The purpose of this investigation was to examine the effects of modeling on sex-role development in first-grade children. It was hypothesized that when male-female pairs of children observed a male-female pair of models each child would probably shift his sex-role behavior toward that of the like-sex model. Sixty children served as subjects. The results did not support the hypothesis. The data supported the opposite effect: each child imitated the actions of the opposite-sex model. These results are explained in terms of identification, rather than modeling, theory. (Author)
THE EFFECTS OF MODELING ON THE DEVELOPMENT OF SEX-ROLE BEHAVIORS IN CHILDREN

Candace S. Garrett
Indiana University

This article is based on a dissertation submitted to Iowa State University at Ames in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to express her thanks to her major professor, Dr. Edwin Lewis, her statistical consultant, Dr. Leroy Wolins, and other members of her doctoral committee for their help. She also wishes to thank the staff and first-graders at Abbie Sawyer Elementary School, the site of the study, the two junior high school students who served as models, and the undergraduates who served as judges and research assistants.

Requests for reprints and for complete copies of tables and figures should be sent to Candace S. Garrett, Center for Innovation in Teaching the Handicapped, Indiana University, Bloomington, Indiana 47401.
Abstract

THE EFFECTS OF MODELING ON THE DEVELOPMENT OF SEX-ROLE BEHAVIORS IN CHILDREN

Candace S. Garrett, Indiana University

The purpose of this investigation was to examine the effects of modeling on sex-role development in first-grade children. It was hypothesized that when male-female pairs of children observe a male-female pair of models each child would shift his sex-role behavior toward that of the like-sex model. Sixty children served as subjects. The results did not support this hypothesis. Instead, the data supported the opposite effect: each child imitated the actions of the opposite-sex model. These results are explained in terms of identification, rather than modeling, theory.
By the time a child is three years old, he (or she) usually can distinguish between the sexes; he also knows which sex he is (Brown, 1958). Soon after this, he begins to exhibit appropriate sex-typed behaviors and attitudes (Sears, Maccoby, and Levin, 1957). The learning of these behaviors and attitudes is an important and pervasive aspect of development which has consequences not only in childhood but also in the adult years. For example, in the last few years a great deal of concern about the role of women in modern society has arisen, concern revolving around the aspirations, occupations, careers, and behavior patterns of women and the conflicts that choices in these areas may produce. The basis for these conflicts is formed during childhood, as children acquire their concepts of what are appropriate sex-role behaviors.

There are several theories concerning sex-role development in children, including the psychoanalytic theory of identification, Brown's theory of identification, and social learning theory. This study is primarily concerned with the latter.

According to Mischel (1966), observational learning from models, live or symbolic, is the first step in the acquisition of sex-typed behaviors. Bandura and Walters (1959) reinforce this point of view by stating that most of the values that will influence a child's behavior are acquired through imitation of the important adults in his life.

Research on the effects of imitation or modeling has shown that children will model aggression (Bandura, Ross, and Ross, 1965a; Hicks, 1965; Kuhn, Madsen, and Becker, 1967; Madsen, 1968), self-reward criteria (Bandura and Kipers, 1964; Mischel and Liebert, 1966), self-imposed delay of reward (Bandura and Mischel, 1965), and moral judgments (Bandura and McDonald, 1963). Modeling takes place even when the children are not
directly reinforced for modeling and when the model is not directly punished or rewarded for his behaviors.

Certain characteristics of the model usually facilitate imitation: prestige (Bandura, 1962), control over the child or over important resources for the child (Bandura, 1962; Grusec and Mischel, 1966; Maccoby, 1959), and adult status (Bandura and Kupers, 1964).

Several research studies have attempted to examine the effects of the sex of the subject and the sex of the model on imitation. In studies examining aggression, especially physical aggression, boys display more modeled aggression than girls do (Bandura, 1965a; Bandura, Ross, and Ross, 1963a; Flanders, 1968; Hicks, 1965; Madsen, 1968). However, other than in studies concerning aggression, no consistent sex differences have been found (Bandura and Huston, 1961; Bandura, Ross, and Ross, 1963a; Flanders, 1968; Mischel and Grusec, 1966; Mischel and Liebert, 1966).

Modeling has been suggested as one of the major mechanisms involved in sex-role development (Bandura and Walters, 1959; Mischel, 1966). This study was designed to examine the effects of modeling on sex-role development in children. Specifically, it was hypothesized that when male-female pairs of children observe a male-female pair of models each child would shift his sex-role behaviors toward those of the like-sex model.
METHOD

Subjects

Sixty first-grade children from a middle class elementary school in a Midwestern town served as subjects. Efforts were made to insure that the two opposite-sex members of each subject pair were relatively unfamiliar with each other: they were taken from different first-grade classes, had not attended the same kindergarten class the previous year, and did not live on the same street. Once these criteria were met, the pairing was random. These children were selected from the group of 73 children (out of a total of 86) whose parents returned permission slips indicating their approval for participation.

Tasks

Four construction-type toys, chosen to encourage interaction between the members of the subject pairs, were used as the tasks in this study. One of these, Tog'1, was used only in the modeling tapes. Since each pair of subjects was observed three times (pre-test, post-test, and follow-up), a different toy was used each time, and task order was balanced across the study. The trade names of these toys are Clink-a-Links (Toy 1), Ringa-Majigs (Toy 2), and Lego (Toy 3); each consists of a set of relatively small pieces that can be hooked together to form many kinds and shapes of objects.

Modeling Tapes

Two junior-high school students, one male and one female, served as models in the two ten-minute modeling video-tapes. This age group
was chosen because it was thought that they would be more effective and manipulable as models than younger children; adults were not used because of the type of task that formed the basis of their interactions. In the male-dominated version, the male model was instructed to be very dominant and verbally aggressive and to initiate the activities of the pair. He was told to decide what to build, what parts of the project the female model should construct, what colors to use, etc.; he was to direct the construction and to freely criticize his partner. The female model was instructed to be compliant: she was to be meek, mild, unconfident, and to ask for help often. In the female-dominated version, the models switched roles.

Ratings

Five female undergraduate students served as judges. All five rated all of the subjects in this study. The judges were given four hours of training; feedback, written and oral, indicated that at the end of the training sessions the judges could reliably use the scales. To insure objectivity and to counteract possible practice effects, they were not told the purpose of this research until the study was completed and they viewed the videotaped records of the subjects' interactions in an order which did not correspond to the participation order of the subjects.

Each child was rated on four scales; initiating activities, dominance, aggression, and compliance. The judges viewed three 15-minute interactions of each subject pair. Each 15-minute videotape was separated into five three-minute segments. At the end of each segment, each judge rated each child on a continuum from 0 to 100 indicating how certain she
was that the child had actually engaged in each of the four types of behaviors: "100" indicated that she was completely certain that the behavior was exhibited, "0" indicated that she was completely certain that the behavior was not exhibited, and "50" indicated that she was not certain whether or not the behavior was exhibited.

To obtain each child's trait scores for a 15-minute segment, the five ratings on each trait were summed. This resulted in one score (range 0 to 500) on each trait for each subject by each rater during each part of the study. A general activity score for each subject was obtained by summing all four trait scores.

Procedure

The research was conducted in a three-room mobile laboratory that was parked on school grounds. The videotape equipment used to record the subjects' interactions was located in one room. The children, in the next room, were videotaped through a one-way mirror in the wall between the two rooms. The children viewed their modeling tape on a television monitor located in the third room.

The experimenter escorted each pair of children from their classrooms to the laboratory. The children were seated at the table in the center room for the pre-test portion of the study and were given one of the construction toys to play with. They were asked if they had ever played with a toy like that one before. If one or both said no, the experimenter showed them various ways that the parts of the toys could be put together. They were then told that they should decide on one thing to build and that they should build it together. They were given 15 minutes to play together with the toy. If the children did not interact
the experimenter encouraged them to do so as many times as was necessary. During this time, their interactions were videotaped.

For half of the pairs of children, one child was then taken back into his classroom and exchanged for a like-sex child; the other half kept the same partners. This was done to determine if there would be an effect due to the children's experience of interacting with each other before viewing the modeling tape. In either case, the pair was next taken to another room in the laboratory and told that they were going to watch a short television show. They then saw one of the two versions of the modeling tape.

After viewing the tape, the children returned to the center room for the post-test portion. They were given a new construction toy, were told to play with it together, and were encouraged when necessary to interact. They were videotaped for another 15 minutes, following which they were taken back into their classes.

One week after their initial participation, the same pairs of children (who saw the modeling tape together) were brought back out to the laboratory for the follow-up portion. They were given a third new construction toy and similar directions to play with it together. They were videotaped for another 15 minutes and then returned to their classrooms.

The principal investigator served as experimenter throughout the study. She remained within the children's sight at all times.

Analysis

Analysis of variance and graphical techniques were used to analyze the data. The ratings for the subjects in each group (same partner-different partners) were analyzed separately. Also the ratings from each
judge within each of these groups were analyzed separately; this was done to avoid the possibility that any one judge would spuriously influence the whole analysis and interpretation. This resulted in ten analysis-of-variance tables. The variables included in the analysis were treatment (male-or female-dominated modeling tape), task order (1, 2, or 3), subject pair (1 or 2), sex of the subject (male or female), time (pre-test, post-test, or follow-up), and trait (initiating activities, dominance, aggression, or compliance).

RESULTS

Table 1 presents the values of the eight mean square terms used as error terms in the $F$ tests for each judge between same-partner and different-partners groups.

Insert Table 1 About Here

Two different types of comparisons can be made using this table. One of these is a comparison of the relative size of error in each error term between the two groups to determine the effect of changing partners during the study. Comparing between groups, most of the errors contained in each error term are about the same size as the corresponding errors in the other group. However, two error terms are not the same size. In the pairs with time error term, the errors in the different-partners group are about seven times as large as those in the same-partner group. In the pairs with trait error term, the errors in the same-partner group are about six times those in the different-partners group. Changing partners introduced much more variance in the pairs with time and pairs with trait error terms.

The second type of comparison that can be made from this table is that of comparing the relative size of each error term with the relative
size of the overall error term to determine which effects contributed large amounts of variance. Most of the error terms are about the same size as the overall error term. However, the errors in each group in the pairs error terms are eight or ten times as large as the errors in the overall error term. As expected, individual differences among pairs contributed a large amount of variance to the experiment.

Table 2 is a summary table of all ten analysis-of-variance tables showing the significant effects found by each judge within each group (same-partner, different-partners).

There are two consistently significant main effects across groups. One is trait. As Figure 1 shows, in the both groups there was more dominance and compliance evident than initiating activities and aggression. Also the profile of trail scores is more extreme in the same-partner group than in the different-partners group. In addition, Figure 1 indicates that within each group the judges agreed on the different amounts of the different traits displayed. This is an indication that the judges actually did use different behaviors to score each trait and hence that the four traits are indeed separate; it also appears that, since so little aggression as defined in this study was observed, aggression was not a very useful trait for this type of measurement.

The other consistently significant main effect is time. A plot of this effect (not included here) indicates that there was more activity in
the post-test and follow-up parts of the study than in the pre-test part.

Since it is clear that switching partners between the pre- and post-test portions of the study introduced more error and resulted in less activity; only the results from the same-partner group were used in the remainder of the analysis.

Table 2 also shows that treatment, task-order, and time entered into several significant interactions. Figure 2 is a summary plot of the treatment with task-order with time interaction in which the ratings of the five judges have been averaged to obtain a mean judge rating. This plot suggests that the male-dominated version of the modeling tape was more successful in inducing behavior change than was the female-dominated (i.e., the effects due to task-order are most clearly evident in the bottom half of Figure 2). It also shows that Toy 3 was not all sensitive to the behavior changes that resulted from exposure to the modeling tapes, while Toy 1 was quite sensitive. Toy 2 was relatively sensitive in the follow-up when it followed Toy 1 in the post-test, but it was not sensitive otherwise. This indicates that a large part of the variance in this study is due to task differences.

If the original hypothesis of this study is correct, then the treatment with sex of subject with time with trait interaction would be significant. Only one judge in one group found this effect significant. Therefore, the hypothesis was not verified.

Two judges in the same-partner group found the treatment with sex of subject with trait interaction significant; if the data had been analyzed
by averaging across judges, the resulting interaction effect probably
would have attained significance. Also, since a large part of the variance
in this study is due to task differences, the fact that this interaction
occasionally reached significance indicates that an examination of this
interaction may produce some interesting trends. Figure 3 is a summary
plot of treatment with sex \(c\) with trait interaction obtained by
averaging across judges. In general, this figure, in conjunction with
the previously discussed effect of time shows, that: (a) boys who viewed
the male-dominated tape became less dominant and initiated fewer activities
than boys who viewed the female-dominated tape; (b) girls who viewed the
male-dominated tape became more dominant and initiated more activities
than girls who viewed the female-dominated tape; (c) girls who viewed the
male-dominated tape became more dominant and initiated more activities
than boys who viewed this same tape; and (d) boys who viewed the female-
dominated tape engaged in more initiation of activities, more dominance,
and less compliance than girls who viewed this same tape.

In general, the ratings made by the judges were quite reliable. Each
of the original graphs in which each judge's ratings were plotted separately,
such as Figure 1, shows that the judges were relatively consistent among
themselves. They generally ranked the subjects the same way on the various
traits and saw the same differences among the subject pairs. Also, Table
1 indicates that the error terms for each judge within each group are
about the same size, again, indicating inter-rater reliability. Inter-
rater reliability coefficients could have been calculated, but since no
single coefficient would adequately describe the data and since the above
two results show that the ratings are quite reliable, further calculation seemed both inefficient and unnecessary.

DISCUSSION

Two interesting main group differences are evident in the results. First, more activity in general took place in the same-partner group than in the different-partners group. In the pairs with time interaction, the errors in the different-partners group were about seven times as large as those in the same-partner group. Each subject in the former group had to meet the demands of a new partner as well as contend with the task in the post-test portion. This is reflected in the data.

The results also demonstrate that the nature of the tasks used in this type of study are highly important. Task differences contributed enough variance to make interpretation of the important effects quite difficult. This finding was highly unexpected since the tasks were similar in an obvious way: they all were construction toys. Therefore, it was assumed that they would be equally sensitive in this study. Obviously, this was not true.

This task effect may have resulted from the subject's familiarity with the toys. Of the 48 subjects who participated in the study, 7 reported that they already were familiar with Toy 1, 40 reported that they were familiar with Toy 2, and 46 reported familiarity with Toy 3. The most unfamiliar task, Toy 1, was also the most sensitive task. The other two tasks were highly familiar to the children, and neither was very sensitive. Perhaps unfamiliar tasks are needed so that the subject does not bring to the study preconceived habits or boredom derived from previously playing with the toys.
In addition to being the most unfamiliar, Toy 1 contained only 40 separate pieces as compared to at least 100 in each of the other two toys. Hence, the children had to interact because there were not enough pieces for each to play alone. This could have contributed to the task's sensitivity.

Table 3 represents the type of sex-role behaviors hypothesized by a modeling approach as well as indicating the actual behaviors which occurred in the study. Obviously, the data do not support the modeling hypothesis.

They can however be interpreted in terms of identification theory. It is possible, for example, that the male subjects may have gained vicarious satisfaction from identifying with the male model in the male-dominated version of the modeling tape and consequently no longer felt that they had to act in a masculine manner. However, this version may have disturbed the girls if they identified with the female model, so that they reacted to their partners by being more "masculine." Similarly, the female subjects may have gained vicarious satisfaction from identification with the female model in the female-dominated tape and hence did not need to act toward their partners in a masculine manner, while the boys may have been dissatisfied and so reacted to their partners by being quite masculine. Such an interpretation emphasizes identification, rather than modeling, as the basic operating process.

It is not clear why the male-dominated version of the modeling tape should have caused children of both sexes to engage in more activity than did the female-dominated version.
Table 1. Values of mean square error terms for each judge between groups (Group 1: children who kept the same partner throughout; Group 2: children who changed partners)

<table>
<thead>
<tr>
<th>Pairs</th>
<th>Pairs X Sex</th>
<th>Pairs X Time</th>
<th>Pairs X Trait</th>
<th>Pairs X Sex X Time</th>
<th>Pairs X Sex X Trait</th>
<th>Overall Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 1</td>
<td>Group 2</td>
<td>Group 1</td>
<td>Group 2</td>
</tr>
<tr>
<td>Judge 1</td>
<td>107,820</td>
<td>28,728</td>
<td>3,673</td>
<td>1,718</td>
<td>2,894</td>
<td>2,507</td>
</tr>
<tr>
<td>Judge 2</td>
<td>74,054</td>
<td>35,317</td>
<td>4,717</td>
<td>1,436</td>
<td>3,640</td>
<td>35,237</td>
</tr>
<tr>
<td>Judge 3</td>
<td>72,073</td>
<td>42,165</td>
<td>5,979</td>
<td>2,868</td>
<td>4,515</td>
<td>28,880</td>
</tr>
<tr>
<td>Judge 4</td>
<td>96,792</td>
<td>45,878</td>
<td>2,952</td>
<td>2,440</td>
<td>4,640</td>
<td>35,237</td>
</tr>
<tr>
<td>Judge 5</td>
<td>65,976</td>
<td>24,236</td>
<td>6,099</td>
<td>982</td>
<td>3,637</td>
<td>35,412</td>
</tr>
</tbody>
</table>

Group 1: children who kept the same partner throughout; Group 2: children who changed partners.
Table 2. Summary table of the significant effects found by each with the same-partner and different-partners groups.

<table>
<thead>
<tr>
<th>Source</th>
<th>Same-Partner</th>
<th>Different-Partners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment X Sex</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>Treatment X Task-Order X Sex</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>Treatment X Time</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>Task-Order X Time</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>Treatment X Task-Order X Time</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>Trait</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>Treatment X Trait</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>Task-Order X Trait</td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>Sex X Time</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Treatment X Sex X Trait</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Treatment X Task-Order X Sex X Time</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Time X Trait</td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td>Treatment X Time X Trait</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Task-Order X Time X Trait</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Sex X Time X Trait</td>
<td>.01</td>
<td>.05</td>
</tr>
</tbody>
</table>


Figure 1. Mean score for each trait for all subjects in Group 1 (same-partner) and Group 2 (different-partners)
INITIATING ACTIVITIES

DOMINANCE

AGGRESSION

COMPLIANCE

MEAN SCORE FOR EACH TRAIT

JUDGE 1

JUDGE 2

JUDGE 3

JUDGE 4

JUDGE 5

GROUP 1

GROUP 2

INITIATING ACTIVITIES

DOMINANCE

AGGRESSION

COMPLIANCE

TRAIT
Figure 2. Summary plot of the mean score of all judges for combined traits for each task-order* by time for Group 1 (same-partner) subjects viewing male-dominated tape and female-dominated tape

*Task-Order 1: Toy 1, Toy 2, Toy 3
Task-Order 2: Toy 2, Toy 3, Toy 1
Task-Order 3: Toy 3, Toy 1, Toy 2
MEAN SCORE OF ALL JUDGES FOR COMBINED TRAITS

GROUP 1

FEMALE-DOMINATED TAPE

MALE-DOMINATED TAPE

PRE-TEST  POST-TEST  FOLLOW-UP  PRE-TEST  POST-TEST  FOLLOW-UP  PRE-TEST  POST-TEST  FOLLOW-UP

TIME

250
200
150
100
50
0
50
100
150
200
250
Figure 3. Summary plot of the mean score of all judges for each trait for Group 1 (same-partner) male and female subjects viewing male-dominated tape and female-dominated tape.
MEAN SCORE OF ALL JUDGES FOR EACH TRAIT

INITIATING ACTIVITIES
DOMINANCE
AGGRESSION
COMPLIANCE

MALE SUBJECTS
FEMALE SUBJECTS
FEMALE-DOMINATED TAPE

MALE-DOMINATED TAPE

GROUP 1

MEAN SCORE OF ALL JUDGES FOR EACH TRAIT
<table>
<thead>
<tr>
<th>Hypothesized Behaviors</th>
<th>Actual Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male model dominant aggressive &amp; initiator</td>
<td>Female Ss</td>
</tr>
<tr>
<td>Feminine</td>
<td>Masculine</td>
</tr>
</tbody>
</table>
REFERENCES


