Investigating the causes of reading comprehension failure: The comprehension-age match design

Kate Cain, Jane V. Oakhill,

University of Sussex

&

Peter E. Bryant

University of Oxford

Abstract

The reading-level (or reading-age) match design has become a widely-used tool for investigating the possible direction of the relation between particular skills and word reading ability: Cause or consequence. This paper outlines an analogous method for identifying candidate causes of reading comprehension failure, the "comprehension-age match design", and discusses the strengths and limitations of this design.

Keywords: reading comprehension, causal relations, word reading accuracy, ability-match designs

Introduction

Children's failure to understand written text can arise at different stages in the processing of language. For example, children with impoverished vocabularies and/or those who experience difficulties in word reading can fail to understand sentences and longer pieces of text (e.g. Liberman & Shankweiler, 1991; Perfetti, 1985). In addition, there exists a group of children who demonstrate reading comprehension problems in the presence of age-appropriate word reading and vocabulary skills (e.g. Oakhill, 1994; Stothard & Hulme, 1996). Oakhill and colleagues have shown that such children experience problems with a range of higher-level reading-related skills, such as inference making, comprehension monitoring, and anomaly resolution (de Sousa & Oakhill 1996; Oakhill, 1982, 1984; Yuill, Oakhill & Parkin, 1989).

However, the finding that skilled comprehenders make, for example, more inferences than less skilled comprehenders is essentially correlational. It does not tell us anything about the direction of the relation between these two skills. It may be that the skilled comprehenders' greater experience of reading and understanding stories has strengthened their inferential skill, relative to that of the less-skilled comprehenders. Alternatively, it may be that the less-skilled comprehenders' poorer inferential skill has (at least in part) caused their poorer comprehension.

Three types of experimental investigation can be used to help to determine whether or not strength in one skill, such as inference making ability, is the result or cause of strength in another skill, such as reading comprehension. These methods are: longitudinal investigations, training/intervention studies, and ability-match comparisons. Although causal hypotheses are best tested by the first two methods, these types of investigation are both time consuming and costly endeavours and, therefore, only worth pursuing once candidate causes have been identified.

The third method of investigation, using an ability-match comparison, is to include a control group that is matched to the less skilled comprehenders on the variable which this group has lagged behind, i.e. on comprehension age. This design

is analogous to the widely used reading-age match design in which the performance of backward readers on the task(s) of interest is compared to that of younger children who are matched with the backward readers for word reading skill (e.g. Bradley & Bryant, 1978). The argument here is that if the reading-age match (or RAM) control group does better than the backward readers on some task, such as rhyme detection, this difference cannot be said to be a product of the two groups' word reading levels since the groups are at the same level of word reading. It is, therefore, more likely that the difference is associated with the cause of the backward readers' delayed reading. The RAM design has proved a popular method for identifying candidate causes of delayed or deficient word reading skills (e.g. Frith & Snowling, 1983; Siegel & Ryan, 1988; Stanovich, Nathan & Vala-Rossi, 1986).

For just the same reason, Cain and Oakhill (1996) have advocated an ability-match design to help establish whether or not particular skills are the result of good reading comprehension (see also Stothard & Hulme, 1992). The comparison of, for example, the inferential skills of less skilled comprehenders to those of a younger group of equivalent comprehension ability, a comprehension-age match (or CAM) group, could help to determine which is the more plausible explanation for the relation between the two skills: Does proficiency in inference making lead to good reading comprehension, or does good reading comprehension result in superior inference making? If the comprehension-age match group make more inferences than the less skilled comprehenders, we can rule out the possibility that superior inference skill is the result of superior reading comprehension in general, because the CAM group and less skilled comprehenders are matched on this variable. This result would identify children's inferential skills as a likely cause of the level of their reading comprehension, a hypothesis which can then be explored further with longitudinal designs and training studies.

In recent work using the comprehension-age match design, we have found that less skilled comprehenders generate significantly fewer inferences than both skilled comprehenders and a comprehension-age match group (Cain & Oakhill, 1998). This

finding suggests that good inference making ability is not the product of reading comprehension level for the reasons discussed above. Instead, we conclude that it is more likely that inference making skill facilitates comprehension development, a hypothesis which we are currently pursuing in a longitudinal study.

The nature of the test that we routinely use in our work (the Neale Analysis of Reading Ability - Revised British Edition, Neale, 1989) requires that additional measures are taken to establish that the groups in this design are adequately matched, an issue that will be discussed further below. In addition, it is necessary to verify a crucial assumption of the CAM design: That the younger comprehenders have not had greater reading experience than the older poor comprehenders. This assumption is obviously more contentious than the analogous assumption in the RAM design where the crucial dependent variable is word reading skill, rather than comprehension skill. Relevant reading experience might also include the experience of listening to stories, which might of course influence comprehension but is unlikely to influence word reading. It is possible that the younger, CAM, group have had greater reading experience, at least in this broader sense. However, we have evidence that this is not the case, discussed below. Using data from our recent work, we first outline how we have selected groups for this design, and then discuss these two points, in turn.

Group selection

In our recent work using the comprehension-age match design (e.g. Cain & Oakhill, 1996, 1998) we have used two tests in the selection process. One was the vocabulary subtest of the Gates-MacGinitie Primary Two Test (MacGinitie & MacGinitie, 1989, Gates & MacGinitie, 1965) which indicates a child's ability to read and understand isolated words. This test was used to screen out exceptional readers, i.e. those children who performed either very badly or very well. The remaining "average" performers were tested on Form One of the Neale Analysis of Reading Ability - Revised British Edition (Neale, 1989). The Neale test consists of a series of short self-contained stories, graded in difficulty. Children read each story aloud and

any words that are misread or refused by the child are supplied by the tester. After each story they are asked a series of questions about the story. Testing stops after the child has made a prescribed number of word reading errors, i.e. no further stories are attempted once this limit is reached. Because word reading errors are corrected, children will have read or had read to them all the words in a story that they are subsequently asked questions about. Therefore, in this test, ability to answer comprehension questions is not as dependent upon word reading skill as it is in other assessments of reading comprehension. The Neale provides separate age equivalent scores for both word reading accuracy, based on the number of words that a child is unable to read or misreads, and reading comprehension, based on the number of questions about the stories that the child answers correctly.

Less skilled comprehenders. In our studies we have usually seen children aged between 7-8 years who have age-appropriate word reading accuracy. However, all of our less skilled comprehenders have comprehension ages that are below their chronological ages and at least 6 months below their word reading accuracy age. The characteristics of typical groups are shown in table 1.

TABLE 1 AROUND HERE

Skilled comprehenders. These children have always been selected from the same classes as the less-skilled comprehenders. Their word reading accuracy ages are also within the "normal" range for their chronological age, and their comprehension scores are either at or above that predicted by their word reading accuracy ages. The skilled and less skilled comprehender groups are always matched on the following variables: chronological age, Neale word reading accuracy, and Gates-MacGinitie vocabulary scores, but are selected to differ in terms of Neale reading comprehension performance. An example of typical group characteristics is shown in table 1. The mean reading comprehension ages of the two groups of skilled and less skilled comprehenders shown in table 1 were significantly different, $t(24) = 10.45$, $p < .001$.

The two groups were matched for Neale word reading accuracy, $t(24) < 1.0$, and Gates-MacGinitie vocabulary scores, $t(24) < 1.0$. Thus, we can assume that the less-skilled comprehenders' performance, relative to that of the skilled group, was not poorer because of some general reading disability, but was a specific text-level comprehension deficit.

Comprehension-age match (CAM) group. This group comprises children aged 6-7 years who are also progressing normally in their reading: Both their word reading accuracy and reading comprehension ages are within 6 months of their chronological age and there is never more than 6 months difference between the two scores. This group is selected so that their mean comprehension age is not significantly different from that of the less-skilled comprehenders ($t(24) < 1.0$, for the groups shown in table 1).

Are the groups adequately matched for comprehension skill?

We turn now to the first point raised above: How the nature of the test instrument (in this case the Neale) may pose problems for such a design. As table 1 illustrates, the less skilled comprehenders read more stories during the Neale test than the younger CAM group: 3.29 compared to 2.50, $t(24) = 4.04$, $p < .001$, because their word reading was superior. This difference was to be expected because all groups had word reading skills which were commensurate with their chronological age and the less-skilled comprehenders were older than the CAM group. Thus, it is possible that the assessment of the CAM group's comprehension skill was constrained by their word reading skill and that if they had been able to read more stories and, thus, attempt a greater number of comprehension questions, their general comprehension ability may have been found to be greater. Additional comparisons were necessary to ensure: i) that the less skilled and CAM groups were adequately matched for comprehension ability and ii) that the older skilled comprehenders were of superior comprehension ability to this younger group and had not simply obtained higher scores as a consequence of having read more stories.

The raw comprehension scores were reanalysed excluding the scores on the stories that had been too difficult for the younger CAM group. Because of unequal sample sizes, we were unable to conduct this analysis on a matched pair basis for the groups shown in table 1. Instead we compared performance in the following way. Four CAM children had successfully completed 3 stories during the Neale assessment. They were 'paired' with four less-skilled comprehenders who were of the same absolute comprehension level and also with four skilled comprehenders of comparable word reading ability to these four less skilled comprehenders. The total number of questions that these children had answered correctly in the first three stories of the Neale were the scores used in the analysis. All of the other CAM children read two stories. Therefore, for all remaining children (skilled comprehenders, less skilled comprehenders and the comprehension-age match children) only the number of questions answered correctly on the first two stories of the Neale test were included in this re-analysis. In this way, we were able to compare comprehension ability between the groups for only the stories that the CAM group had successfully read. The pattern of data was the same as that obtained in the original reading comprehension assessment (using the Neale test): less-skilled comprehenders = 8.71 (sd = 1.90), skilled comprehenders = 12.67 (sd = 3.17), and CAM group = 9.58 (sd = 1.24). As expected, the difference between the less skilled comprehenders and the comprehension-age match group was not significant, $t(24) = 1.36$, but the difference between the skilled comprehenders and the comprehension-age match group was significant, $t(22) = 3.14$, $p < .01$. Thus, we can be confident that, in this sample, the less-skilled and comprehension-age match groups were accurately matched for comprehension skill and, in addition, that the skilled comprehenders and comprehension-age match group differed in comprehension skill. A similar pattern of results has been obtained with different samples of subjects (Cain & Oakhill, 1996, 1998).

An additional check of our initial (Neale) comprehension matching would be to compare performance on a listening comprehension task. Stothard and Hulme

(1992) report comparable patterns of data for listening and reading comprehension among 3 experimental groups selected in a similar manner to our own. We have replicated Stothard and Hulme's findings for the three groups described above. In our study the children listened to the first 4 stories of the Neale Analysis (Form 2) and were asked the set questions after each passage. The mean scores were: less skilled = 8.93 (sd = 3.85), skilled = 16.67 (sd = 3.96), and CAM = 10.0 (sd = 2.17). The skilled comprehenders answered significantly more questions correctly than both the less skilled group, $t(24) = 5.04$, $p < .0001$, and the comprehension-age match group, $t(22) = 5.11$, $p < .001$. There was no significant difference between the less skilled comprehenders and the comprehension-age match group, $t(24) < 1.0$.

Do the comprehension-age match group have greater reading experience than the older less skilled comprehenders?

We now turn to the second issue raised above. It is possible that, in such a design, the younger comprehension-age match group may have greater reading experience than the older less skilled comprehenders. Such a difference may be the cause of positive test differences. For instance, Perfetti (1994) proposes that a possible source of comprehension failure is inadequate knowledge about text structures, which may arise because of insufficient reading experience. Indeed, a relation between children's reading comprehension and the time they spend reading has been found (Anderson, Wilson, & Fielding, 1988; Cipielewski & Stanovich, 1992; Taylor, Frye, & Maruyama, 1990).

It seems unlikely that children a year younger than our less skilled comprehenders would have had greater exposure to print overall and, in view of their inferior word reading skills, we might expect their print experiences to date to be qualitatively different to those of the older less skilled comprehenders. In addition if, as research has shown, exposure to print is related to word reading ability (e.g. Allen, Cipielewski, & Stanovich, 1992; Greaney, 1980; Greaney & Hegarty, 1987; Stainthorp, 1997a) we might speculate that younger children who had greater levels of

print exposure than our older participants would have been better word readers anyway and would therefore have been excluded from our study in the initial selection process because their word reading accuracy would have been above the norm. However, as noted earlier, it is possible that the CAM children had more exposure to print in the form of listening to stories, which may have affected their reading comprehension skills.

In response to these speculations we report data from a proxy measure of print exposure that employed a similar method to the Author and Title Recognition Tests developed by Stanovich and colleagues (e.g. Cipielewski & Stanovich, 1992; Cunningham & Stanovich, 1990; Stanovich & West, 1989). We developed an Author Recognition Test by selecting popular children's writers using information provided by the local public library and local book shops, and from reference books targeted at children's librarians (1). On the basis of pilot work with 6-7 and 7-8-year-olds we selected 32 authors for our experimental list. Each had written at least two books targeted at the 5-12 year range. In addition, each author was known by at least one child from either age group but none were familiar to all. In the experiment proper, we assessed children's recognition of these names taking into account their false "recognition" of foil names, to adjust for response bias (2). The dependent variable was a measure of d-prime.

The test was given to 13 less-skilled, 12 skilled comprehenders, and 15 comprehension-age match children, selected according to the criteria outlined above. The d-prime scores were: less skilled comprehenders = .62 (sd = .37), skilled comprehenders = .73 (sd = .24), and CAM = .33 (sd = .68). The difference between the skilled and CAM group tended towards significance, $t(24) = 1.95$, $p < .065$, but more importantly, there was no indication of a difference between the two older groups ($t(24) < 1.0$) and no evidence that the younger CAM group had greater exposure to print than the less skilled comprehenders. Instead our data followed the pattern of word reading accuracy differences between our groups.

Discussion

In this paper we have outlined the logic behind the comprehension-age match group and the confirmatory checks on reading assessment scores that are necessary to ensure that groups are adequately matched. Furthermore, we have found no evidence that the younger CAM group has greater print exposure than the older less skilled comprehenders. Thus, we conclude that where less skilled comprehenders are found to be poorer on a task than children of comparable comprehension skill the result cannot simply have arisen through differences in reading comprehension performance. Instead, this finding identifies the skill of interest as a candidate cause of good comprehension.

There are obvious limitations with this design. Null results are ambiguous to interpret and there is always the possibility that the pattern of data found is actually the result of a tertium quid, an additional cognitive skill which is mediating the link between two variables (see Bryant & Goswami, 1986). For example, early vocabulary ability may predict later reading comprehension skill because the vocabulary measure may be a surrogate measure of intelligence. Thus, the relation between vocabulary and comprehension skill could be via this variable. It is, thus, advisable to take such factors into account, when exploring further the preliminary findings obtained with designs such as the one described here. A series of planned regression analyses is one technique available to control for the possible mediating influence of other cognitive factors (e.g. Reynolds, 1992; Shirey & Reynolds, 1988).

To conclude, it must be noted that a difference between less skilled comprehenders and a comprehension-age match group does not prove causality. However, it indicates that strengths in the skill being measured are more plausibly the cause of reading comprehension success, than the result of it. Such positive findings (e.g. Cain & Oakhill, 1996, 1998) can then be followed up with more time-consuming

and costly designs such as longitudinal or intervention studies, to provide a more stringent test of this hypothesis.

Footnotes

(1) We found that a Title Recognition Test was not as sensitive as the Author Recognition Test. This may be, as Stainthorp (1997b) has noted, because of the popularity of serialising children's books on television and radio in the UK.

(2) We actually used a 2-Alternative Forced Choice task to reduce subject bias, and the Signal Detection Theory model of memory processes (Macmillan & Creelman, 1991; McNicol, 1972; Snodgrass & Corwin, 1988).

Acknowledgements

Our work was supported by the Economic and Social Research Council in the form of a studentship awarded to Kate Cain and a research grant (R000 23 5438) awarded to Jane Oakhill and Peter Bryant. We are grateful to Bob Carlyon for his advice on the design and analysis of the Author Recognition Test, and to Ralph Reynolds and an anonymous reviewer for their comments on an earlier draft of this article. Finally, we would like to thank the staff and pupils from the Brighton & Hove Schools who participated in this work.

References

Allen, L., Cipielewski, J., & Stanovich, K. E. (1992). Multiple indicators of children's reading habits and attitudes: Construct validity and cognitive correlates. Journal of Educational Psychology, 84, 489-503.

Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. Reading Research Quarterly, 13, 285-303.

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organisation as a possible cause of reading backwardness. Nature, 271, 746-747.

Bryant, P., & Goswami, U. (1986). Strengths an weaknesses of the reading level design: A comment on Backman, Mamen and Ferguson. Psychological Bulletin, 100 (1), 101-103.

Cain, K., & Oakhill, J. (1996). The nature of the relationship between comprehension skill and the ability to tell a story. British Journal of Developmental Psychology, 14, 187-201.

Cain, K., & Oakhill, J. (1998). Comprehension skill and inference making ability: Issues of causality. In C. Hulme & R. M. Joshi (Eds.), Reading and Spelling: Development and Disorder, (329-342). Mahwah, NJ: LEA.

Cipielewski, J., & Stanovich, K. E. (1992). Predicting growth in reading ability from children's exposure to print. Journal of Experimental Child Psychology, 54, 74-89.

Cunningham, A. E. & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading experience. Journal of Educational Psychology, 82, 733-740.

de Sousa, I., & Oakhill, J. (1996). Do levels of interest have an effect on children's comprehension monitoring performance? British Journal of Educational Psychology, 66, 471-482.

Frith, U. & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. British Journal of Developmental Psychology, 1, 329-342.

Gates, A. I., & MacGinitie, W. H. (1965). Gates-MacGinitie Reading Tests. New York: Columbia University Teachers' College Press.

Greaney, V. (1980). Factors related to amount and type of leisure reading. Reading Research Quarterly, 15, 337-357.

Greaney. V., & Hegarty, M. (1987). Correlates of leisure-time reading. Journal of Research in Reading, 10, 3-20.

Liberman, I. Y., & Shankweiler, D. (1991). Phonology and beginning reading: A tutorial. In L. Rieben and C. A. Perfetti (Eds.), Learning to Read: Basic Research and its Implications, (pp. 3-17). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

MacGinitie, W. H., & MacGinitie, R. K. (1989). Gates-MacGinitie Reading Tests. Chicago: Riverside Publishing Company.

Macmillan, N. A., & Creelman, C. D. (1991). Detection Theory: A User's Guide. New York: Cambridge University Press.

McNicol, D. (1972). A Primer of Signal Detection Theory. London: Allen & Unwin.

Neale, M. D. (1989). The Neale Analysis of Reading Ability - Revised. Windsor: NFER-Nelson.

Oakhill, J. V. (1982). Constructive processes in skilled and less- skilled comprehenders' memory for sentences. British Journal of Psychology, 73, 13-20.

Oakhill, J. V. (1984). Inferential and memory skills in children's comprehension of stories. British Journal of Educational Psychology, 54, 31-39.

Oakhill, J. (1994). Individual differences in children's text comprehension. In M. A. Gernsbacher (Ed.), Handbook of Psycholinguistics, (pp. 821 - 848). London: Academic Press.

Perfetti, C. A. (1985). Reading Ability. Oxford: Oxford University Press.

Perfetti, C. A. (1994). Psycholinguistics and reading ability. In M. A. Gernsbacher (Ed.), Handbook of Psycholinguistics, (pp. 849 - 894). London: Academic Press.

Reynolds, R. E. (1992). Selective attention and prose learning: Theoretical and empirical research. Educational Psychology Review, 4, 345-391.

Shirey, L. L., & Reynolds, R. E. (1988). Effect of interest on attention and learning. Journal of Educational Psychology, 80,159-166.

Siegel, L. S., & Ryan, E. B. (1988). Development of grammatical sensitivity, phonological, and short-term memory skills in normally achieving and learning disabled children. Developmental Psychology, 20, 200-207.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. Journal of Experimental Psychology: General, 117, 34-50.

Stainthorp, R. (1997a). The non-word reading skills of precocious readers. Paper presented at the meeting of the Society
for the Scientific Study of Reading, Chicago, March 1997.

Stainthorp, R. (1997b). A children's author recognition test: A useful tool in reading research. Journal of Research in Reading, 20, 148-158.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. Reading Research Quarterly, 24, 402-433.

Stanovich, K. E., Nathan, R. G., & Vala-Rossi, M. (1986). Developmental changes in the cognitive correlates of reading ability and the developmental lag hypothesis. Reading Research Quarterly, 21, 267-283.

Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language comprehension and working memory skills. Reading and Writing, 4, 245-256.

Stothard, S. E., & Hulme, C. (1996). A comparison of phonological skills in children with reading comprehension difficulties and children with word reading difficulties. Journal of Child Psychology and Child Psychiatry, 36, 399-408.

Taylor, B. M., Frye B. J., & Maruyama, G. M. (1990). Time spent reading and reading growth. American Educational Research Journal, 27, 351-362.

Yuill, N. M., Oakhill, J. V., & Parkin, A. J. (1989). Working memory, comprehension skill and the resolution of text anomaly. British Journal of Psychology, 80, 351-361.

Table 1. The mean scores (and standard deviations) of typical groups

	Less skilled comprehenders (N=14)	Skilled comprehenders (N=12)	Comprehension-age match (N=12)
Chronological age	7,7 (4.44)	7,7 (4.04)	6,6 (3.88)
Gates-MacGinitie	37.21 (4.00)	37.42 (3.00)	32.92 (2.91)
Neale word reading accuracy	7,9 (5.17)	7,11 (5.73)	6,7 (4.98)
Neale reading comprehension	6,7 (3.87)	8,1 (5.14)	6,8 (3.11)
Number of stories read in Neale test.	3.29 (0.47)	3.33 (0.65)	2.50 (0.52)

Note. Where appropriate, ages are given as years, months with standard deviations in months. The word reading accuracy and comprehension scores are the age equivalent scores provided in the Neale test, and the number of stories read refers to the stories that were completed during this assessment.