

2005

The effect of pretense on the creativity of storytelling in preschool and kindergarten-age children

Storm R. Ross

Follow this and additional works at: <http://commons.emich.edu/theses>



Part of the [Psychology Commons](#)

Recommended Citation

Ross, Storm R., "The effect of pretense on the creativity of storytelling in preschool and kindergarten-age children" (2005). *Master's Theses and Doctoral Dissertations*. 119.

<http://commons.emich.edu/theses/119>

This Open Access Thesis is brought to you for free and open access by the Master's Theses, and Doctoral Dissertations, and Graduate Capstone Projects at DigitalCommons@EMU. It has been accepted for inclusion in Master's Theses and Doctoral Dissertations by an authorized administrator of DigitalCommons@EMU. For more information, please contact lib-ir@emich.edu.

THE EFFECT OF PRETENSE ON THE CREATIVITY OF STORYTELLING IN
PRESCHOOL AND KINDERGARTEN-AGE CHILDREN

by

Storm R Ross

Thesis

Submitted to the Department of Psychology

Eastern Michigan University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

General Experimental Psychology

Thesis Committee:

Marilyn Bonem, PhD, Chair

Dennis Delprato, PhD

Elliott Bonem, PhD

February 11, 2005

Ypsilanti, Michigan

Abstract

Previous research suggests a link between pretense and the creative behavior of children, such as associative fluency. Most results show that children who engaged in pretense prior to an associative fluency task gave more nonstandard uses for play objects than children who engaged in imitation play or coloring. However, the effect of pretense play on other forms of creativity has not been extensively researched. One such form of creativity is storytelling.

The present study examined the relationship between pretense and storytelling. It was hypothesized that the stories a child tells after engaging in pretense play would be ranked and rated by experts as more creative than the stories told after engaging in non-pretense activities. Teachers were used as subjective judges of creativity. The results of two experiments suggest that engaging in pretense play prior to storytelling does not lead to more creative stories. Suggestions for future research are given.

Table of Contents

Abstract.....	ii
List of Figures.....	v
Review of Literature.....	1
Experiment 1.....	21
Participants.....	21
Stimulus Materials.....	22
General Procedure and Research Design.....	22
Baseline Condition.....	24
Directed Pretense Condition.....	25
Free Play Condition.....	25
Imitation Condition.....	25
The Dependent Measure.....	26
Ratings of Behavior.....	28
Results.....	28
Discussion.....	38
Experiment 2.....	43
Participants.....	43
Stimulus Materials.....	43
General Procedure and Research Design.....	44
Baseline Condition.....	45
Pretense Condition.....	46
Imitation Condition.....	47

The Dependent Measure	47
Ratings of Behavior	49
Results.....	49
Discussion.....	64
General Discussion and Conclusion	66
References.....	72
Appendices.....	77
Appendix A: Preschool Informed Consent Form	78
Appendix B: Preschool Story Rating Scale	82
Appendix C: Activity Behaviors Checklist.....	84
Appendix D: Kindergarten Informed Consent Form	89
Appendix E: Kindergarten Story Rating Scale	93
Appendix F: Highest Single-Ranked Stories	95
Appendix G: Approval for the Use of Human Subjects	101

List of Figures

Figure 1. Creativity rankings of the stories told by Participant 1 in Experiment 1.....	30
Figure 2. Creativity ratings of the stories told by Participant 1 in Experiment 1.....	31
Figure 3. Creativity rankings of the stories told by Participant 4 in Experiment 1.....	33
Figure 4. Creativity ratings of the stories told by Participant 4 in Experiment 1.....	34
Figure 5. Creativity rankings of the stories told by Participant 1 in Experiment 2.....	51
Figure 6. Creativity ratings of the stories told by Participant 1 in Experiment 2.....	52
Figure 7. Creativity rankings of the stories told by Participant 2 in Experiment 2.....	54
Figure 8. Creativity ratings of the stories told by Participant 2 in Experiment 2.....	55
Figure 9. Creativity rankings of the stories told by Participant 3 in Experiment 2.....	57
Figure 10. Creativity ratings of the stories told by Participant 3 in Experiment 2.....	58
Figure 11. Creativity rankings of the stories told by Participant 4 in Experiment 2....	60
Figure 12. Creativity ratings of the stories told by Participant 4 in Experiment 2.....	61

Review of Literature

A preschool-age child stands on the schoolyard during recess, using a tree branch as a dancing partner. His friend is yelling for someone to help her capture a vicious fire-snake, which in reality is a jump-rope. A third child uses a leaf as an oven mitten to capture the fiery beast. Certainly the pretend play of these children could be considered creative. However, could the creativity of their play actually be enhancing the creativity of other activities?

To answer this question, it may be beneficial to consider early theories concerning the cognitive processes that occur during play that highlight the creative nature of play. By the middle of the twentieth century, many cognitive theories concerning play and creativity had arisen, including those by Piaget and Vygotsky (Dansky, 1999). According to Piaget (1962), pretense play is an activity ripe with creative qualities. *Distortion assimilation* is a hypothesized cognitive mechanism that allows for objects to be mentally transformed into whatever the child might want (for example, a stick could be distorted into a person and a leaf into a car). Through *free assimilation*, these seemingly unrelated distorted objects can be combined into a cohesive play episode (the stick/person drives the leaf/car). Vygotsky (1930/1990) also stated that such object transformations take place during pretense and are a source of creativity but stressed that the origins of creative play are rooted in the interactions between the parents and the child (Vygotsky, 1932/1987; Smolucha, 1992). Vygotsky did not place the same emphasis on the individual abilities of children as Piaget (Dansky, 1999).

Bretherton (1989) agreed that pretense has creative properties but defined those properties somewhat differently. Bretherton (1989) stated that there are three components to pretense. The first component, multiple role taking, shows that children can both act in the

playframe and create the playframe simultaneously. The second component, transformations of reality, or counterfactual thinking, showcases the child's ability to create novel plots. The third component, blurring conceptual levels or manipulating the playframe, details how the child can easily step across the line between reality and fantasy at any given time during the play sequence.

Although these cognitively based developmental theories are only a small sample of those concerning play and creativity, they serve to illustrate an important point that is shared by all: many properties of play, especially pretense play, have been theorized to invoke either cognitive processes that produce creative products or creative behavior. However, despite the fact that developmental theories have detailed the creative nature of play for decades, little research has been done to determine whether the creative nature of play actually affects the development of creativity in other areas. Although theorists such as Vygotsky and Piaget hypothesized the importance of creative play on the development of creative problem solving (Dansky, 1999), little empirical work has been done further in this direction. The work that does exist, however, mainly concerns the effect of play on associative fluency.

Associative fluency is defined by Dansky (1999) as "the tendency to produce numerous ideas in response to questions or other situations; associative fluency is an important dimension of the process of creative problem solving" (p. 393). Associative fluency is closely linked to creative problem solving in that associative fluency represents the ability to come up with multiple creative ways to solve a problem. Given this definition, one could conclude that associative fluency represents a particular dimension of creativity. One of the first studies to look at the causal relationship between play and associative fluency was done by Dansky and Silverman (1973).

In the study by Dansky and Silverman (1973), the experimenters hypothesized that engaging in free play would temporarily enhance associative fluency. The authors sampled 90 children between the ages of 4 and 6 from a nursery school and day care center. Three groups were formed: play, imitation, and control. The children in the play and imitation groups were presented with the following materials: 10 paper towels, a screwdriver, a wooden board with five screws set in it, a pile of 30 paper clips, 15 blank cards, 10 empty matchboxes, and a tray containing six wet plastic cups. The play group was given 10 minutes to play with the materials as they wished, whereas the imitation group was told to imitate the experimenter after the experimenter engaged in particular behaviors with the materials. The control group was given a box of crayons and sketches to color and allowed to color for 10 minutes. After each condition, the participants were given an alternate-uses test on how many different ways four of the materials from the play and imitation conditions could be used. Results showed that children in the play group gave more nonstandard uses for the objects than children from the imitation or control groups. To control for the possibility that object familiarity might interact with play, Dansky and Silverman (1975) later replicated their experiment using different materials during the alternate-uses test than the materials used in the play and imitation conditions. Although some research suggests that developmental differences in the type of play behaviors observed during free play exist between children ages 45-59 months and 58-69 months (Rubin, Watson, and Jambor, 1978), this was not seen in either the initial experiment or its replication.

Li (1978) took the experiments of Dansky and Silverman a step further, assessing the difference between unstructured play and being instructed to specifically engage in pretense. She began by randomly assigning 120 kindergarten children to one of four conditions:

control, imitation, free play, and make-believe play. The children in the latter three groups were presented with the following materials: 10 paper towels, 10 empty matchboxes, and 30 paperclips. The imitation group was also given 6 wet plastic cups, a bunch of toothpicks, and 15 blank cards. Participants in the free-play group were presented with the materials and were allowed to play with them as they wished. The make-believe play group was told a brief imaginative story to transition into the activity, after which they were asked to directly engage in make-believe with the stimulus materials. Participants in the imitation group were asked to imitate the experimenter after the experimenter engaged in particular behaviors with the objects. The control group was given colored pencils and sketches to color. After each condition, the participants were given an alternate-uses test on how many different ways three of the materials from the play and imitation conditions (a paper towel, a matchbox, and a paper clip), as well as an item not from the conditions (a screwdriver), could be used. Results showed that participants in the make-believe play group gave significantly more nonstandard responses for the screwdriver than did the free-play and control groups. In addition, participants in both the make-believe play and free-play groups gave significantly more nonstandard responses for the paper clip than did the control group.

Although Dansky and Silverman (1973) noted many of the pretense behaviors of the children in the discussion section of their original study, these behaviors were not systematically recorded. As a result, both pretense and other forms of play (constructive play, preplay behaviors) could have been occurring in the pretense group. Forms of play that do not incorporate make-believe would not be expected to enhance associative fluency. In addition, although Li (1978) noted that associative fluency was enhanced for both the free play and make believe play groups, she did not systematically record the behaviors of the

groups. Because of this, it could not be determined whether the unstructured nature of the play or engaging specifically in pretense was resulting in the enhancement of associative fluency. However, Dansky (1980) conducted another experiment to see whether make-believe, the central component of pretense, was actually a mediator between play and associative fluency.

Dansky (1980) began by sampling 96 preschool children. The children were observed during a free play period, and their behavior was recorded every 15 seconds for two minutes. Whereas many studies on play focus on both social and cognitive levels of play and differentiate between solitary, parallel, and group play (Christie, 1992), Dansky (1980) was only interested in the differences between pretense and other forms of play regardless of the social level of the observed play. Therefore, the following behaviors were recorded: role play, object transformation, verbal communication within the context of role play, nonverbal interaction during role play, make-believe (which includes all of the preceding behaviors), social or solitary preplay (including manipulative, locomotor, or explorative behaviors unrelated to make-believe or constructive play), and constructive play (involves creating or building something out of several objects). On the basis of these observations, children were categorized as either *players* (children likely to engage in make-believe) or *non-players*. A week later, children were seen in pairs and randomly assigned to pretense, imitation, or control conditions. The pretense and imitation conditions were similar to those from the Dansky and Silverman (1973) experiment, whereas the control condition was a convergent problem-solving task that had the children guessing which of the objects the experimenter was thinking on the basis of clues. Afterwards, the children were given the alternate-uses test for the objects in the pretense and imitation conditions. Results showed that only the

children in the pretense group who were initially classified as players displayed enhanced associative fluency. From this result, the author concluded that it is the use of make-believe in a free-play environment, rather than the lack of structure of the environment, that enhances associative fluency.

Not all research has found a link between pretense and associative fluency. Smith and Whitney (1987) hypothesized that the positive findings of Dansky and Silverman (1973, 1975), as well as those of Li (1978), were due to experimenter effects. The authors were critical of the fact that that previous research was conducted directly by the primary investigators; thus, the chance that effects were seen due to unconscious experimenter expectations was higher. The authors hypothesized that a relationship between play and associative fluency would not be seen if experimenter bias were controlled. The authors began by sampling 64 children and assigning them to fantasy play, play, or imitation conditions. Participants in the fantasy-play condition were encouraged to engage in make-believe play with the stimulus materials, whereas the free play and imitation conditions were similar to those used by Dansky and Silverman (1973). The stimulus materials were also similar to those of Dansky and Silverman (1973). The groups were monitored to ensure that the fantasy-play group was engaged in more pretense behaviors than the other groups. After the activity, each participant was taken to a separate room and was administered an alternate-uses test by a different experimenter. Each participant was presented with two familiar objects and two objects not used in the previous condition during this test while a third experimenter wrote down the responses. Participants were tested again a week later by the same experimenter. Results indicated that there was no significant difference between groups for associative fluency. The authors concluded that experimenter biases had

contributed to positive findings for play and associative fluency up until this point. However, Dansky (1999) pointed out that it may be the disruptive nature of the transition between the activity and the alternate-uses test, as well as meeting with a new stranger for the alternate-uses test, rather than experimenter bias, that is having an impact on associative fluency.

Taken as a whole, this research provides evidence that a directional relationship exists between pretense and associative fluency. What has not been addressed, however, is whether pretense can affect other forms of creativity. As stated previously, associative fluency is an important component of creative problem solving (Dansky, 1999), but what of other forms of creativity? Problem solving is only one of many different ways children can express their creativity, yet little research has been conducted to determine what other forms of creativity can be affected by play. Although recent research has suggested that play can affect the creativity of collages produced by children (Howard-Jones, Taylor, & Sutton, 2002) and will be discussed later in this paper, other areas of creativity have yet to be explored. One of those areas is storytelling.

In order to examine the relationship between play and storytelling, it is important to first note the developmental progression of creativity in the storytelling abilities of children. However, it should also be noted that little existing research examines the relationship between the storytelling abilities of children and other forms of creativity; in fact, little existing research looks at the storytelling abilities of children in general. The research that does exist about the storytelling abilities of children examines the relationship between storytelling and many different constructs, including its effect (along with pretend play) on narrative recall (Kim, 1999) and the moral ideals children convey during storytelling (Pramling, Norlander, and Archer, 2001). Despite the lack of existing research and the

variety of topics covered, a few studies have looked at the development of the creative storytelling of children, including those by Magee and Sutton-Smith (1983) and Geist and Aldridge (2002).

Magee and Sutton-Smith (1983) outlined nine stages of development of the storytelling abilities of children from analyzing the content of interactions between a child and her parents during storytelling times while the child was between the ages of 10 and 29 months. In the first stage, *dialogue with picture books*, children repeat elements of a story their parents tell them and are rewarded with positive feedback from the parents. *Role reversal with picture books* occurs when the child takes a more active role, asking questions about what occurs during the story. *Listening to storybooks* comes next, with the child assuming the role of interested listener to the parent's role of storyteller. The fourth stage, *contributing to storybooks*, allows the child to both help in the telling of the story and to express emotions about the events of the story. During the fifth stage, *picture-telling*, the child is now able to recall the events of a simple picture book and can answer simple questions concerning the content of the story. What is considered early storytelling comprises the sixth stage, during which the child attempts to add new material to a previously learned story, usually resulting in the confusion of the adults to whom the child is telling the story. This ability is refined during the seventh stage, *personal narratives*, when the child begins to tell stories based on real-life events without the use of a book, once again frequently producing confusion in his or her audience. *Co-telling stories* occurs in the eighth stage, during which the parents tell an original story with help from the child, and occurs at about the same time as the emergence of pretense play. The final stage, *storytelling*, has the child in complete control of the story, incorporating make-believe in order to construct an

original narrative. Through this case study, Magee and Sutton-Smith outlined the emergence of creativity in a child's storytelling abilities, tracing the path from repeating the elements of a picture book to the creation of all-original stories without the use of outside sources.

Although Magee and Sutton-Smith (1983) used a case study and naturalistic observation to analyze the development of creative storytelling in children, Geist and Aldridge (2002) used multiple participants to examine this topic. The authors were particularly concerned with the developmental differences between children in the following areas: The type of stories (personal narrative, fantasy, realistic fiction), the content of the stories (imaginative elements, focus on peers or parents, personal fears), and the organization of fairy tales that children tell (complexity, egocentric organization, distinction between fantasy and reality). The authors believed that by analyzing these areas of children's storytelling, they would be given some insight into many areas of development, including creativity.

Geist and Aldridge (2002) began by going to four classrooms in an elementary school to conduct four 45-minute lessons on the plots and characters of fairy tales, held over the course of four days. The four classrooms that the researchers visited included one kindergarten, one first-grade, one second-grade, and one third-grade classroom. After the four days, each child made up a fairy tale and told it orally. The stories were tape recorded and transcribed, after which a content analysis was done.

Results of the Geist and Aldridge (2002) study showed a number of findings. The content of the stories of kindergartners was based on previously heard material, whereas first graders based their stories on familiar surroundings, second graders came up with original material, and third graders drew from personal experiences to tell their stories. However, the

type of stories of kindergartners tended to be more fantastical, with a shift towards realism progressing throughout all the grades, peaking at the third grade. Developmental differences in the organization of the stories were also observed. Older children were better at organizing their stories, tended to focus less on the self, and were better at distinguishing the difference between fantasy and reality.

On the basis of these results, Geist and Aldridge (2002) proposed a five-stage developmental model of children's storytelling abilities. During the first stage, children are only able to tell disjointed stories without a cohesive theme. In the second stage, children are able to construct cohesive phrases but match them with other phrases in a way that does not form a cohesive story. At the third stage, children are able to tell cohesive short stories but without much detail. The fourth stage involves the children trying to elaborate on the short stories, resulting in the confusion of the themes of the original short story. The final stage shows the child being able to tell a cohesive yet elaborative original story.

Although the studies of Magee and Sutton-Smith (1983) and Geist and Aldridge (2002) both looked at the development of the creative storytelling abilities of children, it is apparent that there are many differences between the two proposed theories. Perhaps the main cause of the differences between the two theories is how storytelling is defined. Geist and Aldridge (2002) believe that storytelling can only be considered such when the child is able to tell an original story that is both understandable and complex, whereas Magee and Sutton-Smith (1983) only require that the narrative told by the child be original. For Magee and Sutton-Smith (1983), the child's ability to tell a story might be refined as he or she gets older, but as long as the central theme of the story is original, storytelling is occurring. As a result, storytelling is occurring at about age two, according to Magee and Sutton-Smith

(1983), but it is not observed until around the age of eight in the Geist and Aldridge (2002) model.

In addition to the differences between definitions of storytelling, both studies are hampered by their sampling procedures. Magee and Sutton-Smith (1983) used a case study, whereas the sample of the Geist and Aldridge (2002) study was composed entirely of African-American children from low-SES backgrounds. As stated previously, research on the relationship between storytelling and creativity as well as on the storytelling of children is sparse; perhaps these issues can be clarified with future research. However, what both studies show is that creativity is a central component of storytelling and that storytelling in turn can be used to express creativity. Both studies also show that storytelling abilities change with age, and other research has shown this as well (Berman, 1995). Regardless of whether creative storytelling is defined only as having an original narrative or as having a narrative that is both original and complex, both theories highlight the creative nature of children's storytelling.

At this point, it has been established that play has the ability to affect a particular form of creativity, that being associative fluency. Also, the creative components of children's storytelling abilities have been highlighted. However, before research can be done to determine the relationship between play and storytelling, a measure of creative storytelling must be decided upon. Although literature concerning the creativity of children's storytelling is somewhat sparse, research on the general construct of children's creativity is abundant.

Perhaps the most popular way to assess children's creativity is with the Torrance Tests of Creative Thinking, developed by Paul Torrance in 1974. The Torrance Tests of Creative Thinking, or TTCT, have been used in over 2000 published studies. The TTCT uses

two types of tests, each with alternate forms: Thinking Creatively with Words and Thinking Creatively with Pictures. Both the verbal and the figural tests score for fluency, originality, elaboration, and flexibility. An alternate scoring system for the figural tests also scores for resistance to premature closure, abstract reasoning, synthesis, humor, and many other forms of creative thinking (Torrance, 2000).

Torrance (2000) developed the TTCT on the basis of the following eight psychometric criteria: objectivity, reliability, validity, historical record of success, appropriateness of language, cultural appropriateness, cost of administration, and time for administration. Certainly these eight criteria seem good; one can imagine achieving these criteria to be the goal of any good psychological measure. However, testing is not the only indicator of creativity. Other selection criteria might include the history of the child, the opinions of a judging audience, parent nominations, biographical data, professional judgments, performance tests, observations, and personal interviews. Although Torrance acknowledges the importance of these indicators, he is also quick to point out that they should be considered secondary to a more objective and empirically supported measure such as the TTCT (Torrance, 2000).

Decades of research have used the TTCT, and a recent study by Alexander et. al (1994) provides a good example of the TTCT being applied to story problems, a construct very similar in nature to storytelling. The researchers sampled 100 children in a Head Start program and read to the children one realistic and one fanciful story. In each story, the main character wound up in a predicament that required problem solving in order to find a solution. The children gave their solutions to each story, and these solutions were transcribed. The TTCT was then used to measure the fluency, elaboration, flexibility, and

originality of the children's solutions. In addition, the researchers also rated the effectiveness and realism of the children's solutions. The children's story preferences were also recorded. Results indicated that the children's solutions to both stories were realistic and unoriginal and that those who preferred the fanciful story gave more flexible solutions to the fanciful story than those who preferred the realistic story.

The TTCT has also been used by Berretta and Privette (1990) to compare the effect of unstructured and structured play activities on creativity. The authors began by sampling 184 children and assigning them to one of six conditions: structured art (completing color-by-number pictures), flexible art (making pictures using a variety of materials), structured drama (reading and acting out a play), flexible drama (making up and acting out a play based on a TV show), structured playground activities (an imitative game), and flexible playground activities (an acting-out charades game). After each condition, participants were given the TTCT. Results indicated that participants who engaged in the flexible activities showed significantly greater creative thinking than participants who engaged in structured activities. It should be noted that the authors were more interested in the structure of an activity rather than the occurrence of pretense per se. However, the nature of the flexible play activities (acting freely in the drama group, pretending to be other creatures or to be doing other things in the playground activities group) were such that pretense behaviors would naturally occur during the activities. Still, the authors did not systematically record the behaviors exhibited by the children; therefore, it can not be determined whether the occurrence of pretense or the lack of structure facilitated creativity.

Although the TTCT has been used for decades to assess the creativity of children, in recent years critics of the TTCT have proposed alternative ways to assess creativity. The

critics of the TTCT are not challenging the reliability of the TTCT; certainly that has been demonstrated over the years in over 2000 studies. However, the reason some criticize the TTCT is because they believe that it is trying to make objective that which by definition is subjective.

Perhaps the most vocal opponent of the TTCT in recent years has been Teresa Amabile. While the TTCT and Amabile's measure of creativity assess the creative products of children; the difference between them concerns the objective versus subjective nature of creativity and how to generalize what is seen in creative products. Amabile (1982b) argued that tests such as the TTCT claim to follow objective, universal criteria. However, unlike a concept such as intelligence, what is deemed creative is subjective, resulting only from the opinions of others. This is not to say that universal constructs such as novelty do not constitute creativity but that the observable products of creativity are judged in a purely subjective manner. In addition, although it is recognized that people, processes, and products can all be considered creative (Amabile and Tighe, 1993), science at this point can only subjectively measure the one observable construct of these three: products (or the behavior leading to them).

Amabile (1982b) formulated a measure of creativity based on the assertion that looking at the products of creative individuals in a subjective manner, not the individuals themselves, is the only way to assess creativity. Although Amabile (1982) was convinced that science must be relegated to observable phenomena, she also stated that creativity is subjective; there are no universal rules as to what makes something creative. In addition, Amabile (1982b) maintained that because creativity is subjectively defined, the best judges of creativity are those who are considered experts in the field they are judging. Thus, art

critics would be the best judges of creative paintings and literature reviewers would be the best judges of creative stories. These judges do not consciously think of theories of creativity when judging; they simply judge creativity when they see it (Baer, 1994). Thus, judges should not compare the product they are critiquing to a universal definition of creativity but to other products they consider creative. Although this might make conclusions from a study sample specific, it is the only way known thus far to assess creativity without relying on an objective, universal definition. Finally, in the experimental setting, judges can also assess two other factors that might affect creativity: the technical aspects and aesthetic appeal of products, the makeup of which depends on the type of product being assessed. It is these three constructs that people consider when judging products (Amabile, 1982b).

Amabile (1982b) hypothesized that experts of a particular field would independently rate the creativity of a set of products similarly, thus establishing reliability and construct validity and supporting the theory that creativity can be subjectively defined by experts. The author began by sampling 22 young girls and having them construct designs with construction paper and glue. These designs were then rated independently by different groups of experts who were not trained in any way prior to the study. The experts included art teachers, art students, and psychology faculty who rated the designs on 23 dimensions of creativity, technical ability, and aesthetic appeal. Results indicated that interjudge reliability for the art teachers was .88, and 16 of the 23 dimensions had reliabilities greater than .8. In addition, factor analysis revealed two different factors: creativity and technical goodness. Subsequent replications had higher interjudge reliabilities and stronger reliabilities of the dimensions, along with more dimensions with high interjudge reliabilities (Amabile, 1982b). In addition, one replication, which had judges assessing the creativity of poetry, revealed a

third factor: style, which consisted of clarity, appropriateness, and consistency (Amabile, 1982b). On the basis of these studies, Amabile later formulated a model of how creative products are produced (Amabile, 1983; Amabile, 1993). Finally, the long-term stability of such assessments has been demonstrated (Baer, 1994).

Amabile's consensual assessment technique has even been used recently by Howard-Jones, Taylor, and Sutton (2002) to assess the effect of structured and unstructured play on the generation of creative products by children. The authors began by sampling 52 children and assigning them to one of two conditions. Children in the first group participated in a structured activity, which consisted of copying text from the classroom board. Children in the second group participated in an unstructured activity. They were given blue salt dough and told that they could do with it what they wanted. After engaging in an activity for 25 minutes, the participants were given a variety of materials and were asked to make a collage. The next day, children from the structured activity engaged in the unstructured activity and vice versa and afterwards were again asked to make a collage. The collages were then given to 10 expert judges to rate only for creativity. Results indicated that the collages produced by children after the unstructured activity were rated as significantly more creative than the collages produced after the structured play activity. Again, the authors were interested in the structure of the play more than the occurrence of pretense, and therefore, the behaviors of the children were not systematically recorded.

Hennessey and Amabile (1988) have also applied the consensual assessment technique to storytelling. The researchers began by sampling 115 children between the ages of 5 and 10. The researchers showed each child a book entitled *A Boy, a Dog, and a Frog* (Mayer, 1967), which contains black and white drawings with no words. The researchers

first went through the book page by page with the child. Afterwards, the child was asked to go through each page and tell a story about what is happening in the book. These stories were tape recorded and transcribed, after which the transcriptions were given to three elementary school teachers who rated the stories on the following dimensions: creativity, liking, novelty, logic, emotion, grammar, detail, and straightforwardness. These dimensions were rated on a five-point continuous scale, with 1 = low, 3 = medium, and 5 = high. The teachers were also instructed to rate the stories against one another rather than against a universal standard of creativity.

Results indicated the following interjudge reliabilities for the eight dimensions: creativity = .91, liking = .88, novelty = .79, logic = .73, emotion = .77, grammar = .85, detail = .89, vocabulary = .89, straightforwardness = .90, and imagination = .88. In addition, factor analysis revealed three factors that the eight dimensions loaded on: creativity, technical goodness, and factual detail. These results suggest that the experts were able to differentiate between creativity, technical goodness, and factual detail when making their judgements (Hennessey & Amabile, 1988).

Although the TTCT has been used for many decades to assess creativity, perhaps it is time to use another measure to look at this construct. It may be that Amabile's argument is valid, that it is more beneficial to view creativity in a subjective manner. However, it may also be true that the TTCT and expert ratings are *both* valuable in assessing creativity, just different components. After all, Amabile did argue that certain aspects of creativity *are* universal, such as novelty. If this is true, then one could certainly imagine these aspects being measured using objective criteria. It might also be true that both measures are assessing the same aspects of creativity and that it is just a matter of which measure is the

most valid. At any rate, it is beneficial to examine the construct of creativity with the newer measure of expert ratings. As stated before, this might make the results of the present study sample-specific, but such is the nature of subjective rating. Although the level of creativity displayed by each child relative to one another would not generalize, as the underlying constructs such as novelty are not being measured, the effect of play on creativity could generalize. Further evidence for using Amabile's assessment technique is shown by its demonstrated reliability and validity through multiple replications.

Taken as a whole, the literature points to a direction to follow to determine whether play affects the creativity of storytelling in preschoolers. A design similar to that used by Dansky and Silverman (1973) could be used to examine the effect of play. Afterwards, children could tell stories, and these stories could be analyzed in the same manner as that outlined by Hennessey and Amabile (1988). However, there are still very important components to look at in order to determine how the experiment should proceed. Although both the independent and dependent variables have been defined and the measures of the experiment have been determined, the structure of the storytelling session must be carefully planned in order to maximize the chance for creativity to occur.

The chance for creative expression that storytelling provides can be drastically impaired if the environment is not conducive to this. Tavalin (1995), while examining a storytelling group of four middle-aged women, found that creativity could easily be stifled if the storyteller was unable to fully utilize the storytelling session for personal expression. This seems easy enough to rectify; simply allow the storyteller to speak as he or she pleases. However, Tavalin (1995) also stressed the importance of an open audience. Therefore, when

listening to a story, an audience must respect the storyteller and allow the story to be presented without interruption.

Although the study by Tavalin (1995) mainly focused on how to maximize creativity in an adult storytelling session, the simple point of allowing the storyteller to be free to speak could certainly be applied to children. However, there are other factors that can stifle the creativity of individuals that can also be generalized to children. These include such stipulations on creativity as evaluation (Amabile, 1979; Amabile, Goldfarb, & Brackfield, 1990) and surveillance (Amabile, Goldfarb, & Brackfield, 1990). Therefore, in an experimental setting, a child should not feel like he or she is being evaluated or watched but rather that the storytelling session is a fun and natural activity. Other research with children has identified other environmental detriments to creativity, including competition (Amabile, 1982a).

On a related note, one aspect of Amabile's research that has drawn heavy criticism concerns the effect of reinforcement on creativity. Whereas Amabile argued that contracts for reward can inhibit creativity (Amabile, Hennessey, & Grossman, 1986), other researchers claimed that results such as this often occur when experimental procedures are flawed (Eisenberger & Selbst, 1994; Eisenberger & Cameron, 1996). The Amabile et. al. (1986) study used a one-time reward that the children knew about in advance, a condition that critics stated might in fact reduce creativity. However, critics also argued that this result can not be generalized to reinforcement as a whole. Research has shown that large rewards presented multiple times with as little attention as possible paid to them in advance can result in significantly increased creativity (Eisenberger & Selbst, 1994; Eisenberger & Cameron, 1996).

Keeping these points in mind, one can propose an experimental method for researching the link between pretense and the level of creativity in children's stories. Although group designs have often been used to examine pretense and creativity, a single-subject design can also be used to examine this relationship. Rather than comparing different play groups on the rated creativity of their stories, a single child could engage in pretense, imitation, and coloring activities, telling a story after each activity. It was hypothesized that on the basis of previous research concerning pretense and associative fluency, the stories a child told after engaging in pretense activities would be ranked and rated by experts as more creative than the stories he or she told after engaging in imitation and coloring activities. Two experiments were conducted to test this hypothesis. In Experiment 1, preschool children engaged in two pretense activities (free play and directed pretense) and two other play activities (coloring and imitation). Because of problems encountered during the first experiment, Experiment 2 was undertaken to retest the hypothesis while fixing the problems of the first experiment. In Experiment 2, kindergarten children engaged in one pretense activity (directed pretense) as well as coloring and imitation activities. The imitation activity was included in both experiments to compare pretense to a different type of play, and the coloring activity was included to make sure that simply engaging in an activity, regardless of whether or not it is pretense, is not a factor that can contribute to the creativity effect.

Experiment 1

Participants

Participants were four children from a preschool in the Washtenaw County area. Because previous research on the dependent measure has been conducted with children no younger than 5 (Hennessey and Amabile, 1988), preschoolers between the ages of 4 and 5 were used for the experiment. Participant 1 was a female child aged 5 years, 0 months. Participant 2 was a female child aged 5 years, 1 month. Participant 3 was a male child aged 4 years, 12 months. Participant 4 was a male child aged 4 years, 3 months. Participants were selected through a two-step process: First, the director of the children's preschool was contacted, the purpose of the experiment was explained, and permission to conduct the experiment was obtained. Afterwards, informed parental consent was obtained from parents electing to have their children participate (see Appendix A). Consent forms were given to the preschool teacher, who distributed them to parents. Those children whose parents granted permission participated in the experiment.

Judges of the stories were three experts with early childhood education backgrounds working in Southeast Michigan. Judge 1 was a curriculum coordinator with 15 years experience teaching preschool children. Judge 2 was a preschool teacher with two years teaching experience. Judge 3 was the director of an elementary school with a master's degree in early childhood education, three years experience teaching kindergarten children, and three years experience as a program director. Previous research (Amabile, 1982b; Hennessey and Amabile, 1988; Howard-Jones, Taylor, and Sutton, 2002) suggests that these individuals would be considered experts on the creativity of preschoolers. Judges were

obtained by calling local elementary schools. One judge was randomly selected to receive a fifty-dollar gift certificate to a local department store for his or her participation.

Stimulus Materials

The stimulus materials were similar to those used by Dansky and Silverman (1973). The nature of these materials is such that changes in the dependent variable would probably not be due to the stimulating nature of the materials themselves but due to the type of play the children engaged in with these materials. The materials used during the all sessions but baseline included a plastic screwdriver and screws set, a roll of paper towels, a pile of paper clips, a stack of blank cards, 2 empty matchboxes, a pile of small sticks, and 6 wet plastic cups. All activity materials were set on a table that the children could reach. The materials used during the baseline coloring activity were a box of crayons and two sketches of geometric shapes to color. The experimenter also presented each child with a picture from a book entitled *A Boy, a Dog, and a Frog* (Mayer, 1967) during the storytelling portion of each session. Each picture consisted of a black and white drawing with no words.

General Procedure and Research Design

The procedure was similar to that used by Dansky and Silverman (1973) and Li (1978), altered to fit a single-subject design. The experimenter first met with each child individually one week prior to the proper experiment in order to familiarize him or her with the experimenter and the experimental situation. During these sessions, the experimenter and the child sat at the table that would be used to present the stimulus materials during the actual experiment. Each child was invited to color with the experimenter for 10 minutes. Praise was periodically delivered contingent on activity participation; statements were descriptive rather than evaluative in content. Afterwards, the experimenter thanked each child for

coloring with him and stated that he looked forward to seeing the child again the following week.

During the actual experiment, each child met with the experimenter individually for a session lasting approximately 15 minutes, for a total of 15 sessions. One session per day during the weekday occurred for each child; if a child missed a day, the sessions resumed as normal on the following day. For each session, the child met with the experimenter during the school day in a schoolroom that was used by the child's class for activities on a semiregular basis. Each session consisted of the child engaging in one of four activities followed by a storytelling portion.

On the first three days, each child participated in a baseline session. On days 4 through 6, Participants 1 and 2 were scheduled to engage in directed pretense sessions, whereas Participants 3 and 4 were scheduled to engage in imitation sessions. Day 7 consisted of a baseline session for all children. For days 8 through 10, all participants were scheduled to engage in free play sessions. Day 11 consisted of a baseline session for all children. For days 12 through 14, Participants 1 and 2 were scheduled to engage in imitation sessions, whereas Participants 3 and 4 were scheduled to engage in directed pretense sessions. The order of free play and imitation days were counterbalanced; however, because the addition of a fourth condition (directed pretense) occurred late in the design, counterbalancing for all experimental conditions was not achieved. The assignment of counterbalancing was randomly determined. After the activity portion of each session, the children participated in the storytelling portion of the session.

On the basis of prior research concerning play and associative fluency, it was hypothesized that the free play and directed pretense sessions would produce stories that

were ranked and rated as more creative than the stories told during the imitation and coloring sessions. The imitation sessions were included to compare pretense to a different type of play, as previous research suggests that pretense is both more creative and can stimulate more creative behavior than can other forms of play (Dansky, 1973). The coloring sessions were included as a nonplay control condition to make sure that simply engaging in an activity, regardless of whether or not it is pretense, is not a factor contributing to the creativity effect.

Baseline Condition

During the baseline sessions, each child was given the box of crayons and two sketches of geometric shapes and was permitted to color as he or she wished. Comments that concerned the child's coloring and were not evaluative were periodically delivered. The coloring activity lasted for approximately 10 minutes, after which the experimenter and child moved on to the storytelling portion of the session.

The storytelling portion was similar to that in the Hennessey and Amabile (1988) study. First, the experimenter told the child that he was going to show the child a picture from the book. A different picture was shown during each storytelling portion, and the order of pictures used was randomly determined. After the picture was presented, the child was asked to tell a story to the experimenter based on the picture he or she saw. These stories were videotaped. Comments that concerned the child's story and were not evaluative were periodically delivered. The storytelling portion of each session lasted between 1 and 2 minutes.

Directed Pretense Condition

During the directed pretense sessions, each child was first presented with the stimulus materials by the experimenter and told, “Here are our play-things. We can pretend they’re whatever we want them to be. What could we do to play pretend with them?” Each child was then allowed to play with all of the objects. Comments that concerned the pretense behaviors of the child and were not evaluative were periodically delivered. If a child did not interact once with one of the stimulus materials, the child was given one prompt in the 4th and another in the 7th minute of the play period stating that the ignored objects could also be played with. The directed pretense activity lasted for approximately 10 minutes, after which the experimenter and child moved on to the storytelling portion of the session.

Free Play Condition

During the free play sessions, each child was first presented with the stimulus materials by the experimenter and told, “You may play with all of these things. Do whatever you would like to do with them.” Each child was then allowed to play with all of the objects. Comments that concerned the pretense behaviors of the child and were not evaluative were periodically delivered. If a child did not interact once with one of the stimulus materials, the child was given one prompt in the 4th and another in the 7th minute of the play period stating that the ignored objects could also be played with. The free play activity lasted approximately 10 minutes, after which the experimenter and child moved on to the storytelling portion of the session.

Imitation Condition

During the imitation sessions, each child was instructed to watch the experimenter perform four tasks with the stimulus objects: turning the plastic screws with the plastic

screwdriver, putting cards together with the paper clips, wiping the wet cups with the paper towels, and putting small sticks in the empty match boxes. The presentation of these tasks was randomly ordered for each session. After a task was modeled, the child was asked to repeat the actions of the experimenter. Brief praise was delivered contingent upon attempt to imitate, although the child was prompted again to try imitating the actions of the experimenter if he or she did not do so initially. Time spent with each stimulus object was roughly equivalent. The imitation activity lasted for approximately 10 minutes, after which the experimenter and child moved on to the storytelling portion of the session.

The Dependent Measure

After all the stories told by the children were transcribed from recordings, copies of the transcribed stories were given in random order to the three judges. The judges were presented with groups of seven stories and were asked to read each story in the group once. Afterwards, the judges went back and ordered the stories from most to least creative, using their own subjective definitions of creativity. The judges were then given another set of stories and were asked to read them and afterwards fit them into the previously established order. This was repeated until all stories were ordered on the basis of creativity. Once this was accomplished, the judges were asked to go through and write the ranking for each story on the cover page of the story (see Appendix B). Judges were also told which picture each story was about and were given copies of the pictures in order to help establish contexts for the stories.

After the stories were ranked, the judges then went back through the stories and rated the creativity of the stories on a 10-point continuous scale, with 1 = lowest and 10 = highest. Again, the judges were asked to rate the stories using their own subjective definitions of

creativity. In addition, the judges were instructed to rate the stories relative to one another as opposed to rating them against a universal standard of creativity. Therefore, the judges were instructed to rate at least one of the stories a *1* and one of the stories a *10*. Judges were encouraged to spread the ratings out along the entire rating scale. Judges were also told that their ratings did not have to be evenly distributed and that ratings could go in-between boxes.

Once all of the stories were rated for creativity, each judge was asked to rate how “story-like” each story was. This was done using a three-point scale, with 1 = “has no elements of being a story for a 4-5 year old,” 2 = “this is a primitive story for a 4-5 year old,” and 3 = “this is a typical story for a 4-5 year old.” This was implemented due to the brevity of the stories, which usually consisted of only a couple of sentences. Because the stories were so short, it was unclear whether the children were actually telling what could technically be considered stories. Judges were instructed that, unlike with the creativity ratings, they should rate the stories on the basis of a universal standard of how “story-like” each story was. Because of this, judges were able to rate all stories as they wished; no story needed to be rated a 1 or a 3 if the judge felt that it did not deserve such a rating. The judges were also told that they needed to only mark in the boxes and not in-between.

The experimenter met with each judge individually, and judges made their assessments independent of one another. Judges were able to make their assessments during one meeting. The meetings lasted between 2 and 4 hours and took place at a location that was convenient for each judge. Judges were not aware of the purpose of the study, nor were they aware of the conditions and the children the stories were from.

Ratings of Behavior

Two observers used a measure similar to that used by Dansky (1980) to examine the videotapes and determine the type and amount of activity each child engaged in (see Appendix C). The observers were undergraduate psychology majors and were trained over the course of three 2-hour sessions. Observer 1 rated all activity sessions, whereas Observer 2 only rated one of each type of activity session (free play, directed pretense, imitation, and coloring) per child. Observers examined the participants and recorded at 15-second intervals the presence of various activity behaviors, for a total of 40 intervals per session. Initially, the behaviors included three categories of pretense behaviors (role-play, object transformation, and object acts), along with the categories of coloring, other constructive play (involves creating or building something out of several objects), and imitation. However, during training an acceptable overall percent agreement of 0.8 could not be established with these categories. The three pretense behaviors were then collapsed into an overall pretense category, and raters went through training again. Categories in the original Dansky (1980) study that involved coding social interactions between children (verbal communication and nonverbal interaction) were not included in the present study. If a child participating in the imitation or coloring activities engaged in a pretense behavior during 10% or more of the time intervals, the story from that session was not to be used in the analysis of data. As with the judges, the observers were not aware of the purpose of the study. Observers were also unaware of the storytelling portion that took place immediately after each condition.

Results

There were two measures of story creativity: judges' rankings (with a high ranking of 30, reverse scored for graphing purposes) of all stories across participants and judges'

relative ratings for each story on a 10-point scale (1 = low, 10 = high) for each participant. Results of these two measures are presented separately for each participant. In addition, each story was judged on a three-point scale that assessed the extent to which a story resembles a true story. Due to participant attrition, the data for only two subjects are presented here. Participant 2 refused to participate in the study after day 10, whereas Participant 3 failed to return to school after day 8 due to circumstances unknown. Therefore, judges only examined the stories from Participants 1 and 4.

The results of the judges' rankings of creativity for Participant 1 are presented in Figure 1, whereas the results of the judges' ratings of creativity for Participant 1 are presented in Figure 2. According to these measures, the three judges noted a similar increasing trend in creativity across initial baseline, with a mean ranking of 14.67 for Judge 1, 8.67 for Judge 2, and 12 for Judge 3. The mean ratings given by the judges for the initial baseline condition were 6, 5, and 4, respectively. Mean creativity rankings remained at a similar level during the directed pretense condition, with a mean ranking of 13.33 for Judge 1, 12.67 for Judge 2, and 13 for Judge 3. The mean ratings given by the judges for the directed pretense condition were 6, 5.67, and 4.67, respectively. During the baseline probe condition (session 7), judges' rankings and ratings fell within the ranges established during the previous conditions. According to all judges' rankings and ratings, creativity steadily increased throughout the free play condition, with a mean ranking of 16.67 for Judge 1, 20.67 for Judge 2, and 16.67 for Judge 3 and a mean rating of 7 for Judge 1, 8.33 for Judge 2, and 5.33 for Judge 3. Creativity neither increased nor decreased overall during the next baseline probe condition (session 11), with rankings of 22, 24, and 24 and ratings of 9, 9, and 8, respectively. During the imitation condition, creativity rankings initially decreased and then

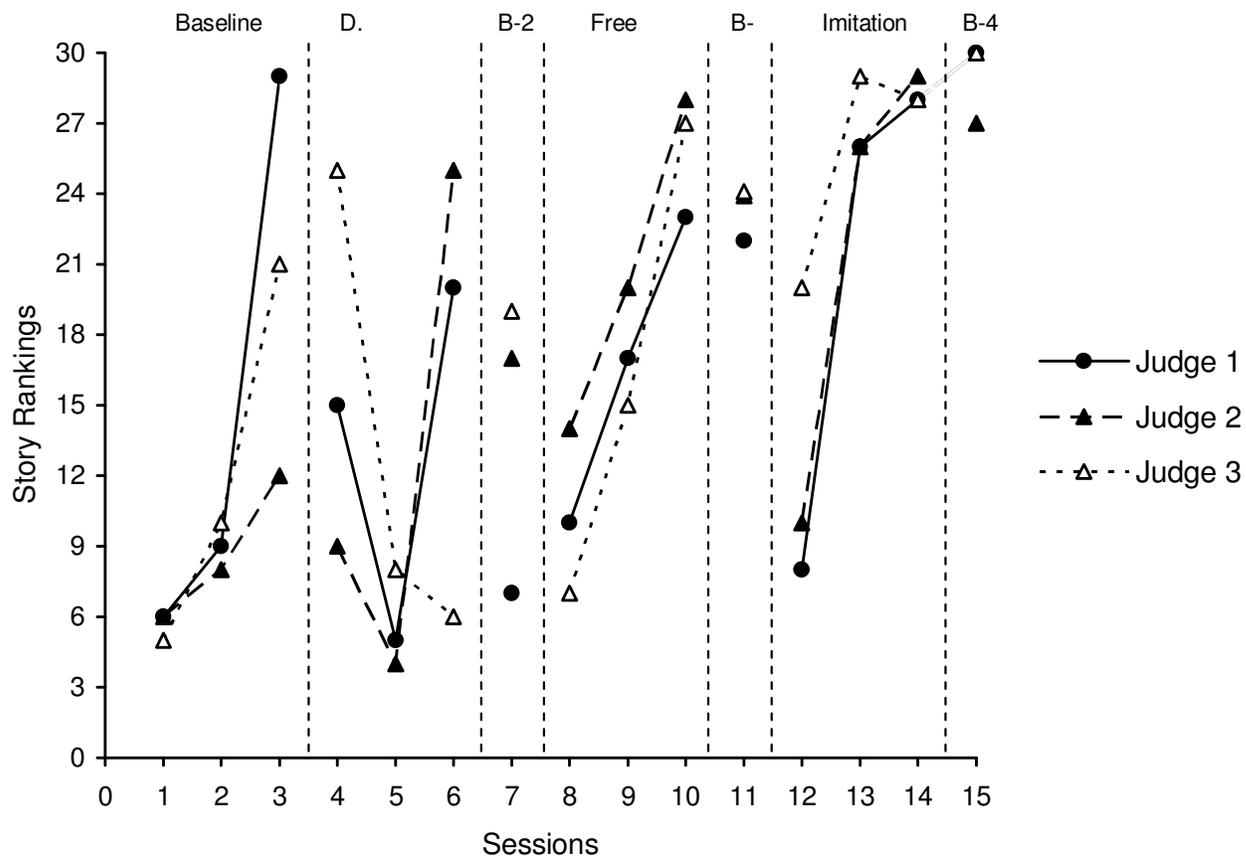


Figure 1. Creativity rankings of the stories told by Participant 1 in Experiment 1. Stories were from 15 separate sessions and were ranked relative to one another and to the stories told by Participant 4. Participant 1 engaged in four separate types of sessions: baseline, directed pretense (“D. Pretense”), free play, and imitation. Three different judges ranked the stories.

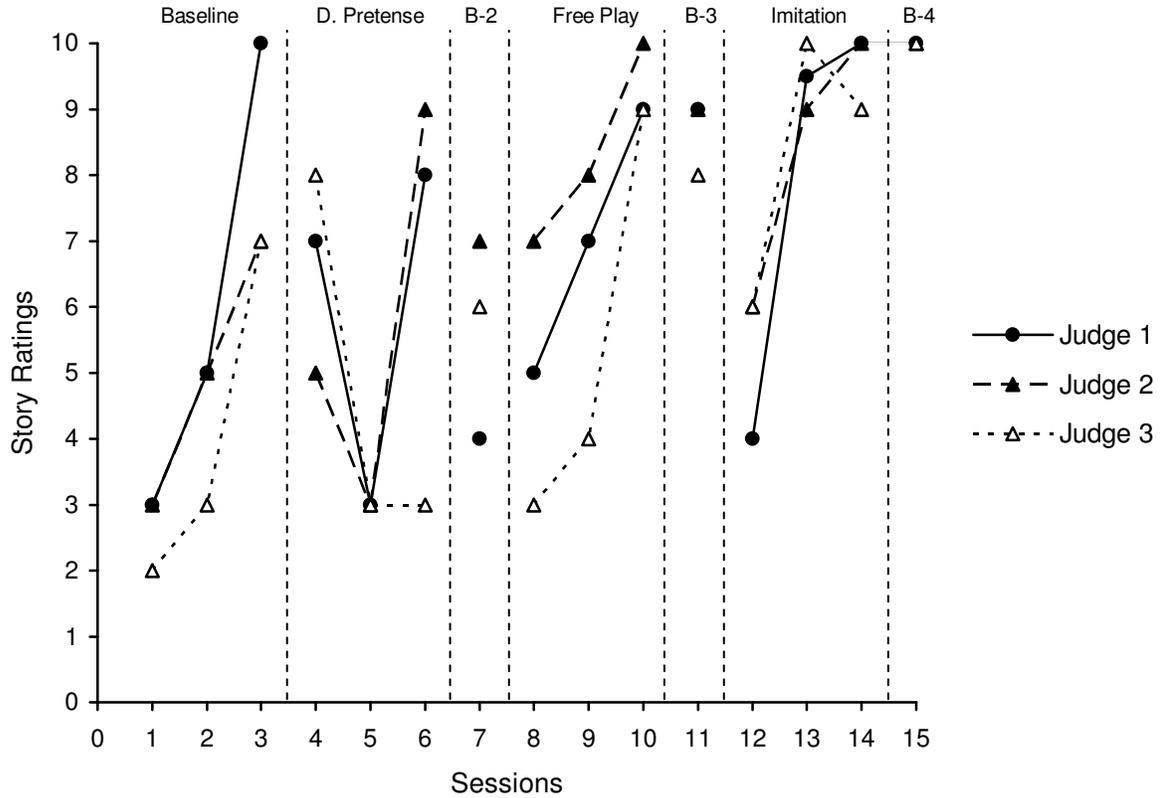


Figure 2. Creativity ratings of the stories told by Participant 1 in Experiment 1. Stories were from 15 separate sessions and were rated relative to one another and to the stories told by Participant 4. Participant 1 engaged in four separate types of sessions: baseline, directed pretense (“D. Pretense”), free play, and imitation. Three different judges rated the stories.

increased dramatically but were still in the range previously established with mean rankings of 20.67, 21.67, and 25.67 for the three judges. The mean ratings given by the judges for the imitation condition were 7.83, 8.33, and 8.33, respectively. During the last baseline probe (session 15), the story was ranked and rated as high or higher than previous stories, with rankings of 30, 27, and 30 and three ratings of 10. Thus, according to the ranking and rating methods, Participant 1 showed no difference between conditions. It appears that an experience effect was occurring for Participant 1, with a tendency for increasing trends both within and across conditions.

The results of the judges' rankings of creativity for Participant 4 are presented in Figure 3, whereas the results of the judges' ratings of creativity for Participant 4 are presented in Figure 4. According to these measures, the three judges noted a similar trend in creativity across initial baseline, with mean rankings of 17.67 for Judge 1, 15.67 for Judge 2, and 16 for Judge 3. The mean ratings given by the judges for the initial baseline condition were 7.33, 7.33, and 4.67, respectively. Mean creativity rankings remained at a similar level during the imitation condition, with mean rankings of 19.67 for Judge 1, 23 for Judge 2, and 17 for Judge 3. The higher mean ranking for Judge 2 is mainly due to the ranking this judge gave to the first imitation session (ranking = 30). The mean ratings given by the judges for the imitation condition were 8, 8.67, and 5.33, respectively. During the baseline probe condition (session 7), the judges' rankings and ratings fell far below the ranges established during the previous conditions. During the free play condition, creativity rankings initially increased and then decreased dramatically, staying within the lower end of the range previously established with mean rankings of 10.33, 6.33, and 9.67 for the three judges. The mean ratings given by the judges for the free play condition were 4.33, 3.67, and 3,

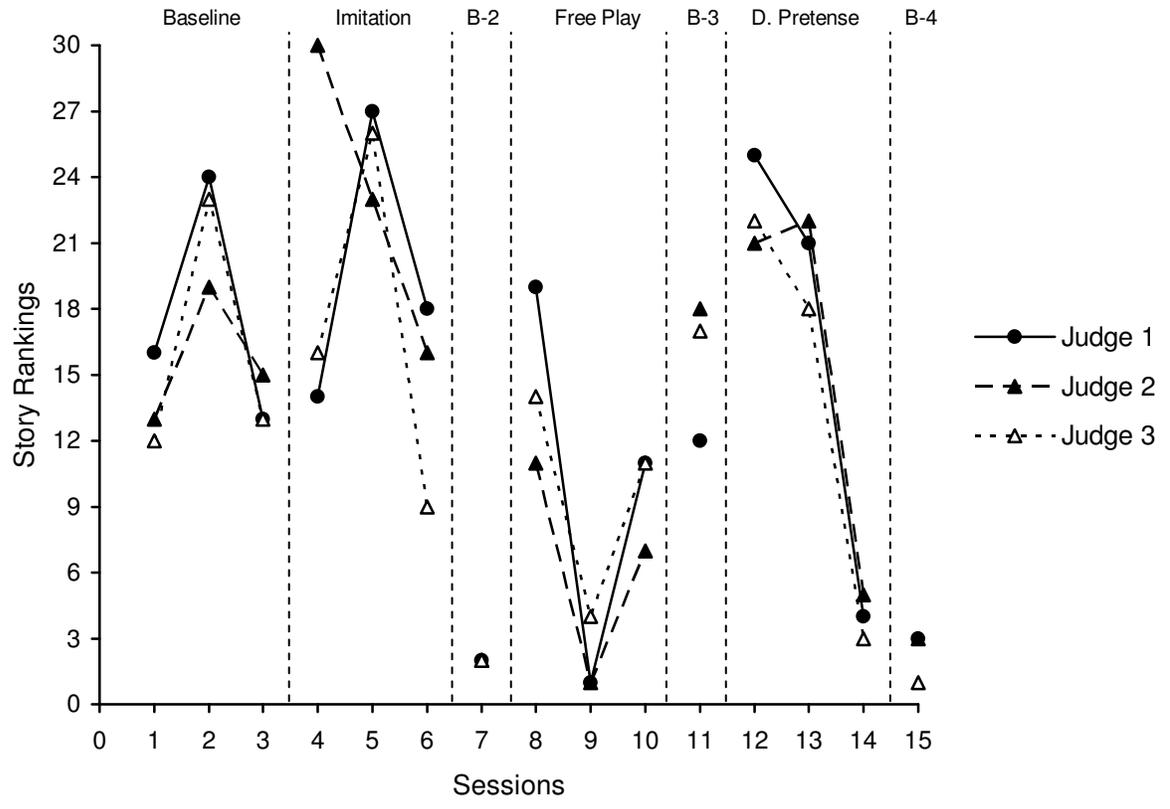


Figure 3. Creativity rankings of the stories told by Participant 4 in Experiment 1. Stories were from 15 separate sessions and were ranked relative to one another and to the stories told by Participant 1. Participant 4 engaged in four separate types of sessions: baseline, imitation, free play, and directed pretense (“D. Pretense”). Three different judges ranked the stories.

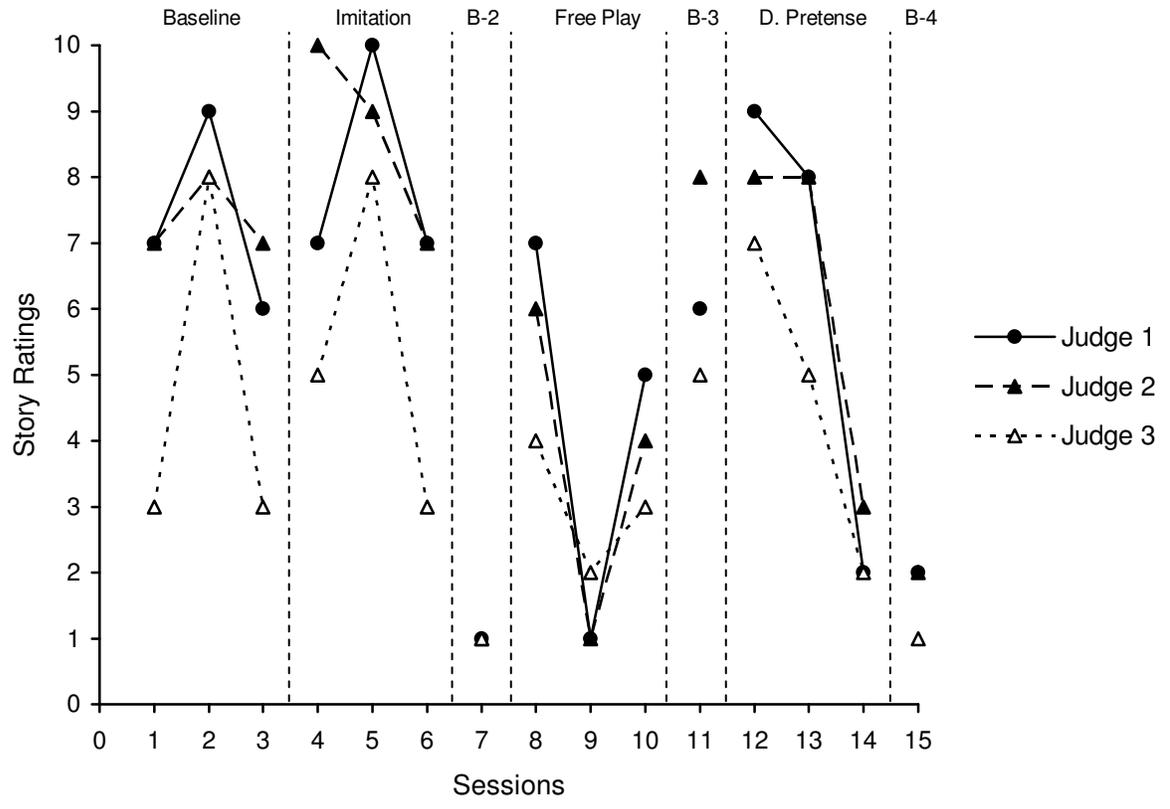


Figure 4. Creativity ratings of the stories told by Participant 4 in Experiment 1. Stories were from 15 separate sessions and were rated relative to one another and to the stories told by Participant 1. Participant 4 engaged in four separate types of sessions: baseline, imitation, free play, and directed pretense (“D. Pretense”). Three different judges rated the stories.

respectively. During the next baseline probe condition (session 11), the judges' rankings and ratings fell within the ranges established during the previous conditions. According to all judges' rankings, creativity decreased throughout the directed pretense condition but was still in the range previously established with mean rankings of 16.67, 16, and 14.33 for the three judges. The mean ratings given by the judges for the directed pretense condition were 6.33, 6.33, and 4.67, respectively. During the last baseline probe (session 15), the story was ranked and rated lower than most of the previous stories, with rankings of 3, 3, and 1 and ratings of 2, 2, and 1, respectively. Thus, according to the ranking and rating methods, Participant 4 showed little difference between conditions. Overall, the creativity rankings and ratings for the imitation condition were slightly higher than those for the other conditions. In addition, the creativity rankings and ratings for the free play condition were lower than those for the other conditions.

To determine the interrater reliability for the rankings given by judges, the coefficient of concordance, an index of agreement, was calculated for each child (Winer, 1962). To do this, the rankings for each child were first entered into separate 3x15 tables, with the rows representing the judges and the columns representing the stories. Because judges ranked the stories for both children together, the data needed to be transformed so that interrater reliability could be determined for each individual child. For example, the highest (reversed) ranking Participant 1 received from Judge 1 was 30, whereas the lowest (reversed) ranking Participant 1 received from this same judge was 5. Transformed, these rankings became 1 and 15, respectively, with all other stories told by that child falling in between. This was repeated for the other two judges, and the data were analyzed. The coefficient of concordance was calculated for each child. These were then added together and divided by

the number of children to determine the mean coefficient of concordance. The mean coefficient of concordance was .853. To determine the mean intercorrelation between the rankings assigned by judges, the intercorrelation between rankings for each child was calculated, added together, and divided by the number of children. The mean intercorrelation was .779.

To determine the interrater reliability for the ratings given by the three judges, the Spearman-Brown prediction formula was used (Winer, 1962). Reliability determined by this formula is known as effective, or aggregate, reliability, and reflects the agreement between all three judges (Rosenthal & Rosnow, 1991). As with the rankings, the data for the ratings of each child were analyzed separately. Ratings were entered into 15x3 tables, with the rows representing the stories and columns representing the judges. The effective interrater reliability was calculated for each child. These reliabilities were then added together and the total was divided by the number of participants to determine the mean effective interrater reliability. The mean effective interrater reliability for the judges' ratings was .914. However, because zero points did not exist on the scale, the interrater reliability was recalculated to adjust for differences in frame of reference for each child. Adjusted for differences in frame of reference, the mean effective interrater reliability for the judges' ratings was .921. To determine the mean reliability of a single rating for the adjusted data, r_1 was calculated (Winer, 1962) for each child, these reliabilities were added together, and the total was divided by the number of participants. The mean reliability of a single rating for the adjusted data was .797.

Also relevant to the validity of these measures of creative storytelling was the rating of how much a story resembled a true story. For almost all stories across all participants, the

judges gave a rating of 3, which indicated a judgement that the story was typical of a story. Out of a total of 90 judgments (15 stories per participant, two participants, and three judges), there were 16 exceptions. Fourteen of these exceptions rated stories as a 2, indicating that the story had at least some of the elements of a true story. Only two stories, according to only one judge, were rated 1, meaning they were rated as having no elements of a story. Thus, there is a fair degree of agreement that judges were ranking and rating the creativity of actual stories.

Behavioral observation data from Observer 1 indicated that participants were usually engaged in the activity prescribed for each condition during that condition. The presence of coloring behaviors during baseline sessions was recorded in an average of 86.25% and 72.083% of the time intervals for Participants 1 and 4, respectively. The presence of imitation behaviors during imitation sessions was recorded in an average of 83.33% and 80.833% of the time intervals for Participants 1 and 4, respectively. The presence of pretense behaviors during directed pretense sessions was recorded in an average of 95% and 70% of the time intervals for Participants 1 and 4, respectively. The presence of pretense behaviors during free play sessions was recorded in an average of 73.333% and 5.833% of the time intervals for Participants 1 and 4, respectively.

Two different formulas were used to calculate interrater reliability between behavior observers. The first method of calculating the overall interrater reliability first involved calculating the overall percentage of agreement between the two observers for each session. This involved adding the number of 15-second time intervals observers agreed on and dividing by the total number of intervals (40) for that session (Foster & Cone, 1986). All eight reliabilities were then added together and divided by 8 to obtain the overall interrater

reliability. When this method was used, the overall interrater reliability was 0.853. The second method of calculating the overall interrater reliability first involved the calculation of Cohen's Kappa for each session. When interrater reliability was calculated this way, the final interrater reliability was corrected for chance agreement (Foster & Cone, 1986). After Cohen's Kappa was calculated for each session, all eight reliabilities were added together and divided by 8 to obtain the overall interrater reliability. The overall interrater reliability using this method was .347.

Discussion

The hypothesis that the stories a child told after engaging in free play and directed pretense activities would be rated by experts as more creative than the stories they told after engaging in imitation and coloring activities was not supported by the present study. Rather, by looking at the graphs of the story rankings and ratings for each child, one can see that the creativity of the stories told by both participants seemed to fluctuate from day to day. Participant 1 appeared to display overall more creativity as time went on, especially within each condition, with the highest levels of creativity occurring in the imitation and last baseline sessions, whereas Participant 4 appeared to display less creativity overall as time went on across conditions, with the lowest levels of creativity occurring during the second baseline, free play, directed pretense, and last baseline sessions. However, for no group of sessions was creativity stable for either participant.

This variability in creativity does not seem to be associated with disengagement from the activities themselves. For almost all of the sessions, the participants were engaged in the activity meant for that session for the majority of time intervals for that session. At no time did Observer 1 record a participant engaged in a pretense behavior during an imitation or

baseline session. Therefore, it is probably not general disengagement from the activities or engaging in other-than-programmed activities that explains the level of creativity of the stories told by the participants.

Although the participants did appear to be engaged in the proper activity for most of the sessions, one group of sessions did pose a problem for this experiment: the free play sessions. As stated previously, these sessions were added to the experiment late in its development to compare the results of the free play sessions to the results of the directed pretense sessions, as Li (1978) did. However, carry-over effects may have occurred because of this. Participant 1 had engaged in the directed pretense sessions prior to the free play sessions; as a result, she exhibited pretense behaviors during the majority of the time intervals of the free play sessions. Participant 4 had engaged in the imitation sessions prior to the free play sessions. As a result, he engaged in behaviors similar to the imitation behaviors during the free play sessions. He spent much of his time wiping cups, clipping cards together, and moving sticks from one box to another. None of these behaviors fell under the coding categories. Also, as stated before, Dansky (1980) later found that it is engagement in pretense, not the lack of play structure, which affects creativity. Because of this, the problem of carry-over effects, and the desire to shorten the experiment in general to guard against participants' becoming bored, the free play sessions were eliminated from Experiment 2.

Boredom of participants was displayed during this experiment, but again it is not an adequate explanation for the failure to support the hypothesis. Although Participant 3 was dropped from the experiment because he did not show up to school for a number of days, Participant 2 showed a lack of interest early on and eventually refused to participate

altogether. All participants expressed boredom at one point or another during the baseline sessions. Although geometric shapes were used to guard against the occurrence of pretense behaviors during the baseline sessions, they could have affected the results by causing a lack of interest on behalf of the participants. Because the current study attempted to collect multiple session data, the problem of boredom may not have been encountered in the previous research. The coloring pictures were changed for Experiment 2, which included a wider variety of pictures for the participants to color.

The interrater reliability of the judges, comparable to that of previous research, was quite high. This is noteworthy because this measure has traditionally been used for group designs. In a single-subject design, one would expect less variability in the stories told, and therefore, there would be a need for more sensitive measures. The rankings and ratings of the judges were also comparable to one another, with reliability a little lower for the rankings. This is to be expected, as judges had more choices at their disposal for the rankings than for the ratings.

The interrater reliability between the behavior coders posed some problems. Because the traditional formula of dividing the total number of intervals by intervals of agreement was used to calculate interrater reliability, interrater reliability was sufficiently high. However, when Cohen's Kappa, which controls for chance agreement, was calculated, reliability was found to be quite low. This is probably due to the fact that the raters were coding behaviors that were very frequent in occurrence. Because of this, the potential occurrence of chance agreement rises considerably. Both raters coded a similar proportion of occurrence of a particular behavior from session to session, however. Also, this result, programmed to ensure the fidelity of the intervention, does little to explain the general finding that pretense

did not seem to enhance creativity in this experiment. If such an effect had been found, an acceptable interrater reliability for the fidelity check would give confidence that the effect was due to pretense. However, because no such effect was found, the interrater reliability between the behavior coders is of little consequence.

There was much concern about the stories told by the children; specifically, the stories were very short on average. Most stories consisted of only one to four sentences, the least creative often being a simple one-sentence description of the picture. For many of the sessions, the participants appeared to have difficulty formulating stories to go with the pictures, often having to think in silence for a bit to even come up with a story of a few sentences. Two of the stories were rated by the judges with a 1, or “has no elements of being a story for a 4-5 year old.” This could very well be because the single picture was not a sufficient stimulus during some of the sessions. Single pictures from the storybook used in previous research were used to accommodate the single-subject design format. Because the experiment was not a group design, more than one stimulus for the storytelling portion was needed. Single pictures from the same book were chosen because they were relatively equivalent in content. However, previous research suggests that young children who are asked to tell a story with pictures have difficulty doing so unless there are multiple pictures with which to do so and the pictures are unambiguous in both situation and sequence (Bornens, 1990; Shapiro & Hudson, 1991). Therefore, sets of sequential pictures were used for Experiment 2.

The results of Experiment 1 are also inconclusive because only two participants finished the experiment. With so little data, it is hard to draw any solid conclusions about the results of the experiment. The design problems previously addressed also make it hard to

draw any conclusions about the results. Therefore, a second experiment was undertaken, this time with kindergarten children from the same school as were the preschool children from Experiment 1. Experiment 2 was designed to control for the design problems seen during Experiment 1, specifically providing multiple pictures for the storytelling activity, enhancing the engagement of the baseline activity (coloring), eliminating the possibility of carryover effects from free play, and controlling for developmental limitations on creative storytelling. Also, Experiment 2 was a simpler design, eliminating predictions regarding the effects of free play. This was decided upon because the experimental variable of interest, pretense, appeared to be captured adequately during the directed pretense sessions. As with Experiment 1, the hypothesis of Experiment 2 was that the stories a child told after engaging in pretense activities would be ranked and rated by experts as more creative than the stories they told after engaging in imitation and coloring activities.

Experiment 2

Participants

Participants were 4 kindergarten-age children from the same school as were the preschool-age children in Experiment 1. Participant 1 was a male child aged 5 years, 4 months. Participant 2 was a male child aged 5 years, 8 months. Participant 3 was a female child aged 5 years, 9 months. Participant 4 was a female child aged 5 years, 4 months. As with Experiment 1, informed parental consent was obtained from parents electing to have their children participate. Consent forms were given to the kindergarten teacher, who distributed them to parents (see Appendix D). Those children whose parents granted permission participated in the experiment.

Stimulus Materials

The stimulus materials were mostly the same materials as in Experiment 1. As with Experiment 1, the materials used during the all sessions but baseline included a plastic screwdriver and screws set, a roll of paper towels, a pile of paper clips, a stack of blank cards, 2 empty matchboxes, a pile of small sticks, and six wet plastic cups. All activity materials were set on a table that the children could reach. The materials used during the baseline coloring activity consisted of a box of crayons and four sketches to color. The sketches were taken from a children's coloring book and consisted of common objects the participants would recognize (food items, buildings, toys, etc.). In order to minimize the chance of pretense occurring during the baseline sessions, sketches that contained people were not used.

During the storytelling portion of each session, the experimenter also presented each child with a set of pictures from one of six children's books. The books were *A Boy*, *a Dog*,

and a Frog (Mayer, 1967), *Frog, Where Are You?* (Mayer, 1969), *A Boy, a Dog, a Frog, and a Friend* (Mayer & Mayer, 1971), *Frog on His Own* (Mayer, 1973), *Frog Goes to Dinner* (Mayer, 1974), and *One Frog Too Many* (Mayer & Mayer, 1975). Each picture consisted of a black and white drawing with no words. Each set consisted of 11 sequential pictures, with two sets taken from each book. The exception to this was *A Boy, a Dog, and a Frog* (Mayer, 1967); only one set of pictures was taken from this book. Therefore, there were a total of 11 sets of pictures, one for each session. This series of books was used because the number and types of characters, as well as the number of situations and dynamic events in each book, were roughly equivalent.

General Procedure and Research Design

The procedure was similar to that used during Experiment 1. The experimenter first met with each child individually one week prior to the proper experiment in order to familiarize him or her with the experimenter and the experimental situation. During these sessions, the experimenter and the child sat at the same table that would be used to present the stimulus materials during the actual experiment. Each child was invited to color with the experimenter for 10 minutes. Praise was periodically delivered contingent on activity participation; statements were descriptive rather than evaluative in content. Afterwards, the experimenter thanked the child for coloring with him and stated that he looked forward to seeing the child again the following week.

During the actual experiment, each child met with the experimenter individually for a session lasting approximately 15 minutes, for a total of 11 sessions. One session per day during weekdays occurred for each child; if a child missed a day, the sessions resumed as normal on the following day. For each session, a child met with the experimenter during the

school day in a schoolroom that was used by the child's class for activities on a semiregular basis; this was the same room used during Experiment 1. Each session consisted of the child engaging in one of three activities followed by a storytelling portion.

On the first three days, each child participated in a baseline session. On Days 4 through 6, Participants 1 and 3 engaged in imitation sessions, whereas Participants 2 and 4 engaged in pretense sessions. Day 7 consisted of a baseline session for all children. For Days 8 through 10, Participants 1 and 2 engaged in imitation sessions, whereas Participants 3 and 4 engaged in pretense sessions. After the activity portion of each session, the children participated in the storytelling portion of the session.

On the basis of prior research concerning play and associative fluency, it was hypothesized that the pretense sessions would produce stories that were ranked and rated as more creative than the stories told during the imitation and coloring sessions. As with Experiment 1, the imitation sessions were included to compare pretense to a different type of play. The coloring sessions were again included as a nonplay control condition to make sure that simply engaging in an activity, regardless of whether or not it is play, is not a factor contributing to the creativity effect.

Baseline Condition

During the baseline sessions, each child was given the box of crayons and four sketches from a coloring book and was permitted to color as he or she wished. Comments that concerned the child's coloring and were not evaluative were periodically delivered. The sketches were grouped prior to the start of the experiment, and the set of sketches a child received during each baseline session was randomly determined. The coloring activity lasted

for approximately 10 minutes, after which the experimenter and child moved on to the storytelling portion of the session.

The storytelling portion was similar to that in the Hennessey and Amabile (1988) study. First, the experimenter told the child that he was going to show him or her a set of pictures and would like the child to think of a story to go with the pictures while they were being shown. A different set of pictures was shown during each storytelling portion, and the order of picture sets used was randomly determined. The experimenter then went through the set of pictures one picture at a time in silence, changing pictures when the child stated that he or she was ready to see the next one. Once all of the pictures were shown, the child was asked to tell a story to the experimenter on the basis of the picture he or she saw by stating “one thing” about each picture, as in the Hennessey and Amabile (1988) study. The experimenter then went back to the beginning of the picture set and went through the pictures again, with the child telling his or her story to the experimenter. These stories were videotaped. Nonevaluative comments concerning the child’s story were periodically delivered. The storytelling portion of each session lasted between 5 and 10 minutes.

Pretense Condition

The pretense sessions were similar to the directed pretense sessions in Experiment 1. During the pretense sessions, each child was first presented with the stimulus materials by the experimenter and told, “Here are our play-things. We can pretend they’re whatever we want them to be. What could we do to play pretend with them?” Each child was then allowed to play with all of the objects. Comments that concerned the pretense behaviors of the child and were not evaluative were periodically delivered. If a child did not interact once with one of the stimulus materials, the child was given one prompt in the 4th and another in

the 7th minute of the play period stating that the ignored objects could also be played with. The pretense activity lasted for approximately 10 minutes, after which the experimenter and child moved on to the storytelling portion of the session.

Imitation Condition

During the imitation sessions, each child was instructed to watch the experimenter perform four tasks with the stimulus objects: turning the plastic screws with the plastic screwdriver, putting cards together with the paper clips, wiping the wet cups with the paper towels, and putting small sticks in the empty match boxes. The presentation of these tasks was randomly ordered for each session. After a task was modeled, the child was asked to repeat the actions of the experimenter. Brief praise was delivered contingent upon attempt to imitate, and the child was prompted again to try imitating the actions of the experimenter if he or she did not do so initially. Time spent with each stimulus object was roughly equivalent. The imitation activity lasted for approximately 10 minutes, after which the experimenter and child moved on to the storytelling portion of the session.

The Dependent Measure

After all the stories told by the children were transcribed from recordings, copies of the transcribed stories were given in random order to the three judges from Experiment 1. As with the preschool stories, the judges were presented with groups of seven stories and were asked to read each story in the group once. Afterwards, the judges went back and ordered the stories from most to least creative, using their own subjective definitions of creativity. Once this was completed, the judges were given another set of stories and were asked to read them and afterwards fit them into the previously established order. This was repeated until all stories were ordered on the basis of creativity. Once this was accomplished, the judges were

asked to go through and write the rankings for each story on the cover page of the story (see Appendix E). Judges were also told which set of pictures each story was about and were given copies of the picture sets in order to help establish context for the stories.

After the stories were ranked, the judges then went back through the stories and rated the creativity of the stories on a 10-point continuous scale, with 1 = lowest and 10 = highest. This was the same scale as in Experiment 1. Again, the judges were asked to rate the stories using their own subjective definitions of creativity. In addition, the judges were instructed to rate the stories relative to one another as opposed to rating them against a universal standard of creativity. Therefore, the judges were instructed to rate at least one of the stories a 1 and one of the stories a 10. Judges were encouraged to spread the ratings out along the entire rating scale. Judges were also told that their ratings did not have to be evenly distributed and that ratings could go in-between boxes.

Once all of the stories were rated for creativity, each judge was asked to rate how “story-like” some of the stories were. This scale was used for one story from each child from each condition, for a total of 12 stories. This was done using a three-point scale, with 1 = “has no elements of being a story for a 4-5 year old,” 2 = “this is a primitive story for a 4-5 year old,” and 3 = “this is a typical story for a 4-5 year old.” This was implemented in order to compare the kindergarten stories to the much shorter preschool stories. Judges were instructed that, unlike with the creativity ratings, they should rate the stories on the basis of a universal standard of how “story-like” each story was. Because of this, judges were able to rate all stories as they wished; no story needed to be rated a 1 or a 3 if the judge felt that it did not deserve such a rating. The judges were also told that they needed to only mark in the boxes and not in between.

The experimenter met with each judge individually, and judges made their assessments independently of one another. Judges were able to make their assessments during one meeting, which was the same meeting in which they ranked and rated the preschool stories from Experiment 1. The meetings lasted between two and four hours and took place at a location that was convenient for each judge. Judges were not aware of the purpose of the study, the experimental conditions, or which stories were told by which children.

Ratings of Behavior

The two observers from Experiment 1 examined the videotapes and used the same measure from Experiment 1 to determine the type and amount of activity each child engaged in. As with Experiment 1, Observer 1 rated all activity sessions, whereas Observer 2 only rated one of each type of activity session (pretense, imitation, and coloring) per child. Observers examined the participants and recording at 15-second intervals the presence of various activity behaviors, for a total of 40 intervals per session. The behaviors included pretense (including role-play, object transformation, and object acts), coloring, other constructive play, and imitation. If a child participating in the imitation or coloring activities engaged in a pretense behavior during 10% or more of the time intervals, the story from that session was not to be used in the analysis of data. As with the judges, the observers were not aware of the purpose of the study. Observers were also unaware of the storytelling portion that took place immediately after each condition.

Results

As with Experiment 1, there were two measures of story creativity: judges rankings (with a high ranking of 44, reverse scored for graphing purposes) of all stories across

participants and judges relative ratings for each story on a 10-point scale (1 = low, 10 = high) for each participant. Results of these two measures are presented separately for each participant. In addition, one story from each child from each condition was rated on a three-point scale that assessed the extent to which a story resembled a true story. Unlike Experiment 1, there was no participant attrition.

The results of the judges' rankings of creativity for Participant 1 are presented in Figure 5, whereas the results of the judges' ratings of creativity for Participant 1 are presented in Figure 6. According to these measures, Judge 1 and Judge 3 noted a similar trend in creativity across initial baseline, with a mean ranking of 30.67 for Judge 1 and 30 for Judge 3, whereas Judge 2 noted an increasing trend in creativity, with a mean ranking of 15.33. The mean ratings given by the judges for the initial baseline condition were 8.67, 5.33, and 7, respectively. During the imitation condition, the creativity rankings of all three judges initially decreased and then increased dramatically, with overall lower mean rankings of 24.67, 19.67, and 23, respectively. The mean ratings given by the judges for the imitation condition were 7.17, 6, and 5.33, respectively. During the baseline probe condition (session 7), judges' rankings and ratings fell within the ranges established during the previous conditions. According to the rankings and ratings of Judge 1 and Judge 2, creativity steadily decreased throughout the pretense condition, with a mean ranking of 18 and a mean rating of 5.83 for Judge 1 and a mean ranking of 16 and a mean rating of 5.83 for Judge 2. According to the rankings and ratings of Judge 3, creativity decreased then increased throughout the pretense condition, with a mean ranking of 21 and a mean rating of 5.33. During the last baseline probe (session 11), judges' rankings and ratings fell within the ranges established during the previous conditions. Thus, according to the rankings and ratings of Judge 1 and

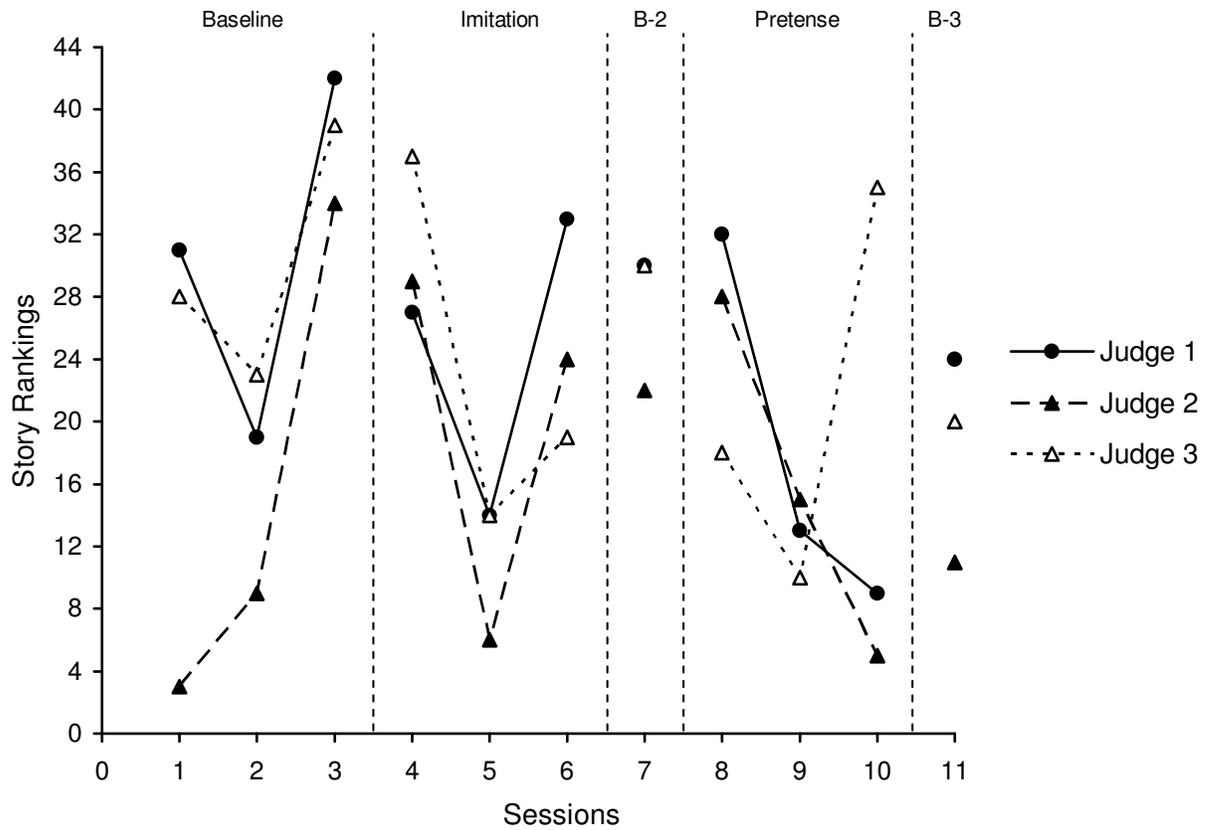


Figure 5. Creativity rankings of the stories told by Participant 1 in Experiment 2. Stories were from 11 separate sessions and were ranked relative to one another and to the stories told by the other three participants. Participant 1 engaged in three separate types of sessions: baseline, imitation, and pretense. Three different judges ranked the stories.

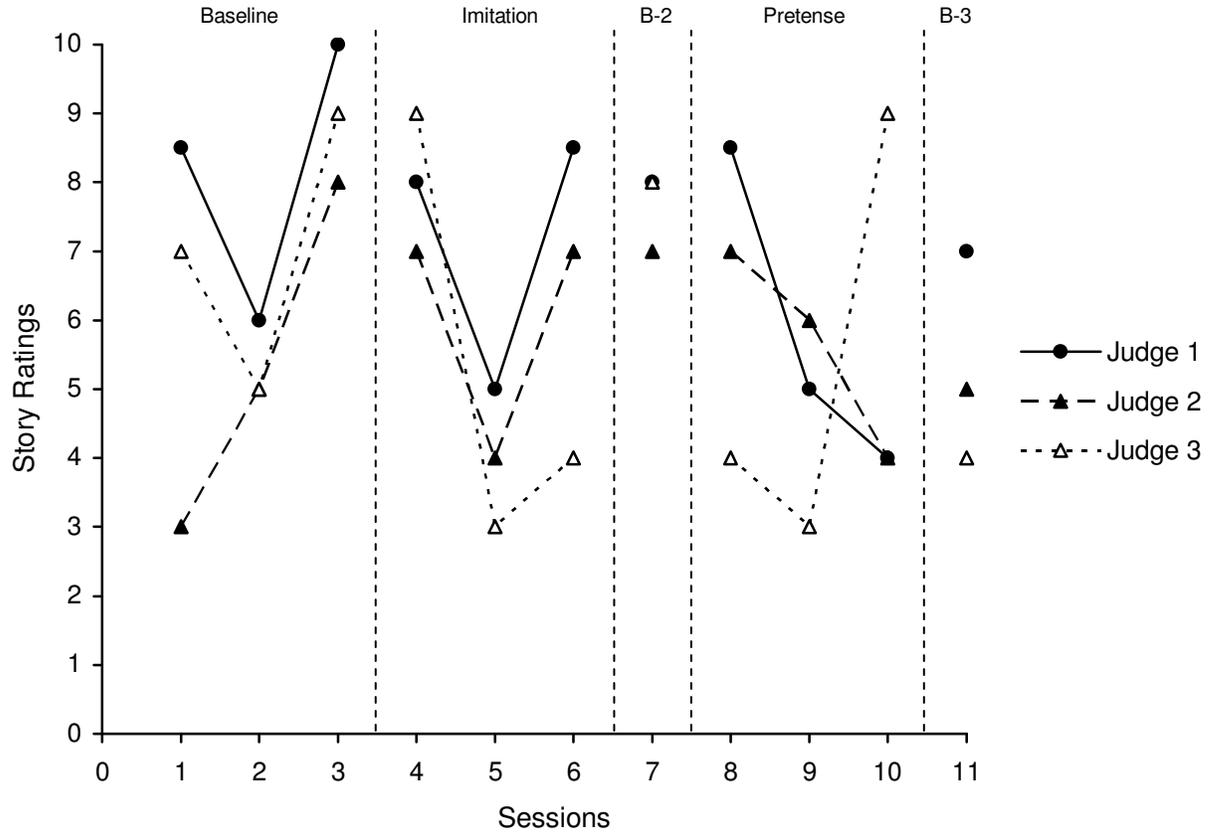


Figure 6. Creativity ratings of the stories told by Participant 1 in Experiment 2. Stories were from 11 separate sessions and were ranked relative to one another and to the stories told by the other three participants. Participant 1 engaged in three separate types of sessions: baseline, imitation, and pretense. Three different judges rated the stories.

Judge 3, the stories told by Participant 1 during the initial baseline condition were more creative than the stories told during the other conditions. According to the rankings and ratings of Judge 2, Participant 1 showed little difference between conditions. No judge ranked or rated the stories told during the pretense condition as overall higher in creativity than the stories told during the other conditions. Rather, two of the three judges ranked and rated the creativity of the stories told during the pretense condition lower overall than the stories told during the other conditions.

The results of the judges' rankings of creativity for Participant 2 are presented in Figure 7, whereas the results of the judges' ratings of creativity for Participant 2 are presented in Figure 8. According to these measures, all three judges noted a gradual decrease in creativity during initial baseline, with mean rankings of 18.67, 25.67, and 20, respectively. The mean ratings given by the judges for the initial baseline condition were 5.67, 7, and 5, respectively. During the pretense condition, the creativity rankings of all three judges increased and then decreased again, with overall lower mean rankings of 14, 14, and 9, respectively. The mean ratings given by the judges for the pretense condition were 5, 5.67, and 2.67, respectively. During the baseline probe condition (session 7), judges' rankings and ratings were lower than the ranges established during previous conditioning. According to the rankings and ratings of Judge 1 and Judge 3, creativity continued to decrease throughout the imitation condition, with a mean ranking of 3 and a mean rating of 1 for Judge 1 and a mean ranking of 2 and a mean rating of 1.67 for Judge 3. According to the rankings and ratings of Judge 2, creativity increased and then decreased throughout the imitation condition, with a mean ranking of 6 and a mean rating of 3.67. During the last baseline

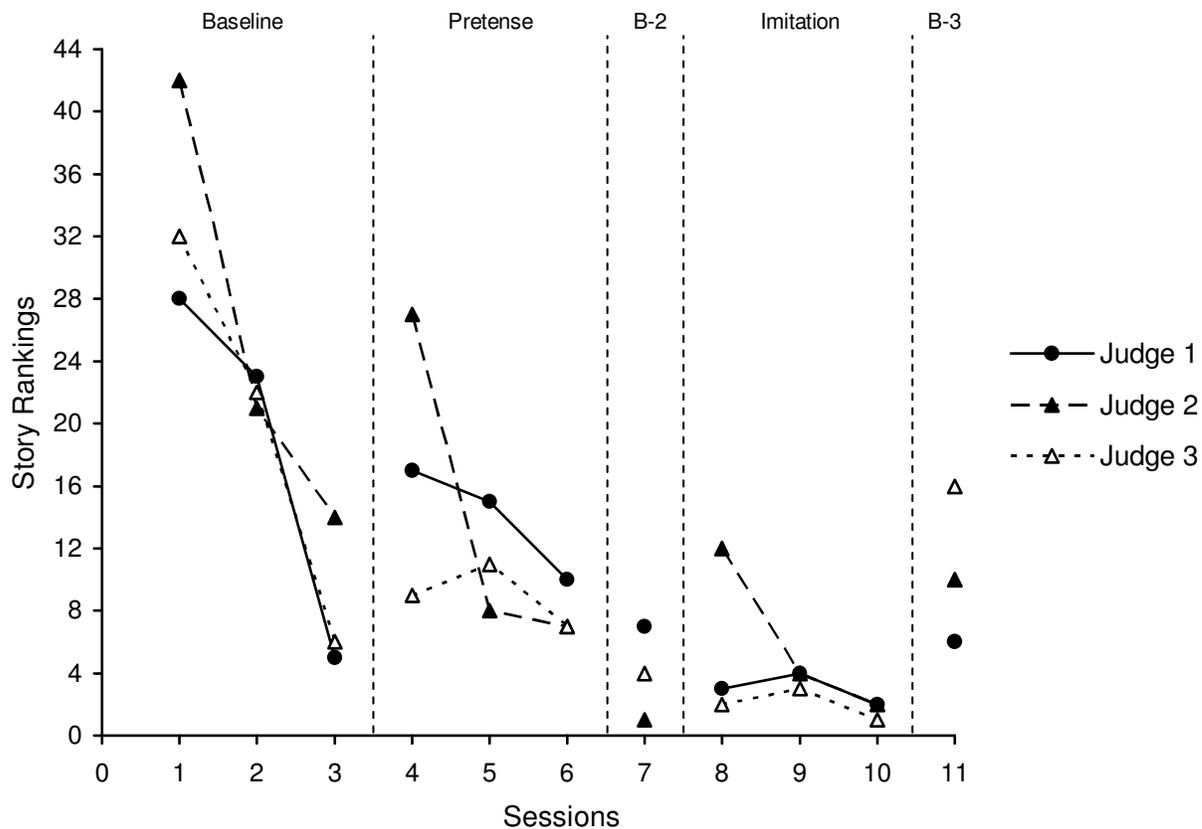


Figure 7. Creativity rankings of the stories told by Participant 2 in Experiment 2. Stories were from 11 separate sessions and were ranked relative to one another and to the stories told by the other three participants. Participant 2 engaged in three separate types of sessions: baseline, pretense, and imitation. Three different judges ranked the stories.

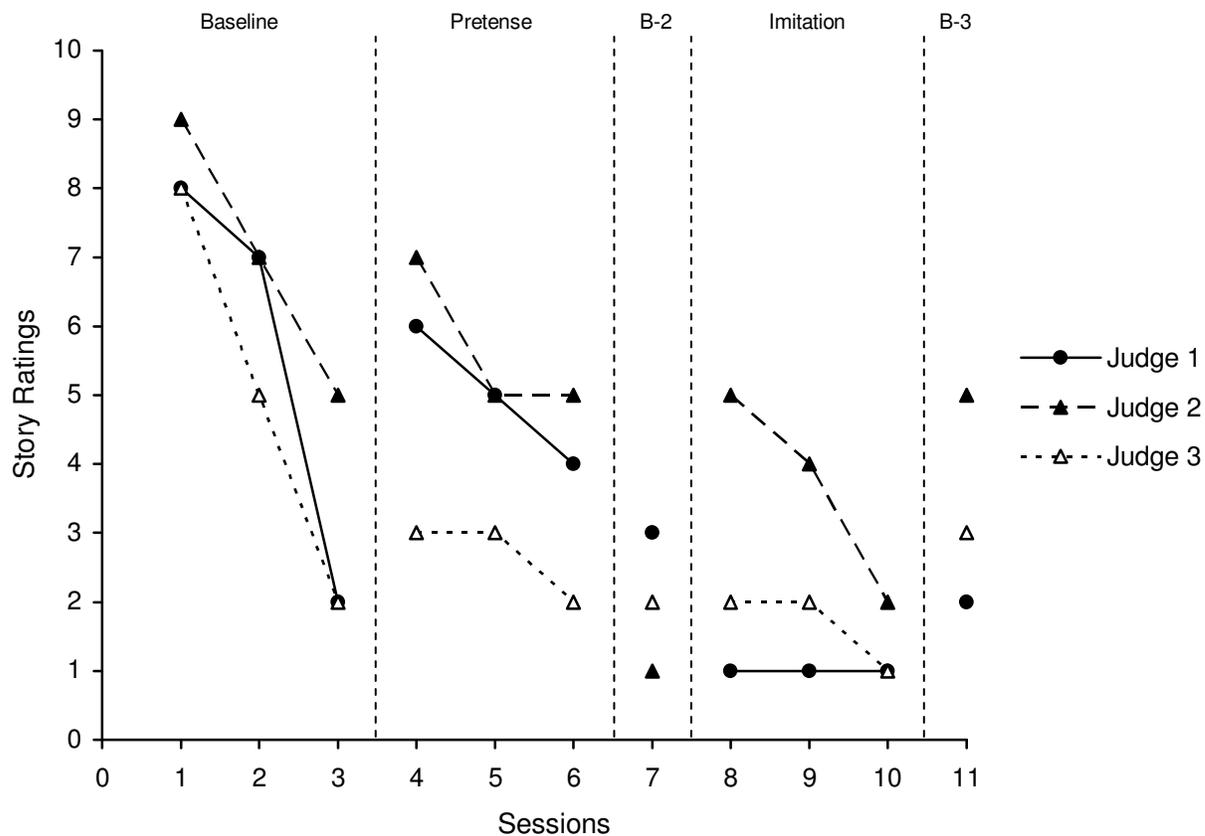


Figure 8. Creativity ratings of the stories told by Participant 2 in Experiment 2. Stories were from 11 separate sessions and were ranked relative to one another and to the stories told by the other three participants. Participant 2 engaged in three separate types of sessions: baseline, pretense, and imitation. Three different judges rated the stories.

probe (session 11), judges' rankings and ratings fell within the ranges established during the previous conditions. Thus, according to the rankings and ratings of all three judges, the stories told by Participant 2 during the initial baseline condition were more creative than the stories told during the other conditions. In addition, an overall decrease in creativity within and across conditions was observed.

The results of the judges' rankings of creativity for Participant 3 are presented in Figure 9, whereas the results of the judges' ratings of creativity for Participant 3 are presented in Figure 10. According to these measures, the three judges noted a similar trend in creativity across initial baseline, with a mean ranking of 18.33 for Judge 1, 24.33 for Judge 2, and 26.67 for Judge 3. The mean ratings given by the judges for the initial baseline condition were 5.67, 6.67, and 6.63, respectively. Mean creativity rankings remained at a similar level during the imitation condition, with a mean ranking of 25.33 for Judge 1, 28.33 for Judge 2, and 21 for Judge 3. The mean ratings given by the judges for the imitation condition were 7.33, 7, and 5, respectively. During the baseline probe condition (session 7), judges' rankings and ratings fell within the ranges established during previous conditioning. During the pretense condition, creativity rankings decreased overall, with mean rankings of 16, 20, and 15 for the three judges. The mean ratings given by the judges for the pretense condition were 5.33, 6.33, and 3.33, respectively. During the last baseline probe (session 11), judges' rankings and ratings were overall lower than during previous conditions. Thus, according to the ranking and rating methods, Participant 3 showed little difference between conditions. Overall, the creativity rankings and ratings for the pretense condition and final baseline session were slightly lower than those for the other conditions.

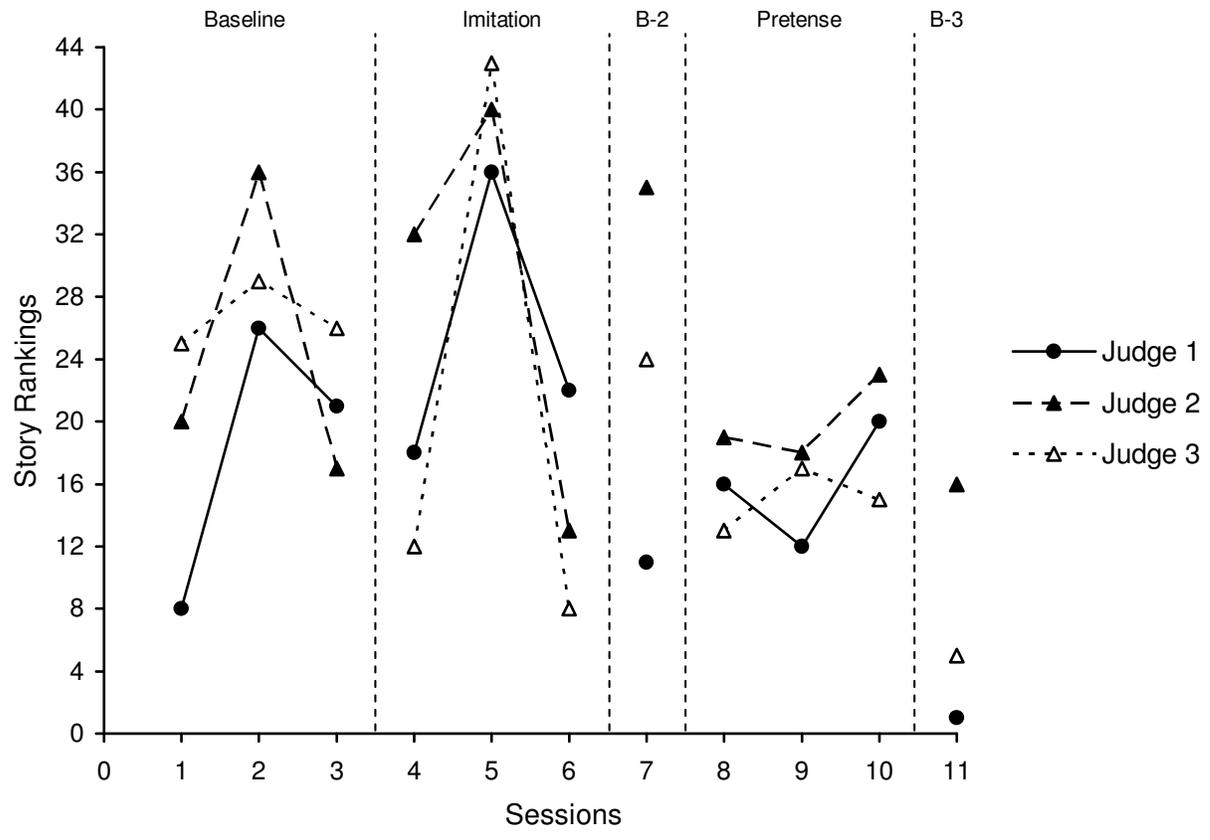


Figure 9. Creativity rankings of the stories told by Participant 3 in Experiment 2. Stories were from 11 separate sessions and were ranked relative to one another and to the stories told by the other three participants. Participant 3 engaged in three separate types of sessions: baseline, imitation, and pretense. Three different judges ranked the stories.

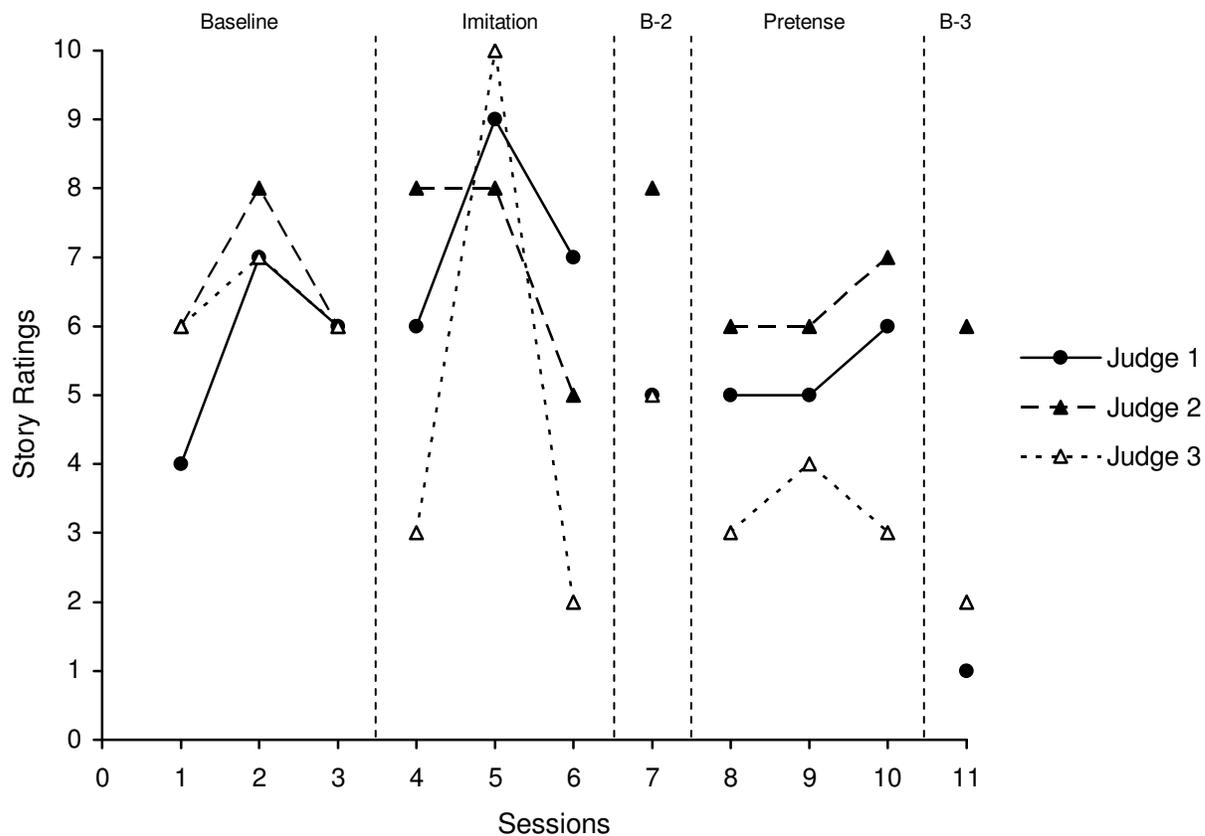


Figure 10. Creativity ratings of the stories told by Participant 3 in Experiment 2. Stories were from 11 separate sessions and were ranked relative to one another and to the stories told by the other three participants. Participant 3 engaged in three separate types of sessions: baseline, imitation, and pretense. Three different judges rated the stories.

The results of the judges' rankings of creativity for Participant 4 are presented in Figure 11, whereas the results of the judges' ratings of creativity for Participant 1 are presented in Figure 12. According to these measures, the three judges noted a similar high level in creativity across initial baseline, with a mean ranking of 32.67 for Judge 1, 36.33 for Judge 2, and 37.33 for Judge 3. The mean ratings given by the judges for the initial baseline condition were 8.17, 8, and 9, respectively. Mean creativity rankings remained at a similarly high level during the pretense condition, with a mean ranking of 40.67 for Judge 1, 29.67 for Judge 2, and 34.33 for Judge 3. The mean ratings given by the judges for the pretense condition were 9.33, 7.33, and 8.67, respectively. During the baseline probe condition (session 7), judges' rankings and ratings fell within or slightly below the ranges established during previous conditioning. Creativity rankings and ratings were high throughout the imitation condition, with a mean ranking of 39.67 for Judge 1, 42 for Judge 2, and 34.33 for Judge 3 and a mean rating of 9.33 for Judge 1, 9.33 for Judge 2, and 7.67 for Judge 3. Judge 3 did note a large decrease in creativity during the last imitation session (session 10), giving that story a creativity ranking of 21 and a rating of 4. During the last baseline probe (session 11), judges' rankings and ratings fell within the ranges established during the previous conditions. Thus, according to the ranking and rating methods, Participant 4 showed little difference between conditions. Despite the fact that each judge gave the highest overall rankings and ratings of story creativity to a different condition, no judge ranked or rated the overall level of creativity of the stories told during a particular condition as higher than those of the stories told during the other conditions.

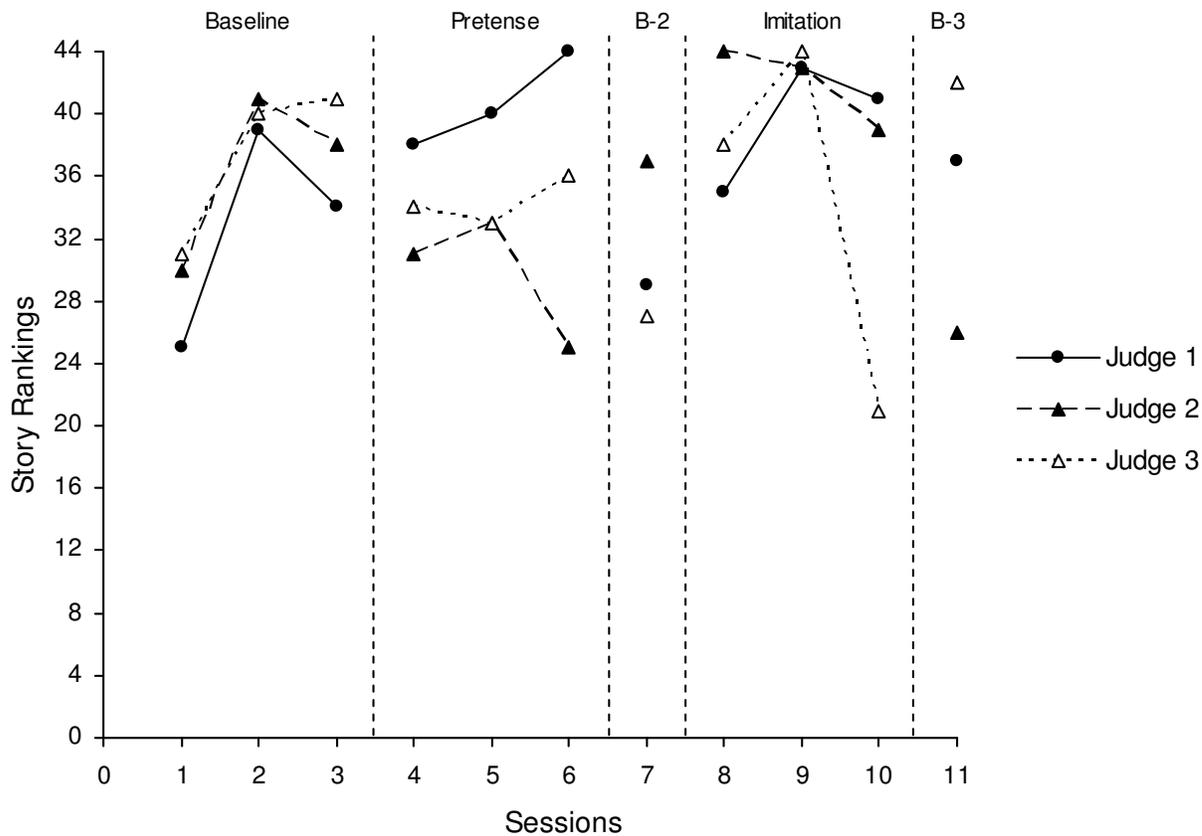


Figure 11. Creativity rankings of the stories told by Participant 4 in Experiment 2. Stories were from 11 separate sessions and were ranked relative to one another and to the stories told by the other three participants. Participant 4 engaged in three separate types of sessions: baseline, pretense, and imitation. Three different judges ranked the stories.

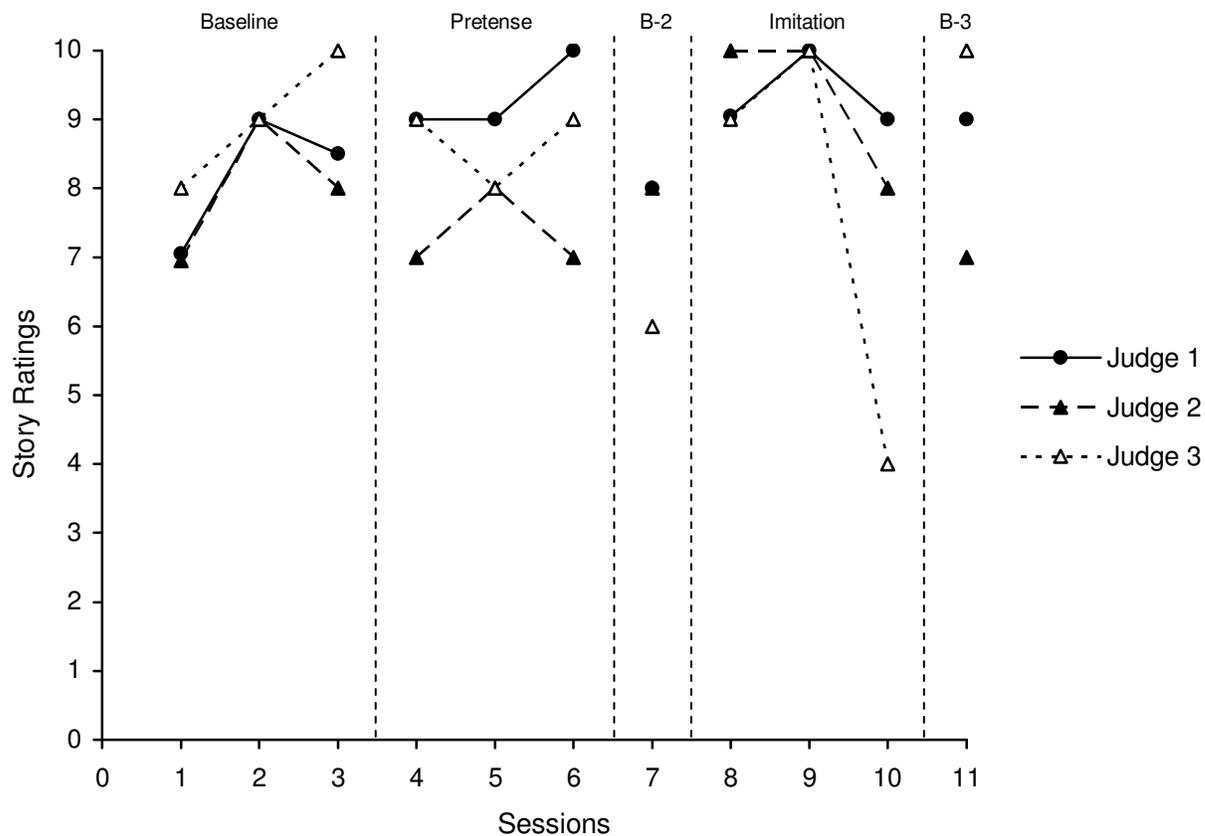


Figure 12. Creativity ratings of the stories told by Participant 4 in Experiment 2. Stories were from 11 separate sessions and were ranked relative to one another and to the stories told by the other three participants. Participant 4 engaged in three separate types of sessions: baseline, pretense, and imitation. Three different judges rated the stories.

To determine the interrater reliability for the rankings given by judges, the coefficient of concordance, an index of agreement, was calculated for each child (Winer, 1962). To do this, the rankings for each child were first entered into separate 3 x 11 tables, with the rows representing the judges and the columns representing the stories. The data were transformed so that interrater reliability could be determined for each individual child. Transformed, rankings were between 1 and 11, with all stories told by that child falling in between. This was repeated for the other two judges, and the data were analyzed. The coefficient of concordance was calculated for each child. These were then added together and divided by the number of children to determine the mean coefficient of concordance. The mean coefficient of concordance was .616. To determine the mean intercorrelation between the rankings assigned by judges, the intercorrelation between rankings for each child was calculated, added together, and divided by the number of children. The mean intercorrelation was .424.

To determine the interrater reliability for the ratings given by the three judges, the Spearman-Brown prediction formula was used (Winer, 1962). As with the rankings, the data for the ratings of each child were analyzed separately. Ratings were entered into 11 x 3 tables, with the rows representing the stories and columns representing the judges. The effective interrater reliability was calculated for each child. These reliabilities were then added together and the total was divided by the number of participants to determine the mean effective interrater reliability. The mean effective interrater reliability for the judges' ratings was .587. Again, because zero points did not exist on the scale, the interrater reliability was recalculated to adjust for differences in frame of reference for each child. Adjusted for differences in frame of reference, the mean effective interrater reliability for the judges'

ratings was .651. To determine the mean reliability of a single rating for the adjusted data, r_1 was calculated (Winer, 1962) for each child, these reliabilities were added together, and the total was divided by the number of participants. The mean reliability of a single rating for the adjusted data was .429.

As with Experiment 1, how much a story told by a child represented a true story was relevant to the validity of the measures of creative storytelling. For almost all stories, the judges gave a rating of 3, which indicated a judgement that the story was typical. As with Experiment 1, there were a few exceptions to this. However, all three of these exceptions were ratings of 2, indicating that even these stories had at least some elements of a story. Thus, there is a fair degree of agreement that the judges were ranking and rating the creativity of actual stories.

Behavioral observation data from Observer 1 indicated that participants were usually engaged in the activity prescribed for each condition during that condition. The presence of coloring behaviors during baseline sessions was recorded in an average of 77.5%, 95%, 91%, and 90% of the time intervals for the respective participants. The presence of imitation behaviors during imitation sessions was recorded in an average of 89.167%, 74.167%, 83.333%, and 80.833% of the time intervals for the respective participants. The presence of pretense behaviors during pretense sessions was recorded in an average of 66.667%, 87.5%, 90.833%, and 80% of the time intervals for the respective participants. In addition, Participant 3 displayed pretense behaviors in 2.5% of the time intervals of one of the imitation sessions (session 4).

Again, two different formulas were used to calculate interrater reliability between the two behavior observers (Foster & Cone, 1986). The overall interrater reliability was

found when the overall percentage of agreement for each session was calculated, the resulting values were added together, and the total was divided by 12 sessions. When this method was used, the overall interrater reliability was 0.879. To take chance agreement into account, Cohen's Kappa was then calculated for each session, the values were added together, and the total was divided by 12 sessions. Interrater reliability using this method was .487.

Discussion

The hypothesis that the stories a child told after engaging in pretense activities would be ranked and rated by experts as more creative than the stories they told after engaging in imitation and coloring activities was not supported by the present study. One can see that, similar to the two participants from Experiment 1, the creativity of the stories told by all participants seemed to fluctuate from day to day. This can be seen by simply looking at the graphs of the story rankings and ratings for each child. Participant 1 did not seem to ever achieve a stable level of creativity, with his overall level of creativity appearing to decrease slightly as time went on. Participant 2 did not achieve a stable level of creativity within any condition, but did tell stories that were judged as less creativity as time went on both across and within conditions. Overall, Participant 3 appeared to tell her least creative stories during the pretense sessions, with a variable level of creativity throughout the experiment. Regardless of the condition, the stories of Participant 4 were all rated high in creativity, with only one story judged by one judge falling below the upper half of creativity rankings and ratings for all the children. Overall, the level of creativity of the stories told by the participants was variable and no higher during the pretense condition than during any other condition.

Again, this variability in creativity does not seem to be associated with disengagement from the activities themselves. For all of the sessions, the participants were engaged in the activity meant for that session for the majority of time intervals for that session. Observer 1 only recorded a participant engaged in a pretense behavior during an imitation or baseline session once; this was for Participant 3 during session 4 for only one time interval. Also, the difficulties with potential carryover effects during the free play sessions of Experiment 1 were not seen in Experiment 2, as these sessions were eliminated for the second experiment. Therefore, it was probably not disengagement from the activities that affected the level of creativity of the stories told by the participants.

Boredom of participants did not appear to be a problem during this experiment. Participants appeared to enjoy all activities presented during all sessions and were actively engaged during all activities. At no point during the baseline sessions did the participants express dissatisfaction with coloring the pictures. Also, unlike Experiment 1, the participants did not appear to have difficulty formulating stories to go with the picture series. After going through a set of pictures in silence once, moving from picture to picture at the child's pace, each participant was able to go back and tell his or her story with little hesitation and produced much longer stories than the children in Experiment 1. Not even the least creative stories were rated a 1 on the "story-like" scale by any judge. The shorter length of Experiment 2 also may have contributed to guarding against the boredom of the participants.

The interrater reliability of the judges was somewhat lower than what was found in Experiment 1. Although variability in the creativity of participants was present for three of the participants, Participant 4 did not have a high level of variability in the creativity of the stories she told. Therefore, although all three judges judged the stories told by Participant 4

as high in creativity, interrater reliability for this participant was very low for both the rankings (interrater reliability = .422) and the ratings (adjusted effective interrater reliability = .412). Of the four participants, an acceptable level of interrater reliability was only seen with Participant 2 for both the rankings (interrater reliability = .808) and the ratings (adjusted effective interrater reliability = .911). This result suggests that the measure proposed by Amabile may not be appropriate or sensitive enough for a single-subject design, that the judges used were not appropriate, or that more judges were needed. However, Amabile's measure has been found to yield an acceptable interrater reliability in a single-subject design when two judges are used to judge the creativity of children's drawings (Baker & Winston, 1985). Despite the low interrater reliability between judges, no judge judged the stories told by a participant during the pretense sessions as more creative than the stories told during baseline and imitation sessions.

The interrater reliability between the behavior coders again posed some problems. As before, the interrater reliability found when Cohen's Kappa was used was much smaller than the reliability found with the traditional formula of dividing total number of intervals by intervals of agreement. Again, this is probably due to the fact that the raters were coding behaviors that were very frequent in occurrence. As in Experiment 1, both raters coded a similar proportion of occurrence of a particular behavior from session to session.

General Discussion and Conclusion

Neither experiment of the present study showed that the stories a child tells after engaging in pretense activities would be rated by experts as significantly more creative than the stories they tell after engaging in imitation and coloring activities. The level of creativity for each of the six participants during the pretense sessions was not significantly higher for

either the rankings or the ratings given by any of the judges. In fact, the level of creativity during all of the sessions was fairly variable for all of the participants.

The lack of support for the hypothesis is a finding that is similar to that of Smith and Whitney (1987) but is contradictory to the findings of most of the past research (Dansky, 1980; Dansky & Silverman, 1973; Dansky & Silverman, 1975; Li, 1978; Howard-Jones, Taylor, & Sutton, 2002). There could be multiple reasons for this. First, the type of creativity behavior measured, storytelling, is quite different from other tasks that have been shown to be affected by pretense or free play, such as associative fluency and collage making. The associative fluency studies had children first play with objects and then give verbal responses regarding the uses for either the same objects or different objects. The Howard-Jones, Taylor, and Sutton (2002) study had children first engage in a very hands-on creative activity (playing with salt dough) and then engage in an activity with a comparable amount of object manipulation (collage making). In both these instances, the training and testing situations were quite similar. Storytelling, on the other hand, is an activity that is very different from the testing activities in the previous studies. The children of the present study did not engage with objects similar to those in the first part of the experiment during the storytelling sessions, nor did they move from one hands-on condition to another similarly hands-on condition. Rather, they were asked to move from a very hands-on activity to a less hands-on activity and to engage with stimuli dissimilar to what they had encountered in the training situation. Taking these points into account, one could argue that engaging in a short period of pretense does not lead to a short-term higher level of creativity in all situations. Instead, perhaps engaging in a short period of pretense only sets the stage for creative rehearsal of objects. Thus, the results are going to hold as far as the creative activity is

similar to the creative measure. The more similar the training in the first condition is to that of the testing condition, the more creative the behavior seen will be. Because of this, it could be further argued that the use of objects is an artifact of the methodology in previous experiments looking at the relationship between pretense and creativity. Having children engage only in pretense involving role-play and no object manipulation and afterwards testing them on associative fluency or other creative expressions could test this hypothesis.

A second possibility could be that the transition between the two conditions is too disruptive. This is a similar point to Dansky's (1999) criticism of Smith and Whitney (1987). It may not be that storytelling is too different of an activity from associative fluency or collage making. However, the transition between the pretense activity and the storytelling activity may need to be carried out in a different fashion than the transition between pretense activities and other expressions of creativity. For the kindergarten storytelling activity, the children were asked to first look at all the pictures and then go back and tell their story. It may be that any short-term effect pretense would have on creative storytelling is lost during the time that the children are initially reviewing the storybook pictures. One way to resolve this issue would be to make the training and testing situations more similar. Perhaps incorporating the pictures into the coloring, imitation, and pretense conditions would make for a smoother transition to the storytelling portion of each session. However, as stated previously, one could argue that by structuring the experiment this way, one is only setting up a situation that would be conducive to the creative rehearsal of objects and not conducive to creativity per se.

A third possibility might be that events were occurring outside of the experiments that were affecting the results. It appeared that the level of creativity of the stories told by the

children changed day to day regardless of the experimental condition the children were engaged in. Although the experimenter met with each child at approximately the same time each day, the school curriculum did change from day to day. It might be that the activity a child was engaged in prior to meeting with the experimenter was influencing the creativity of the story the child told that day. However, if pretense is indeed the potent variable it has been proposed to be, its effects should be large enough to compensate for the impact of extraneous variables.

The present study has a number of flaws. First, a multiple baseline design could have been a more experimentally sound design to test the hypothesis. If a multiple-baseline design were used, the effects of events outside of the experiment might be better controlled. However, there were concerns, highlighted in Experiment 1, that a more lengthy experiment might lead to participant boredom and eventual attrition. When the hypothesis was tested with the current experimental design, it appeared that all participants of Experiment 2 were still actively engaged and content after 11 sessions. Although Participant 4 in Experiment 1 did finish all of the sessions, he was vocal about his general boredom with the activities during the last few sessions. Therefore, any more than 11 sessions using the present activities might result in boredom and eventual dropout from the experiment.

As stated before, the interrater reliabilities between the judges, particularly in Experiment 2, were not as high as one would have hoped. This would perhaps be more of a problem if the general trend of the data supported the hypothesis. However, despite a lower level of interrater reliability, no one judge found that the pretense condition led to stories that were more creative. Still, the interrater reliability for the judges in Experiment 2 was not at a desirable level. Perhaps using more judges, as suggested by Amabile (1982b), would result

in an acceptable level of interrater reliability being obtained. However, the practicality of using more judges is an issue. It was very difficult to obtain the three judges used in the study; obtaining more judges might be even more difficult. In addition, if so many judges are needed to obtain a high interrater reliability, both the utility and the face validity of the measure must be questioned. In real-life situations, the number of judges of creativity in competitive settings is usually limited to only three or so. These judges are expected to be consistent with one another. If they are not consistent with one another, more judges are not added in order to make the overall interjudge reliability higher; the problem judges are simply replaced for the next competition. Perhaps a better measure can be developed to assess the creativity of products in a subjective manner.

A third flaw lies with the interrater reliability of the behavior raters. Using a modified version of Dansky's (1980) checklist of behaviors, the raters ended up coding behaviors that were very frequent. This in turn affected the reliability when Cohen's Kappa was calculated. Because the behaviors of interest were so frequent, the likelihood of chance agreement between the two observers rose dramatically. Perhaps a different measure, using shorter interval lengths or more particularly defined behaviors, would result in higher interrater reliabilities that control for chance agreement. Despite this flaw, it should be noted again that the low interrater reliability of the behavior raters had little impact on the actual results of the experiments. The children still exhibited the appropriate behaviors for the appropriate training situations, and engaging in such behaviors did not seem to have an effect on the creativity of their storytelling.

The present study was unable to find a link between pretense and the creativity of stories told by children. This is not to say that such a link does not exist; the present study

has a number of flaws that make drawing conclusions difficult. However, even if interrater reliabilities were at an acceptable level, the variability of creativity during session types still makes for a lack of definitive conclusions. This is also not to say that there is no link between pretense and other creative products; it may be that storytelling is a unique creative product in its relation to pretense. However, it may also be the case that previous studies have had methodological artifacts in the form of objects used during the training and testing procedures. Perhaps by using other creative products, as well as different measures and experimental designs, the relationship between pretense and creativity in children will become clearer.

References

- Alexander, P. A., Tamara, L. J., White, S. H., Parson, J. L., Cotropia, K. H., Liu, H. C., & Ackerman, C. M. (1994). Young children's creative solutions to realistic and fanciful story problems. *Journal of Creative Behavior*, 28(2), 89-106.
- Amabile, T. M. (1979). Effects of external evaluation on artistic creativity. *Journal of Personality and Social Psychology*, 37(2), 221-233.
- Amabile, T. M. (1982a). Children's artistic creativity: Detrimental effects of competition in a field setting. *Personality and Social Psychology Bulletin*, 8(3), 573-578.
- Amabile, T. M. (1982b). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43(5), 997-1013.
- Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, 45(2), 357-376.
- Amabile, T. M. (1993). What does a theory of creativity require? *Psychological Inquiry*, 4(3), 179-181.
- Amabile, T. M., Goldfarb, P., & Brackfield, S. C. (1990). Social influences on creativity: Evaluation, coercion, and surveillance. *Creativity Research Journal*, 3(1), 6-21.
- Amabile, T. M., Hennessey, B. A., & Grossman, B. S. (1986). Social influences on creativity: The effects of contracted-for reward. *Journal of Personality and Social Psychology*, 50(1), 14-23.
- Amabile, T. M., & Tighe, E. (1993). Questions of creativity. In J. Brockman (Ed.), *Creativity* (pp. 7-27). New York: Simon & Schuster.
- Baer, J. (1994). Performance assessments of creativity: Do they have long-term stability? *Roeper Review*, 17(1), 7-11.

- Baker, J. E., & Winston, A. S. (1985). Modifying children's creative drawing: Experimental analysis and social validation of a self-instructional procedure. *Education and Treatment of Children, 8*(2), 115-132.
- Berman, R. A. (1995). Narrative competence and storytelling performance: How children tell stories in different contexts. *Journal of Narrative & Life History, 5*(4), 285-313.
- Berretta, S., & Privette, G. (1990). Influence of play on creative thinking. *Perceptual & Motor Skills, 71*(2), 659-666.
- Bornens, M. T. (1990). Problems brought about by "reading" a sequence of pictures. *Journal of Experimental Child Psychology, 49*(2), 189-226.
- Bretherton, I. (1989). Pretense: The form and function of make-believe play. *Developmental Review, 9*, 383-401.
- Christie, J. F., & Enz, B. (1992). The effects of literacy play interventions on preschoolers' play patterns and literacy development. *Early Education & Development, 3*(3), 205-220.
- Dansky, J. L. (1980). Make-believe: A mediator of the relationship between play and associative fluency. *Child Development, 51*, 576-579.
- Dansky, J. L. (1999). Play. In M. Runco & S. Pritzker (Eds.), *Encyclopedia of Creativity* (pp. 393-408). New York: Academic Press.
- Dansky, J. L., & Silverman, I. W. (1973). Effects of play on associative fluency in preschool-aged children. *Developmental Psychology, 9*(1), 38-43.
- Dansky, J. L., & Silverman, I. W. (1975). Play: A general facilitator of associative fluency. *Developmental Psychology, 11*(1), 104.

- Eisenberger, E., & Cameron, J. (1996). Detrimental effects of reward: Reality or myth? *American Psychologist, 51*(11), 1153-1166.
- Eisenberger, E., & Selbst, M. (1994). Does reward increase or decrease creativity? *Journal of Personality and Social Psychology, 66*(6), 1116-1127.
- Foster, S. L., & Cone, J. D. (1986). Design and use of direct observation procedures. In A. R. Ciminero, K. S. Calhoun, & H. E. Adams (Eds.), *Handbook of behavioral assessment* (2nd ed., pp. 253-324). New York: Wiley-Interscience.
- Geist, E. A., & Aldridge, J. (2002). The developmental progression of children's oral story inventions. *Journal of Instructional Psychology, 29*(1), 33-39.
- Hennessey, B. A., & Amabile, T. M. (1988). Story-telling: A method for assessing children's creativity. *Journal of Creative Behavior, 22*(4), 235-246.
- Howard-Jones, P. A., Taylor, J. R., & Sutton, L. (2002). The effect of play on the creativity of young children during subsequent activity. *Early Childhood Development and Care, 172*(4), 323-328.
- Kim, S. Y. (1999). The effects of storytelling and pretend play on cognitive processes, short-term and long-term narrative recall. *Child Study Journal, 29*(3), 175-191.
- Li, A. K. (1978). Effects of play on novel responses in kindergarten children. *Alberta Journal of Educational Research, 24*(1), 31-36.
- Magee, M. A., & Sutton-Smith, B. (1983). The art of storytelling: How do children learn it? *Young Children, 38*(4), 4-12.
- Mayer, M. (1967). *A boy, a dog, and a frog*. New York: Penguin Putnam Inc.
- Mayer, M. (1969). *Frog, where are you?* New York: Penguin Putnam Inc.
- Mayer, M. (1973). *Frog on his own*. New York: Penguin Putnam Inc.

- Mayer, M. (1974). *Frog goes to dinner*. New York: Penguin Putnam Inc.
- Mayer, M., & Mayer, M. (1971). *A boy, a dog, a frog, and a friend*. New York: Penguin Putnam Inc.
- Mayer, M., & Mayer, M. (1975). *One frog too many*. New York: Penguin Putnam Inc.
- Piaget, J. (1962). *Play, dreams, and imitation in childhood*. New York: Norton.
- Pramling, P. N., Norlander, T., & Archer, T. (2001). Moral expressions in 6- to 7-year-old children's stories in Sweden, Hungary, and China: A phenomenological study. *Childhood: A Global Journal of Child Research*, 8(3), 361-382.
- Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of Behavioral Research: Methods and Data Analyses* (2nd ed.). New York: McGraw-Hill.
- Rubin, K. H., Watson, K. S., & Jambor, T. W. (1978). Free-play behaviors in preschool and kindergarten children. *Child Development*, 49(2), 534-536.
- Shapiro, L. R., & Hudson, J. A. (1991). Tell me a make-believe story: Coherence and cohesion in young children's picture-elicited narratives. *Developmental Psychology*, 27(6), 960-974.
- Smith, P. K., & Whitney, S. (1987). Play and associative fluency: Experimenter effects may be responsible for previous positive findings. *Developmental Psychology*, 23(1), 49-53.
- Smolucha, F. (1992). A reconstruction of Vygotsky's theory of creativity. *Creativity Research Journal*, 5(1), 49-67.
- Tavalin, F. (1995). Context for creativity: Listening to voices, allowing a pause. *Journal of Creative Behavior*, 29(2), 133-142.

- Torrance, P. E. (2000). Preschool creativity. In B. A. Bracken (Ed.), *The Psychoeducational Assessment of Preschool Children* (3rd ed., pp. 349-363). Needham Heights: Allyn & Bacon.
- Vygotsky, L. S. (1987). Imagination and its development in childhood. In R. Rieber, A. Carton (Eds.), *The Collected Works of L. S. Vygotsky, Vol. 1: Problems of General Psychology* (pp. 339-349). New York: Plenum Press. (Originally a lecture presented in 1932)
- Vygotsky, L. S. (1990). Imagination and creativity in childhood. *Soviet Psychology*, 28(1), 84-96. (Original work written in 1930)
- Winer, B. J. (1962). *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

Appendices

Appendix A: Preschool Informed Consent Form

1) Purpose of the Study: The purpose of the study is to determine if the type of play of preschool children has an effect on the creativity of stories they tell. Specifically, we are looking to see if pretending, imitating, and coloring have different effects on the creativity of the stories preschool children tell.

2) Length of Participation: Your child will meet with the researcher for about 30 minutes a day for 11 consecutive school days. In addition, he or she will meet with the researcher for approximately fifteen minutes a week before the experiment to become acquainted with the researcher.

3) Refusal to Participate and Withdrawal from the Study: Your child is not required to participate in this study. If you do allow your child to participate, you may withdraw them from the study at any time without any penalty to you or your child.

4) Description of Procedures: Your child will first meet with the researcher a week prior to the experiment for about fifteen minutes. During this time, the researcher will become acquainted with the child while they color. During the actual experiment, each child will meet individually with the researcher for about 30 minutes a day. The children will first either use objects and pretend with them, use objects to imitate the actions of the researcher, or color sketches. After ten minutes of activity time, the child will then tell a story to the researcher based on a wordless picture book entitled A Boy, a Dog, and a Frog. On subsequent days, this sequence will be repeated with the same activity or the remaining activities (either pretend, imitation, or coloring). The stories will be videotaped and transcribed, then later given to two preschool teachers to rate on various aspects of creativity.

Videotapes will be kept in a locked cabinet when not in use, and your child's name will be removed from the transcribed stories to ensure anonymity.

5) Description of Procedures that May Result in Risk or Discomfort: There are no known procedures utilized in this study that may result in discomfort for you or your child or put either of you at risk of any sort. In fact, these activities are similar to those that parents and preschool teachers typically engage in with children. If any procedures result in discomfort for you or your child, the experiment will halt immediately and you may suspend participation or withdraw your child from the study if you see fit. You may discuss any issues concerning your participation at any time with either the researcher or thesis chair.

6) Expected Benefits of the Study: Your child's participation will help determine if a particular activity can enhance the creativity of preschool children in other realms, such as storytelling. Your child should enjoy participating in the enrichment activities and may display enhanced creativity.

7) Use of Research Results: Information provided by you and your child will be entered into a statistical software package. Such information will be recorded using a research number to ensure anonymity. The results of this study will be published in psychological journals and presented at conferences to people specializing in creativity and play. Your child's identity will be completely anonymous.

8) Approval of Research Project: This project has been approved by the Eastern Michigan University Human Subjects in Research Committee. If you have any questions, please contact

the Human Subjects Review Committee, College of Arts and Sciences, at (734) 487-4348.

Research Participants Rights: I have read or had read to me all of the above. Any questions I have regarding the study have been answered by Storm Ross or Dr. Marilyn Bonem. I have been informed of any risks or discomfort that may occur for my child or myself, and have also been informed of the possible benefits of participating in this study for my child or myself. I understand that participation in this study is strictly voluntary, and that I may refuse participation without penalty. I am also aware that if I choose to allow my child to participate, I can withdraw my child from the study at any time without penalty. I also understand that while the data of the study may be published, any personal information about my child or myself will not be revealed unless required by law. I also understand that steps have been taken to ensure confidentiality.

If my child or I experience emotional reactions that are difficult to manage, I understand that Storm Ross can be contacted for referral to a mental health agency or notification to the Eastern Michigan University Psychology Clinic may also be made. I also understand that I should notify Storm Ross if my child is experiencing emotional distress as a result of the experiment.

I understand my rights as the parent or guardian of a participant and the rights of my child as a research participant and I voluntarily consent to my child's participation in this study. I understand what this study involves and how and why it is being done. I will receive a signed copy of this consent form.

Participant's name (Print)

Date

Name of Participant's Parent or Guardian

Date

Signature of Parent or Guardian

Date

Signature of Witness or Research Assistant

Date

Signature of Principal Investigator

Date

Appendix B: Preschool Story Rating Scale

Story Number _____

Preschool Story Rating Scale

Instructions

You will begin by reading a set of stories provided to you by the researcher. At the top of each story is a reference number to either a picture (for the preschool stories) or a set of pictures (for the kindergarten stories) that the story is based on; these numbers refer to pictures that will be provided to you to help with your understanding of the story. After reading all of the stories in the set, please rank the stories in order of creativity, using your own subjective definition of creativity. Once you have ordered all the stories in the set, please tell the researcher and he will provide you with the second set of stories. Again, please read all the stories in the second set. After you have read each story in the second set, please insert the stories into the previously completed ordered set. Please repeat this process until all of the stories have been read and ranked 1 through 30, with 1 meaning “most creative” and 30 meaning “least creative.” Please make sure you rate the stories relative to one another, as opposed to universal standards of creativity. Also, please make sure that the number assigned to the story you are rating matches the number on the rating sheet you are using.

Once you have ordered all 30 stories, please write down the creativity ranking for this story in the space below.

- A) What is the creativity ranking for this story, with 1 meaning “most creative” and 30 meaning “least creative”?

- B) Once you have finished ordering all of the stories from most creative to least creative, please go back and rate each story on the 10-point scale provided below. You should rate at least one story a 1 and at least one story a 10, and you should spread the others out along the entire rating scale. However, the ratings do not have to be evenly distributed. You may give the same rating to as many stories as you want. Ratings may also go in between the boxes if you feel that a particular story does not fit into a particular box.

Creativity- Using your own subjective definition of creativity, please rate the degree to which the story is creative:

1	2	3	4	5	6	7	8	9	10
<input type="checkbox"/>									
Lowest									Highest

C) After rating this story on the ten-point creativity scale, please rate it on how “story-like” it is on the 3-point scale provided below. Unlike the creativity rankings and ratings for parts A and B, please rate this story against a universal standard for how much it resembles an actual story, as opposed to only rating the stories against one another. Therefore, you may rate the stories as you wish.; no story needs to be rated a 1 or a 3 if you feel that none deserve these ratings. Also, unlike the creativity rating scale used in part B, please only mark in one of the three boxes on the scale, not in between boxes.

Considering the age (4-5 years old) of these children, how do you consider this story? :

1	2	3
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Has no elements of being a story for a 4-5 year old.	This is a primitive story for a 4-5 year old.	This is a typical story for a 4-5 year old.

Appendix C: Activity Behaviors Checklist

Instructions

Observations of Activity

One child is observed during each ten-minute observational session. Each ten-minute session is divided into 40 15-second intervals. The observer notes the occurrence of each of the behaviors listed below if they occur at any time during the 15-second interval by placing a slash (/) in the appropriate box on the attached data sheet. If two or more behavior categories are observed during a given interval, each should be noted. If a behavior can not be coded using the six provided categories, leave that interval blank. The categories are:

1) ROLE PLAY: The child pretends to be someone or something they are not. Role-play can be expressed by means of an action and/or verbalization. Pretending to be people, animals, and objects are included in this category. If a child indicates by word or gesture that they are portraying a role, this category should continue to be scored so long as subsequent behaviors are being performed within the context of the role. For example, if a child declares that they are a “mommy” and then proceeds to set a table with toy dishes, those intervals during which the child is setting the table are scored as “role-play,” irrespective of whether or not the child declares that they are a “mommy” during each of those intervals. In addition, any behavior in which the child is using an object in an object-appropriate pretend way, but also in a manner in which the child is portraying a role, should be coded as “role play.” For example, if the child pretends that they are using the plastic screwdriver to turn pretend/invisible screws, this behavior should be coded as “role play.” If the child were simply turning real screws with the screwdriver, the behavior would go unrecorded. In addition, if the child were using the object in an object-inappropriate way, but also in a manner in which the child

is portraying a role (for example, the child pretends to wrestle the screwdriver, which they say is a dragon), then “object transformation” and/or “object acts” would be coded in addition “role play.” (See definitions below).

2) OBJECT TRANSFORMATION: The child pretends that an object is something other than what it really is. This may be indicated by verbalization and/or gesture. For example, if a child pretends that a box is a house, this would be coded as “object transformation.”

However, if the child pretends an object is something that has an active role (for example, pretending that a stick is a car, and then proceeding to drive the car around), then the behavior would be coded as both “object transformation” and “object acts.”

3) OBJECT ACTS: This category is similar to role playing except that, rather than assuming the role themselves, the children pretend that some inanimate object is performing the role. In this case, the object might be object-appropriate (for example, the child states that their pencil is mad) or object-inappropriate (for example, the child pretends that the plastic screwdriver is a dragon). In the latter case, the behavior would be double-coded as both “object transformation” and “object acts.”

4) COLORING: Includes behaviors related to the coloring of a picture. However, if the occurrence of another codable behavior happens during a 15-second interval, that behavior is also coded in its proper category. For example, if while coloring a child states that they are a princess, the behavior for the 15-second interval would be coded as role-play and as coloring.

5) OTHER CONSTRUCTIVE PLAY: Constructive play involves the creation of something. For example, dropping paper clips would result in a blank scoring, but building a structure out of paper clips would be scored as constructive play. Making designs out of objects are also score as constructive play. The quality of the construction is not a factor in determining

whether or not play is constructive. As soon as it becomes evident that the child is making something, the behavior should be coded as other constructive play-unless the construction is a part of, or involves, another codable activity. For example, if the child makes reference to or uses their construction in a manner that indicates object transformation or object acts, the latter categories would be scored for those 15-second segments in which the behavior occurred, and “other constructive play” would not be coded.

6) IMITATION: Includes all behaviors that involve the child imitating someone in the immediate vicinity. Even if the imitated behavior is a make-believe behavior, it would be coded as imitation and not role-play. For example, if a child states that they are a dragon, the behavior would be coded as role-play. However, if the child states they are a dragon after someone in the room states they are a dragon, the behavior would be coded as imitation and not as role-play. Similarly, if the child throws a cup, the behavior would not be coded. However, if the child throws a cup to imitate someone else in the room who just threw a cup, the behavior would be coded as imitation.

Child #:
Sex:
Observer:

Session #:
Date:
Age:

	Role Play	Object Transformation	Object Acts	Coloring	Other Constructive Play	Imitation
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						

	Role Play	Object Transformation	Object Acts	Coloring	Other Constructive Play	Imitation
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						

Appendix D: Kindergarten Informed Consent Form

1) Purpose of the Study: The purpose of the study is to determine if the type of play of kindergarten children has an effect on the creativity of stories they tell. Specifically, we are looking to see if pretending, imitating, and coloring have different effects on the creativity of the stories kindergarten children tell.

2) Length of Participation: Your child will meet with the researcher for about 30 minutes a day for 11 consecutive school days. In addition, he or she will meet with the researcher for approximately fifteen minutes a week before the experiment to become acquainted with the researcher.

3) Refusal to Participate and Withdrawal from the Study: Your child is not required to participate in this study. If you do allow your child to participate, you may withdraw them from the study at any time without any penalty to you or your child.

4) Description of Procedures: Your child will first meet with the researcher a week prior to the experiment for about fifteen minutes. During this time, the researcher will become acquainted with the child while they color. During the actual experiment, each child will meet individually with the researcher for about 30 minutes a day. The children will first either use objects and pretend with them, use objects to imitate the actions of the researcher, or color sketches. After ten minutes of activity time, the child will then tell a story to the researcher based on a wordless picture book. On subsequent days, this sequence will be repeated with the same activity or the remaining activities (either pretend, imitation, or coloring). The stories will be videotaped and transcribed, then later given to two preschool teachers to rate on various aspects of creativity. Videotapes will be kept in a locked cabinet

when not in use, and your child's name will be removed from the transcribed stories to ensure anonymity.

5) Description of Procedures that May Result in Risk or Discomfort: There are no known procedures utilized in this study that may result in discomfort for you or your child or put either of you at risk of any sort. In fact, these activities are similar to those that parents and kindergarten teachers typically engage in with children. If any procedures result in discomfort for you or your child, the experiment will halt immediately and you may suspend participation or withdraw your child from the study if you see fit. You may discuss any issues concerning your participation at any time with either the researcher or thesis chair.

6) Expected Benefits of the Study: Your child's participation will help determine if a particular activity can enhance the creativity of kindergarten children in other realms, such as storytelling. Your child should enjoy participating in the enrichment activities and may display enhanced creativity.

7) Use of Research Results: Information provided by you and your child will be entered into a statistical software package. Such information will be recorded using a research number to ensure anonymity. The results of this study will be published in psychological journals and presented at conferences to people specializing in creativity and play. Your child's identity will be completely anonymous.

8) Approval of Research Project: This project has been approved by the Eastern Michigan University Human Subjects in Research Committee. If you have any questions, please contact the Human Subjects Review Committee, College of Arts and Sciences, at (734) 487-4348.

Research Participants Rights: I have read or had read to me all of the above. Any questions I have regarding the study have been answered by Storm Ross or Dr. Marilyn Bonem. I have been informed of any risks or discomfort that may occur for my child or myself, and have also been informed of the possible benefits of participating in this study for my child or myself. I understand that participation in this study is strictly voluntary, and that I may refuse participation without penalty. I am also aware that if I choose to allow my child to participate, I can withdraw my child from the study at any time without penalty. I also understand that while the data of the study may be published, any personal information about my child or myself will not be revealed unless required by law. I also understand that steps have been taken to ensure confidentiality.

If my child or I experience emotional reactions that are difficult to manage, I understand that Storm Ross can be contacted for referral to a mental health agency or notification to the Eastern Michigan University Psychology Clinic may also be made. I also understand that I should notify Storm Ross if my child is experiencing emotional distress as a result of the experiment.

I understand my rights as the parent or guardian of a participant and the rights of my child as a research participant and I voluntarily consent to my child's participation in this study. I understand what this study involves and how and why it is being done. I will receive a signed copy of this consent form.

Participant's name (Print)

Date

Name of Participant's Parent or Guardian

Date

Signature of Parent or Guardian

Date

Signature of Witness or Research Assistant

Date

Signature of Principal Investigator

Date

Appendix E: Kindergarten Story Rating Scale

Story Number _____

Kindergarten Story Rating Scale

Instructions

You will begin by reading a set of stories provided to you by the researcher. At the top of each story is a reference number to either a picture (for the preschool stories) or a set of pictures (for the kindergarten stories) that the story is based on; these numbers refer to pictures that will be provided to you to help with your understanding of the story. After reading all of the stories in the set, please rank the stories in order of creativity, using your own subjective definition of creativity. Once you have ordered all the stories in the set, please tell the researcher and he will provide you with the second set of stories. Again, please read all the stories in the second set. After you have read each story in the second set, please insert the stories into the previously completed ordered set. Please repeat this process until all of the stories have been read and ranked 1 through 44, with 1 meaning “most creative” and 44 meaning “least creative.” Please make sure you rate the stories relative to one another, as opposed to universal standards of creativity. Also, please make sure that the number assigned to the story you are rating matches the number on the rating sheet you are using.

Once you have ordered all 44 stories, please write down the creativity ranking for this story in the space below.

B) What is the creativity ranking for this story, with 1 meaning “most creative” and 44 meaning “least creative”?

B) Once you have finished ordering all of the stories from most creative to least creative, please go back and rate each story on the 10-point scale provided below. You should rate at least one story a 1 and at least one story a 10, and you should spread the others out along the entire rating scale. However, the ratings do not have to be evenly distributed. You may give the same rating to as many stories as you want. Ratings may also go in between the boxes if you feel that a particular story does not fit into a particular box.

Creativity- Using your own subjective definition of creativity, please rate the degree to which the story is creative:

1	2	3	4	5	6	7	8	9	10
<input type="checkbox"/>									
Lowest									Highest

C) After rating this story on the ten-point creativity scale, please rate it on how “story-like” it is on the 3-point scale provided below. Unlike the creativity rankings and ratings for parts A and B, please rate this story against a universal standard for how much it resembles an actual story, as opposed to only rating the stories against one another. Therefore, you may rate the stories as you wish.; no story needs to be rated a 1 or a 3 if you feel that none deserve these ratings. Also, unlike the creativity rating scale used in part B, please only mark in one of the three boxes on the scale, not in between boxes.

Considering the age (4-5 years old) of these children, how do you consider this story? :

1	2	3
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Has no elements of being a story for a 4-5 year old.	This is a primitive story for a 4-5 year old.	This is a typical story for a 4-5 year old.

Appendix F: Highest Single-Ranked Stories

Participant 1, Experiment 1: highest ranked pretense story (session 10)

“He seems like he tripped over a log, dropped his things, and he’s falling. The dog’s hurrying to get into the water so he can have safety. The frog is surprised because he’s seeing that a boy’s gonna fall in his direction. He thought he saw there was a lily pad on the other side, and so he was gonna jump all the way over here (to the lily pad). So he (the boy) wouldn’t get on the frog.”

Participant 1, Experiment 1: highest ranked non-pretense story (session 15)

“He seems like he wants to catch the frog when the frog’s not looking. The dog’s talking to him (the frog) “Do not bother him and he won’t do it!” Cause that’s his friend, and he doesn’t want his friend to get caught by the net. So he’s (the dog) telling him (the frog) to not bother him (the boy); and then he (the boy) might catch him (the frog), (but) if he (the frog) doesn’t bother him (the boy), he won’t. He (the frog) jumps off the thing (log) and goes inside the water so he can have a few sips of water so he (the boy) won’t get him; ‘cause then he (the boy) won’t know where he (the frog) is.”

Participant 4, Experiment 1: highest ranked pretense story (session 12)

“This is the dog, and the boy’s kinda lookin’, and he pulls out the net. Then he’s trying to catch him (the frog). ‘Hop!’ Then the frog was goin’ Up! Up! All the way to the sky! And then he lands on the bucket! Then the boy drowns, and the dog stays up. He (the dog) goes under (the water). He’s (the frog) just safe.”

Participant 4, Experiment 1: highest ranked non-pretense story (session 4)

“The frog is on the rock and he’s trying to hop over there (the lily pad), then here (the water), then here (the grass), then here (the tree). (Then to the) grass. Then he’s gonna hop

over here (the water), and then “splash!” (Then) he bumps right into the tree. Then he hops on the water and then he goes up in here (the hole in the tree).”

Participant 1, Experiment 2: highest ranked pretense story (session 10)

“He, like, gets the net and goes jumping in! He got the net and the pail behind the dog. He ran with the net into the water. And the fish (lily pad) sinks a little bit more. He jumps, and his net and pail went flying down with the frog, and the fish (lily pad) sinks a little bit more. Now the fish (lily pad) is big-time sinking. This time up instead of down! And then the net and the pail and he (the boy) jumps into the water. He (the frog) jumps on the pail! You can just see the fish (lily pad) there. That whole frog, and the pail’s on the boy. He’ll (the boy) jump over the frog. He (the boy) goes on here (the log); he gonna get the dog. He got the dog! The fish (lily pad) is now more big-time. He (the boy) just gets the dog, but the fish (largest lily pad) is on that piece of ice, just cracking! There’s the bait (flower on big lily pad), and, like, that’s the mother fish (largest lily pad) and that’s the baby fish (second largest lily pad). The frog’s mad.”

Participant 1, Experiment 2: highest ranked non-pretense story (session 3)

“There’s the boy and the dog and the frog. There’s the turtle, and there’s a gift. They’re gonna get a pencil and then they’re gonna write ‘from’ and ‘to’ on it. And then they have to write the names of who it’s going to, and who it’s from. They open the gift, and there’s the tag and the tape, and there’s the box that the gift came in. And there’s the box that the gift came in again, and there’s the tag and the tape, the frog and the turtle. (There’s) the boy, the frog, another frog, and the turtle and dog. They’re inviting the baby to come to the momma frog. And then the boy, the frog, the other frog, the turtle, and the dog, they’re resting. There’s the boy, the dog, the turtle, and the two frogs. They’re wrestling, the two

frogs. Maybe the frog, the dog, and the momma frog, and the turtle, and the boy are all angry, 'cause look at all their faces. Maybe the turtle and the two frogs, the boy, and the dog, they're going for a walk. And now, the momma frog is on the turtle, and then the baby frog, he goes on the ice. And then the boy, and the dog, and the momma frog, and the turtle are all angry again. There's the frog, and the boy, and the dog, and the turtle, and the other frog. They're telling the frog, 'go away!'

Participant 2, Experiment 2: highest ranked pretense story (session 4)

"The frog is hiding in the salad. And when she takes a bite, she finds the frog. Then he hops out of the salad into the drink. Then the frog kisses him when he goes to drink. The frog's saying 'goodbye' and this guy's (the waiter) gonna grab him. They're (the couple) like 'sigh.' The little boy's pointing here; he's pointing to the frog. The guy's sending him (the boy) out. Now they're going home, and everybody's angry at him (the boy). And they're pointing him to go to his room. All the friends are in the room, and they're happy!"

Participant 2, Experiment 2: highest ranked non-pretense story (session 1)

"They're on a ship and there's two frogs, and this (big) frog doesn't want the little frog to get on. So he kicks it off. That's not a really nice thing. Then the little boy's wondering where the little frog could be. And he's looking. And this is he's crying 'cause he can't find the (little) frog, and the dog is blaming it on the frog. Now everybody's sad, and look at the turtle! He's in his shell. The dog's looking, so...and here comes the little frog! And he bounces on his (big frog's) head. 'Boing!' Ouch! Now that would be a problem. They're happy."

Participant 3, Experiment 2: highest ranked pretense story (session 10)

“The boy was looking in the mirror. The dog was sad ‘cause he didn’t get to go. He was really sad, even the turtle. They were already there. And the dog didn’t know. They were sitting at the table. The frog jumped up into the saxophone. He (the saxophonist) can’t blow out because the frog was stuck inside there and he couldn’t play it. He shook it, and then the frog jumped out onto his face. Then he landed right into the drum. Now he broke the drum, and he (the frog) was gonna eat. The froggy, he was jumping on the plate.”

Participant 3, Experiment 2: highest ranked non-pretense story (session 5)

“A big frog jumped onto the raft with the littler frog. He kicked the little frog off. That wasn’t very, very nice. He looks like a mean ol’ frog! I’m guessing the boy really likes the littler frog better than him. The boy was very sad, and went looking for him (the little frog). But he looked under a lily pad, but nothing there. He was crying when he was walking home. He ran to his bed and cried. He opened his eyes, and the little frog jumped right onto the big frog’s back. The boy was very happy! And the big frog and the little frog loved each other!”

Participant 4, Experiment 2: highest ranked pretense story (session 6)

“That’s the dog saying ‘Ouch! That will burn my tail!’ That’s the boy trying to catch the frog. And that’s the boy surprised with the frog. And that’s the doggy saying ‘Ouch! That would burn my table-head!’ That’s the boy trying to catch the frog, and that’s the dog saying, ‘I am brave! I am brave! I am brave! I am brave!’” And that’s the dog also saying ‘Ouch! That would burn my doody-head!’ That’s the boy not on safety road! And that’s the dog saying ‘Ouch! That will burn my doody-foot!’ That’s the boy saying ‘Ouch! That would burn my hand!’ And that’s the dog saying ‘Ouch! That would burn my little tail!’

That's the boy saying 'Ouch! Where is my frog?' He's saying 'ouch' 'cause he has a bucket over his eyes and he can't see. And that's the dog saying 'Ouch! That would burn my little doody-ear!' That's the frog saying 'Ha ha! I tricked him! (the boy)' That's the dog saying 'Ouch! That will burn my little doody-nose!' That's the frog saying 'Ha ha! I tricked the dog! Ha, ha, ha-ha, ha!' And that's the boy mad at the frog. And that's the dog saying 'Ouch! I am brave! I am brave!' And the doggy's also saying 'Ouch! That would burn my doody-self!' That's the dog trying to help the boy to capture the frog. And that's the dog saying, 'I am brave! I am brave!' The dog's also saying 'I will catch you! I will catch you!' That's the boy thinking that he dog is the frog. And that's the dog saying 'Ouch! Get this thing off of me!' And that's the dog saying 'Ouch! That would burn my little doody-head!' And that's also the boy saying 'There's a bucket on my head. You (the dog) must have do that!' That's the frog saying 'I do not like to be caught! I do not like to be caught!'"

Participant 4, Experiment 2: highest ranked non-pretense story (session 9)

"That's the boy saying, 'Is this gift for me?' And that's the dog saying 'Ouch! That would burn my little doody-nose!' while sliding like a snake. And that's boy saying '(Gasp)! It's my very own frog!' And the dog is saying 'Yay! Yay! Hooray! Hooray! Hooray! That frog is for me!' And that's him (the boy) getting his baby frog out. And the dog is saying '(Sigh). When am I getting my very own frog?' That's the boy saying, 'There you go! You (the baby frog) can play while I play with my doggy!' And that's the dog also saying 'Ouch! That would burn my little doody-eyes!' while putting his whiskers straight. And that's the boy saying, 'Play nicely.' And that's him (the big frog) just singing a song, but he sounded like a frog being squeezed by a boa constrictor. And that's the dog saying, '(Sigh). When will they play nicely?' And that's the lady frog trying to eat the baby frog. And that's the

boy and the dog and the turtle mad at the frog that was about to eat the baby frog. That's the dog saying, 'Grr! I'm not scared of you!' That's the frog about to push the baby frog out of the boat (turtle). And that's the frog that just pushed the baby frog out of the boat. And that's the boy saying 'You (the big frog)! You are in time out!' And that's frog saying, 'Hmph! Then I don't need them!' And that's the fly saying (in a southern accent), 'Hmm! Why is that little guy sitting by himself?' The end."

Appendix G: Approval for the Use of Human Subjects

January 27, 2004

Dear Mr. Ross:

The CAS-Human Subjects Committee has considered your application, #2153, "The Effect of Pretense on the Creativity of Storytelling in Preschool-Age Children", and consider it EXEMPT. This means that the proposal does not need further consideration by the University Human Subjects Committee and you may proceed with your research.

This letter should be presented with your thesis draft to the Graduate School as proof that you met the guidelines for research involving human subjects in the College of Arts and Sciences. Good luck with your endeavors, and your career.

Sincerely,
Michael J. Brabec, Chair
CAS-HSC