A Computerized Diagnostic Test for Dyslexia

Matthew R. Sorenson
A Computerized Diagnostic Test For Dyslexia

Matthew R. Sorenson
WICAT Education Institute

The Department of Health, Education and Welfare has estimated (HEW 1969) that 15% of the nation's schoolchildren suffer from specific reading disability, or dyslexia. Most of them remain undiagnosed during the early years when the foundation skills and emotional outlook towards learning are established. Because reading is important for studying all subjects, dyslexia can be a particularly unfortunate learning disability.

There has been some disagreement on how to define dyslexia (Benton and Pearl 1978, Boder and Jarrico 1983, Kinsbourne 1982, 1983, Liberman 1983). If the definition is too broad, dyslexics may not receive the specialized attention they need. Yet if the definition is too narrow, many dyslexic kids will remain undetected and untreated.

The lack of agreement on even the basic concepts has been characteristic of dyslexia research. The field has suffered from poor tests, poor research design, poor models, overspecialization by related disciplines, and a general lack of linguistic sophistication.

Problems with reading tests

Educators and researchers have been using standard reading tests for evaluating dyslexics (appendix A). These tests are not diagnostic, since they don't indicate specific problem areas. They are normative and therefore uninterpretable for non-normal readers such as dyslexics. They do not address the necessary linguistic issues involved in a language deficit like dyslexia. And finally they do not yield the fine-grained information needed to set up a useful remediation program.

Problems in reading research

Vellutino's (1979) review of dyslexia research pointed out the inadequate and inconsistent methods that have been used in sampling, experimental controls, data analysis, and interpretation. The result is a muddle of basic concepts and findings.

'Although competent researchers have recently become more actively invested in the study of dyslexia in young children, most descriptions of the disorder are based either upon clinical studies and informal observations or upon loosely designed experimental contrasts that have typically yielded equivocal and conflicting results.' (Vellutino 1979:3)

Models of reading

One reason for and one result of this lack of productive research is the lack of any useful models of either reading or reading disability (Doehring et al 1981:26, Calfee 1982, Kinsbourne 1982:209). Without a theoretical model, experiments cannot be fruitfully constructed, nor experimental results fruitfully compared.
Too many specialists

There are many disciplines interested in dyslexia and each one has their own focus to pursue. In neuropsychology questions have arisen about the role of hemispheric dominance patterns (Satz 1976, Aaron 1982), localization of functions (Benson 1983), high level integration of information from the various association areas (Geschwind 1979), and the cytoarchitectural pathology of the cortex areas involved in reading (Galaburda 1983).

Vellutino (1979) and Liberman (1978, 1983) have argued convincingly that the behavioral problems observed among dyslexics can be best explained in terms of a verbal/decoding deficit.

That approach was been taken a step further by those who would explain low level symptoms by positing higher level cognitive and metalinguistic dysfunction (Orton Society Bulletin 1980, Pirozzolo and Wittrock 1981, Tunmer and Bowery 1984).

Studies on monozygotic and dizygotic twins, as well as family histories, have shown a significant genetic factor in the distribution of dyslexia (DeFries and Baker 1983, Decker and DeFries 1981, Finucci 1978).

Also, delays in the basic maturational processes have been blamed for dyslexia (Fletcher and Satz 1980, Elkind 1976, DeHirsch 1984). Although the longitudinal studies for investigating the role of maturation are difficult to pursue, they are necessary; both for a clear understanding of the disability itself, and for creating and evaluating remediation programs.

A narrow view of language

A general problem in this multidisciplinary landscape is that the redundant, recursive and abstract nature of language is unappreciated. The linguistic aspects of reading are usually taken to amount to a linear coding between sound and print. Even those who consider dyslexia to be primarily a language problem have been reluctant to look beyond word-level decoding. With very few exceptions (Vogel 1975, Kean 1984), the extent to which dyslexics have problems in morphology, syntax, lexical structures, and discourse processing have remained unexamined.

The sorts of cognitive tasks that are required for dealing with language have also been slighted in favor of memory and sensory processing. For example, the skills we have included under the label classification need more attention. The ability to class elements into abstract sets must be involved in establishing a phonology, in making sense of morphological alternations, in relating words into semantic and syntactic word classes, in paraphrasing sentences, and in associating language tasks with the appropriate situations.

Yet the dyslexia literature and the tests used by dyslexia researchers have not been sensitive to these issues.
Figure 1. WRAP's three dimensional measurement space.
It appears to the WICAT Education Institute that there is a clear need for a fine grained and comprehensive diagnostic test for dyslexia. We have undertaken the development of a computerized test to fill this need. For the moment, we will call it the WICAT Reading Abilities Profile (WRAP).

In establishing a conceptual basis for our computerized diagnostic test, we have tried to be as eclectic and empirical as possible. A framework that addresses the important issues and variables in the field is schematized in figure 1. Here we have a matrix defined by three dimensions: the level of language organization, the complexity of the cognitive task, and the input/output sensory modalities.

**Language level**

The language level that has commanded attention in past research has been the word. The literature has dealt with word-attack skills, word-level phonics, blending of phonemes into words, sight word recognition, and vocabulary. The instruments used have primarily been structured word lists. There was some early attention to letter reversals but it was found that the phenomenon was constrained by context in the word. The use of Berko's (1958) closure test of inflectional morphology has been increasing lately, but only in a few circles. Concern with higher levels of language organization has been largely limited to very gross measures of oral reading of passages.

Yet we know that reading involves a delicate coordination of all the various language levels, from **segments**, to **affixes**, to **words**, to **phrases**, to **sentences** and into **discourse structures** (cf. the variety of views in Spiro, Bruce and Brewer 1980). The size of the chunk that can be processed may be different for kids with different kinds of deficits.

**Cognitive tasks**

Along the second axis are the operations that must be applied to the different sized chunks of language.

The first thing a reader must do is direct **attention**. The units must be **identified**, then they can be **discriminated** from each other. Similarities and differences are compared so as to **classify** the units into functionally equivalent sets. They each must be **remembered**, and then put into or pulled out of strings, which demands **analysis**. And we need to find out how readers **learn** to manipulate and **create** these meaningful strings.

All of these cognitive tasks have been implicated as contributing to reading and must be involved in reading deficits. In dyslexia research, three have been emphasized: identification (recognition), memory, and analysis (decoding).
Figure 2. Types of items to test cognitive tasks.

<table>
<thead>
<tr>
<th>Attention</th>
<th>Identification</th>
<th>Classification</th>
<th>Discrimination</th>
<th>Memory</th>
<th>Analysis</th>
<th>Learning</th>
<th>Creating</th>
</tr>
</thead>
<tbody>
<tr>
<td>match pictures to words</td>
<td>recognize items embedded in a matrix</td>
<td>match two things the same or different?</td>
<td>match words to pictures</td>
<td>identify the part that changed</td>
<td>are two things the same or different?</td>
<td>closure / fill in the blank</td>
<td>grammaticality judgement</td>
</tr>
<tr>
<td>use two or more screens</td>
<td>use two or more screens</td>
<td>use two or more screens</td>
<td>use two or more screens</td>
<td>use two or more screens</td>
<td>use two or more screens</td>
<td>use two or more screens</td>
<td>use two or more screens</td>
</tr>
<tr>
<td>building whole from parts</td>
<td>building whole from parts</td>
<td>building whole from parts</td>
<td>building whole from parts</td>
<td>building whole from parts</td>
<td>building whole from parts</td>
<td>building whole from parts</td>
<td>building whole from parts</td>
</tr>
<tr>
<td>free description of pictures</td>
<td>free description of pictures</td>
<td>free description of pictures</td>
<td>free description of pictures</td>
<td>free description of pictures</td>
<td>free description of pictures</td>
<td>free description of pictures</td>
<td>free description of pictures</td>
</tr>
</tbody>
</table>

- contexts
- discourse
- sentences
- phrases
- syntactic
- words
- affixes
- lexicial
- letters
- sounds
- segmental
Figure 3. Traditional categories of reading subskills.
Item types

Of course, the problem in constructing a test is to isolate the variables to be measured. This can be done by using certain presentation formats for eliciting a particular cognitive task. Figure 2 shows the types of items that address the respective cognitive tasks. Then various sized language units can be inserted into that invariant presentation format. Using this 'slot-and-filler' approach will allow us to isolate the variables and will also make the test more simple and compact.

Both of these primary dimensions are hierarchical. The simpler components contribute to the more complex. Given this relation between variables, we would expect that the effects of a deficit in one area will spread to others.

If the spread is from complex to simple, then the disrupted processes are most likely of the top-down sort. A case of this sort arise if a kid cannot interpret words without knowing the meaning of the sentence that contains them. On the other hand, a spread from the simple into the complex would indicate problems in bottom-up processing (cf. Haber and Haber 1981, Rummelhart 1977).

Sensory modality

The third dimension of our conceptual framework would allow us to investigate which sensory modalities were either a problem or a crutch for a dyslexic. I will defer discussion on this point because the actual test will not implement this dimension right away. The kinds of technology and interface devices that will allow manipulation of how subjects receive the items and how they input responses to the computer are going to be emerging in the near future.

Within this framework we can identify where traditional categories of reading subskills would lie, as in figure 3. For example, the problem of perceptual reversals would be tested by items that fall in the slot defined at the *letter* level of language organization and involving the *identification* level of cognitive operations. Word attack skills would be located by crossing the *word* level with *analysis*. Storytelling would show up in the slot where *texts* and *creating* intersect, etc.

Items and subtests relevant to these areas are being collected from the best of the existing tests. However, these tests address only small parts of the model. So where good items do not exist, they are being written. The range of variables examined by existing tests is shown in figure 4.
Figure 4. Existing tests used in dyslexia research.
WRAP logic

Once we have collected or created items that can test the combinations of variables we want, they will be computerized. Item sequencing will be governed by intelligent and computer-adaptive diagnostic logic. The test will proceed to search for areas in the measurement space that seem to present problems to the student. Problems are indicated by slow response latency, inaccuracy, and inconsistency. All of these can be monitored and manipulated by the computerized test.

Computer-adaptive tests, like the WRAP, can adapt to the subjects' responses, administering only those items that are appropriate in terms of content and difficulty. If they miss an item, an easier one follows; if they pass an item, a harder one is used. This process continues until the correct level of ability is estimated. Besides being more efficient and accurate, computer-adaptive testing avoids the problems of giving every item to every student.

There is more to a test items than just right or wrong. By using informative distractors, we can find out not only where the dyslexic has problems, but what compensatory strategies are being used to deal with them. That is, if a kid consistently keys in on sound patterns when confused (or on word shape, or on semantic similarities) that information can help to establish a subtype.

Where there is evidence of a problem, the focus of the test would shift to that particular subspace. Related items would be chosen and a search would begin for the boundaries of the problem. The test would end at the point where the problem area, or areas, were well defined.

Every time the test is administered, we could add the result to a cumulative data base, which would tell us about which configurations of problems can be expected, which variables are correlated, which items are the most predictive, and which diagnostic strategies are the most productive. As a result, then, of its experience over time the test could provide the basis for its own refinement. The refinements themselves would take the form of a structured body of inferences, called an 'inference engine', which would take past results and present subject responses to make judgements about what items to present next.

Such a combination of tests, diagnostic logic, user-adaptivity, inference engine, and growing knowledge base constitutes an expert system (Hayes-Roth et al 1983, Winston and Pendergast 1984). Another essential feature of expert systems that we need to include in the WRAP is a flexible output interface. Different people are going to require different explanations for what has been determined by the test.

The results could be expressed in the language of reading pedagogy for teachers, in psycholinguistic jargon for researchers, or in very general terms for the student or parent. In any of these cases, the test should also be able to respond to user queries about procedures that were used and conclusions that were reached.

An overview of the structure of the WRAP is shown in figure 5.
users or consumers

explainer
for queries & justification

inputs
cognitive operation
language level
sensory modality
difficulty level

tester or researcher

output
inaccuracies
inconsistencies
latencies
strategy

student or dyslexic

inference engine
from findings
to hypotheses

response record
array

knowledge base
deficit patterns
grammar

knowledge engineer

Figure 5. Overview of WRAP
Subtypes

There is a consensus in the field that there are several subtypes of dyslexia that have specific deficit patterns and demand specific kinds of remediation. However, there is no consensus on what the subtypes are, and classification systems abound (Boder and Jarrico 1983, Mattis et al 1978, Lyon 1983, Malatesha and Dougan 1982, Rosenthal et al 1982, Doehring et al 1981, Satz and Morris 1981).

Because of the multiplicity of interests and biases, these typologies are built upon different sets of assumptions and findings. It is difficult to compare or synthesize them. What is needed to make sense out of the many schemes is consistent test results within a comprehensive model. This is the contribution that the WRAP can make.

Using the conceptual framework of the WRAP, some comparison can be made of several of the most prominent subtype classifications, as in figure 6.

Although admittedly there is some oversimplification in this comparison, we can see that in spite of the plethora of labels, there is some overlap. The first group of subtypes concerns problems of visually processing shapes of words. The second grouping clusters around tasks of analyzing phonological sequences. The third cluster seems to involve problems in meaning, naming, and interpreting sentences. A fourth grouping, which should be plotted on the third dimension of our measurement space, implicates problems of the hand and mouth, in the output process.

These four subtypes of reading deficit are very idealized. Most kids have been found to suffer from a mixture of problems, and there are likely more subtypes than we have been able to measure.

A complicating factor in subtyping is the issue of compensation. It has been noted that kids with sight problems tend to rely on phonics, and vice versa. This factor can only be controlled for if we have a sensitive enough test, apply it early in the acquisition process, and follow a stable population longitudinally with consistent measurements.

Progress

In pursuing the goal of constructing the WRAP test we have reviewed the research literature and consulted with the top experts in the country. The result of this groundwork is the conceptual framework we have described. The meat on this skeleton is the pool of subtests and items that we have either created or culled from existing tests. We are now on the verge of creating a prototype that will be used for trial tests in the following year, with subjects, both dyslexic and normal, provided by local school districts.
Figure 6. Comparison of several subtyping schemes.
Initially the WRAP test will be implemented on WICAT computers, using our established test development software. Later, we will make updated versions of the test available on IBM and Apple microcomputers.

In the future, there are several enhancements and extensions that we would like to make. First, using the knowledge of what dyslexic kids do wrong and what compensatory strategies they use, we can construct an effective computerized remediation program (cf. Clancey 1984). Second, we can exploit new developments in voice recognition and synthesis, new manual interface devices, and graphics and animation to present items in a variety of formats that test the participation of each sensory mode in the reading process. Third, the diagnostic and remediation programs need to be integrated into the normal reading curriculum. The materials used in classes can be parsed and coded to allow for automatic construction of new test items that will indicate problems and progress in that particular reading program.
References


