ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

PRODUCTION NOTE

University of Illinois at
Urbana-Champaign Library
Large-scale Digitization Project, 2007.

TECHNICAL REPORTS

Technical Report No. 17

CHILDREN'S COMPREHENSION OF HIGH- AND LOW-INTEREST MATERIAL AND A COMPARISON OF TWO CLOZE SCORING METHODS

Steven R. Asher, Shelley Hymel,
and Allan Wigfield

University of Illinois at Urbana-Champaign

November 1976

Center for the Study of Reading

UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN
1005 West Nevada Street
Urbana, Illinois 61801

BOLT BERANEK AND NEWMAN INC.
50 Moulton Street
Cambridge, Massachusetts 02138

CENTER FOR THE STUDY OF READING*

Technical Report No. 17

CHILDREN'S COMPREHENSION OF HIGH- AND LOW-INTEREST MATERIAL AND A COMPARISON OF TWO CLOZE SCORING METHODS

Steven R. Asher, Shelley Hymel,
and Allan Wigfield

University of Illinois at Urbana-Champaign

November 1976

University of Illinois
at Urbana-Champaign
1005 West Nevada Street
Urbana, Illinois 61801

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02138

The research reported herein was supported in part by the National Institute of Education under Grant No. NE-6-00-3-0060 and Contract No. MS-NIE-C-400-76-0116, and in part by research support from the University of Illinois Research Board. The study was conducted with the cooperation of Robeson Elementary School, Champaign, Illinois. Marcia Anderson assisted in preparation of materials, Richard Markell served as one of the experimenters, and R. L. Hinkle helped analyze the data.

*Designation as a Center pending approval.

This page is intentionally blank.

Abstract

The present study investigated children's comprehension of reading material which was of either high- or low-interest. Previous research indicates that children comprehend more of high- than low-interest material when each child is given a mixture of both types of material. This effect could be due to a contrast effect whereby children selectively respond to the more appealing passages in their set of passages. In the present study each child received either all high-interest passages or all low-interest passages but not both. Fifth grade children's interests were assessed using a picture rating technique. One week later each child read cloze passages corresponding to the child's highest or lowest rated topics. Children's cloze responses were scored by the typical exact replacement method and by a method which included synonyms as correct. Results were that children comprehended more of high- than low-interest material, suggesting that the interest effect is not dependent on a contrast phenomenon. Synonym production data indicated that high- achieving children generated more synonyms than low-achieving children and that boys produced more synonyms than girls. In contrast to the oft-stated conclusion in the literature, there appears to be some informational value of scoring synonyms as correct.

Children's Comprehension of High- and Low-Interest Material and a Comparison of Two Cloze Scoring Methods

Recent evidence indicates that children's reading comprehension is affected by their level of interest in the content of the material (Asher and Markell, 1974). Fifth-grade children's interests were individually assessed using a picture-rating technique. Children rated the interest value of each of 25 photographic slides. One week later each child received six passages, three of which corresponded to his or her highly-rated topics, and three of which corresponded to topics that were rated low. All passages were presented in cloze format (Taylor, 1953) with every fifth word deleted. The child's task was to read the passage and replace each of the missing words.

Asher and Markell's findings indicated that girls' reading performance was slightly affected by their interest in the reading material and that boys' performance was strongly affected by the interest level of the material. On low-interest material and on a school-administered reading achievement test, boys performed significantly poorer than girls. However, on the high-interest material the sex difference was eliminated. These results have potentially important implications for the assessment of children's reading ability; giving children passages of low interest may seriously underestimate some children's ability to gain information from written material.

The research methodology in the Asher and Markell study represents an advance over earlier investigations of the effect of interest on reading comprehension. One element is the individualized assessment of children's interests independent of any particular reading material. In one study (Schnayer, 1967), children's interest in a topic was measured after they read a passage on that topic and had been tested for comprehension. This procedure confounds the reading comprehension measure with the interest assessment procedure. Children's interest reports could be a function of either their comprehension of the material or its topic appeal. In other studies, normative data on children's interests have been used to select passages (Bernstein, 1955; Klein, 1969; Stanchfield, 1967). Since individual children's interests differ from group norms (e.g. not all boys like basketball), using group norms instead of individual assessment introduces considerable experimental error.

Second, the Asher and Markell study used a large number of passages sampled from a wide array of reading topics. In much of the previous research only two passages have been used, one which is supposedly high-interest and the other which is intended to be of low-interest (e.g. Dorsel, 1975; Klein, 1969). This restricted sampling increases the likelihood that results obtained are limited to the particular passages employed.

Finally, the use of the cloze procedure as a measure of comprehension has several advantages: (1) it provides objective and replicable procedures for creating test items on any given sample of reading material; (2) it produces reliable scores; and (3) cloze scores correlate highly with standardized reading achievement test scores (Bormuth, 1967; 1968; Rankin and Culhane, 1969). Previous studies of interest effects often have used reading achievement tests specifically developed for each study with no prior demonstration of test reliability or validity (Bernstein, 1955; Stanchfield, 1967). In many cases item selection appears to have been arbitrary.

The present study focuses on two issues. First, the generality of the Asher and Markell findings is examined. In their study each child received both high-interest and low-interest passages. It is conceivable that the results obtained were dependent on a contrast effect whereby children selectively responded to the more interesting passages in their set of materials. Rarely in the school day are children assigned reading material that provides such clearly identifiable variation in topic appeal. In the present study, the possibility that a contrast effect would operate was eliminated by employing a between-subjects design in which each child was given either all high-interest or all low-interest passages. No effect of interest would be expected if the effect is dependent on a contrast phenomenon. If, however, the interest effect is not dependent on a contrast effect then the original findings should be replicated with a between-subjects design.

The second issue examined in the present research is whether the validity of the cloze procedure is increased by accepting synonyms as well as exact replacements of the deleted words. Asher and Markell (1974) scored a response as correct only if it was an exact replacement or a misspelled exact replacement of a deleted word. This procedure follows the oft-cited conclusion that accepting synonyms does not increase the validity of the procedure and only increases inefficiency and subjectivity of scoring (Bormuth, 1965; Jongsma, 1974; Taylor, 1953).

There may be reason to question the generality of this conclusion. Most studies favoring the use of exact replacement scoring systems have used passages rather than individual readers as the unit of analysis. Cloze scoring methods are compared in terms of how they discriminate passages which vary in reading difficulty level. Two most frequently cited studies are by Taylor (1953) and Bormuth (1965). Using a small sample of readers (N = 12) and passages (N = 3), Taylor (1953) compared an exact scoring method with a weighted scoring method in which partial credit was given for synonym replacements. The weighted scoring method raised scores for each of the 3 passages but did not change the ranking of the passages in terms of difficulty. In a more extensive study, with 50 readers and 20 passages, Bormuth (1965) found that the exact replacement method discriminated among the passages slightly better than the exact plus synonym scoring method.

Although the exact method best discriminates among passages, the exact plus synonym method might best disciminate between individual

readers. For example, readers who comprehend more of a passage might produce more synonyms than poorer readers. A few studies have compared various cloze scoring systems using the individual as the unit of analysis. The approach has been to correlate cloze scores with achievement test scores to determine which scoring method produces the highest correlation. Bormuth (1965) found that the exact replacement score correlated .82 with achievement scores while a score based on the number of synonyms correlated .64 with achievement scores. Unfortunately, an exact plus synonym scoring category was not included in these analyses. The correlation of an exact plus synonym scoring system with achievement scores is the critical test; a scoring system based on synonyms alone would not likely be used.

One study which compared exact and exact plus synonym scoring systems was done by Ruddell (1964). Six different passages were used. For all six passages the split-half reliability coefficients were higher with the exact plus synonym scores than for the exact replacement scores. For two of the passages the difference was significant. Ruddell found no significant differences in the validity of the exact and exact plus synonym scoring methods as measured by correlations of cloze scores on each passage with achievement scores. However, 5 of the 6 correlations with exact plus synonym scores were slightly higher than those with exact scores.

In another study which compared scoring systems, Schoelles (1971) found higher correlations between exact plus synonym scores and achievement test scores than between exact replacement scores and achievement scores. For second grade children, the correlation for exact plus synonym scores was .94, and for the exact scores, .89. For sixth graders, the exact plus synonym scores correlated .82 with achievement scores and the exact scores correlated .38.

It seems, then, that scoring synonyms as correct does not alter the distribution of passage scores but it may influence the way in which the scores of individual readers are distributed. However, interpretation of previous research is made particularly difficult, since only Ruddell (1964) has provided information on how synonyms were defined and none of the studies have reported data on how reliable judges are in deciding whether a response is a synonym.

The present study examined whether the correlation of children's cloze scores and their reading achievement test scores increases by accepting synonyms as correct. The present study also investigated whether the production of synonyms is differentially affected by the interest level of the material. If children are more motivated on high-interest material, then they might produce more synonyms as well as more exact responses. Alternatively, children might be less familiar with the vocabulary on low-interest material, resulting in the production of more imprecise but near-correct responses of this type of material. Findings regarding the production of synonyms, then, could be suggestive

of the type of processes that are operating when children read high- and low-interest material.

Method

Subjects

The study was conducted eighteen months after Asher and Markell's in the same school and grade level. Seventy-five children participated. They constituted the entire fifth-grade population of the school except for four children who were repeatedly absent. Achievement test data from the school-administered Scholastic Testing Service Educational Development Series reading achievement test were available for 71 of the 75 children. Accordingly, four children were eliminated for the sample. Another child assigned to the low-interest condition was eliminated because even her five lowest rated topics averaged above the midpoint of the scale. Of the final sample of 70 children, 38 were girls and 32 were boys. The children's average IQ on a school-administered STS Educational Development Series ability test was 107.

Materials

Interest Slides

Twenty-five color slides were used to assess interests. Each slide represented a single theme or topic and the topics covered a wide range of interest areas. The topics are listed below in the randomly selected order in which they were presented to children.

1. Forest
2. Jet Airplane
3. Priest
4. Dogs
5. Astronauts
6. Bride
7. Calf
8. Basketball
9. Butterflies
10. Marionettes
11. Monkey
12. Flowers
13. Bullfighting
14. Skiing
15. Food
16. Living Room
17. Maps
18. Painting
19. Circus
20. Race Cars
21. Canoe
22. Model Trains
23. Mother and Child
24. Insects
25. Cats

Reading Materials

Twenty-five passages from the Britannica Junior Encyclopedia (1970) were used. This source was originally selected by Asher and Markell (1974) because it is written for elementary school children in the fourth grade or above (Walsh, 1973) and provides a wide range of topics in a more consistent style than would be obtained from diverse sources. The passages corresponded in topic to the 25 photographs. Each passage was transformed into a ten item cloze passage by deleting the tenth word and every fifth word thereafter. An entire sentence followed the last deletion. Each deletion was replaced with a 15-space line on which children could print their replacements.

Procedure

The interest assessment and the reading comprehension task were administered in two separate sessions one week apart. The children were tested in their classrooms during their reading period. Different experimenters administered the two sessions to minimize the possibility that children would perceive the connection between the interest assessment and the reading activity.

Interest Assessment

Experimenter I told the children, "I'd like to find out about what kids are interested in. I'm going to show you 25 slides. For each slide I'd like you to mark, on the sheets we'll give you, how interesting the picture is to you. Who knows what 'interesting' means?" After a few children had responded, Experimenter I summarized their comments by saying, "So, something is interesting when you like it and would like to find out more about it." Experimenter I then distributed to each child a form with twenty-five 1-7 rating scales, and drew a 1-7 scale on the blackboard. At the low end of each scale were the words "not at all interesting" and, at the high end, "very interesting." The nature and use of the rating scale were explained:

"If a picture is very interesting to you--if you like it very much and want to know more about it--mark a number at this end of the scale. (The experimenter pointed to Numbers 5, 6, and 7 of the scale on the blackboard.) You can mark it with a circle, an X, a check, or whatever you want. If a picture is not at all interesting to you--if you don't like it and wouldn't care to find out more about it--mark a number at the low end of the scale. (The experimenter pointed to the Numbers 1,2, and 3 of the scale.) If the picture is of medium interest to you--if you like it but don't like it a lot-mark a number here.(The experimenter pointed to Numbers 3,4, and 5.) Let's try an example for practice. If I showed a picture of a pile of dollar bills, what number would you choose? (The experimenter called on several students.) If I showed a picture of a piece

of dirt, what number would you choose? (The experimenter again called on several students.) So you can see that different people are interested in different things. If anyone has any questions, raise your hand and I'll try to answer them. (Experimenter 1 then presented the slides announcing the number of each one as it was projected.) Here's Picture Number 1 .. Here's Picture Number 2 ..., etc."

The slides were presented at the rate of approximately one every 10 seconds. When all pictures had been rated, the children were asked to write their names on their rating sheet.

Reading Comprehension Task

One week after the interest assessment, Experimenter 2 gave the children five cloze passages to read. Children were randomly assigned to either the high- or low-interest conditions. Those in the high-interest condition received cloze passages that corresponded to their five highest-rated pictures. Those in the low-interest condition received cloze passages that corresponded to their five lowest-rated pictures. When topics shared the fifth highest or fifth lowest ratings they were randomly selected from those topics sharing equal ratings. Five passages were used because the sixth lowest-rated topics, on the average, tend to be moderately rated. Using six passages per child would therefore have weakened the high-versus low-interest manipulation.

Each of the five cloze passages, appropriately titled in upper-case letters, was mimeographed on 8 1/2 X 5 1/2-inch paper and enclosed in a legal-size envelope. The envelopes were numbered from one to five to

specify the order in which the passages were to be read. The particular order of the five passages was randomly assigned for each child. In addition to these five envelopes each child received an additional envelope which contained a reading enjoyment scale. The purpose of this scale was to learn whether children in the high-interest condition enjoyed their material more than children in the low-interest condition. The item was: "I enjoyed the paragraphs I just finished reading." Below the item was a 1-7 scale with "disagree" under the "1" and "agree" under the "7".

Before children were given the envelopes, Experimenter 2 gave the following instructions:

"I am going to show you a reading game. (Experimenter 2 gave each child a sample paragraph.) This is a paragraph with some words missing. The idea is to read the paragraph and decide what words are missing. Each paragraph has 10 missing spaces. Take a minute to look at the paragraph. (The experimenter paused.) OK. Now I'll read the paragraph with all of the words in it. You follow along with me. (The experimenter read the sample paragraph aloud, collected the sample paragraph from each child and then gave each child the test envelopes.)".

" You now have six envelopes. Five have paragraphs in them. Start with the first paragraph and try to fill in all the missing words. When you are done with a paragraph, put it back in the envelope and put it aside on your desk. Then you can go on to the second envelope; then the third, fourth, and fifth. Once you put a paragraph back in the envelope you can't go back. Do you have any questions? OK. Read each paragraph carefully and try to fill in the missing words. I can't help you read any of the

words, but if you have trouble spelling any words raise your hand and I will help. Spelling doesn't count in this game. If you are having trouble, don't get stuck. Go on to the next blank or a new paragraph. You have 40 minutes for the five paragraphs. That should be plenty of time. Any questions?"

"When you are done with the five paragraphs, open the next envelope. It contains a question about how much you enjoyed reading the paragraphs. If you enjoyed it very much circle one of the higher numbers. If you didn't enjoy it, circle one of the low numbers. You can circle one of the numbers in the middle if that is how you feel. Got the idea? Any questions? OK. You can begin."

When each child was finished, Experimenter 2 collected the material and unobtrusively recorded the time. The average time for completing the task was 18 minutes.

Cloze scoring method

Each child received two scores, one based on the number of exact replacements supplied and one based on the number of exact replacements plus synonyms. Supplied words were counted as correct despite spelling errors if the supplied word was clearly recognizable. Since there were five passages with ten deletions each, a child's score could range from 0-50.

A rule was needed for deciding whether a response was a synonym. A Thesaurus offers an objective basis for deciding about synonyms; however,

many passage words were not listed there. The scoring procedure used here was as follows: One person, blind to the experimental hypotheses, scored the passages for exact replacements only, (spelling errors were allowed), and then listed on separate sheets of paper the incorrect responses that children provided for each item. These responses were listed directly under a heading which was the correct response. These lists were then given to three judges (1 male and 2 female college students) who were also unaware of the purpose of the experiment. Judges were instructed to put a check mark next to each of the responses that were synonymous with the correct response.

The instructions given to the judges for identifying words as synonyms were as follows:

"The following definition of synonyms will be used: A is a synonym of B if A and B have the same meaning. The task here requires your judgments on what we call 'synonyms in context'. A and B are synonyms in the context of a passage if they maintain the meaning of the sentence an of the passage intended by the author. Therefore in judging words as to whether or not they are synonyms in the context of the passage you should check that part of the passage immediately before and immediately after the blank. Read the sentence in which the blank is found and the sentence following it so you get the idea the author is trying to get across."

"In judging words as synonyms, remember that spelling errors don't count unless they lead to a change in tense or number (e.g., choose and boy should not be thought of as incorrect spelling of chose and boys). The synonym in context must be of the same tense and number as the original word; thus, was and is are not synonyms in context nor are is and are. The words must also be of the same grammatical case; thus we and us are not synonyms in context. The words must be grammatically correct with respect to the sentence, thus a and an would not be synonyms."

About an hour of training was provided on completely unrelated passages to ensure that judges understood the task and the definition of a synonym. Interjudge agreement (number of agreements divided by the total number of items) averaged 93% between each pair of judges (Judge A and B = 93%; Judge B and C = 95%; Judge A and C = 92%). However, synonyms occurred infrequently (children averaged only 2.8 synonyms out of 50 deletions). Most of the non-exact responses that children produced were clearly not synonyms. Because of the high number of non-synonym judgments, a strong agreement among judges would be expected by chance. Thus the 93% figure may not be an accurate indication of reliability of synonym judgments. A more appropriate measure of agreement for these data is kappa, K (Cohen, 1960; Light, 1971), which calculates the proportion of joint judgments in which there is agreement after chance agreement is excluded. The k value obtained for this data was .615, $Z = 2.730$, $p<.01$. Table 1 presents observed and expected agreements and disagreements of the synonym judgments. As can be seen in

the table, all agreements among judges concerning both synonyms (S) and non-synonyms ($\bar{S}$) are higher than would be expected by chance. All disagreements are less than would be expected by chance. Thus, synonym judgments were found to be reliable. Each child received credit

---

Insert Table 1 here

---

for a synonym if his or her non-exact replacement was judged to be a synonym by at least two of the three judges.

Results

Standardized Reading Achievement

The STS Educational Development Series reading achievement test had been administered by the school prior to the study. Data from the test were analyzed to learn whether males and females differed in their standardized test performance and whether children randomly assigned to high- and low-interest conditions were of similar reading ability. The reading comprehension score for boys was 29.3 and for girls was 28.5. The difference is not statistically significant, t (68) = .32. Further inspection of achievement data indicated that despite random assignment to condition, boys in the low-interest condition ($\bar{X}$ = 31.7) tended to be higher achievers than boys in the high-interest condition ($\bar{X}$ = 26.6), t (30) = 1.55, $p < .15$. Accordingly, analyses performed here on the effect of interest on children's performance used standardized reading scores as a covariate. In this way any potentially confounding effects of reading ability on performance in high- and low- interest conditions were statistically removed.

Preference Ratings

The post-reading preference ratings made by boys and girls in the high- and low-interest conditions were first compared. A 2 x 2 (Sex x Interest) analysis of covariance was performed with the preference rating as the dependent variable and standardized reading achievement scores as the covariate. The adjusted ratings are presented in the top spanner of Table 2. Results indicated that children who received the high-interest material expressed significantly more enjoyment than children who received low-interest material, F (1,65) = 4.18, $p < .05$. Boys and girls rated the reading material similarly, F (1,65) = .15, and the interaction between sex and interest was not significant, F (1,65)= 1.43. Both boys and girls, then, preferred the passages that corresponded to their high-interest areas. These results validate the use of the picture rating technique since the picture ratings predicted the reading preferences for both sexes.

Insert Table 2 about here

Reading Comprehension

Next, the effects of sex and interest on children's reading comprehension were examined. To facilitate comparison with Asher and Markell's findings an analysis was first performed on cloze performance based on exact replacements only. A 2 x 2 (Sex x Interest) analysis of covariance was performed with standardized reading achievement scores as the covariate and cloze scores as the dependent variable. Results of this analysis

(middle spanner, Table 2) indicated that the effect of interest was significant, F (1,65) = 4.41, $p < .05$, with children comprehending more of high- than low-interest material. A second result was that the effect of sex was not significant, F (1,65) = .38. This finding parallels the lack of a significant sex difference on the standardized reading achievement test. Finally, the interaction of sex and interest was not significant, F (1,65) = .02. The lack of interaction resulted from the fact that both boys and girls did better on the high-interest than low-interest material. This contrasts with the findings of Asher and Markell (1974) who found that boys' performance was strongly influenced by the interest level of material but girls' performance was not.

Another x interest covariance analysis was performed with reading achievement scores as the covariate and the exact plus synonym scores as the dependent measure. The adjusted exact plus synonym cloze scores are presented in the bottom spanner of Table 2. This analysis produced findings similar to those using exact replacement scores. The effect of interest was marginally significant, F (1,65) = 3.65, $p < .06$, the effect of sex was not significant, F (1,65) = 1.31, and the interaction of sex x interest was not significant, F (1,65) = .02. The effect of scoring synonyms as correct was to increase the average correct cloze score for the total sample from 14.0 to 16.8 and to increase the standard deviation from 7.2 to 8.5. The fact that the interest effect is slightly weaker here is probably due to the increased variability associated with including synonyms as correct.

The slightly-elevated standard deviation suggests that the effect of scoring synonyms was to somewhat widen the gap between good and poor readers. To examine directly whether higher achievers produced more synonyms than low achievers the sample was divided at the median achievement test score and a 2 x 2 x 2 (Sex x Interest x Achievement Level) analysis of variance was performed with number of synonyms as the dependent measure. These data are presented in Table 3. As expected, the effect of achievement level, was significant, F (1,62) = 23.21, $p < .001$, with higher achieving children producing more synonyms than lower-achieving children. The effect of sex was also significant, F (1,62) = 6.06., $p < .05$ with boys producing more synonyms than girls. The effect of interest was not significant, F (1,62) = .27; children produced a similar number of synonyms in the high- and low-interest conditions. None of the interactions between the main effects were significant or approached significance.

Insert Table 3 about here

Finally, analyses were performed to compare the relationship of standardized reading achievement test scores to cloze scores when only exact replacements were accepted versus exact replacements plus synonyms. For the total sample the correlation between standardized achievement test scores and cloze scores based on exact replacements only was r (68)= .49, $p < .05$. The correlation when cloze scores included synonyms was r (68) = .53, $p < .05$. This small change in the correlation follows from the fact that the correlation between the two types of cloze scores is

very high, r (68) = .98, $p < .001$. Synonyms were low frequency responses. Even though certain children produced more synonyms than others the occurrence of synonyms was too infrequent to alter greatly the distribution of cloze scores.

Discussion

Results of this experiment partially replicate and partially qualify earlier findings. Using the exact replacement scoring system traditionally employed, children were found to read better on high- than low-interest material. This result indicates that the effect of interest is not dependent on contrast effects that might be part of a within-subjects design. Children performed better on high-interest passages even when they were unaware of the range of topics available in the experiment. In this sense, the present experiment, with a between-subjects design, replicated the original findings of Asher and Markell (1974).

The results qualify the original findings insofar as sex and interest did not interact; both boys and girls did better on high- than low-interest material. Asher and Markell found that boys' performance was strongly affected by the content of reading material while girls' performance was only slightly affected. One difference between the two experiments is the type of research design employed. However, the potential role of this factor is probably minimal given some other recent data. In a study examining interest effects on black and white children's reading comprehension, boys and girls were tested with a within-subjects design (Asher, 1976).

Boys and girls of both races performed better on high- than low-interest material. Apparently, then, the effect of interest on comprehension for girls is not a function of the type of research design employed.

Another explanation has to do with possible secular or cultural changes. Girls in the past may have been more willing to work hard on a task even though it was uninteresting. Perhaps changing definitions of sex-roles are leading girls to be less oriented toward meeting external standards and more concerned with internal criteria such as their interest in the task. The rather short time interval between the original study and the present experiment makes this interpretation highly speculative but nonetheless possibly correct.

Perhaps the most plausible interpretation is that the type of interest effects obtained is a function of the developmental level of the children tested. The boys and girls in Asher and Markell's study were significantly different on the school-administered reading achievement test, with girls achieving higher scores. In contrast, boys and girls in the present study and the other recent study (Asher, 1976) did not significantly differ on the same test. All three studies were conducted with fifth grade children. From previous literature it appears that sex differences in reading comprehension are in a transitional phase about this age level (Gates, 1961; Hughes, 1953; Stroud and Lindquest, 1942) with sex differences in comprehension rather consistently obtained with younger children, non-existent at later ages, and inconsistently obtained in fifth grade.

One intriguing possibility is that the gap in achievement test performance narrows in the later years not simply because boys "catch up" but because girls become motivated to excel only if the task is appealing to them. In this case, girls, like boys, would begin to show effects of interest on their reading comprehension. They, too, would be particularly motivated when the material is appealing to read. If this interpretation is correct then studies at grade levels where sex differences typically occur (e.g. fourth grade) should find strong interest effects for boys and weaker effects for girls. Studies at grade levels where boys and girls typically perform alike (e.g. sixth grade) should show similar effects of interest for both sexes. Thus further research is needed to establish the conditions under which the interest effect is obtained for both sexes versus boys only.

Another purpose of the present experiment was to assess the relative correlations of cloze performance with standardized test performance when synonyms are or are not accepted as correct responses. The data support the findings of Ruddell (1964) that including synonyms only slightly increases the correlation of cloze scores with standardized achievement test scores. However, inspection of the performance of different groups of children suggests that scoring synonyms as correct does have differential effects across children. Children who are higher achievers produced more synonyms as well as more exact responses. In addition, boys were found to produce a significantly greater number of synonyms than girls.

These findings suggest that, depending on the purpose of an experiment, it may be useful to score synonyms as correct. The general bias in the literature against accepting synonyms may be leading to the loss of valuable information when the individual rather than the passage is the unit of analysis.

A final purpose of the present study was to examine whether children produce more or less synonyms on high-interest material. The finding was that a similar number of synonyms was produced on high- and low-interest material. Thus it seems unlikely that the inclusion of synonyms will substantially alter the pattern of results obtained in studies of interest effects. Still, the issue might be re-examined in future research. The encyclopedia material used in the present research, although written for children in the fourth grade or above (Walsh, 1973), is quite challenging (Asher and Markell, 1974). Material of a less difficult nature might yield a different pattern of synonym production.

Part of the bias against accepting synonyms as correct replacements results from the decreased scoring objectivity. In the present study, the interjudge reliability using coefficient Kappa was quite satisfactory. The scoring of synonyms as correct did not lead to a serious loss of objectivity. As a further caution, a response was scored as a synonym if two out of three judges independently so decided. Given the low probability that a judge will classify a response as a synonym, this is basically a conservative procedure that results in only a small increase in each child's score. Still, the scoring of synonyms produced some interesting

findings across groups of children. Further research will indicate whether the gains in new information outweigh the costs of using a somewhat more complicated scoring procedure.

Further research is also needed to learn why children read better on high- than low-interest material. One possibility is that children are more motivated on high-interest passages and attend more, work harder, etc. Another possibility is that children comprehend more of high-interest material because they are more knowledgeable about the content. One approach to evaluating these explanations would be to provide a strong external incentive for trying hard on both types of reading material. This could indicate whether children are able to comprehend as much of low- as high-interest material when they are motivated to do well on both. Whichever explanation of the interest effect is ultimately supported, it appears that researchers or teachers seeking to assess children's competence in reading comprehension have reason to consider carefully their selection of passage topics. Assignment of passages based on an individualized assessment of children's interests appears to facilitate children's reading comprehension.

Test Reference

Educational Development Series. Elementary level Form A. Scholastic Testing Service. Bensenville, Illinois: Scholastic Testing Service, 1971.

References

Asher, S. R. The effect of interest on reading comprehension for black children and white children. Unpublished manuscript, University of Illinois, 1976.

Asher, S. R. and Markell, R. A. Sex differences in comprehension of high- and low-interest material. Journal of Educational Psychology, 1974, 66, 680-687.

Bernstein, M. R. The relationship between interest and reading comprehension. Journal of Educational Research, 1955, 49, 283-288.

Bormuth, J. R. Validities of grammatical and semantic classifications of cloze test scores. Proceedings of the International Reading Association, 1965, 10, 283-286.

Bormuth, J. R. Comparable cloze and multiple-choice comprehension test scores. Journal of Reading, 1967, 10, 291-299.

Bormuth, J. R. Empirical determination of the instructional reading level. Proceedings of the International Reading Association, 1968, 13, 716-721.

Britannica Junior Encyclopedia. Chicago: Encyclopedia Britannica, 1970.

Cohen, J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 1960, 20, 37-46.

Dorsel, T. N. Preference--success assumption in education. Journal of Educational Psychology, 1975, 67, 514-520.

Gates, A. I. Sex differences in reading ability. Elementary School Journal, 1961, 61, 431-434.

Hughes, M. Sex differences in reading achievement in the elementary grades. In H. M. Robinson (Ed.), Clinical studies in reading II. Chicago: University of Chicago Press, 1953.

Jongsma, E. A. The cloze procedure: A survey of the research. Occasional Papers in Reading, Indiana University, School of Education, 1974.

Klein, H. A. Interest and comprehension in sex-typed materials. Paper presented at the International Reading Association Conference, Kansas City, May 1969. (ERIC Document Reproduction Service No. ED 030 551)

Light, R. J. Measures of response agreement for qualitative data: Some generalizations and alternatives. Psychological Bulletin, 1971, 76, 365-377.

Rankin, E. F., and Culhane, J. W. Comparable cloze and multiple-choice test scores. Journal of Reading, 1969, 13, 193-198.

Ruddell, R. A study of the cloze comprehension technique in relation to structurally controlled reading material. Proceedings of the International Reading Association, 9, 1964, 298-302.

Schoelles, I. S. Cloze as a predictor of reading group placement. Paper presented at the meeting of the International Reading Association, Atlantic City, April 1971. (ERIC Document Reproduction Service No. ED 053 868)

Shnayer, S. W. Some relationships between reading interests and reading comprehension. Unpublished doctoral dissertation, University of California, Berkeley, 1967.

Stanchfield, J. M. The effect of high-interest materials on reading achievement in the first grade. National Reading Conference Yearbook, 1967, 16, 58-61.

Stroud, J. B., and Lindquist, E. F. Sex differences in achievement in the elementary and secondary schools. Journal of Educational Psychology, 1942, 33, 657-667.

Taylor, W. L. "Cloze procedure:" A new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

Walsh, S. P. General Encyclopedias in Print 1973-1974: A Comparative Analysis. New York: R. R. Bowker, 1973.

Table 1

Agreements and Disagreements of

Synonym (S) and Non-Synonym ($\bar{S}$) Judgments

Observed

		Judge A	
		$\bar{S}$	S
Judge B	$\bar{S}$	1106	40
	S	19	71

		Judge A	
		$\bar{S}$	S
Judge C	$\bar{S}$	1052	28
	S	73	83

		Judge B	
		$\bar{S}$	S
Judge C	$\bar{S}$	1068	18
	S	74	76

Expected

		Judge A	
		$\bar{S}$	S
Judge B	$\bar{S}$	1043	103
	S	82	8

		Judge A	
		$\bar{S}$	S
Judge C	$\bar{S}$	983	97
	S	142	14

		Judge B	
		$\bar{S}$	S
Judge C	$\bar{S}$	1001	79
	S	145	11

Table 2

Mean Adjusted Reading Preference Ratings, Exact Cloze Scores, and Exact Plus Synonyn Cloze Scores

	Sex	
Interest Level	Boys	Girls
	Reading Preference Ratings	
High	4.55	4.17
Low	3.13	3.77
	Exact Cloze Scores	
High	15.99	15.30
Low	13.10	11.96
	Exact Plus Synonym Cloze Scores	
High	19.74	17.53
Low	16.18	14.41

Table 3

Mean Number of Synonyms Produced

		Sex	
		Boys	Girls
Interest Level			
High	High Achievers	5.33	2.88
	Low Achievers	2.33	1.70
Low	High Achievers	4.30	3.56
	Low Achievers	2.00	1.45

CENTER FOR THE STUDY OF READING

TECHNICAL REPORTS

No. 1: Halff, H. M. Graphical Evaluation of Hierarchical Clustering Schemes, October 1975.

No. 2: Spiro, R. J. Inferential Reconstruction in Memory for Connected Discourse, October 1975.

No. 3: Goetz, E. T. Sentences in Lists and in Connected Discourse, November 1975.

No. 4: Alessi, S. M., Anderson, T. H., & Biddle, W. B. Hardware and Software Considerations in Computer Based Course Management, November 1975.

No. 5: Schallert, D. L. Improving Memory for Prose: The Relationship Between Depth of Processing and Context, November 1975.

No. 6: Anderson, R. C., Goetz, E. T., Pichert, J. W., & Halff, H. M. Two Faces of the Conceptual Peg Hypothesis, January 1976.

No. 7: Ortony, A. Names, Descriptions, and Pragmatics, February 1976.

No. 8: Mason, J. M. Questioning the Notion of Independent Processing Stages in Reading, February 1976.

No. 9: Siegel, M. A. Teacher Behaviors and Curriculum Packages: Implications for Research and Teacher Education, April 1976.

No. 10: Anderson, R. C., Pichert, J. W., Goetz, E. T., Schallert, D. L., Stevens, K. V., & Trollip, S. R. Instantiation of General Terms, March 1976.

No. 11: Armbruster, B. B. Learning Principles from Prose: A Cognitive Approach Based on Schema Theory, July 1976.

No. 12: Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. Frameworks for Comprehending Discourse, July 1976.

No. 13: Rubin, A. D., Bruce, B. C., & Brown, J. S. A Process-Oriented Language for Describing Aspects of Reading Comprehension, November 1976.

No. 14: Pichert, J. W., & Anderson, R. C. Taking Different Perspectives on a Story, November 1976.

No. 15: Schwartz, R. M. Strategic Processes in Beginning Reading, November 1976.

No. 16: Jenkins, J. R., & Pany, D. Curriculum Biases in Reading Achievement Tests, November 1976.

No. 17: Asher, S. R., Hymel, S., & Wigfield, A. Children's Comprehension of High- and Low-Interest Material and a Comparison of Two Cloze Scoring Methods, November 1976.

No. 18: Brown, A. L., Smiley, S. S., Day, J. D., Townsend, M. A. R., & Lawton, S. C. Intrusion of a Thematic Idea in Children's Comprehension and Retention of Stories, December 1976.

No. 19: Kleiman, G. M. The Prelinguistic Cognitive Basis of Children's Communicative Intentions, February 1977.

No. 20: Kleiman, G. M. The Effect of Previous Context on Reading Individual Words, February 1977.

No. 21: Kane, J. H., & Anderson, R. C. Depth of Processing and Interference Effects in the Learning and Remembering of Sentences, February 1977.