25

DO CHILDREN READ WORDS BETTER IN CONTEXT OR IN LISTS?

A classic study revisited

Tom Nicholson


The aim of this study was to reevaluate the research carried out by Goodman (1965), in which it was found that children made 60%-80% fewer errors when reading words in context, as compared with reading words in an isolated list. This study has been cited at least 85 times in the literature (see Staff, 1986). It has also been reprinted in *Theoretical Models and Processes of Reading* (Singer & Ruddell, 1985), a standard reference in the field of reading. Such indicators suggest classic status.

The reason for the popularity and durability of the study may be that it has given support to methods of teaching reading in which use of context is strongly encouraged, as in the “whole language” approach to teaching reading. For example, in New Zealand the use of context clues is regarded as a major factor in the reading process. Thus, when children encounter a difficult word, they are encouraged to guess what the word might be, to look at the first letter and guess, or to read through the end of a sentence and find other context clues to help them guess the word.

In short, children are expected to use context clues as a major strategy in identifying words and to give only secondary attention to letter-sound analysis. In fact, contrasting strategies, such as phonics, are downplayed: “It is very seldom that every letter in a word has to be identified and its related sound blended, because the reader bears in mind and uses the available syntactic cues” (Department of Education, 1985b, p. 48).

This emphasis on context, however, has not solved prevalent reading problems in that many New Zealand children (at least 15%) do not seem to make much progress and require individual reading tutelage in their second year of schooling (Nicholson, 1987, 1989).
In short, there are practical reasons for revisiting Goodman’s (1965) classic study. The dramatic findings of the study suggested that context clues were an important part of the reading process. However, there is quite a bit of evidence to suggest that the study may have overestimated the effect of context cues in reading. If this is the case, then perhaps educators need to reassess the evidence in favor of context as a strategy for reading words.

In the classic study, 100 children from Grades 1 to 3 in an inner city school were given increasingly difficult lists of words to read until a level of difficulty was reached at which the lists were neither too easy nor too hard. The students were then given text material to read, which was taken from a graded reading series that included the same words the students read in the lists. Children’s reading errors, or miscues, were noted in both list and context versions and relative improvement in context was calculated.

The results showed a dramatic reduction in the number of errors made in the condition in which children read the words in context. Specifically, first-grade children averaged 9.5 errors in lists but only 3.4 errors in context, a gain of more than 60%. Second graders averaged 20.1 errors in lists but only 5.1 errors in context, a gain of 75%. Third graders averaged 18.8 errors in lists but only 3.4 errors in context, a gain of slightly more than 80%.

However, the findings of the study may have given a misleading impression for two reasons. First, there was no comparison of individual differences between good and poor readers. In other words, it was not clear whether the results applied mainly to poor readers, to good readers, or to both. Second, there was no allowance for the effect of order; thus, it could not be determined whether the results were due to context or to the effect of having had a second opportunity to read the words. The latter effect often occurs in everyday reading, in which children make initial errors but correct them on a second attempt.

The question of an order effect was first raised by Pearson (1978). He reported that when children were given the same list twice, there was a 20% improvement without any context help. Pearson also reported that efforts to replicate Goodman’s (1965) classic study had only produced gains of 40% among first graders and 50%–60% among third graders. Thus, taking into account the effect of reading the same list twice, Pearson suggested that the actual extent of improvement was only in the region of 20%–30% (rather than 60%–80%). In reply to this evidence, Goodman (1980) argued that such results, although more moderate, were nevertheless consistent with his 1965 study.

The question of individual differences was raised by Allington (1978) who found, in a study of fourth graders, that when he gave the same task as in Goodman’s (1965) classic study but in counterbalanced order context
benefited poor readers but made no difference for good readers. Since then, Stanovich (1980, 1986) and Nicholson (1986) have summarized many other studies that have found similar individual differences in which poor readers gained more from context than did good readers. In regard to the effect of context itself, Rayner and Pollatsek (1989) summarized a number of experimental studies that indicate that context is not a significant factor in normal word identification, or, as phrased by Rayner and Pollatsek, “the effects of context on word identification (whether due to lexical or postlexical processes) are usually pretty modest” (p. 236).

This is not to say that context does not help word identification. For example, no one would deny that context helps the child to guess that saucer is the missing word in the sentence. “The cat drank milk from the ——”. Yet Gough (1984) stated that such contextual richness is not typical of normal reading:

It seems, then, that while the effects of context can be dramatic, those effects are diminished with the skilled reader reading clear text. It is even possible that they are normally absent, for the contexts studied to date are hardly representative. In every single study, the targets have been nouns, in sentence final position, and highly predictable.

(p. 245)

However, Goodman (1987), in reference to his 1965 study and in response to some of the issues raised thus far, stated that he continues to perceive value in the findings of that study, arguing that studies with different findings may be ecologically invalid. Goodman (1987) suggested that these studies may have involved “contrived” contexts such as paragraphs, sentences, and cloze procedures rather than actual texts that are likely to be read by children in a familiar environment, such as a school setting:

Now here’s a puzzle. How is it possible, that Phillip Gough, P. David Pearson and Tom Nicholson, among a host of other respected researchers could have found in their research that children are more successful reading isolated words? The answer, in every case, is that they did not use real texts in their studies. These are researchers who believe that they must control the variables in their studies. So they contrived contexts.

(Goodman, 1987, p. 6)

For the most part, this criticism is correct. Some of the research reported above did involve contrived material, although the research was conducted in this way to enable a more precise analysis of what happens
when children read words in context and in lists. (However, Gough, 1984, as reported above, criticized the same research because much of it was contrived in favor of context.)

An illustration of a contrived context is a study in which good and poor readers read the same text materials. Although this procedure controls the variable of text difficulty, it lacks ecological validity because good and poor readers usually read materials that differ considerably in difficulty. Another example of a contrived context is a situation in which a researcher reads some difficult material aloud to help provide a similar amount of contextual help for the poor reader. This procedure also differs from the normal situation. In short, contrived contexts have ecological problems.

To avoid the problem of contrived contexts, it is necessary to relax some of the controls that are typical of the experimental situation. Only by doing so is it possible to deal with the aforementioned issues related to order effects and individual differences.

Such attempts were made in a replication of Goodman’s (1965) classic study by Nicholson, Lillas, and Rzoska (1988) with a sample of 6- and 8-year-old children that included both good and poor readers. The procedure was similar to that used in Goodman’s study. The children read short passages, in context and list form, from an informal prose test. The context passage was the actual prose material in the test; the list passage was the same material typed backward, from the last word to the first.

The only difference in Nicholson et al.’s (1988) procedure from that of Goodman’s (1965) study was that children were given the context and list forms in counterbalanced order. Nicholson et al. found that the younger good and poor readers and the older poor readers did better in context than with lists, whereas the older good readers did slightly better on the list than with words in context. Thus, even with an ecologically valid replication study, the older good readers showed a different pattern of results from that of the other readers.

In a similar replication effort, Nicholson, Bailey, and McArthur (1991) obtained results indicating that the findings of Goodman’s (1965) classic study were related to the order in which children were given the reading tasks. The results showed that poor readers did better in context, regardless of whether they read the list first or the context passage first. On the other hand, Nicholson et al. found that good readers, even the younger ones, did better in context only when they read the list first.

However, a limitation of the above two studies was that only 32 children were used in each sample. Another drawback was that errors were counted for just one passage per child. In short, the corpus of errors was small.

With these limitations in mind, the purpose of the present study was to carry out a more substantial replication, with more children and with a
larger corpus of errors, based on children’s reading of several passages at different levels of accuracy. In this way, the study approximates as closely as possible Goodman’s (1965) classic study in which 100 children were studied and in which children read extended text material from graded readers.

In Experiment 1 the purpose was to evaluate the effects of reversing the original order of testing, as used in Goodman’s (1965) classic study. In other words, children were given words in a context passage first, then in a list form. Experiment 2 was a straight replication of the classic testing procedure, in which children read the list first, then the context passage. The purpose of the second experiment was to provide a check on the first experiment, to see whether the order of testing had an effect on the number of errors children made in context, as compared with the number of errors they made in lists.

In summary, the basic aim of the following two experiments was to replicate and extend the two smaller studies discussed above (Nicholson et al., 1988; Nicholson et al., 1991) by using new subjects and by collecting new data.

**Experiment 1**

**Method**

*Subjects.* One hundred children, who were selected from two suburban schools in a large regional city of New Zealand, participated in the study. The schools drew on a wide socioeconomic base. There were 32 6-year-olds (mean age = 6 years, 6 months), 34 seven-year-olds (mean age = 7 years, 6 months), and 34 eight-year-olds (mean age = 8 years, 5 months). There were 53 girls and 47 boys in all. Within each age group, there were similar numbers of good, average, and poor readers, as determined on the basis of reading levels provided by the schools.

The school’s senior staff selected the children for poor, average, and good reading groups using test data that were confidential to them. The children’s reading levels were then checked by comparing the children’s relative performance on the prose-reading test used in the actual study. The prose test combined sets of informally graded passages from three different prose tests (Department of Education, 1981, 1983, 1985a).

The combining of the passages was necessary to provide a wide range of graded material that would be suitable for children with a range of reading levels, from 6-year-old poor readers to 8-year-old good readers. However, I recognized that these passages had not been formally normed. Thus, the age levels for each passage were regarded as only roughly indicative of the children’s actual reading levels. This lack of standardization did not interfere with the main use of the passages, however, the purpose of which was
to establish whether there were differences among the poor, average, and good readers selected by the schools.

The checking procedure involved asking each child to read progressively more difficult passages. The cutoff point for each child was the most difficult passage at which he or she read above the 90% accuracy level. For the 6 year olds, the mean cutoff scores (reading levels) were the following: 12.3 years for good readers, 8.0 years for average readers, and 5.7 years for poor readers. For the 7 year olds, the mean cut off scores were as follows: 13.5 years for good readers, 8.1 years for average readers, and 5.5 years for poor readers. For the 8 year olds, the mean cut off scores were the following: 15.4 years for good readers, 12.3 years for average readers, and 6.3 years for poor readers.

It should be noted, however, that the aforementioned checking procedure was supplementary to the initial grouping, which was determined by teacher judgment on the basis of test data confidential to the school. That is, the groups used in the study were the same as those selected by the school staff. The checking procedure was simply a way of ascertaining, with a broadly graded informal prose test, what the school regarded as good, average, and poor. Although the scores of some of the groups seem rather high, this may not be the case, given that the prose test had not been graded on the basis of standardized testing.

**Design and procedure.** The design of the first experiment involved asking children to first read passages in context and then read the same material in list form. This testing order is opposite that used in Goodman’s (1965) classic study. As previously mentioned, the aim of this design was to determine whether the results of the classic study would stand up if children were given the same tasks in the reverse order. If the results of the classic study were correct, then it would seem likely that children would read better in context, regardless of the order of testing used, given that it is widely believed that children read words more easily in context.

The lists were the passages typed backward, from the last word to the first, in vertical columns. Although this procedure was simple, children very rarely noticed that the list form was the story typed backward.

Children read the passages in context first. They were tested individually, beginning with the easier stories and continuing until their accuracy dropped below 90%. Whenever possible, in the context task, children were given the actual readers from which the test passages were taken.

Except for the most advanced passages, these materials were usually illustrated. This is similar to the normal school situation, in which illustrations are more frequent in easier materials, less so in harder ones. In addition, because many teachers regard illustrations as part of the normal reading environment, especially for beginning readers, no attempt was made to prevent children from looking at the pictures. This was also consistent with the procedure used in Goodman’s (1965) classic study, in
which children read stories from basal texts. Such texts are usually illustrated at the easier levels.

With such a wide range of reading ability in the sample, this meant that some children read different passages from those read by others, and some children read more passages than did others. In fact, the average number of words read by children of each ability level ranged from about 100 among the poorest readers to about 1,000 among the best.

A few days after the initial testing, the same children read the same stories in list form, except for those stories in which their reading accuracy had dropped below 90%.

Scoring involved marking all errors except self-corrections, including omissions in which words were skipped because of their difficulty or because they were not noticed and substitutions in which text words were replaced by other words. Addition errors were counted, but they were infrequent. In short, virtually all errors were counted. This procedure again seemed similar to that in the classic study.

In addition, in line with the classic study, children were asked to retell each passage in their own words after reading it. The purpose of the retelling procedure was to encourage children to read for comprehension. The retellings were informally scored on an A–E scale, mainly to give children the impression that comprehension was also being tested. The retellings were not a formal part of the study itself, but they ensured that children were attending to context.

I analyzed the data, specifically, all passages that children were able to read at the 90%–99% accuracy level, which is meant to include easy and instructional material but not difficult material (Department of Education, 1985b).

The procedure for comparing children’s errors in context and in list was similar to that used in Goodman’s (1965) classic study, in which differences in error scores were compared in terms of raw score differences and percent gains. For example, in the classic study, the second graders had a mean error score of approximately 20.0 for the lists and 5.0 for context. The difference in mean scores was 15.0, which was scored as a 75% gain in context: \[ \frac{(20.0 - 5.0)}{20.0} \times 100 \].

In addition, the number of words attempted by each child was calculated on the basis of the number of passages attempted in which he or she was able to read with between 90% and 99% accuracy. Children attempted the same words in list and in context, so the number of words attempted in each condition was the same. In short, the mean number of words was based on all passages in which error data were analyzed. As previously mentioned, the fact that children read different passages, of different levels of difficulty and of different lengths, meant that the mean number of words read differed considerably among the groups of children at each age level and at each reading level.
Because there was wide variation among groups in the difficulty levels of materials attempted, the number of words attempted, and the number of errors made, the results were analyzed separately for each reading level within each age group. This was done by calculating nine one-sample \( t \) tests for the raw score gains, as well as nine one-sample \( t \) tests for the percent gains. Although there was some danger of inflated Type I error with such a large number of \( t \) tests, this seemed to be the best procedure to follow. As a cautionary note with this procedure, the results should be interpreted in terms of their general pattern, in addition to evaluating the specific findings.

**Results and discussion**

As can be seen in Table 1, the poor readers at all age levels and the 6- and 7-year-old average readers generally showed significant gains with context (although there was no significant percent gain for the 8-year-old poor readers). However, the 6- and 7-year-old good readers and the 8-year-old average readers showed no reliable gains, and the 8-year-old good readers actually gained significantly with list.

In short, when given material at a comparable level of difficulty (that is, a level at which all children scored 90%–99%, which was similar to the cutoff of the classic study), only the poor readers at each age level and the younger average readers made gains in context. This finding is not the same as that reported in the classic study (Goodman, 1965, cited in Singer & Ruddell, 1985) in which “the children in this study found it harder to recognize simple words than to read them in stories” (p. 134).

If in Goodman’s (1965) classic study words in lists were harder to read than the same words in stories, then there may have been an order effect in which the apparent improvement of the good readers was due to the effect of reading the same words twice, once on the list and then a second time in context. To find out whether this explanation is correct, I carried out a second experiment. In this experiment, as in the classic study, children read words in list form before reading them in context. If there was an order effect, then the good readers in this experiment would read better in context, as occurred in the classic study.

**Experiment 2**

**Method**

*Subjects.* There were 97 children in the follow-up study, who were selected from a different suburban school in the same city as in Experiment 1. Students attending this school were drawn from a predominantly middle-class base. There were 33 six-year-olds (mean age = 6 years, 5 months), 34
Table 1: Experiment 1: mean error scores, standard deviations, raw gains, percent gains, and mean number of words attempted

<table>
<thead>
<tr>
<th>Age/reading level</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>df</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>11</td>
<td>196</td>
<td>81</td>
<td>11.09</td>
<td>4.78</td>
<td>22.82</td>
<td>12.62</td>
<td>11.73</td>
<td>8.17</td>
<td>4.75**</td>
<td>10</td>
<td>47</td>
<td>14</td>
</tr>
<tr>
<td>Average</td>
<td>10</td>
<td>390</td>
<td>121</td>
<td>19.40</td>
<td>8.20</td>
<td>32.50</td>
<td>16.10</td>
<td>13.10</td>
<td>10.77</td>
<td>3.84**</td>
<td>9</td>
<td>32</td>
<td>34</td>
</tr>
<tr>
<td>Good</td>
<td>11</td>
<td>748</td>
<td>169</td>
<td>31.55</td>
<td>13.90</td>
<td>28.00</td>
<td>16.94</td>
<td>-3.55</td>
<td>11.14</td>
<td>-1.06</td>
<td>10</td>
<td>-33</td>
<td>74</td>
</tr>
<tr>
<td>7 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>12</td>
<td>135</td>
<td>96</td>
<td>8.33</td>
<td>7.09</td>
<td>16.75</td>
<td>15.92</td>
<td>8.42</td>
<td>10.37</td>
<td>2.81*</td>
<td>11</td>
<td>47</td>
<td>25</td>
</tr>
<tr>
<td>Average</td>
<td>11</td>
<td>315</td>
<td>136</td>
<td>17.46</td>
<td>7.61</td>
<td>27.36</td>
<td>15.91</td>
<td>9.91</td>
<td>12.29</td>
<td>2.67*</td>
<td>10</td>
<td>26</td>
<td>30</td>
</tr>
<tr>
<td>Good</td>
<td>11</td>
<td>794</td>
<td>178</td>
<td>38.55</td>
<td>12.99</td>
<td>31.73</td>
<td>15.76</td>
<td>-6.82</td>
<td>11.27</td>
<td>-2.01</td>
<td>10</td>
<td>-63</td>
<td>110</td>
</tr>
<tr>
<td>8 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>11</td>
<td>203</td>
<td>150</td>
<td>12.64</td>
<td>10.97</td>
<td>20.82</td>
<td>20.71</td>
<td>8.18</td>
<td>11.40</td>
<td>2.38*</td>
<td>10</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>Average</td>
<td>11</td>
<td>737</td>
<td>146</td>
<td>32.09</td>
<td>13.12</td>
<td>31.36</td>
<td>18.16</td>
<td>-0.73</td>
<td>10.17</td>
<td>-0.24</td>
<td>10</td>
<td>-14</td>
<td>34</td>
</tr>
<tr>
<td>Good</td>
<td>12</td>
<td>954</td>
<td>493</td>
<td>34.67</td>
<td>18.34</td>
<td>25.25</td>
<td>10.90</td>
<td>-9.42</td>
<td>13.80</td>
<td>-2.36*</td>
<td>11</td>
<td>-40</td>
<td>56</td>
</tr>
</tbody>
</table>

Note

*p < .05 **p < .01.
seven-year-olds (mean age = 7 years, 6 months), and 30 eight-year-olds (mean age = 8 years, 4 months). There were 47 girls and 50 boys in all.

As in the first experiment, children were selected for poor, average, and good reading groups on the basis of senior staff judgment with test data confidential to the school. The reading levels of the children were then checked by comparing children’s results on the informal prose test that was used in Experiment 1.

The cutoff point for estimating reading levels was the most difficult story level at which the child was able to score at 90% accuracy or better. For the 6-year-olds, the mean cutoff scores (reading levels) were the following: 12.9 years for good readers, 6.7 years for average readers, and 5.3 years for poor readers. For the 7 year olds, the mean scores were as follows: 15.3 years for good readers, 9.7 years for average readers, and 6.0 years for poor readers. For the 8 year olds, the mean cutoff scores were the following: 14.9 years for good readers, 11.9 years for average readers, and 7.1 years for poor readers.

*Design and procedure.* The reading materials were the same as those used in the first experiment. The children were also tested in the same way as in Experiment 1, except that they read list before context. As previously mentioned, this design was directed toward determining whether the results obtained in Goodman’s (1965) classic study were due to an order effect. If there was an order effect, then the gains made in list by the good readers in Experiment 1, in which the participants were given context before list, would be expected to be reversed in this experiment, in which participants were given list before context.

The only difference in procedure was that the 6 year olds were given both tasks within 1 week, whereas the other two age groups had to wait longer to be retested. However, for the 7- and 8-year-old groups, the time gap between list and context was about the same.

As in the first experiment and as in Goodman’s (1965) classic study, data were analyzed as raw gain and percent gain scores. As in Experiment 1, mean numbers of words read was also calculated.

*Results and discussion*

Unfortunately, 6 of the original 11 6-year-old poor readers were unable to cope with even the easiest list. When faced with words such as *book*, *mother*, and *father*, these students floundered. When given the same words in context, these children improved dramatically, but still not enough to reach the 90% cutoff. Thus, their scores were excluded from the analysis, reducing the sample from 97 to 91.

The results are shown in Table 2. A simple summary of the findings is that most groups showed significant gains in context. This result was similar to that of the Goodman’s (1965) study, although the significant
Table 2: Experiment 2: mean error scores, standard deviations, raw gains, percent gains, and mean number of words attempted

<table>
<thead>
<tr>
<th>Age/reading level</th>
<th>No. of words</th>
<th>Context errors</th>
<th>List errors</th>
<th>Raw gain</th>
<th>t test</th>
<th>% gain</th>
<th>t test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>6 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>5</td>
<td>62</td>
<td>22</td>
<td>4.80</td>
<td>2.59</td>
<td>15.40</td>
<td>4.22</td>
</tr>
<tr>
<td>Average</td>
<td>8</td>
<td>273</td>
<td>105</td>
<td>8.00</td>
<td>5.76</td>
<td>28.13</td>
<td>11.86</td>
</tr>
<tr>
<td>Good</td>
<td>14</td>
<td>928</td>
<td>151</td>
<td>35.86</td>
<td>18.28</td>
<td>50.07</td>
<td>22.31</td>
</tr>
<tr>
<td>7 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>10</td>
<td>190</td>
<td>137</td>
<td>8.00</td>
<td>5.38</td>
<td>21.10</td>
<td>12.01</td>
</tr>
<tr>
<td>Average</td>
<td>11</td>
<td>617</td>
<td>218</td>
<td>26.64</td>
<td>14.50</td>
<td>41.00</td>
<td>18.71</td>
</tr>
<tr>
<td>Good</td>
<td>13</td>
<td>796</td>
<td>137</td>
<td>29.92</td>
<td>21.28</td>
<td>32.77</td>
<td>13.11</td>
</tr>
<tr>
<td>8 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td>9</td>
<td>298</td>
<td>115</td>
<td>10.00</td>
<td>3.74</td>
<td>19.22</td>
<td>5.97</td>
</tr>
<tr>
<td>Average</td>
<td>11</td>
<td>592</td>
<td>127</td>
<td>18.64</td>
<td>10.72</td>
<td>26.09</td>
<td>15.03</td>
</tr>
<tr>
<td>Good</td>
<td>10</td>
<td>714</td>
<td>194</td>
<td>17.90</td>
<td>12.26</td>
<td>27.40</td>
<td>18.67</td>
</tr>
</tbody>
</table>

Note

*p < .05, **p < .01.
gains were made only by the poor and average readers and by the 6-year-old good readers. The 7- and 8-year-old good readers did not show significant gains.

In summary, the results of this second experiment showed a pattern similar to that of Goodman's (1965) classic study, that is, children generally read words better in context than in isolated lists. However, the results also showed that, when children were given materials of comparable difficulty, that is, materials on which all children scored in the 90%-99% accuracy range, the benefits of context went to the poor and average readers and to the 6-year-old good readers rather than to the 7- and 8-year-old good readers.

**General discussion**

It appears that the findings of Goodman's (1965) classic study may have exaggerated the effects of context. If context actually helped children to read better, then they would have read better in context, regardless of whether they read the words in list form first or in context form first.

Instead, the present study showed that the results of Goodman's (1965) classic study held up consistently only for poor readers at each age level and for the 6- and 7-year-old average readers, who read better with context, regardless of whether they read the words in context first or in list first. In contrast, the 6-year-old good readers and the 8-year-old average readers improved with context only when they read the words in list form first. Finally, the 7- and 8-year-old good readers did not improve with context, whether they read the list first or the context passage first. In fact, when given the context passage first, the 8-year-old good readers did better with the list. If good readers really were able to read better in context, then this should not have occurred.

The present findings were consistent with two previous replications, discussed above, by Nicholson et al. (1988) and Nicholson et al. (1991), even though these studies used smaller samples of children and a smaller corpus of errors. The only unusual finding of the present study, in comparison with these previous replications and with results reported by Stanovich (1980, 1986), was that the younger good readers did not improve in context when given the context-list order, whereas the usual finding has been that poorer and also younger readers gain in context. Inspection of the results of Nicholson et al. (1988) showed that the younger good readers did gain in context, even in the context-list order, although this did not occur in the Nicholson et al. (1991) study.

A possible explanation for this apparent discrepancy is that the good readers in this latest replication were, on average, better readers than those in the previous replications, in that they attempted more difficult passages. Experience in carrying out these previous replications suggests
that this was the case. This may mean that it would make sense to consider
children’s reading development and their relative reliance on context clues
as a matter of how well they can read, regardless of age.

Taken together, the results of the present replication study and the two
previous ones (Nicholson et al., 1988; Nicholson et al., 1991) provide a
weight of evidence to suggest that the classic study may have overempha-
sized the positive effects of context. Instead, the results of these various
replications suggest that Stanovich’s (1980) interactive-compensatory
model of reading may be a more appropriate explanation of what happens
when children read in context. The model claims that poor readers rely on
context to compensate for their poor decoding skills, whereas good
readers, who are good at decoding, have less need to do so.

If this model proves to be valid, then the whole-language approach, as
is practiced in New Zealand schools, may be overstating the case for
context. For example, the following remarks about the reading process
may well be a prescription for reading failure for many children rather
than reading success.

Reading may be thought of as a constantly repeated process of
sampling, predicting, confirming and self-correcting. At the outset,
the text is sampled, and significant visual features are searched for
and picked out, some words and/or letters being instantly recog-
nised.

On this basis, predictions of meaning and text are made. This
enlightened guessing makes use of the reader’s background know-
ledge and oral language, and the semantic, syntactic, and grapho-
phonic cues which the sampling has brought into focus.

(Department of Education, 1985b, pp. 25–26)

However, in defense of “enlightened guessing,” this practice does not rule
out the possibility that good readers use context clues to help them
become good readers, that is, by using such clues to acquire decoding skills
(Tunmer, 1990). But, as Tunmer pointed out, this is not the same as
enlightened guessing in the sense intended above:

It is important to distinguish this type of contextual facilitation
from that associated with the views of Goodman and Smith.
Goodman and Smith argue that the use of context to predict
words is a major feature of ongoing sentence processing, whereas
the view proposed here is that the ability to reflect on sentence
structures (i.e. syntactic awareness) in combination with emerging
phonological recording skills is essential for acquiring word recogni-
tion skills.

(Tunmer, 1990, p. 101)
I believe that reliance on enlightened guessing, of the kind described above, as part of the whole-language approach, only confuses children. As Gough and Hillinger (1980) pointed out, children start well but soon reach a plateau at which words become hard to guess so that only the acquisition of decoding skills will help them to read the many words they already know and understand in their listening vocabulary. Although, as Tunmer (1990) pointed out, children can use enlightened guessing to become better decoders, Tunmer cautioned that many children may not use it solely for this purpose and may continue to rely on this strategy, even when it breaks down.

Therefore, if enlightened guessing has the potential to cause reading problems, then perhaps children should be given more instruction in phonics, which would give them an alternative to reliance on guesswork. Although this method of instruction is not perfect and requires children to be phonologically sophisticated for it to succeed, phonics instruction at least confronts children directly with the decoding task.

In contrast, the whole language approach does not do this. As previously mentioned, at least 15% of New Zealand children require remedial instruction in their second year of school, at 6 years of age, after experiencing this approach. Therefore, there may be practical benefits from insisting on at least some early instruction in phonics (Thompson, Tunmer, & Nicholson, in press) to determine whether this method of instruction can reduce the incidence of reading difficulties in children. Perhaps this idea should also be given attention in other countries in which enlightened guessing is strongly emphasized. A controlled study, in which some children receive mandatory phonics instead of, or in addition to, the mainstream whole-language approach, while others continue with the regular approach would be a way of investigating this idea.

Although phonics is difficult to teach well, phonics skill may be an important advantage for good readers. It seems clear that students who have strong phonics skills are better at using context than poor readers, when given the same words to guess, in easier material (Nicholson & Hill, 1985). However, with more difficult material, in which words are harder to guess, phonics skills would become extremely helpful to good readers, and this is probably what enables these readers to read just as well when given words in lists as when given words in context.

To conclude, the findings of Goodman's (1965) classic study appear to be problematic. Although the classic results give the impression that all children read dramatically better in context, the evidence now suggests that only poor and younger average readers clearly read better in context. Good readers, on the other hand, even the younger ones, seem to be less reliant on context clues. This may mean that the enlightened guessing idea is incorrect, especially as a description of how good readers read.
Acknowledgement

I would like to thank Pam Halls and Owen Sutherland for their help in making this study possible, Sue Dymock and Leanne Lim for help with data collection, John Wells and Alan Lee for help with data analysis, and an anonymous reviewer for helpful advice.

Correspondence concerning this article should be addressed to Tom Nicholson, who is now at the Education Department, University of Auckland, Private Bag, Auckland, New Zealand.

References


Department of Education. (1985b). Reading in junior classes with guidelines to the revised Ready to Read series. Wellington, New Zealand: Author.


Staff. (1986, August). This week's citation classic. *Current Contents/Social and Behavioral Sciences*, 18, 12.
