Identifying the Causes of Reading Disability

Donald Shankweiler,† Stephen Crain,† Susan Brady,†† and Paul Macaruso†

1. INTRODUCTION

Everyone is aware of the hazards in determining cause and effect from observations that are essentially correlational. This thorny problem has surfaced recently in discussions of the interpretation of research on reading disability. It has been argued that the issue of whether causation can be inferred from obtained differences between good and poor readers turns on the selection of an appropriate control group. A much discussed proposal advocates the abandonment of designs that compare groups of children who differ in reading ability but who are matched in chronological age (the CA match control group). Instead, the critics would substitute comparison groups of children who are at the same reading level (RL) but who differ in age (Bacalan, Mamen, & Ferguson, 1984; Bryant & Bradley, 1985; Bryant & Goswami, 1986). Although we grant that the RL match control group allows one to determine the direction of causation in some cases, we do not agree that it is always the best approach to understanding reading disability. Comparisons of groups matched on chronological age can also yield knowledge of causation if certain conditions are met.

In our opinion, the debate in the reading research literature over the conditions that must be satisfied in order to permit inferences about causation has focused too narrowly on research method. Previous discussion has overlooked two critical ingredients that jointly constrain the direction of causation and thereby enable correlational data obtained from CA match designs to address questions about the causes of reading disorder. One ingredient is an articulated set of hypotheses that ties reading to an explicit model of language processing. The other prerequisite is a convergent pattern of results. The findings of individual studies, viewed piecemeal, are often ambiguous. Because the criticisms of research using age matched groups have generally been directed to studies in isolation, they have sometimes failed to take into account the full weight of the evidence. Our response to the criticisms is based on discussion of an interlocking set of research findings that support a specific model of language processing and a set of assumptions about the special demands of reading.

Here is the way we will proceed. In Section 2 we summarize the arguments advanced by Bryant and his colleagues against matching on chronological age. Then, in Section 3, we present the case for the general usefulness of the CA match control group in testing hypotheses about the causes of reading disorder. We outline some assumptions about the reading process and we show how CA match comparisons
were used to narrow the range of causal factors. This section also considers the possibility, raised by Bryant and his colleagues, that differences between good and poor readers may be caused by differences in reading experience. The remainder of the paper is a general discussion of the problems of subject selection in research on reading. Section 4 discusses the role of the control group in assessing the significance of individual variation among poor readers, and Section 5 considers the uses of intelligence tests and educational achievement tests in selection of appropriately matched groups.

2. THE CASE AGAINST MATCHING ON CHRONOLOGICAL AGE

At issue is whether the CA match control group can, in principle, allow positive inferences to be drawn about the causes of reading disability. In considering this issue, our first task is to summarize the case that has been made against the CA match control group and to examine the alternative that has been recommended.

In a CA match design, good and poor readers (selected by standard reading tests) are matched in chronological age and, in addition, on measures such as IQ, oral vocabulary, etc. In an RL match design, poor readers are compared with younger normal readers who are performing at the same reading level. Bryant and Goswami (1986) conclude that positive and negative results have different values depending on which control group is used. Their argument goes as follows: Consider first the interpretation of positive results obtained with a CA match design (i.e. where poor readers are inferior to good readers on some cognitive ability). Here it is impossible to distinguish cause from effect: the differences on the criterion measure(s) may have been a consequence, not a cause, of the differences in reading ability. That is to say, differences in reading experience might be the cause of obtained differences on a variety of abilities that could conceivably benefit from such experience. Thus, a positive result with a CA match comparison group is, in principle, uninformative. On the other hand, if the two groups were matched on reading level at the outset, differences in level of reading skill could not be the cause of any obtained differences between the groups on other measures. Therefore, obtaining a positive result with the RL match can identify causes of poor reading.

It remains to consider the interpretation of negative results (i.e., when no between-group differences are obtained). Here, too, according to Bryant and Goswami, interpretation of the findings hinges on the choice of control group. In this case, negative results with CA match controls can be interpreted in causal terms, but only in a limited way, by ruling out the causal significance of any factor that failed to distinguish subjects who differ sharply in reading level. With an RL match control, negative results are inherently ambiguous. The failure to find differences could mean that the tested ability is irrelevant. Alternatively, the factor could be relevant to reading acquisition, but the older poor reader group might compensate for its absence by adopting a substitute strategy.

The principal implication that Bryant and Goswami draw from their analysis is that the CA match can be used only to rule out a factor's causal role in poor reading; it cannot affirmatively establish that a factor plays a causal role. In order to make such an inference, positive results with an RL match must be obtained. If this argument is valid, it would undermine many established hypotheses in the field—in effect, all hypotheses based on positive results with the CA match design.

An example chosen by Bryant and Goswami to illustrate the argument against the CA match design is the hypothesis that special limitations in working memory can cause reading problems. They reject the inference, based on use of the CA match design, that poor working memory is one of the causes of reading disability. They claim that causation could just as well go in the opposite direction: "...the successful Shankweiler et al.
readers could have better STMs because of their greater experience of reading" (1986, p. 102). We will refer to this as the reading experience hypothesis. On this hypothesis, the problem with CA match designs is that they allow two things to vary: reading experience and working memory. While we agree that there is nothing in the empirical data per se that argues for the direction of causation, we contend that the direction of causation becomes readily traceable when the experimental findings are viewed from the perspective of a set of assumptions about reading and its relation to language acquisition. We illustrate how this theoretical framework constrains inference making, by ruling out the reading experience hypothesis in some cases in which it might be invoked.

3. THE USE OF AGE MATCHED CONTROLS IN TESTING HYPOTHESES ABOUT READING

We begin by stating the assumptions that have guided our search for causative factors. Several considerations led us to believe that the origins of reading disorder should be sought in the language domain, rather than in some other cognitive system or in a general disruption of cognitive function. First is the observation that reading is largely parasitic upon primary language acquisition. The child who is learning to read does not have to acquire a new communication system, but can rely on preexisting language structures that have long been exploited in spoken communication by the time instruction in reading begins. To be sure, reading experience may modify the language of the reader, but we contend that such modifications are acquired only by skilled readers and are limited to such secondary aspects of language as the enrichment of lexical knowledge and the enhancement of metalinguistic skills. In Section 4, we indicate why we reject the possibility that the primary grammatical structure of language, including phonology, syntax, and semantics, could be acquired through reading (for discussion of primary and secondary language abilities, see Liberman, 1983; Mattingly, 1984).

Many empirical findings support the contention that the source of the difficulties of reading is in the language domain, and does not reflect a general deficit in perceiving visual or auditory patterns, or in forming analytic strategies. For example, poor readers have been found to be impaired relative to good readers in identification of acoustic stimuli masked by noise, but only when the stimuli are speech. Other kinds of sounds presented in noise are as accurately perceived by poor readers as by good readers (Brady, Shankweiler, & Mann, 1983). Further evidence indicates that poor readers have deficiencies in language-specific tasks. For instance, poor readers are reliably worse in memory for pictures of familiar objects, letters, nonsense syllables, and strings of unrelated words, but they are equivalent to good readers in memory for unfamiliar faces and nonsense designs (Katz, Shankweiler, & Liberman, 1981; Liberman, Mann, Shankweiler, & Werfelman, 1982; Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977).

From the earliest of these studies the same basic design was used: a group of poor readers was compared to a CA matched control group on parallel verbal and nonverbal tasks. Incorporating this design, these studies revealed positive differences between reader groups that could be attributed to poor readers' limitations in processing verbal materials, but negative findings when they were tested on materials that do not lend themselves to linguistic coding. Thus, we see a consistent pattern of positive and negative results indicating that the deficits of poor readers are language-related.

From further consideration of the internal organization of the language faculty, and the nature of alphabetic writing, we and our colleagues at Haskins Laboratories were led to expect that some limitation in the management of phonological
structures might be the specific source of reading difficulty. It seemed obvious that mastery of reading requires the learner to discover how the segments of the orthography represent the phonological segments of the language. It also seemed obvious that until the learner has explicit awareness of phonological structure, it would prove impossible to grasp the orthographic code, let alone become skilled in its use.

In view of this, we asked whether explicit awareness of phonological structure is a natural consequence of knowing a language. In research beginning in the early 1970s it was shown that the answer is no; preschool children, and even adult illiterates, find it difficult to abstract and manipulate phonological segments of spoken words (Liberman, 1973; Liberman, Shankweiler, Fischer, & Carter, 1974; Morais, Cary, Alegria, & Bertelson, 1979; Rozin & Gleitman, 1977). These facts have been confirmed repeatedly in research based on readers of several languages. Moreover, it has been found that children who are poor readers typically have greater difficulty in analyzing spoken words into their constituent phonemes than those who are good readers. Indeed, measures of phonological awareness have emerged time and again as the best predictors of reading success (see Bryant & Bradley, 1985; Liberman & Shankweiler, 1985; Stanovich, 1986, for reviews). Evidence that the source of the difficulty is linguistic and does not stem from the general cognitive demands of these phonological tasks comes from several studies that fail to find reader group differences on nonlinguistic counterparts to these tasks (Morais, Cluytens, & Alegria, 1984, and Pratt & Brady, in press).

In addition to phonological awareness, we hypothesized that another prerequisite is needed in order to forge a link between the orthography and phonological structure: an efficient verbal working memory. If the fault in reading disorder is in phonological processes, one natural consequence would be that the affected individual would have special limitations in working memory, since it has been known for many years that this form of memory relies heavily on coding based on phonological structure. We refer here to the classical findings of Conrad (1964; 1972; Baddeley, 1966) and others that even confusions of visual stimulus items in short-term memory typically have a phonological basis and are not based on visual or semantic similarity. Converging evidence from several laboratories finds that poor readers are inferior to their age-matched peers on memory tests that require them to retain the order of occurrence of a series of words or objects that can be named (Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979; Wagner & Torgesen, 1987). As we saw, differences in recall have been obtained with a variety of verbal materials, including spoken sentences, but not with materials that cannot readily be coded phonologically.¹ These findings were interpreted, at the time, as reflecting deficiencies of poor readers in verbal short-term memory buffer capacity (our more recent view is discussed below). Given deficient performance on various measures of verbal memory, a beginning reader with a severe limitation in storage capacity would have a double handicap. An abnormal limitation of capacity would impede learning the orthographic code, and would also hobble other on-line reading processes that make heavy demands on phonological storage (see also Baddeley, 1986; Perfetti, 1985).

So far we have reviewed the guiding assumptions and some of the major findings of our research. We have indicated how designs that employ CA match controls were used to narrow down the possible explanations of reading disability. The most basic application of the CA match was to show that the problems of the poor reader are in the language domain and not in visual processing or in some other cognitive domain. A second application was to show that the source of the problem lies at the phonological level of language. Children who fail to make expected progress in reading typically have difficulties in two phonologically-related areas: they have

¹ Shankweiler et al.
difficulty in becoming aware of the internal structure of words and they have special limitations in verbal working memory. Other findings, however, have led some researchers to question whether all the problems of disabled readers can be explained in these terms. In addition to several differences between good and poor readers that clearly implicate the phonological component of language, the reader group differences also extend to spoken sentences as well. These findings have been confirmed repeatedly in experiments in which subjects are asked to act out sentence meanings with toy figures (e.g., Byrne, 1981; Mann, Shankweiler, & Smith, 1984).

3.1 Reading and Sentence Comprehension: A Problem of Interpretation

Since good and poor readers have been found to differ in comprehension of some spoken sentences, it became an important priority of our research to find out whether the observed differences were related to their differences in phonological processing. The conclusion we have reached is that both problems are tied together—the comprehension failures of poor readers are derived from their limitations in phonological processing. In reaching this conclusion, we directly confront the claim of Bryant and his colleagues that positive differences between age matched good and poor readers do not permit us to draw inferences about the direction of causation. Our research shows that this claim is unwarranted in light of the evidence we have gathered for the view that the differences between reader groups in spoken language comprehension are best explained as a response by the language processing system to a deficit at the phonological level. If this explanation is correct, the direction of causation has been assigned: the poor readers in our studies were handicapped in several aspects of the reading process from the very beginning, because the phonological processing capabilities that they brought to the reading task were inadequate. We call this proposal the processing limitation hypothesis (PLH).

There is another point of view, however. Poor readers may lag behind good readers of the same age in the acquisition of certain linguistic structures, and they may fail to comprehend those structures in tests of spoken language understanding. We refer to this as the structural lag hypothesis (SLH). This hypothesis supposes that children who are poor readers are delayed relative to good readers in the acquisition of critical language structures. The structural lag hypothesis is tied to an implicit assumption about the course of language acquisition, as well as to an assumption about linguistic complexity. It supposes that some linguistic structures develop before others, with the course of development determined by the relative complexity of the structures. Our comments are directed to a version of the structural lag hypothesis that holds that some poor readers suffer from a developmental lag in syntactic knowledge. This version of the SLH appears to draw support from some classical studies in language acquisition that find the late emergence of certain constructions, for example, relative clauses, temporal terms, and adjectives with exceptional control properties, such as easy, in sentences like The bear is easy to reach. (See C. Chomsky, 1969; Clark, 1970; Sheldon, 1974.)

On the SLH, the differences between good and poor readers on spoken language comprehension are completely separate from their differences on phonological processing. If this position were upheld, the direction of causation would remain an open question. It would then have been appropriate to ask, further, whether the poor readers had failed to acquire some of the structures needed for comprehension, or whether, instead, the good readers had advanced beyond their age-matched classmates because their greater experience in reading has enhanced their knowledge of critical grammatical structures. Either eventuality poses a challenge to our efforts to tie together the observed differences between good and poor readers at the sentence level and at the level of the phonology.

We will now sketch how we have countered these challenges to the processing limitation hypothesis (PLH). First we explain how the architecture of the language
apparatus relates failures of spoken language comprehension to limitations in processing arising at the phonological level. Then we show how our research has succeeded in ruling out the alternative account (the SLH), which proposes that the differences between good and poor readers in spoken language are unrelated to their differences in phonological processing. The last step in the argument takes up the broader issue of the possible role of reading experience in explaining reader group differences.

To begin, let us explain how the failure of poor readers to interpret spoken sentences correctly is expected, given the conception of the language apparatus that underlies the PLH. We have long held the position that the language apparatus forms a biologically coherent system that is isolated from other parts of the cognitive and perceptual apparatus. In contemporary terms, language forms a “module” (Fodor, 1983). Our present view extends the notion of modularity, by differentiating subcomponents of the language faculty. As we conceive of it, the language module is composed of a hierarchy of structures and processors, each of which functions according to its own properties. The structures include the phonology, lexicon, syntax, and semantics. Each level of structure is served by a special-purpose parsing mechanism. A parser consists of algorithms for accessing rules used in assigning structural representations, as well as mechanisms for resolving ambiguities. Although the operations of the various components are interleaved in time, permitting the system as a whole to function “on-line,” the flow of information between levels is tightly regulated. This control process is one of the chief functions of the working memory system, in our view. The strongest position, and the one that we adhere to in the absence of counterevidence, is that information transfer within the language apparatus is unidirectional, beginning at the lowest level with phonological processing and proceeding upward to the syntactic and semantic parsers (for discussion, see Crain & Steedman, 1985; Fodor, 1983; Forster, 1979; Shankweiler & Crain, 1986).

Since information flows bottom-up through the system, the first task of the “executive component” of working memory is to relay the partial results of the phonological analysis of the linguistic input upward through the system, thereby freeing this component for analysis of subsequent material. A low-level deficit in processing phonological information would create a bottleneck that would impede the transfer of information to higher levels in the system (for related discussion, see Crain, Shankweiler, Macaruso, & Bar-Shalom, 1987; Perfetti, 1985; Shankweiler & Crain, 1986). Our current view, then, gives less weight to limitations of buffer capacity, as such, than to limitations in the efficient transfer of phonological information to the higher components of the verbal working memory system in on-line language processing.

Poor readers, on this processing limitation view, would be expected to have difficulties in spoken language comprehension on sentences that place heavy demands on working memory, but should be the equals of good readers on all other sentences, regardless of inherent complexity. Appealing to modularity principles, the PLH supposes that the acquisition of primary language structures is essentially complete by the time instruction in reading and writing begins. Early mastery of complex syntax is seen to be a consequence of the innate specification of many syntactic principles that either come “prewired” or are subject to rigid system-internal innate constraints on grammar construction (see, e.g., Chomsky, 1965, 1981). Since syntactic structures are largely built into the blueprint for language acquisition, it follows that inherent complexity of grammatical structures, as such, will not be a source of reader group differences (Crain & Shankweiler, 1987). Poor readers will be at a disadvantage, however, in contexts that stress verbal working memory. Whether the materials are presented by eye or by ear, their performance

Shankweiler et al.
will reflect their working memory liabilities (Liberman & Shankweller, 1985; Mann et al., 1984; Shankweller & Crain, 1986).

3.2 Testing between the SLH and the PLH

As we saw, if the PLH is correct, the direction of causation is assigned, but the issue is open if the SLH is correct. It is essential to show, then, how we have used CA matched controls to test between the PLH and its rival, the SLH.

We have seen that the SLH and the PLH are broad hypotheses about the factors that make reading difficult to learn and about the characteristics that distinguish children who are making normal progress from those who are experiencing reading failures in the early grades. Both hypotheses view reading as building on earlier language acquisition, and both predict that children who are disabled in reading will also manifest language deficits in certain instances in comprehension of spoken language. Despite these features in common, the hypotheses ultimately diverge. First, on the SLH, the putative syntactic deficits are independent of further deficits the poor readers may have with phonological structures. By allowing at least two basic deficits in poor readers, this hypothesis abandons a unitary explanation of reading disability. The PLH, on the other hand, attempts to tie together the entire symptom complex of poor readers as a consequence of deficient phonological processing.

A second difference between the hypotheses concerns the possible role of reading experience. The PLH does not envision a role for reading experience in the acquisition of syntactic knowledge. In contrast, the SLH can readily accommodate the supposition that reading and writing contribute the experience needed for complete mastery of late-emerging syntactic structures. As we saw, there is evidence that several of these late emerging structures are more difficult in spoken language tasks for poor readers than for good readers (e.g., Byrne, 1981; Stein, Cairns, & Zurif, 1984). This evidence has been interpreted as indicating late maturation in poor readers of spoken-language competence, as befits the proposal that poor readers have not yet encountered the requisite data for complete acquisition of the finer points of syntax. The PLH explains these phenomena in other terms.

In the foregoing sections, we demonstrated how a combination of positive and negative results in studies using a CA match design allowed us to isolate the source of several symptoms of reading disability, most generally, to the language domain, and subsequently, to the phonological component of language. We discussed how the sentence comprehension failures of poor readers could also occur as a by-product of deficiencies in phonological processing. We saw, though, that the SLH places in jeopardy the suggestion that all of the symptoms of reading disability arise from a deficit in phonological processing.

It is important to appreciate the role that CA match comparisons have played in recent research to test between the alternative explanations of the sentence comprehension problems of poor readers. The studies we have conducted called for pairs of tasks that vary memory load, while keeping syntactic structure constant. If reading disability stems from a structural lag, then children who have reading problems should perform poorly on both tasks. But according to the PLH, poor readers should have greater difficulty than their age-matched controls only in tasks that place heavy demands on working memory, whatever the inherent complexity of the linguistic structure being investigated. When the same test materials are presented in tasks that minimize processing load, poor readers should do as well as good readers. In short, confirmation of the PLH depends on obtaining both positive and negative results.2

In several studies, this pattern has been identified. We find that on tasks that are demanding of memory resources poor readers tend to make more errors than good
readers. Positive differences have been obtained with sentences containing relative clauses (Mann et al., 1984), sentences containing adjectives with special control properties (Byrne, 1981), and sentences with the temporal terms before/after (Crain et al., 1987). However, poor readers have proven to be the equals of good readers on each of these constructions when the same materials are presented in ways that lessen memory load (Crain, 1987; Crain et al., 1987; Smith, 1987). Contrary to the expectations of the SLH, in these contexts both reader groups achieve such a high level of accuracy that competence with the constructions under investigation would seem guaranteed. In short, the pattern of positive and negative findings clearly favors the PLH. In light of this research, it is apparent that CA match designs can successfully distinguish between these competing hypotheses about the source of spoken language comprehension problems in poor readers.4

3.3 A Critique of the Reading Experience Hypothesis

We return to the criticisms that have been levied against the CA match comparison. The argument against CA match designs turns, as we saw, on the premise that only negative results with this control group can lead to definite conclusions about the sources of reading disability. Positive results, on this argument, cannot justify definitive statements about causation, since the possibility must be entertained that limited reading experience was the cause of breakdowns in processing, instead of being caused by them (as Bryant and his colleagues suggest). It seems to us that the possible contribution of reading experience is at the heart of the criticism against CA match designs, so it deserves closer scrutiny. Since we have chosen to allow reading experience to vary, we must respond to this criticism by showing that this factor could not account for differences between reader groups on the dependent measure.

In the preceding section, a pattern of results was presented that defies explanation on the SLH, a hypothesis that was shown to accommodate the view that reading experience contributes to the acquisition of syntax. It remains to generalize the argument by showing how the theoretical framework we have adopted can overcome the challenge posed by other versions of the reading experience hypotheses. No one, to our knowledge, has explicitly worked through the various possible implications of the idea that reading experience creates the between-group differences identified by research. Therefore we limit discussion to possibilities that, in our view, are the most plausible ways in which reading experience could add to the primary language structures and processors involved in spoken language comprehension (keeping in mind, as we acknowledged earlier, that secondary language abilities can surely be expected to profit from reading experience).

First we consider the possibility that reading experience could enhance language understanding in the way Bryant and Goswami (1986) suggested, by improving the efficiency of verbal working memory. We contend, to the contrary, that it is ad hoc to invoke reading experience as an explanation of the working memory differences between good and poor readers, once these differences are viewed in conjunction with impairments in language processing found in other populations with working memory deficiencies.

We view it as noteworthy that the experimental results we have summarized were predicted by a theoretical framework that is itself supported by a substantial literature in other areas of psycholinguistic investigation, including language acquisition, language breakdown in aphasia, and normal sentence processing. We were guided by this theoretical and empirical base in selecting materials and tasks for our studies of reading disability. Thus, in order to provide support for the supposition that the cause of performance failures in poor readers lies within the working memory system, we chose structures that were known to be problematic for Shankweiler et al.
other groups with working memory limitations: for example, very young children (Tavakolian, 1981), mentally retarded adults (W. Crain, 1986), aphasics (Caramazza & Berndt, 1985) and, in some cases, even normal adults, when spoken sentences are presented in a way that stresses memory (Crain & Fodor, 1985). As we noted, the findings of studies using a CA match design revealed, in each case, that good and poor readers did not differ in their linguistic knowledge but did differ in their ability to use that knowledge when the task situation stresses working memory. This brings the data from poor readers into line with other populations with notable limitations in verbal working memory.

Impoverished reading experience cannot explain the difficulties that all these groups encounter with the problem structures. In light of the proposal that working memory underlies comprehension failures in each population, it would be arbitrary to invoke reading experience as a causal factor for the poor reader group alone. On the contrary, since these diverse groups share in common severe limitations in verbal working memory, it is reasonable to infer that this is the common cause of their problems in comprehension. In short, the reading experience hypothesis offers an explanation for the observed pattern of positive and negative results in poor readers that is logically possible, but implausible; it can be invoked only at the expense of a unitary account of the processing deficits that afflict diverse populations.5

Now we are prepared to deal with a version of the reading experience hypothesis that we sketched earlier—the proposal that differences exist between good and poor readers in grammatical competence, and that these arise from differences in reading experience. Though this idea might have seemed supportable, we have already given empirical grounds to reject it. To these we now present additional grounds to bolster the argument, by considering this hypothesis in the light of theoretical and empirical findings from language learnability. These findings make it highly implausible to suppose that reading experience can implant new syntactic knowledge. It follows that differences in sentence understanding between good and poor readers are not likely to be the product of reading experience.

To see this, suppose (counterfactually) that we had obtained the pattern of results predicted by the SLH, with the poor readers performing less well both on the task that stresses working memory and on the task that minimizes memory demands. This result would imply that the poor readers were missing critical syntactic structures. It could not reasonably be interpreted, on analogy with Bryant and Goswami’s argument regarding working memory, as evidence that reading experience could have been the cause of good readers’ superior performance. To draw that conclusion would undermine one of the basic tenets of the modularity perspective, namely the hypothesis that syntactic principles are in large part specified by an innate “Universal Grammar.”6

The proposal that reading experience can implant primary syntactic knowledge hinges on the assumption that reading experience is better suited than spoken language to promote advancement through the later stages of language acquisition. In contrast, the theory of Universal Grammar requires that readily available data suffice for children’s acquisition of syntactic knowledge (Lasnik & Crain, 1985). The necessary linguistic data must be abundant to ensure that each child encounters all experience needed in grammar construction. This means that input from speech sources alone is sufficient for completion of grammar acquisition. If we take the theory of Universal Grammar seriously, it makes little sense to suppose that the acquisition of primary syntactic structures could be enhanced by reading experience, especially at the early stages of reading. And this, in turn, means that it makes little sense to invoke reading experience to explain any disparity between good and poor readers in the acquisition of syntax.
Of course, children’s convergence on the adult grammatical system sometimes requires extensive input from the environment. For example, specific experience is required to learn the properties associated with particular lexical items. So it is conceivable that reading experience could instill useful knowledge in these cases. However, for syntax, there is yet another reason to question the hypothesis that reading experience plays a fundamental role. It is widely held by proponents of Universal Grammar that mastery of syntax requires exposure to a particular kind of input from the environment, in the form of sentence-meaning pairs (e.g., Pinker, 1984; Wexler & Culicover, 1980), or sentence-intonation pairs (Morgan, 1986). This is buttressed by a formal demonstration that “surface” data alone (e.g., isolated written sentences) are insufficient for grammar formation (Hamburger & Wexler, 1973). But since acts of reading occur outside of normal social-communicative contexts, reading offers only surface data. Therefore, reading experience cannot supply the requisite data for the acquisition of syntactic principles.

It is time to take stock. We have illustrated the profitable use of CA match designs in testing between specific proposals about the causes of reading disorder. It proved meaningful to infer the direction of causation on the basis of a pattern of converging results that conforms to the predictions of the PLH. The greatest virtue of the PLH, in our view, is its power to explain a host of apparently diverse language deficits that are characteristic of poor readers. Although all of these problems are arguably phonologic in their origin, they manifest themselves as disturbances of both lower-level (phonologic) processes and higher-level (syntactic and semantic) processes. The SLH, in contrast, raises the possibility that poor readers suffer from more than one kind of deficit within the language domain. We noted evidence that at first seemed to support this position, but, ultimately, the PLH gave a better account of the data than the SLH. Finally, we turned to the narrower question of the possible role of reading experience in explaining the observed differences between reader groups. We rejected two versions of the reading experience hypothesis, the notion (suggested by Bryant and Goswami) that reading enhances working memory, and the notion that reading experience is needed to gain full mastery of syntactic structures.

The existence of individual variation among poor readers poses, for many workers in the field, another challenge to the PLH. As we noted, the findings we have reviewed led us tentatively to reject the hypothesis that there are poor readers who meet our selection criteria and who suffer from a syntactic deficit in addition to their deficiencies in phonological processing. However, different selection criteria could be adopted, and the groups that are subsequently formed could present different patterns of deficit. To acknowledge the possibility of multiple basic deficits in poor readers and complex dissociations among them is to question, from another point of departure, the adequacy of a unitary explanation for reading disability and to open the issue of subtypes. In the following section, we discuss the role that RL and CA match control groups have played in assessing the significance of individual variations in the symptom picture associated with poor reading.

4. ROLE OF THE CONTROL GROUP IN ANSWERING QUESTIONS ABOUT SUBTYPES

In any investigation of poor readers, it is readily apparent that they do not all look alike, even after steps have been taken to exclude from the experiment cases with sensory deficits, behavior disorder, and low IQ. What one makes of variations in the symptom picture depends on one’s basic assumptions about reading and where the problems ultimately lie. The data we have reviewed so far are in keeping with the phonologic hypothesis we have advanced. It could be argued, however, that the methods of investigation employed in this research have buried the variation by Shankweiler et al.
treat as "noise" differences that are potentially causally relevant (see Boder, 1973; Mattis, French, & Rapin, 1975). In order to maintain a unitary explanation in the face of individual differences, it must be shown that the differences reflect variations along a continuum.

In this section we confront the problem of individual variation within the population of disabled readers who meet our selection criteria. In doing this, we take a new tack. Whereas in the studies we considered previously, the spoken language problems of poor readers were the subject of investigation, in this section we focus on the reading process itself and examine how the pattern of errors bears on the question of whether some poor readers employ mechanisms that are not used by normal readers at any level of proficiency.

It is well known that damage to the dominant cerebral hemisphere can result in specific disturbances of reading and writing as well as disturbances of spoken language. When problems in reading and writing predominate in the symptom picture, the term "dyslexia" has standardly been used to designate the condition. Ever since the pioneering investigations of Hinshelwood (1917), resemblances have been reported in the clinical literature between the reading behavior of certain individual children who are experiencing inordinate difficulty in reading and that of brain-damaged adult dyslexics. It is easy to see why such parallels arouse keen interest. Their existence would support the idea that at least some children who fail to learn to read are not only retarded in the rate at which they learn, but they actually are deviant in the approach they adopt. The belief that there are qualitative signs that set the dyslexic child apart from other poor readers has been a recurring theme in the clinical neurological literature on reading problems. It is claimed that the adult syndromes of "deep dyslexia," "phonological dyslexia," and "surface dyslexia" find their counterparts in children (see Coltheart, Patterson, & Marshall, 1980). If this claim is valid, it would imply that childhood reading problems, like the adult dyslexias, are surely causally diverse. An urgent priority for research, therefore, would be to characterize the major syndromes of developmental dyslexia (in much the way they have been characterized in adult dyslexia).

The suggestion that syndromes of post-traumatic dyslexia in adults are mirrored in a closely corresponding set of developmental disorders is challenged in a recent paper by Bryant and Impey (1986). These authors employed the RL match design to investigate the question of whether there are special patterns of reading errors associated with some cases of developmental reading problems that set them apart from the normal course of development of reading and align them with particular varieties of acquired dyslexia (as claimed, for example, by Temple & Marshall, 1983, and Coltheart, Masterson, Byng, Prior, & Riddoch, 1983). Bryant and Impey maintain that the errors in question could only be used as a diagnostic for subtyping poor readers if they were found to be truly distinctive; that is, if they are absent in normal, younger children who are reading at the same overall level of proficiency. This comparison had not been made previously. That is to say, the comparisons between the reading error pattern of brain damaged adults and those of certain children and adolescents with severe reading difficulties had failed to include an additional comparison with normal, but immature and only partially competent readers.

Bryant and Impey supplied the appropriate comparison. They studied a group of normal ten-year-old readers, matched in reading level with a 16-year-old reading-disabled adolescent whom Coltheart et al. (1983) had presented as a paradigm case of surface dyslexia. The normal children, like the reading-disabled adolescent, made many errors in reading orthographically irregular words. Moreover, when Bryant and Impey compared the adolescent's errors in reading and performance on a same-different test of word recognition with those of the normal ten-year-olds reading at
grade level, they found much the same pattern of performance. A similar conclusion applied to a case described by Temple and Marshall (1983) as meeting the criteria for phonological dyslexia. Because reading errors have the same distribution in normal younger children, the pattern of errors evinced by each of the "dyslexic" cases cannot be regarded as qualitatively deviant. Bryant and Impey concluded (rightly, in our view) that reports in the literature of resemblances in error pattern between individual children with reading difficulty and reputedly distinct types of acquired dyslexia in adults have not yet brought us closer toward identification of separate causes for different reading problems among children. This is not to deny, of course, the existence of individual differences in reading strategies, which, as Bryant and Impey stress, are evident in both normal and reading-disabled children.

The lesson to be drawn from this application of the RL match bears emphasizing: that variation among cases of reading disability can only be evaluated against a background of variation in normal readers. It is worth noting that research strategies that use a CA match control group can lead to the same conclusion. One type of error to which diagnostic significance has been attached is the tendency to read letter sequences in reversed order and to reverse the orientation of individual letters. The purpose of the studies we are about to review was to evaluate an influential hypothesis concerning directional reversals.

It was claimed by Orton as early as 1925, and often reaffirmed in the literature, that reversals are pathognomonic of a syndrome of developmental reading disturbance that has a visual perceptual basis. But the question of the diagnostic significance of reversals could not be meaningfully evaluated until data were obtained on the kinds of errors normal children make at different stages of learning to read. To this end, an early study by researchers at Haskins Laboratories explored the occurrence of reversal errors in an entire class of second-year pupils in a suburban elementary school (Liberman, Shankweiler, Orlando, Harris, & Berti, 1971) and in a special group of children who had been selected by members of the staff of a learning disability clinic as the result of screening the school population in a large metropolitan area (Fischer, Liberman, & Shankweiler, 1978). The latter children met psychoeducational and medical-genetic criteria for "developmental dyslexia." Although the "dyslexic" children were somewhat poorer in word recognition than the worst of the poor readers recruited in the earlier second grade sample, the groups did not differ in the incidence of reversal errors. Moreover, in both groups, reversals represented only a small proportion of the total number of reading errors. Vowel errors and errors on nonreversible consonants constituted the bulk of the misreadings.

The findings on the "dyslexic" children might have led to a different conclusion had the investigators failed to take the further step of assessing reversals in an unselected school sample. Together, the findings of the two studies lend support to many other indications that the common error pattern is determined primarily by the structure of the language and the orthography and not by visual form. The occurrence of reversal errors among readers of some alphabetic orthographies is essentially a transient normal phenomenon in the beginning stages of reading acquisition (Mann, Tobin, & Wilson, in press; Simner, 1982). Reversals cannot justifiably be regarded as a symptom of a distinctive type of visual dyslexia. These findings, together with those of Bryant and Impey, support the contention that reading disorder in children is on a continuum. We therefore share with Bryant and his colleagues (Bryant & Bradley, 1985; Bryant & Impey, 1986) a skeptical attitude regarding the existence of subtypes of reading disorders.

The findings discussed in this section illustrate that the nature of the research question may influence the choice of control group. As Backman et al. (1984) remark, when questions are raised about the nature of the reading process itself, as opposed to

Shankweiler et al.
its cognitive underpinnings, the RL design has been successfully employed. Although
the RL design may be especially useful for this purpose, the CA match design can also
be used, as the foregoing research studies illustrate. Both sets of studies discussed
above are alike in their insistence that error patterns in reading disabled individuals
must be interpreted with reference to the distribution of errors among normal
learners. Research studies employing both RA and CA match designs find the
empirical basis for subtypes to be unconvincing. For us, the significance of these
negative findings is that they permit us to maintain a unitary hypothesis about the
causes of reading disability—that the various manifestations arise from a single
source. In the absence of counterevidence, either from studies of spoken language
comprehension or from analysis of the error pattern in reading, we are compelled to
continue to pursue the empirical consequences of this hypothesis.

In the following paragraphs, we take up several practical concerns that arise in
selecting an appropriate control group, and consequent problems of interpretation.

5. ACHIEVING MATCHED GROUPS: IN SEARCH OF PRINCIPLES

Granted that control groups are essential in testing hypotheses about the nature of
reading disorder, it remains to ask whether any guidelines can be proposed for the
selection of controls and for matching subjects in research on reading disorders.
Obviously, not every objective can be met in a single study. It follows that there will
be trade-offs in the use of one or another control measure, and that no absolute
prescription can be given. Therefore it is appropriate to examine the consequences of
adopting different criteria for matching subjects in research comparing good and
poor readers, with some examples from the recent literature.

5.1 Controlling for Intelligence

The need for selection criteria in addition to age and reading ability is widely
acknowledged, but which measures should we use? The choice of control measures in
the research we have conducted has been guided by the aim of distinguishing cases of
specific reading disability from reading problems that arise as only one of the
manifestations of low intellectual function or insufficient motivation. Since it has
been found that a host of problems in addition to reading are associated with low IQ,
any analytic goal would be defeated at the outset if uncontrolled variation in relevant
intellectual abilities were allowed (see Jorm, Share, Maclean, & Matthews, 1986;
Rutter & Yule, 1975). By equating groups on intelligence as well as on age we
minimize the possibility that poor readers are failing to achieve normal progress
because of generally poor learning ability. Although reading ability is certainly not
independent of general intelligence, it has long been recognized that a significant
segment of the population within the normal IQ range has special reading difficulties
(see Stanovich, Cunningham, & Feeman, 1984, for a contemporary review).

Having decided to use IQ as a control measure, we are confronted with a range of
choices. Should we use a Verbal or Performance IQ measure, or more specialized tests
of verbal or nonverbal abilities, such as a vocabulary test or figure drawing? Of
course there is no single answer, but the choices we make do influence the results we
obtain (see Satz & Fletcher, 1980). On the one hand it could be argued that
performance tests are the most appropriate control measures because children with
reading disability commonly have additional problems in the language domain that
could depress verbal IQ, and thus could introduce bias into the selection procedure.
On the other hand, the overlap in IQ measures between disabled and normally
achieving readers is so great that it has proved relatively easy to match good and poor
readers, at least poor beginning readers, on IQ measures, even verbal IQ measures.
The choice of IQ measure depends on the research question at hand. For example, if we wish to assess abilities that may be affected by vocabulary knowledge, then use of an IQ measure that tests vocabulary knowledge is essential in order to ensure that the poor reading group does not simply comprise children who have impoverished vocabularies, which could lead to spuriously low performance on the comparison measure. Thus, if we wish to use speech perception measures in our comparisons of the phonological abilities of good and poor readers, where it is well known that word knowledge affects performance, the inclusion of a vocabulary test like the Peabody Picture Vocabulary Test is a critical control (Brady et al., 1983). It is equally critical in studies comparing people of different reading ability on lexical retrieval tasks, such as object naming (Katz, 1985; Wolf, 1981). Further benefits of controls for vocabulary are apparent: they guard against the possibility that the poor readers suffer from a general depression in linguistic processing. In addition, a vocabulary measure addresses, in part, the need to distinguish cases of specific reading disability from those that arise as one manifestation of low intelligence. Clearly, any factor that sets a floor or ceiling on spoken language comprehension will have the same effect on what can be interpreted in print. Lastly, measures like the Peabody bear a high correlation with other linguistic and nonlinguistic indicators of intellectual function, for example, the Wechsler Verbal and Performance tests (as in the WISC-R). Controlling for vocabulary, then, holds these other factors in check, to some degree.

Ideally, one would elect to match pairs of subjects or to equate groups on some measure or measures of intelligence. In our own research, we have usually not attempted to match individual subjects or to select groups in such a way as to equate mean scores for good and poor readers. Instead, exclusionary criteria have been used in the formation of groups by setting minimum standards for inclusion in the study, thereby excluding subjects whose IQs fell above or below the average range. In many of the studies we have done, in which good and poor readers were selected from ordinary classrooms in suburban schools, the resulting groups did not differ significantly in mean score. It should be said, however, that in most of these studies the direction of the IQ difference, though nonsignificant, favored the better readers. Wolford and Fowler (1984), having surveyed the literature, find this to be a quite general phenomenon. This being so, Crowder (1984) has urged that any effects of IQ on the criterion measures be removed by statistical means, such as partial correlation, covariance analyses, or repartitioning subjects into groups based on IQ and comparing differences on the criterion measures for high and low IQ with differences based on reader groups. 

The use of any test from which an IQ can be derived has been criticized by those who point out that IQ tests are omnibus devices, and should not be employed when we have the option of using analytic tests that are more homogeneous in content. The point of these criticisms is well taken. That is why we have typically used a specific verbal IQ measure (e.g., PPVT) that controls for level of achievement in a relevant facet of spoken language comprehension (see above). Whether we use general IQ measures or special purpose devices, standardized tests have a powerful advantage that should not be overlooked: they alone give us benchmarks that make it possible to compare research findings from different laboratories, and, to a degree, across languages and writing systems (see Liberman, Liberman, Mattingly, & Shankweiler, 1980).

The problems of controlling for intelligence, not inconsiderable when comparing children of the same age, assume added dimensions of complexity when persons of different ages are being compared, as in the RL match design. If the subjects are children at different maturational and educational levels, equating for IQ means equating on some normalized standard score, which is of course not to equate on absolute level of performance. To return to an earlier example, suppose we wish to
compare children of different ages on a test of perception of spoken words: we would very likely be forced to use different test words for the two groups, even though both were matched on PPVT score, because their actual vocabulary knowledge would differ. This difficulty is insuperable with the RL match control group and would surely exclude its use for investigations that seek to identify underlying cognitive problems. An important virtue of the CA match design, overlooked in the recent enthusiasm for the RL match, is that it allows us to employ more stringent criteria in selecting subjects.

5.2 Controlling on Tests of Achievement

Some of the critics of IQ measures would substitute measures of classroom achievement in their place. Given that the intent is to identify cases of specific reading disorder, these critics maintain that if we do not use achievement test criteria, we run the risk of including children who are generally low achievers ("under-achievers" perhaps, if IQ is high). Tests of mathematical achievement have been suggested for this purpose.

We agree that the objective has some merit, but, again, there are trade-offs. Indeed, use of an achievement test may backfire by introducing a very definite selection bias of its own. A study by Hall, Wilson, Humphreys, Tinzmann, and Bowyer (1983) is a case in point. The intent of this study was to investigate the hypothesis that poor readers are deficient in verbal working memory and in use of the phonological codes that are believed to support this memory system. These investigators elected to compare good and poor readers who were matched in math achievement on two subtests of the Woodcock-Johnson battery (Woodcock, 1977), one of which was a test of performance on orally presented word problems. This is surely a bad choice in any study whose objective is to test whether poor readers have abnormal limitations in verbal working memory, since word problems inevitably introduce heavy memory demands, particularly when presented orally. Thus, the effect of this selection procedure is to exclude children with special working memory limitations. In view of this it is hardly surprising that Hall et al. failed, in the main, to obtain differences between groups of good and poor readers in letter string and word string recall, and failed to confirm the results of many other investigations that implicate failures of working memory as a causal factor in reading disability.

5.3 Assessment of Reading

The success of any study of reading, whatever its design, depends on valid assessment of reading abilities. Here, again, as in the case of language and cognitive measures, we are confronted with choices. Given the complexity of reading, we cannot expect to tap all the relevant skills with a single measure. Indeed, as Backman et al. (1984) rightly stress, to equate groups on one measure of reading may result in great inequalities on other measures. For example, individuals matched on a test of reading isolated words may differ markedly on a test of comprehension. Less obviously, matching on a test of reading words does not guarantee that the subjects will be matched on decoding skills. In standardized word reading tests in common use, the items are drawn from graded reading texts. This results in the inclusion of words chiefly from the "sight word" vocabulary. Given this manner of construction, such tests do not require decoding skills for successful performance. Two children with the same score may have achieved it by use of very different strategies. Thus we cannot know, without further exploration, what skills contributed to performance.

Since extended testing is often impractical, it is important to consider which measures give the most information. Should we use comprehension measures or decoding measures, or some combination of the two? Comprehension measures are often chosen because they are presumed to have greater ecological validity. While we can agree that the goal of reading is comprehension, we maintain that
comprehension measures are unsatisfactory for many purposes. For one thing, measures of reading comprehension are uninterpretable until we have also assessed comprehension of comparable materials presented in spoken form. Since competence in spoken language sets a ceiling on reading, we cannot infer, without further investigation, whether failure to comprehend some portion of printed text can be viewed as a reading problem and not a general language problem.

Decoding measures, on the other hand, are strictly measures of reading. Properly constructed, they assess the ability to read new words never encountered in print before, a skill that is surely necessary for mastery of reading. Moreover, they show a moderately high correlation with more global reading measures, including comprehension measures (Perfetti & Hogaboam, 1975; Shankweiler & Liberman, 1972; Tunmer & Nesdale, 1985). Research on reading disorder has been handicapped by the absence of an analytic test of decoding. The Decoding Skills Test (Richardson & DiBenedetto, 1986) may prove to fill a long-standing need. It includes both words and pseudowords that sample all the spelling patterns of English.

One of the purported advantages of the RL match design is that it permits one to exclude the possibility that differences in reading experience could account for criterion performance differences between good and poor readers. But, obviously, to equate individuals on some measure or measures of reading is not to equate them in reading experience. Unfortunately, no pure measures of reading experience are available. Experience in reading might be expected to differ for normal and poor readers. At the very least we might expect that disabled readers engage in more (remedial) text reading and less reading for pleasure. But, even for young children from homogeneous family backgrounds, no measures of experience with printed material are available. The potential for individual differences in reading experience increases as the age spread between subjects and their RL matched controls widens. In view of this, it seems wisest to avoid the assumption of equivalent experience, and instead to follow the more conservative course of restricting claims to comparisons based on measures of decoding performance.

6. ASSIGNING CAUSES: SOME CONCLUSIONS

Contemporary discussions of research on reading problems have correctly emphasized the importance of the decision an experimenter makes in choosing a control group. When inquiry focuses on reading behavior itself, the RL match comparison has been used to good effect in testing claims about isolable syndromes of reading disorder. The CA match, on the other hand, is often preferable in research aimed at isolating underlying processes. On the practical side, we offered some criteria for ruling out extraneous factors that may be spuriously associated with differences in reading ability, and we pointed to some of the difficulties in achieving suitably matched groups in either design.

Our defense of the CA match comparison has dealt chiefly with the claims of Bryant and his colleagues because, while agreeing with many of their conclusions about the nature of reading problems and cognizant of their important contributions to the research literature, we disagree with the implication that the matter of inferring cause and effect is chiefly a matter of selection of control group. We argued that, on the contrary, what enables us to infer causation is chiefly an articulated set of hypotheses about the reading process.

We began our defense of CA match comparisons by stating our basic theoretical assumptions that relate reading to the design features of the language apparatus. The manner in which reading is erected on preexisting language acquisitions led us to predict that the causes of reading disability would lie within the language domain. As predicted, seemingly normal school children who fail to make the expected
progress in learning to read were found to have language-related difficulties, including problems in metaphonological awareness and unusual limitations in verbal working memory. Since both of these problems are arguably grounded in phonology, one of our central concerns has been to determine if all the language-related difficulties evinced by poor readers might stem from a single deficit in processing phonological information.

The observation that poor readers have difficulties in correctly interpreting some spoken sentences seemed, at first, to threaten the phonological deficit account. However, in the context of our assumptions about the architecture of the language apparatus, and in view of the phonological nature of the working memory code, we argued that a processing deficit might explain this problem too. If so, it would obviate the need to attribute the comprehension difficulties of reading disabled children to a developmental lag in structural competence, over and above their well-attested deficiencies in phonological processing. In order to tease apart the alternatives, tasks were constructed that stress processing in varying degrees, with syntactic structure held constant. It was found using CA match designs that good and poor readers could be distinguished only under task conditions that were most consuming of working memory resources. Poor readers attained the same high level of performance as good readers in contexts that reduced processing demands. This pattern of positive and negative results defies explanation on the hypothesis that poor readers have less syntactic knowledge. Moreover, any attempt to invoke the reading experience hypothesis to explain the pattern of positive and negative results obtained in tests of sentence interpretation would seem ad hoc, since the same pattern of breakdown in spoken language understanding has been observed in other populations with severe memory limitations.

The view that reading experience inculcates syntactic competence also proved untenable in the light of language learnability considerations. These considerations provided us with a clear-cut example of how theory constrains the direction of causation. It was argued that spoken language experience plays a special role in syntactic development. In contrast, the input supplied by reading is too limited in kind and comes too late to promote basic grammar formation.

The conclusion that the problems of poor readers do not arise at the syntactic component of the language apparatus confirmed our expectation that the source of impairment is at a level of processing “below” the syntactic level; that is, at the level of phonological processing. Since this component feeds information either directly or indirectly to the syntactic component, and since poor readers are known to suffer from abnormal limitations in phonological processing, it is reasonable to infer that this is the source of their difficulties in interpreting both printed text and spoken sentences. The PLH alone seems capable of handling the whole network of empirical findings. Thus a unitary account of reading disability can still be entertained. The unitary account has been assailed also from the standpoint of variation in the symptom picture. We gave our reasons for considering that the arguments for subtypes are weak and inconclusive, a skepticism that is shared by others (Bryant & Bradley, 1985; Perfetti, 1985; Stanovich, 1986).

In sum, the available evidence does not sit well with the SLH or with any version of the idea that reading experience could supply the abilities that poor readers may critically lack, either in working memory or in syntactic competence. On the other hand, there is considerable empirical support for imputing a causal role to the phonological processing limitations of poor readers. The determination of causation rests on two critical ingredients: (i) a research strategy that makes use of both positive and negative results, and (ii) a set of guiding hypotheses about the reading process and its relation to an articulated model of spoken language processing. Since it is impossible to infer causation from correlation alone, both

Identifying the Causes of Reading Disability
Ingredients are necessary. And crucially, both are sufficient to counter the claim that causation can be inferred only when positive results are found with an RL match design. We argued that, taken together, positive and negative results based on CA match comparisons can be interpreted in terms of causation. In our view, the search for causation comes down to the search for the correct set of hypotheses. But, as with any adequate theory, it is not enough to explain the available facts: other consequences of the theory must be pursued, which, if confirmed, lend additional weight to inferences about cause and effect, even when these inferences are based on correlational data.  

ACKNOWLEDGMENT

The authors' research was supported in part by a Program Project Grant (HD-01994) to Haskins Laboratories from the National Institute of Child Health and Human Development. We are grateful to several colleagues and students who provided helpful comments on earlier drafts of this paper. We thank them all, but in particular we should mention Brian Byrne, Leonard Katz, Isabelle Y. Liberman and Diane Lillo-Martin.

REFERENCES


**FOOTNOTES**


†Also University of Connecticut

‡‡Also University of Rhode Island

1Subsequent research investigated implications of the view that the problems of the largest group of poor readers are phonologic in origin. Examination of the nature of errors on verbal memory tasks revealed that both good and poor readers use phonetic coding in preference to some other strategy, but poor readers do so less accurately (Brady, Mann, & Schmidt, 1987). Further studies have supported the phonologic deficit hypothesis. For example, poor readers were found to be less accurate than good readers in articulation of dictated words. Performance on the articulation task correlated with measures of verbal memory, but not with measures of nonverbal spatial memory (Brady, 1986). In addition, studies of object naming uncovered deficits that, again, reflect deficient phonological processing (Denckla & Rudel, 1976; Katz, 1985; Wolf, 1981). For instance, even when the subjects were matched in vocabulary knowledge, differences between the reader groups in access to the phonological properties of particular words in the mental lexicon were evident (Katz, 1985). In these investigations, we find, again, that a combination of positive and negative results has succeeded in further pinpointing the source of poor readers' deficits.

2To address the question of a processing limitation versus a structural deficit, it is also useful to examine the pattern of errors across reader groups. Elsewhere we propose that a processing limitation, and not a structural deficit, can be inferred if (i) there is a decrement in performance by poor readers, as compared to good readers, but (ii) both reader groups reveal a similar pattern of errors across sentence-types, and (iii) poor readers manifest a sufficiently high rate of correct response on a subset of the sentences.

3It is worth noting that the differences between good and poor readers are greatest on temporal term sentences in which there is a conflict between the order in which the events are mentioned and the order in which they should be executed (e.g., "Move the biggest car after you move the red
This is as expected, since this condition requires the subject to maintain information in memory while awaiting subsequent material.

The research strategy we have discussed has been followed in other studies of good and poor readers. As a further example, we mention an experiment by Fowler (in press), which tested several predictions of the PLH using a CA match design. The pattern of children's responses once again revealed clear-cut intercorrelations between measures of reading and various language-related measures: metathenological awareness, working memory, and spoken sentence understanding. Children were compared on two additional tasks. In the first, subjects performed grammaticality judgments, a task that is presumed to place minimal demands on working memory. A second task asked subjects to change some of the ungrammatical sentences from the judgment task, to make them grammatical. Clearly, correcting grammatical anomalies requires the ability to hold sentences in memory for some time. According to the PLH, both good and poor readers should do equally well on the grammaticality judgment task, but differences should occur on the correction task. This is exactly what was found. Reading ability was not significantly correlated with grammaticality judgments (a negative result) but it was correlated with the ability to make appropriate corrections (a positive result).

We readily concede that a case can be made, in principle, for the notion that reading experience could improve the efficiency of working memory for language, including spoken material. One possibility (suggested to us by Brian Byrne) is that orthographic imagery, which may be evoked in skilled readers during perception of spoken language, may reinforce the phonological representations evoked by the sound pattern. To the best of our knowledge, however, the critics of the CA match design, who contend that this design cannot establish a causal role for memory deficit in reading disorder, have not sought to argue their case along these lines, or, indeed, to offer any argument for the reading experience hypothesis as applied to working memory. In any case, we have seen that reading experience cannot explain the entire complex of language-related problems of the poor reader.

Without postulating innate constraints it would be impossible, in our view, to account for the uniformity of natural languages, or to explain the efficiency of language acquisition in the absence of relevant linguistic experience. Further, there is direct empirical support for several specific properties of Universal Grammar, viz. evidence of young children's adherence to explicit structural constraints despite the "poverty of the stimulus" (e.g., Crain & McKee, 1985; Crain & Nakayama, 1987).

When each error type was tabulated as a proportion of the total opportunities for an error of that type to occur, it was found that reversals of letter sequence occurred with an incidence of 8 percent among the dyslexic group and 7 percent among the school group. Similarly, reversals of letter orientation occurred with an incidence of 12 percent and 13 percent, respectively (Fischer et al., 1978).

It has been our experience that when statistical controls were used to remove the effects of IQ variation, differences between good and poor readers persisted (e.g., Brady et al., 1983; Fowler, in press). Similarly, when we recombined groups according to IQ, the differences obtained for groups based on reading ability disappeared (see Mark, Shankweiler, Liberman, & Fowler, 1977; Pratt & Brady, in press). The absence of major between-group differences on criterion measures attributable to IQ is surely a reflection of the subject selection criteria used, and, in our view, is a further vindication of the CA match control group.

The development in recent years of causal modeling techniques has potentially expanded the hypothesis-testing capacities of correlational studies (see Lomax, 1982). We have not attempted to evaluate these approaches here, but we note that promising applications to reading research have been made by Lundberg, Oloffson, and Wall (1982) and Tunmer and Nesdale (1985) in addressing the question of how phonological awareness may promote reading acquisition. Other approaches that combine longitudinal study and training have also been used to good purpose, but discussion of these, too, is beyond the scope of this paper.