge often lags behind their ability to understand (cf., Brainerd, 1973; Bullock, 1984). How subjects sort or organize a set of items is used to infer their underlying understanding of the principles guiding their organization. For example, Chi, Feltovich, and Glaser found that physics novices tended to sort physics problems based on their surface features, such as whether the problem used a pulley or an inclined plane. Experts, on the other hand, sorted problems based on their underlying conceptual structure, such as whether the problem pertained to the balance of forces.

In our task children were presented with two to four pictures of devices that were constructed in the previously viewed *CRO* episode (see Appendix D). In the show the devices were constructed to solve a particular problem and each device was a revision and improvement upon the previous device. The children's task was to order the devices from best to worst for solving the particular problem (e.g., on the interview assessing understanding of catapults children were asked to order the pictures according to "which one was the best way to throw a rock really far"). Although this is the basic sorting task procedure used by Chi, Feltovich, & Glaser, we modified it in two critical ways: 1) we provided children with criteria for sorting the items (i.e., from best to worst), and 2) we presented only a small number of items to sort. These modifications reduced our ability to infer children's underlying conceptual frameworks, but they provided an external stimulus and representation of the children's thinking that they could refer to when they were asked to explain the devices.

Explanation Task

Children's sorting of the devices was used as the stimulus for their explanations. Children were asked to explain why each device received its rank and why it received a higher or lower rank than another picture. Children's explanations thus provided an in-depth measure of their explicit understanding of the devices and provided the justifications for their rankings.

Explanations were coded based on the conceptual frameworks that children used to justify their ratings of the devices. Children's conceptual frameworks are often based on surface features of situations or devices and fail to include underlying causal mechanisms, or they contain misconceptions about the underlying process. With increased knowledge and familiarity, children's conceptual frameworks begin to focus more on the causal mechanisms underlying devices, relying on principles rather than on superficial properties of the situation. (The coding scheme for their explanations is given in Appendix F).

PROCEDURE

The study consisted of three parts: Baseline Assessments and Pretest Assessments, which were conducted on one or two days; Intervention Period, which was conducted across a four-week period; and Posttest Assessments, which were conducted on the final day of the Intervention Period.

Baseline Assessments and Pretest Assessments

The first session was used to administer the Baseline Measures (Science Achievement Assessment, Hands-on Activities and Reading Interest Assessment) and pretest paper-and-pencil assessments (TV Show Interest, Learning & Doing Interest, and Club Interest). Assessments were administered to children individually or in small groups. All questions were read aloud to the children and additional assistance was given when necessary. After completing the assessments children were shown the materials that would be available during the activity period, and the general structure of the study was explained to them. They then returned to their regularly scheduled activities.

Treatment Period

Across the four-week period researchers visited each site three times a week: on the 1) Show and Activity Day, 2) Show Only Day, and 3) Interview Day. The Show and Activity

Day occurred first, followed two days later by the Show Only Day. The Interview Day was conducted the day immediately after the Show Only Day.

Show and Activity Day

The Show and Activity Day lasted approximately 90 minutes. The first 30 minutes were spent viewing and rating the show and the last 60 minutes were spent on the activities. On the first Show and Activity Day, children were first assigned to either the *CRO*-group or *CSD*-group, where they watched the first episode of their respective shows and then filled out the Show Appeal Ratings sheet. Children were then given their "library cards" with a list of available activities and were asked to indicate the activity that they wanted to do first (see Appendix E for the list of activities). Children then lined up at either the Book Stall, the Game Table, or the Computer Area, and requested their activity. If a child's first choice was not available, their preference was recorded and they were asked for their second choice. Once a child had finished playing or reading, they returned the item and were allowed to pick a new one. Except for the computer games, there were no time limits for having an item. Because there was only one computer in each room, and it was a popular activity, a 15-minute time limit was set, and children could only get a second turn after everyone who signed up to play on the computer had their turn.

Show Only Day

Children viewed the next show in the sequence and filled out the appeal rating. They then returned to their regular afterschool activities.

Interview Day

Children in each condition (*CRO*-group and *CSD*-group) were placed into same sex, same age-group pairs. Each pair was then given the interest and comprehension interview

corresponding to the shows viewed that week. Children without partners were interviewed individually. All interviews were audio-tape recorded, and took between 15 and 25 minutes. After the interview children were sent back to their regular afterschool activities.

Posttest Assessments

After the final interview, the TV Show Interest, Learning & Doing Interest , and Club Interest questionnaires were re-administered. Children were then thanked for their participation, and given both a *CRO* comic book and a certificate of participation.

RESULTS

The results are divided into four major sections: Baseline Science Achievement, Appeal of *CRO*, Impact of *CRO* on Children's Interest in Science and Technology, and Impact of *CRO* on Children's Comprehension of Science and Technology. In each section the impact of *CRO* is examined across four factors: Condition (*CRO*-group, *CSD*-group), Sex, Age-Group (Young, Middle, Old) and Science Achievement (Low, Average, High). Graphs depicting the main effect for condition are included in the text. Graphs depicting the effects for the other variables are included in Appendix G.

ANALYSES

Most analyses were conducted using ANOVAs. A repeated measures ANOVA was used when the same measurement was taken more than once (e.g., the three interest measures taken before and after treatment). A factorial design ANOVA was used when comparing responses on measures taken once (e.g., the baseline science achievement and the comprehension measures from each interview). Chi-square tests were used to examine frequency data (e.g., number of children recalling vs. not recalling educationally relevant content from show episodes). For the activity period data, McNemar tests were used to test for the significance of behavior changes over time and Pearson chi-square tests were used to compare two independent samples.

BASELINE SCIENCE ACHIEVEMENT

To ensure that differences in outcome measures were not due to initial differences in prior science achievement between the groups, we compared the *CRO*- and *CSD*-groups by their baseline science scores. Three science achievement groups (Low, Average, High) were constructed based on scores on the science achievement assessment. Children scoring below 50% were categorized as Low, children between 50% and 64% were categorized as Average, and children above 64% were categorized as High. Table 4 shows the science achievement

scores by condition, for boys and girls. No significant differences or interactions in science achievement scores for condition or gender were found, $p > .1$. Thus, the *CRO*- and *CSD*-groups were similar in their science achievement prior to the study.

SCIENCE ACHIEVEMENT	CONDITION			
	CRO-Group		*CSD*-Group	
	Boys (n)	**Girls (n)**	**Boys (n)**	**Girls (n)**
Low	36% (10)	37% (6)	40% (8)	31% (9)
Average	55 (8)	57 (8)	58 (6)	57 (7)
High	77 (4)	74 (8)	72 (7)	69 (5)
Mean Score	**53 (22)**	**58 (22)**	**53 (21)**	**48 (21)**

Table 4. Mean Score on Science Assessment by Condition & Gender

APPEAL OF *CRO*

Show Appeal Ratings

As shown in Figure 1, all *CRO* episodes were rated higher than the *CSD* episodes, $p < .001$. The mean appeal rating for *CRO* vs.*CSD* was 4.4 and 3.5, respectively (Maximum = 5), $p < .001$. This effect held for boys and girls, all age groups, and all levels of science achievement, p's $< .001$.

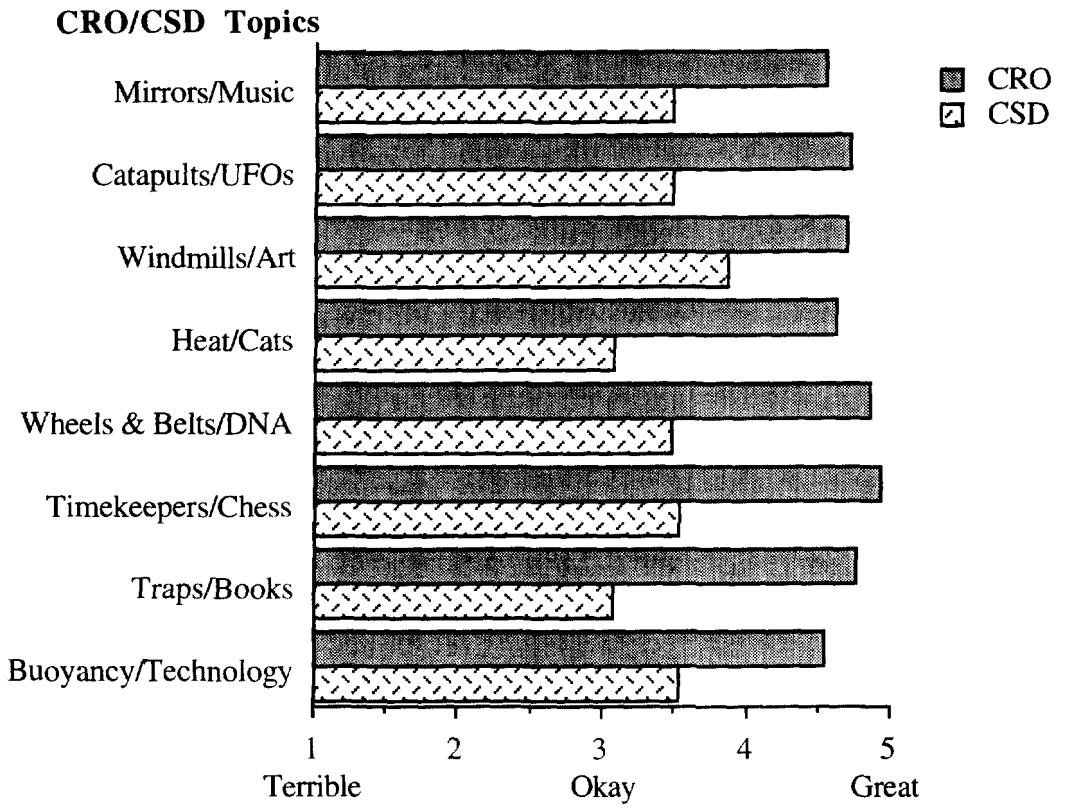

Figure 1. Appeal Ratings for *CRO* & *CSD* Episodes

To more closely examine the appeal of *CRO*, an episode by episode comparison was conducted. Between 79% and 92% of the children rated each of the eight shows as good or great, with the mean appeal of the show increasing from the first four episodes to the second four episodes, $p < .05$. Except for the first episode, which girls rated slightly lower than the boys, all episodes were rated equally highly by both sexes. There was no difference in the appeal of any of the shows based on any other demographic factors; that is, *CRO* appealed highly to all age groups, children, and all levels of science achievement.

Show Appeal - Interview measure

Another measure of the appeal of the show derives from children's responses on the first part of the interviews. When asked if they liked the shows, between 88% and 95% of the *CRO*-group reported liking the episodes, which is consistent with their appeal ratings.

Children were then asked why they liked or disliked the shows. Their reasons were coded into two categories, Specific and Global. Specific statements reflected an appeal for particular aspects of the show, such as the educational value, and the inventive, problem solving process used by Cro and his friends. Examples of specific statements included:

"They teach you in a fun way." Nicholas, 9

"Cro has a big brain, he solves all these problems." Ben, 5

"They put learning into fun." Nick, 10

"It makes learning fun...I like to see how they figured things out." Nick, 9

"Cro does all these things and he was such a little guy!" April, 7

"Sometimes I pretend I'm Cro and I'm solving the problems too." Zack, 6

Children also gave global statements when evaluating why they liked the show. These statements refer to overall evaluations or attitudes toward the show or characters in the show, as illustrated in the following examples:

"I loved it!" Brandon, 7

"Great! I love CRO." Kira, 8

"It's neat." Teeara, 7

"They're good shows and they're fun to watch." Ben, 5

"They're cool!" Seth, 7

In summary, based on both their appeal ratings and their comments during the interviews, children found all the episodes of *CRO* highly appealing, and many of them mentioned the educational value and problem-solving process as reasons for liking the shows. Given these findings, the next step was to examine whether viewing *CRO* affected their interest in doing and learning about the activities and topics that were illustrated in the show.

IMPACT OF *CRO* ON CHILDREN'S INTEREST IN SCIENCE AND TECHNOLOGY

In this section we examine the impact of *CRO* on children's interest in science and technology. The section is divided into three parts. In the first part we investigate the impact of *CRO* on children's interest in viewing various TV shows about science. In the second part we investigate the impact of *CRO* on children's interest in doing science and technology-related activities, and in the third part we examine the impact of *CRO* on children's interest in learning about science and technology.

TV Viewing Interest

To see whether *CRO* affected children's interest viewing other science programs, we compared children's rating of 13 TV shows, both on the pretest and posttest. The shows were categorized as either Science (*CRO*, *Beakman's World*, *Bill Nye the Science Guy*) or Non-Science (e.g., *The Little Mermaid*, *Where on Earth Is Carmen Sandiego*, *The X-Men*).

Figure 2 shows children's interest ratings in watching Science and Non-Science TV shows. Prior to the study, children in both groups rated the programs similarly, with Non-Science shows rated higher than Science shows by all children. However, the *CRO*-group indicated an increased interest in seeing all three Science shows in the posttest ($p < .01$) whereas the *CSD*-group displayed no change in interest. This effect was due to increases in the interest ratings of both boys and girls and low and high science achievers.

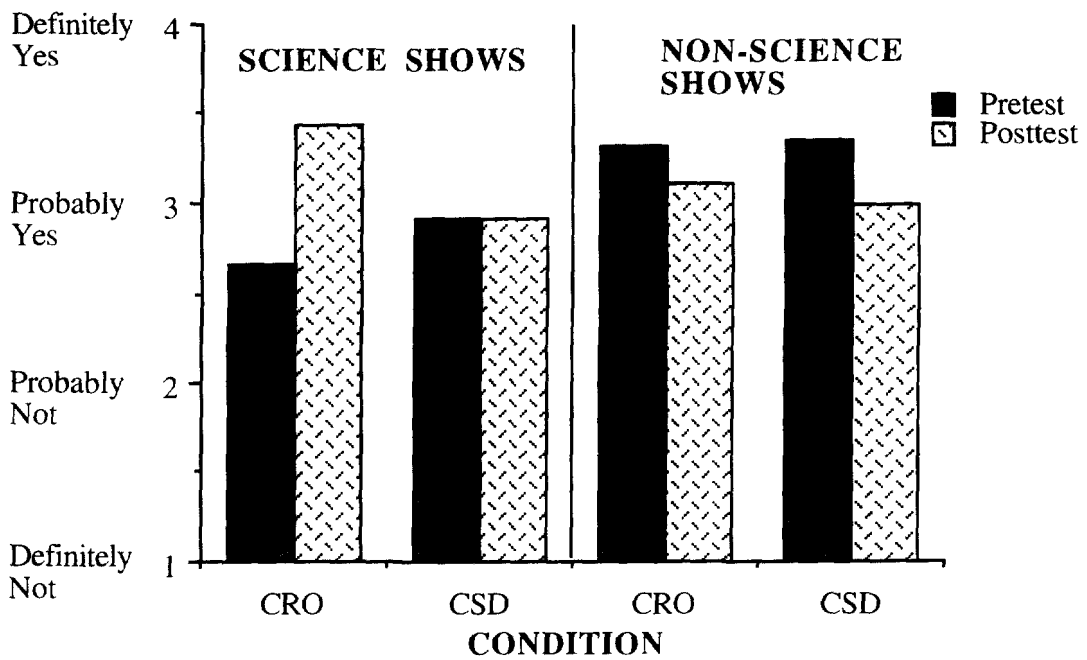

Figure 2. Interest in Viewing Science & Non-Science TV Shows Pretest vs. Posttest

Because children in the *CRO*-group rated the appeal of *CRO* highly, one might think that the impact on their interest in viewing science shows may be limited to *CRO* itself. Indeed, the *CRO*-group showed a significant increase in interest in viewing *CRO* on the posttest, as seen in Figure 3. However, as shown in Figure 4, a similar effect was found for the other two science shows, *Bill Nye the Science Guy* and *Beakman's World*. The *CRO*-group demonstrated a significant increase in interest in viewing these shows compared to the *CSD*-group, with both girls and low science achievers showing the greatest increases in interest. Thus *CRO* had a positive influence on children's interest, not only in viewing *CRO*, but also in viewing other science shows.