



WILEY-
BLACKWELL

SR Society for
CD Research in
Child Development

Some Issues in the Measurement of Children's Comprehension of Metaphorical Language

Author(s): Ralph E. Reynolds and Andrew Ortony

Source: *Child Development*, Vol. 51, No. 4 (Dec., 1980), pp. 1110-1119

Published by: [Blackwell Publishing](#) on behalf of the [Society for Research in Child Development](#)

Stable URL: <http://www.jstor.org/stable/1129551>

Accessed: 01/10/2010 00:57

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=black>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Blackwell Publishing and Society for Research in Child Development are collaborating with JSTOR to digitize, preserve and extend access to Child Development.

<http://www.jstor.org>

Some Issues in the Measurement of Children's Comprehension of Metaphorical Language

Ralph E. Reynolds and Andrew Ortony

University of Illinois at Urbana-Champaign

REYNOLDS, RALPH E., and ORTONY, ANDREW. *Some Issues in the Measurement of Children's Comprehension of Metaphorical Language*. CHILD DEVELOPMENT, 1980, 51, 1110-1119. Elementary school children ranging in age from 7 to 12 years read several short stories and selected (from a set of 4 alternatives) the sentence they thought best completed the story. The correct alternatives were related to the stories in either a figurative (simile or metaphor) or a literal manner. In 2 experiments, subjects selected the correct alternatives significantly more often when they were similes than when they were semantically equivalent metaphors. They also made more correct selections when the alternatives specifically denoted the referent of the metaphorical comparison than when the identity of the referent had to be inferred. The data were interpreted as supporting the view that measures of metaphor comprehension often confound general language performance variables with metaphoric ability.

Interest in the cognitive processes underlying the comprehension of metaphors has grown rapidly during the last few years. It has manifested itself in a few empirical studies conducted with adult subjects and in a rash of developmental studies. Many of these have attempted to establish that there are distinguishable levels of metaphoric comprehension progressing toward fully mature comprehension in early adolescence (e.g., Asch & Nerlove 1960; Winner, Rosenstiel, & Gardner 1976). There have also been numerous attempts to show that the development of the ability to understand metaphors is tied to Piagetian stages (e.g., Billow 1975; Cometa & Eson 1978). In addition, there have been studies aimed at demonstrating that children can understand metaphorical uses of language at much younger ages than the bulk of the available evidence implies (e.g., Gentner 1977; Honeck, Sowry, & Voegtle 1978).

Interesting as such studies are, most of them suffer from one or more of a variety of difficulties—difficulties that frequently relate to the inadequacy of the underlying theoretical account of metaphor per se and, consequently,

to the way in which the comprehension of metaphor is measured. One such difficulty is exemplified in studies (e.g., Asch & Nerlove 1960; Gardner 1974) investigating the comprehension of dual-function terms (terms like *hot*, *hard*, *bright*, etc., that can be applied in two or more domains, such as those of physical objects and of “psychological” characteristics). Results with children tend to show poorer comprehension of such terms when applied to psychological characteristics than when applied to physical objects, a finding that has been taken to show that the comprehension of metaphors is late in developing. However, dual-function terms hardly seem sufficiently representative of metaphorical language to warrant many important generalizations about such language. Furthermore, studies of this kind tend to confound metaphor comprehension with domain familiarity and knowledge of the world. So, for example, poorer performance on the “metaphorical” uses of words like *hard* might merely reflect a less well developed sensitivity to, and knowledge about, psychological characteristics as opposed to physical ones (see, e.g., Flavell 1977, p. 137).¹

The research reported herein was supported in part by the National Institute of Education under contract no. US-NIE-C-400-76-0116, and by a Spencer Fellowship awarded to the second author by the National Academy of Education. We wish to thank Richard Brummer, Debbie Bluder, Joe Campione, Glenn Kleiman, Larry Shirey, Sally Standiford, Rand Spiro, and Nancy Young for their help on various aspects of this work. We would also like to thank Mrs. Edna Horne, principal of Eastlawn Elementary School, and Mr. Ralph Dixon, principal of Pleasant Acres Elementary School, both of Rantoul, Illinois, for their cooperation. Send reprint requests to: Andrew Ortony, Center for the Study of Reading, University of Illinois, 51 Gerty Drive, Champaign, Illinois 61820.

¹ It should be noted that recent research by Sicone, Gardner, and Winner (Note 1) suggests that this may not be the source of the poorer performance.

Other approaches, such as the one reported by Gentner (1977), encourage the inference that, since very young children can perform certain tasks that show evidence of one kind of skill—say analogical reasoning—they have the cognitive wherewithal to understand metaphors. However, such an inference depends on the validity of certain theoretical assumptions—in this case, assumptions about the relationship between analogical reasoning and metaphor comprehension. Although the view that metaphors are based on the principles of analogy has been promulgated at least since the time of Aristotle, that does not mean that it is correct; in fact, there are reasons to suspect that it is not (see Ortony 1979).

In the normal course of events, figurative language, like literal language, occurs in a rich linguistic and physical context. It is now widely accepted that context is a major factor in comprehension. Yet, influential literature on the comprehension of metaphors and other figurative uses of language continues to report investigations of performance on stimuli that are presented with little or no context (e.g., Winner, Rosenstiel, & Gardner 1976). This seems to impose unreasonable and unrealistic demands on children. Ortony, Schallert, Reynolds, and Antos (1978) found that, with adults, the removal of adequate contextual support for an expression had a particularly detrimental effect if that expression required a metaphorical interpretation. There is no reason to believe that children are any less dependent on context than are adults. Accordingly, the present experiments investigated the comprehension of metaphorical language occurring against a reasonably realistic contextual background.

Given such differences in what is to count as metaphorical comprehension, it is hardly surprising that the available evidence concerning its development is inconclusive. In fact, the evidence is inconsistent. Some studies (e.g., Billow 1975; Gardner 1974; Gentner 1977; Pollio & Pollio 1974) suggest that quite young children, aged 5 years or younger, can use and understand metaphorical language, while others (e.g., Asch & Nerlove 1960; Matter & Davis 1975; Schaffer 1930) suggest that the ability to comprehend and use such language does not develop until early adolescence. These and other studies are reviewed in more detail in Ortony, Reynolds, and Arter (1978).

If the existing research is indeed based on differing conceptions of what metaphorical language is, inconsistent findings could result from the fact that measures of metaphor com-

prehension are sometimes confounded with measures of other, uncontrolled, variables. The purpose of the present research was to determine whether this might be the case, and, if so, to identify the kinds of variables involved. Any such enterprise requires its own account of what makes a metaphor a metaphor. The account presupposed in the present research is based on that presented in Ortony (1979) and in Ortony, Reynolds, and Arter (1978). The most important aspect of this view is that similes are metaphorical (as opposed to literal) statements of similarity. It has long been thought that metaphors (most transparently, predicative metaphors) are based on comparisons; so, for example, when we assert that someone, say John, is a snake, the statement is based on the comparison *John is like a snake*. However, this comparison is itself metaphorical (i.e., a simile): John is not *really* like a snake (perhaps eels are really like snakes), he is only like a snake *metaphorically*. The point is that, insofar as metaphors can be reduced to comparisons, the comparisons to which they reduce are themselves metaphorical. Thus, nothing about metaphoricity is explained by observing the connection between metaphors and their corresponding similarity statements. It follows from this that the difference between a (predicative) metaphor and its corresponding simile lies not in the fact that one is metaphorical and the other not, but in the fact that one is an *indirect* statement of the other. Thus, *John is a snake* is an indirect way of asserting that *John is like a snake*, but *both* are metaphorical. In both cases, understanding the assertion involves relating the terms from disparate domains in the appropriate way.

If one is interested in whether metaphorical language as such is a cause of comprehension difficulties for children, it becomes important to distinguish metaphoricity from indirectness. It might be that the child's ability to understand metaphorical language is adversely affected, or even totally obscured, by indirectness. Experiment 1 was designed to determine whether this is so. If metaphorical language itself is a principal source of comprehension difficulties, then children should gain no benefit from receiving stimuli in the form of similes rather than their corresponding metaphors. In such a case, one might say that the limitation on a child's performance was more likely to be a genuine limitation of competence, because simplifying the task by eliminating a general language-related variable would not help the child. One could, with much greater confidence, attribute serious comprehension defi-

1112 Child Development

ciencies to an inability to relate the two domains appropriately. If, on the other hand, an ability to understand metaphorical language were hampered or masked by the indirectness of metaphors, then children might do better on a simile task than on a metaphor task, because similes contain an explicit syntactic signal that a comparison is to be made. Of course, these predictions only make sense if the child has sufficient knowledge about the domains involved. In developing the materials for the present experiments, care was taken to ensure that children were likely to have enough of the requisite knowledge.

In the experiments, children read short stories and then selected what they judged to be the most appropriate of four presented continuation sentences. These sentences were constructed so that in critical cases the correct response involved a metaphorical comparison. Sometimes the comparison was explicit, in the form of a simile, and sometimes it was implicit, in the form of a (corresponding) metaphor. Henceforth the term *figurative* will be used to refer to either metaphor or simile conditions in the experiments. Thus *figurative* is to be contrasted with *literal*.

Experiment 1

Method

Subjects.—The subjects were 240 second-through sixth-grade children from a rural elementary school. Children who were judged by their teachers to be unable to read sufficiently well to perform the task were excluded from the subject pool. Approximately half the children were girls and half boys. Mean ages were: second grade, 8-2 ($N = 50$), third grade, 9-0 ($N = 56$), fourth grade, 10-1 ($N = 46$), fifth grade, 11-2 ($N = 44$), and sixth grade, 12-4 ($N = 44$).

Design and materials.—The design was a four-way factorial design with grade, type of figurative target, list, and block order as between-subjects factors. In addition, there was a small external control group.

The task was to read a short story and then to select the most appropriate continuation sentence (hereafter called the target) in a four-alternative forced-choice test. Each story was accompanied by a color drawing that illustrated its main idea. The manipulation of in-

terest was the type of figurative use employed by the target. In one experimental condition correct selection of the targets involved (ideally) the comprehension of metaphors, whereas in the other condition correct selection involved the comprehension of semantically matched similes.² In addition to the selection of figurative targets, all subjects received items in which they were required to select literal targets.

The experimental texts were eight titled short stories (average length, 70 words) about topics that were considered to be familiar to young children. For each story three sets of four alternatives were constructed, a literal-target set, a metaphor-target set, and a matched simile-target set. In each case, the target sentence was supposed to be the most natural extension of the story; it described what might be expected to happen next. (For ease of discussion we often refer to a story followed by a set of alternatives as an "item.") An example of a complete set of materials for one story will help illustrate the different types of alternative sentences:

The Old Race Horse

Jack Flash had been a great race horse when he was young, but now he was too old to race. His owner thought Jack Flash wasn't good for anything anymore. None of the other people who worked at the ranch where Jack lived paid any attention to him. No one wanted to ride an old broken-down horse. The owner decided that he did not want Jack around where people could see him.

Literal-target set

Jack was sent to one of the pastures in the back of the farm. (T)

The owner of the ranch played with Jack everyday. Jack was given the best stall on the ranch to stay in.

Jack hated eating oats for breakfast.

Metaphor-target set

The worn-out shoe was thrown into the trash. (T)

The saddle was polished and shiny. (A)

The race was going to begin. (A)

The raincoat was new. (R)

Simile-target set

It was like a worn-out shoe that had been thrown into the trash. (T)

It was like a saddle that was polished and shiny. (A)

It was like a race which was going to begin. (A)

It was like a new raincoat. (R)

² The claim that the similes and metaphors were "semantically" matched is intended to imply that the transformations for mapping the one into the other (e.g., *like* deletion/insertion) did not interfere with the basic ideas expressed.

In this example, the first member (T) of each set is the target. In the literal-target set it is the only sentence that, when interpreted literally, makes good sense in the context; none of the sentences is amenable to reasonable metaphorical interpretations. In the metaphor-target set, none of the alternatives makes sense if interpreted literally, but there is a ready metaphorical interpretation that can be given to the target. In the simile-target set, nothing in the story was literally like any of the things mentioned, but, metaphorically speaking, Jack Flash was like a worn-out shoe. Targets varied randomly in location and length with respect to the distractors.

For figurative-target sets, distractors containing no obvious thematic relation to the story were used. Two of them (A) always contained a word or words with high associative relationships to words in the story. The third (R) was a sentence with a superficial resemblance to the target. For instance, if the target had an animal as its subject, the distractor superficially resembling it might also be about an animal. In literal-target sets the three distractors were each related to the story in only a superficial manner.

All the alternatives used normal English sentences. In the metaphor condition the targets required a metaphorical interpretation only because they occurred in the context of the story: The sentence, "The worn-out shoe was thrown into the trash," is not in itself metaphorical, nor is it likely to need a metaphorical interpretation in most contexts in which it would normally be encountered. However, in the story about Jack Flash it must be interpreted metaphorically if it is to make sense at all (for further discussion, see Ortony, Reynolds, & Arter 1978; Ortony, Schallert, Reynolds, & Antos 1978). Metaphors were converted to similes by adding the word *like* together with the appropriate syntactic transformations where necessary. The generic pronoun *it* was used to refer to the topic of the simile that appeared in the story.

As well as grade and type of figurative usage, two other independent variables were included. The first, a list factor, resulted from two different random orders of the eight stories as a safeguard against possible story-sequence effects. The block-order factor was concerned with counterbalancing subjects' exposure to literal and figurative items. The sets of alternatives in the response booklets were arranged so that half of the subjects received booklets

in which a block of four figurative-target sets was followed by a block of four literal-target sets, and the other half received booklets in which the converse was true. The figurative block always contained either only metaphor-target sets or only simile-target sets.

Each list of eight stories was preceded by four practice items. These always appeared in the same order, with the first two being literal and the second two being figurative (both were metaphors for subjects receiving metaphor-target sets and both were similes for those receiving simile-target sets in the figurative block). Since the type of alternatives each subject received in the response booklet defined what condition he or she was in, it was possible for subjects from all conditions to be present in each experimental session.

Finally, as a precaution against the possibility that the correct selection of targets could be reliably accomplished independently of reading and understanding the stories, a separately run external control group received the forced-choice test after seeing only the title of the stories together with the picture. Subjects in this group received no feedback.

To confirm our intuitions as to the appropriateness of the figurative targets, the items were given to 20 students in an introductory psychology class at the University of Illinois. These students worked through the experimental booklets exactly as the experimental subjects did. Eighteen of these subjects completed the booklet without error, the other two made one error each (on different items). On the basis of this evidence, the targets were judged to be reasonable extensions of the stories, at least from the perspective of adults—a reasonable criterion for "mature" comprehension.

Procedure.—Subjects participated in 30 groups ranging in size from four to 10. Students were taken from their classrooms and randomly assigned to one of two treatment groups or to the control group. One treatment group received the first list of experimental stories, the other the second. Response booklets were distributed as the subjects entered the experimental area. Subjects were seated in individual seats facing a projector screen. Each response booklet contained a cover sheet and a page of instructions. The instructions, which were read aloud as the subjects read to themselves, directed subjects to read each story silently as it was shown on the screen. The story was presented via overhead projector and was read aloud by the experimenter. It was

removed and the picture representing the main theme of the story was shown. Then, with the picture still visible, the subjects were told to open their booklets and circle the sentence that best "completes" or "fits with" the story they had just read. When the subjects all acknowledged that they understood the instructions, the four practice items were completed. Subjects were given the correct responses for these practice items. Since subjects were in different figurative conditions, the correct responses on the practice items were identified by the experimenter in general terms such as "it was the one about the robber." Thus, the feedback given to subjects was appropriate regardless of whether they had a simile-target set booklet or a metaphor-target set booklet. Subjects were allowed to ask any questions they wished about the instructions or the task. The eight experimental items were then presented without interruption.

Results and Discussion

Upon interviewing subjects and examining their protocols, it became obvious that the children viewed one of the distractors as a very reasonable continuation of the story. The item was answered incorrectly on 70% of the protocols, with the vast majority of the errors resulting from the selection of this attractive distractor. The item was dropped from all analyses of figurative responses. Table 1 shows the mean proportions of correct responses for both literal and figurative conditions, collapsed across list and block order.

Although five grade levels were tested, the responses of second graders were excluded from all ANOVAs. This was because the subjects made available to us excluded a large proportion (almost 33%) of the second-grade children, namely, those deemed by their teachers to be unable to read sufficiently well to per-

form the task. In other words, those second-grade children who did participate represented a nonrandom sample. Thus, although the second-grade data are included in table 1, they should not be regarded as being representative of second-grade performance overall.

Separate analyses were performed on figurative and literal responses. This was done because, whereas in the two figurative conditions the distractors were semantically matched, and therefore comparable, distractors in the literal condition were not matched (i.e., not semantically related) to those in the figurative conditions, and therefore not comparable.

A 4 (grade) \times 2 (figurative type) \times 2 (list) \times 2 (block order) analysis of variance was performed on figurative responses. Main effects for grade, $F(3,128) = 10.77$, $p < .01$, figurative type (metaphor or simile), $F(1,128) = 20.14$, $p < .01$, and block order, $F(1,128) = 14.55$, $p < .01$, were significant. The grade result was due to increased correct responses by older subjects. The figurative-type effect was due to more correct responses by subjects in the simile condition than in the metaphor condition. The block-order effect was due to generally better performance on the figurative items when they appeared in the first block rather than in the second. Significant interactions were found for grade \times figurative type, $F(3,128) = 3.07$, $p < .05$, grade \times block order, $F(3,128) = 2.76$, $p < .05$, and list \times block order, $F(1,128) = 6.31$, $p < .05$. The grade \times figurative type interaction was due to a reduction in the advantage of similes over metaphors for the older children. The grade \times block order interaction resulted from an increasing advantage of figurative items in the first block for older children. Finally, the list \times block order interaction was due to superior performance by subjects on one of the lists when the figurative items were presented in the first block of items. No other results reached significance.

An identical analysis was performed on literal responses. The only significant finding was a main effect for block order, $F(1,128) = 9.61$, $p < .01$. This was due to superior performance on the literals when they occurred in the first block of four rather than in the second.

The performance of the control group confirmed that, although the color pictures helped subjects retain the stories' main ideas, they did not assist them on the figurative items. With-

TABLE 1

PROPORTION OF CORRECT RESPONSES COLLAPSED
ACROSS LIST AND ORDER CONDITIONS,
EXPERIMENT 1

CONDITION	GRADE				
	2	3	4	5	6
Literal80	.83	.87	.93	.91
Simile39	.34	.56	.57	.66
Metaphor21	.11	.21	.58	.38

NOTE.—Approximately 33% of available second graders had to be excluded from the study because of reading difficulties; hence the second-grade scores represent subjects of better-than-average reading ability.

out the stories subjects averaged only about 6% correct on the simile items and 5% correct on the metaphors (chance $\leq 25\%$). Even in the literal condition, considerable advantages accrued from understanding the stories themselves, subjects in the control group only averaging about 45% correct. Scores were collapsed across grade, list, and block order to obtain these figures.

These results, especially the main effect for figurative type, lend support to the notion that measures of the comprehension of metaphorical language can easily be contaminated by variables having nothing specifically to do with the metaphorical nature of such language. Since there was no difference in the semantic content of the metaphors and the similes, differences in performance must have been due to differences in the surface structure of the comparisons. If subjects had lacked some cognitive process required to relate the disparate domains involved in the figurative targets (e.g., of an old race horse and a worn-out shoe), there would be no reason to expect an overall superior performance on similes than on metaphors. Nor would there be any reason to expect this superiority to be greater for younger children. The evidence for this last result (i.e., the grade \times figurative type interaction), however, needs to be treated with some caution because it seems to depend rather heavily on the performance of only one of the grade levels (fifth or sixth grade).

If the metaphors are viewed as being indirect similes, then the figurative-type effect must have been primarily due to indirectness. Presumably, other variables not specifically and necessarily related to the metaphorical nature of the targets could produce similar effects. For example, a factor that may have added to the difficulty of the figurative items in Experiment 1 was the specificity (or lack of it) of the referring expressions in the response alternatives. Thus, in the metaphor condition, a noun would appear in subject position together with the definite article even though there had been no previous reference to such an object. In other words, again using the Jack Flash example, there was no shoe, saddle, race, or raincoat in the story to which the words in the alternatives could refer. In the simile condition, the generic pronoun *it* was used to refer to the referent. It may well be that children find the generic use of *it* to be quite difficult. Thus, it is possible that, in the experiment, children's ability to understand figurative lan-

guage may still have been partly obscured by the difficulty of identifying the referents of superficially misleading or difficult referring expressions. In Experiment 2 specificity of reference was manipulated by including a specific reference to the topic (i.e., explicit mention of the name of the referent) in some of the experimental conditions. If our general claim is valid that measures of metaphor comprehension tend to be confused with measures of other, theoretically unrelated, performance factors, then manipulating a variable like specificity of reference ought again to result in changes in the overall level of performance. Such changes would tend to support our general claim independently of the theoretical preposition that similes are themselves metaphorical.

A second goal of Experiment 2 related to the developmental trend suggested by the fact that the grade \times figurative type interaction was statistically significant. It can be seen from table 1 that there was a considerable change in the trend of the data from fifth to sixth grade in the metaphor condition. This raises the possibility that the interaction does not reflect a real developmental trend. The procedure was changed in Experiment 2 so as to give greater power. Since Experiment 1 had shown that children could all perform well on the literal items, less emphasis was placed on them in Experiment 2. Subjects received all eight stories in a figurative condition followed by the same eight stories in the literal condition.

Experiment 2

Method

Subjects.—The subjects were 171 students from a rural elementary school, different from the school used in Experiment 1. Children unable to read sufficiently well to perform the task were excluded. Approximately half of the children were girls and half boys. Mean ages were: second grade, 7-6 ($N = 22$), third grade, 8-7 ($N = 37$), fourth grade, 9-8 ($N = 38$), fifth grade, 10-6 ($N = 37$), and sixth grade, 11-6 ($N = 37$).

Design and materials.—The basic design was a three-way factorial design, with grade, figurative type, and reference type as between-subjects factors.

The stories were the same as those used in the first experiment. The item that was dropped from the analysis in Experiment 1 was reused with a slight modification to the distractor that had proved to be defective. The

1116 Child Development

alternatives were identical with those used in Experiment 1 in the two nonspecific-reference conditions, and were appropriately modified in the other, specific-reference, conditions. Alternatives in the specific (reference) metaphor condition were constructed by introducing the identity of the referent in subject position using only expressions that specifically and literally denoted it. An example will demonstrate the differences between the two sets of materials:

The New Baseball Glove

Johnny's old ball glove was ruined. One of his friends had borrowed it and left it out in the rain. Johnny's parents knew how much he liked to play ball, so they gave him a new glove. They told him that he should take better care of this new glove. If he let someone ruin it like the last one, they would not buy him another one. Johnny decided that he would not let his friends even see his new glove.

Literal-target set

Johnny hid the glove in his closet.
Johnny needed a new pair of shoes.
Johnny's father was a baseball umpire.
Johnny's mother drove him to school each day.

Metaphor-target set (nonspecific)

The dog buried the bone in the back yard.
The father dropped a bowl of soup.
The batter missed the ball.
The kitten played with a ball of yarn.

Metaphor-target set (specific)

Johnny was a dog burying a bone in the backyard.
Johnny was father dropping a bowl of soup.
Johnny was a batter missing the ball.
Johnny was a kitten playing with a ball of yarn.

In the nonspecific condition, the simile sets were identical with those in Experiment 1. In the specific condition, the simile sets were derived from the specific metaphor sets by introducing the word *like* after the main verb. The response booklets were constructed slightly differently than those in the first study to accommodate the differences in design. Again, the booklets contained a cover sheet and written instructions. The instructions were the same as in the first study, suggesting that each child read the story silently as it was read aloud and then circle the alternative that best fitted or completed the story. The same four practice items (the two literals followed by the two figurative examples) were used in Experiment 2 as were used in Experiment 1. These items were followed first by the eight figurative items and then by the (same) eight literal items.

Procedure.—The procedure was similar to that of Experiment 1. Subjects participated in 30 groups ranging in size from four to seven. The

students were taken from their classrooms and escorted to the experimental areas, where response booklets were randomly assigned to them. All of the experimental conditions except grade were represented in each session. The rest of the procedure was identical with that of the first study except that all eight stories were administered twice.

Results and Discussion

Informal interviews with subjects after the experiment revealed that the item found to be defective in Experiment 1 still had a highly appropriate distractor. The item was answered incorrectly 78% of the time, again because of the attractiveness of this distractor. Apparently we had misjudged the source of the problem. The item was dropped from all further figurative analyses. Table 2 shows the proportion of correct responses in the various conditions.

Again, as in Experiment 1, the data from the second graders were not included in the two (separate) ANOVAs. A 4 (grade) \times 2 (figurative type) \times 2 (reference type) unweighted-means analysis of variance was performed on the figurative responses. Significant main effects were found for grade, $F(3,133) = 12.99$, $p < .01$, figurative type (metaphor or simile), $F(1,133) = 4.07$, $p < .05$, and reference type (specific or nonspecific), $F(2,133) = 20.07$, $p < .01$. The grade main effect was due to the higher number of correct responses by the older subjects. The figurative-type main effect was due to students doing better on similes than on metaphors. The reference-type effect reflected more correct responses with specific referring expressions than with nonspecific referring expressions.

TABLE 2
PROPORTION OF CORRECT RESPONSES,
EXPERIMENT 2

CONDITION	GRADE				
	2	3	4	5	6
Literal88	.82	.89	.97	.97
Specific simile57	.61	.64	.69	.79
Nonspecific simile54	.40	.49	.71	.70
Specific metaphor43	.45	.69	.75	.80
Nonspecific metaphor29	.14	.49	.51	.57

NOTE.—Approximately 33% of available second graders had to be excluded from the study because of reading difficulties; hence the second-grade scores represent subjects of better-than-average reading ability.

An identical analysis performed on literal responses revealed a main effect for grade, $F(3,133) = 9.98$, $p < .01$. This was due to more correct responses recorded by the older subjects. No other results were significant.

As expected, making the reference specific had a marked effect on the overall level of performance. In the metaphor condition the mean gain across grades resulting from making the reference specific was 25%. It was 12% in the simile condition.

In the present experiment there was no evidence of an interaction between grade and figurative type, $F(3,133) < 1$. Table 2 reveals that while sixth-grade performance increased over the level in Experiment 1, there was also a drop in fifth-grade performance, accompanied by an increase in the performance of the fourth graders. Nor was there an interaction between grade and reference type, again, $F(3,133) < 1$.

General Discussion

Converging evidence from the two experiments suggests that measures of children's ability to understand metaphorical language can all too easily be confounded with measures of other general language variables that have no particular connection to metaphorical language. The present research examined the interfering effects of two such variables, indirectness (metaphors being regarded as indirect similes) and specificity of reference. Both were found to have a significant impact on performance. Similes were understood more easily than corresponding metaphors, and metaphorical language involving specific referents was understood more easily than metaphorical language involving nonspecific referents.

Consider, specifically, the effect of specificity of reference in the metaphor condition of Experiment 2. Referring again to the Jack Flash example, in the metaphor condition the target was either, "The worn-out shoe was thrown into the trash," or "Jack Flash was a worn-out shoe thrown into the trash." Both are metaphors, but in the first case the intended referent of the "worn-out shoe" has to be inferred, whereas in the latter case it is explicitly stated (as being Jack Flash). When the referent of the subject term of the target sentence was explicit—that is, when the target sentence contained a metaphorical predicate—subjects in all grades tested showed evidence of being able to understand the metaphor. By

contrast, when the whole sentence called for a metaphorical interpretation so that children had to infer that the "worn-out shoe" referred to Jack Flash, performance in all grades tested was 20–30 percentage points poorer. Since the structure of very many metaphors encountered in ordinary discourse is that of a metaphorical predicate attached to a (literally) specified subject, the specific referent condition is very representative of normal performance. The relatively high level of performance at all grades with these more familiar metaphoric forms is an important finding.

The sensitivity of measures of metaphor comprehension to distortion through theoretically unrelated variables receives further confirmation by considering a probable reason for the significant block-order effects found in Experiment 1. Subjects performed worse on the second block of four items than on the first block of four, as indicated by both the figurative and literal analyses. A reasonable explanation of this finding is that the first block of experimental items produced an expectation in subjects that subsequent items would be similar in character (there was no break between the two blocks of items). According to this account, subjects would always approach the second block of items with an inappropriate set, resulting in poorer performance. This problem would not arise for items in the first block, where performance may even have benefited from their proximity to practice items of the appropriate type. If this analysis is correct, it would suggest that the expectation of encountering language of a particular type (i.e., literal or figurative) might constitute yet another performance-related variable that could contaminate a metaphor comprehension measure.

While the results of both experiments are consistent in showing the influence of general language processing variables on children's comprehension of metaphorical uses of language, evidence of any interesting developmental trends is less clear. It was anticipated that both experiments might reveal an interaction between grade and figurative type, with performance on metaphors finally converging with that on similes in the later grades. This interaction was significant in Experiment 1 but not in Experiment 2. Furthermore, there was no evidence of an interaction between grade and reference type in Experiment 2. The absence of these interactions remains something of a puzzle—a puzzle whose resolution will have to await further research.

1118 Child Development

Given the inconclusive state of the existing research, an important motivation for this study was to determine how best one might measure a child's ability to understand metaphorical language. Experiment 2 shows that if one were to select the nonspecific metaphor condition as a representative test of metaphoric comprehension, then the average performance across grades would be at only about the 40% level. By contrast, if one were to select the specific simile condition as representative, the average performance level would be close to 70%. This latter measure is the theoretically purest measure of the four, and the data it provides leave little doubt that young children can understand metaphorical uses of language. Of course, like everyone else, and perhaps more so, children can and do make mistakes in interpretation.

Even our purest measure of comprehension still provides a rather conservative test—first, because other variables such as thematic relatedness and general world knowledge may have contributed to comprehension difficulties, and second, because of the way in which the distractors were constructed. It was assumed that 25% correct was the level of chance performance. However, if a child did not perceive the target as being the appropriate choice in the figurative conditions, it is reasonable to suppose that he or she would be drawn toward one of the distractors containing high associates of the theme of the story rather than select an alternative at random. Evidence in support of this supposition comes from the results of the picture-only control group in Experiment 1, where the targets were selected only about 5% of the time. This suggests that children were using a strategy of selecting something that was superficially related if they did not select the target. Additional evidence is provided by an analysis of errors. A genuinely random selection strategy would result in each incorrect alternative being selected with more or less equal frequency. Thus, of the three incorrect responses, the two containing words that were highly associated with the theme of the story should represent about 67% of the erroneous responses, while the unassociated distractor should account for about 33% of the errors. In fact, however, the associated distractors accounted for 88% of the errors in Experiment 1 and 83% in Experiment 2, both significantly higher than 67%. This suggests that subjects were drawn toward a response that possessed at least some superficial relationship to what

they had read. So, the probability of subjects correctly selecting a target while not realizing it was the correct response was less than .25 in the figurative conditions.

These experiments go some way toward explaining the inconsistent findings of previous research. The most optimistic interpretation of the data from the nonspecific metaphor condition in Experiment 2 could not establish metaphoric competence until about age 9½, whereas a comparably optimistic interpretation of the data from the specific simile condition shows a high level of performance as early as age 7½. To study performance with still younger children would necessitate the adoption of an experimental paradigm that did not require the children to read, since the present research shows that this can be a problem even for second-grade children.

Reference Note

1. Sicone, M.; Gardner, H.; & Winner, E. Understanding the psychology in psychological metaphors. Unpublished manuscript, Harvard University, 1979.

References

- Asch, S. E., & Nerlove, H. The development of double-function terms in children. In B. Kaplan & S. Wapner (Eds.), *Perspectives in psychological theory*. New York: International Universities Press, 1960.
- Billow, R. M. A cognitive developmental study of metaphor comprehension. *Developmental Psychology*, 1975, **11**, 415–423.
- Cometa, M. S., & Eson, M. E. Logical operations and metaphor interpretation: a Piagetian model. *Child Development*, 1978, **49**, 649–659.
- Flavell, J. H. *Cognitive development*. Englewood Cliffs, N.J.: Prentice-Hall, 1977.
- Gardner, H. Metaphors and modalities: how children project polar adjectives onto diverse domains. *Child Development*, 1974, **45**, 84–91.
- Gentner, D. On the development of metaphorical processing. *Child Development*, 1977, **48**, 1034–1039.
- Honeck, R. P.; Sowry, B. M.; & Voegtli, K. Proverbial understanding in a pictorial context. *Child Development*, 1978, **49**, 327–331.
- Matter, G. A., & Davis, L. A. A reply to metaphor and linguistic theory. *Quarterly Journal of Speech*, 1975, **61**, 322–327.
- Ortony, A. Beyond literal similarity. *Psychological Review*, 1979, **86**, 161–180.

- Ortony, A.; Reynolds, R. E.; & Arter, J. A. Metaphor: theoretical and empirical research. *Psychological Bulletin*, 1978, **85**, 919-943.
- Ortony, A.; Schallert, D. L.; Reynolds, R. E.; & Antos, S. J. Interpreting metaphors and idioms: some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1978, **17**, 465-477.
- Pollio, M., & Pollio, H. R. The development of figurative language in children. *Journal of Psychology Research*, 1974, **3**, 138-143.
- Schaffer, L. F. Children's interpretation of cartoons. In *Contributions to education*. New York: Teachers College, Columbia University, 1930.
- Winner, E.; Rosenstiel, A.; & Gardner, H. The development of metaphoric understanding. *Developmental Psychology*, 1976, **12**, 289-297.