EDWARD FRY
Rutgers University

PRESIDENTIAL ADDRESS

TOOLS FOR READING RESEARCHERS

Reading research is in an epistemological crises; we don't know what we don't know. Researchers are in a situation not unlike the allegorical blind men in India. Upon discovering an elephant one grabbed ahold of his tail and maintained the elephant was like a rope, another put his arms around a leg and maintained the elephant was like a tree, while a third ran a hand along his side and assured his fellows that the elephant resembled a wall.

Among reading researchers and model builders we have the rope group, the tree trunk schools, and the wall advocates. Staunchly clinging to their own data, they have developed even finer methods of measuring the rope's index of flexibility, the wall's height, and the tree trunk's mean diameter, based upon randomly selected samples according to a theoretical model, refined by analysis of variance, and including a standard error of measurement significant to $p <$ .001. Broad-minded, interdisciplinary researchers have correlated leg diameter and tail diameter and found them to be .73.

This paper is a plea for a little more data before we build an elephant; it is a request that we stop building models until we get more data. I am aware of the argument that models generate and synthesize isolated research studies, and I am also aware that by asking model builders to stop, I am not going to change their behavior one whit. Man's need for closure or perhaps his ego drive for achieving *the* $E = MC^2$ of reading is too strong. However, perhaps I can encourage those who are rightly skeptical of existing models, those who seek hard data from a variety of sources, and those who try to piece together smaller bits of data into firm concepts, to open up the system into wider areas with an increasing refinement of techniques.

I might ask the model builders to look at the recent history of human learning in psychology. What ever happened to the "laws of learning" that we used to teach a few decades ago. They were built on too narrow a data base and toppled over in the light of new information.

With that concept in mind I will discuss some research tools at hand. These tools are not totally new, but by reviewing some of them, it might stimulate thinking for alternate forms of knowledge generation. Most of us wear several hats: we disseminate knowledge about reading through teaching and writing, we are practitioners trying to improve reading acquisition skills, and we either produce or are closely related to, the production and interpretation of new knowledge by research.

Eye Movement

Eye movement research is certainly not new. Javel started using it around in 1879, and Tinker almost exhausted its use in the 1930's and 1940's. But it clearly fits my definition of a tool, and it still has relevance. It is a nice tool. It can easily be observed by anyone (and hence has a lot of surface validity), and it can be objectified, quantified, refined, and subjected to powerful computer analysis.

On the Javel level you can have a class of undergraduates pair themselves off: one student reads and the other student watches the eye movements of the reader to bring meaning to terms like fixation, saccadic movement, return sweep, and regression. On the Tinker level you can purchase eye movement recording instruments which give you graphic records of eye movements and enable you to generate data showing such things as a growth pattern between first grade and college in decreasing duration of fixation, decreasing number of regressions, and increased span of recognition.

Techniques for observing eye movements have progressed from lying a mirror alongside of a book so someone standing behind the reader can observe eye movement, to electro-oculogram devices which can sense eye movement by treating the eye ball as a very delicate flashlight battery which emits an electrical potential. Shifts in the + /- pole can be sensed by placing an electrode on the nose and forehead. These and improved camera devices can study eye movement in greater detail permitting such insights as Rayner's (1975) observation about rather long fixations on infrequent words such as "rendezvous" and "cache". Thus, we can see that duration of fixation is perhaps related to word difficulty. Since computers can be programmed to summate fixation durations, and since, through other research we know that word difficulty is related to readability, it can be seen that readability research in the near future could be enhanced by an electro-oculogram plus a small computer. Since readability is related to, or in some aspects is the obverse side of comprehension, we can see how eye movement might be a measure of comprehension.

In traditional comprehension research and evaluation, questions following the reading of a passage are used, but this raises a problem: "Is the student failing to comprehend the passage or failing to comprehend the question?" The cloze technique avoids this problem. Another way of avoiding it, however, might be with eye movement or *fixation duration* scores. This is what I mean by using research tools — developing new tools and the facility to use them, to pry further into such simple and central problems as "What is involved in comprehension?" or "Why is passage A harder to read than passage B?" Reading efficiency or rate might better be improved or at least studied by looking at eye movements.

Perception

The perception of letters, words, phrases and larger segments has occupied a great deal of research time. As a series of tools, I find it rather boring; it has been around so long and parts of it seem so little changed. In 1908 Huey discussed a study by Messmer in which he found that the longer letters which projected above the line are usually the dominating ones. In 1968 Gibson and her co-workers were

interested in distinctive features and:

> employed a same-different judgment of two letters exposed simultaneously by projecting on a small screen. If the subject thought they were identical, he pressed one button; if he thought they were different, he pressed another. His latency, the time he took to respond, was the index used as well as the few errors that were made.
>
> (Gibson and Levin, 1975)

Huey and other turn-of-the-century researchers were also interested in deleting parts of words. They found, for example, that the top half deletion made reading harder than bottom half deletion and that the first half of a word was more important than the second half. Deletions can also be accomplished by a variety of techniques, such as placing various screens over words or parts of passages. They offer us another technique of determining word difficulty or familiarity.

The tachistoscope, that overworked workhouse of verbal learning research, has enjoyed an amazing longevity. Dodge was using it at the turn of the century, and it is still a staple in experiments that require a short exposure and standard viewing conditions.

Some modern techniques of looking at perception related to the reading task involve embedding words or phrases in different matricies or in running words together to see if the subject can detect the word boundaries (Klein and Klein, 1972).

Eye Voice Span

The span between where the voice is and where the eye is when reading aloud can be easily observed by simply sliding a card over the page while the subject is reading aloud and counting the number of words he continues saying. In 1897 Quartz noted that the eye voice span was longer at the beginning of a line than toward the end with averages varying from 7.4 words at the beginning to 3.8 words at the end of a line. He also noted that the span decreased to zero when an unfamiliar word was encountered. Somewhat more interesting, he noted a close correlation between the increase of eye-voice span and an increase in rate of reading. He felt that "a considerable distance between eye and voice is a condition of intelligent and intelligible reading" (Huey, 1908).

Eye voice span research is still with us. Geyer (1968) looked at it in terms of elapse time rather than number of words, but modern laboratory equipment might be able to provide a whole new range of eye voice span data by continuously linking eye fixation point with voice. Perhaps greater refinements can link eye movement to subvocalization.

Vocalization

There is a lot of overlap between speech research and reading research and many of the tools are the same. To a large extent we share a common history. Reading researchers have long been interested in subvocalization (or inner speech) while reading. It can be reported subjectively by nearly everyone, but trying to objectify it is difficult. In 1900 Curtis studied covert oral activity by mechanically sensing larynx movement; modern researchers tend to use an elec-

tro-myograph or (EMG), a recording device for minute electrical changes associated with muscle activation. Sensors are placed on the throat or mouth areas and slight electrical changes can be noted and associated with silent reading activity long before muscle movement can be detected.

Hardyck (1968) used an EMG as a biofeedback device to lower subvocalization. This lowered EMG did not result in rate increase, but did result in reports of less fatigue with reading for periods of one to three hours. The lowered fatigue seems to be related to breathing pattern; a higher subvocalizer breathes more like a person talking (Hardyck and Petrenovich, 1969). However, lowering subvocalization may lower comprehension in poor readers.

An interesting tool for the study of speech is the voice spectrograph. This device, which is also called the sound or speech spectrograph, is essentially a visual display of speech or other sounds with time moving along the horizontal axis, pitch along the vertical axis, and intensity by degrees of darkness (Prestigiacomo, 1957). Words can literally be shown graphically. Among other interesting things, it shows no temporal pause between phonemes or, in many instances, between words. Spoken speech is a continuous flow. It calls into question what we are really talking about when we talk about a "word." In terms of sound production, a word is not a unit. Perhaps it is a writing convention. Perhaps it derives its definition from a cognitive base. At least, if you are interested in demonstrating what a word really is, you will get precious little help from a spectrograph, and this little bit of data could force some changes in traditional thinking about the reading process, or perhaps even something practical like, how we set about teaching children to read. If words are not inherent in speech, it could be that words are not the most efficient unit of reading instruction.

Electroencephalography

If we are interested in minute muscle changes associated with reading, the next logical area of progress is to study brain activity. EEG's, or electroencephalograms, for all their ability to detect gross activity like epileptic seizures, or generalized states of consciousness like sleep, seem to be a bit too crude for specific reading research. But knowing about this reminds us that a good bit of psycho-physiological research is ultimately related to thinking and thinking to reading. Detecting electrical activity associated with the brain was done with animals in the 1870's and external detection of brain waves was started in the 1920's. Today it is a very active field. Reading researchers have most recently shown an interest in it seeking an illusive phantom called dyslexia in an attempt to explain reading failure. The relationship between reading failure and brain wave irregularities has not been shown in most instances. However, this does not mean that in the future neurological research does not have promise for understanding both reading failure and more importantly, the normal reading process.

Heart Rate

Wark (1971) reported results of a small but unique study utilizing a digital cardiotachometer, a relatively simple, off-the-shelf medical device, that almost instantly reports heart rate from electrodes attached to the arms. A base heart

rate is obtained by having the subjects rest with their eyes closed. When asked to read, heart rate decreased. When they were asked to answer questions, the heart rate went above base rate. Wark concluded, "heart rate seems to be implicated in the reading process in a way not previously suspected."

Should continued investigations show relationships between heart rate and aspects of reading it is possible that some reading clinicians may be interested in affecting reading performance by biofeedback training. Physiological investigations have shown a remarkable ability on the part of yogis to reduce heart rate from 63 to 24 beats per minute. Some biofeedback training has worked but has shown less dramatic ability to affect heart rate (Brown, 1972).

Biofeedback

A number of the devices that we have just been discussing such as cardiotachometers, EEG and EMG have been used in the recent biofeedback movement. Biofeedback is the process of enhancing information about body functions so that they can be put under conscious control or conditioned or trained to more desirable functioning. Early enthusiasm and promise have been somewhat dampened by inexactitude, the discovery that often cheap instruments are unsatisfactory, and the unpleasant fact that much of it is plain hard work. The results are less than miraculous, but I doubt that it will disappear from either medical or learning research. We are prisoners of our own bodies and are at times dominated by our unconscious. The desire for greater freedom and the perennial intrigue of the mind-body problem is not too distant from the subject of this paper: epistemology.

SOFTWARE TOOLS

A research tool need not be a piece of hardware employing electronics; a good paper and pencil test can be every bit as important as a heart beat counter or a cardiotachometer.

Ability Tests

In reading, researchers and practitioners often seek to find either special or generalized abilities that predict reading achievement. Predicting future ability to read often involves some type of perceptual discrimination. I welcome this in the research area, but I am horrified at the erroneous judgments made by practitioners, such as first grade teachers who withhold reading instruction based on a readiness instrument that only correlates with success at about .50. We must all be cautious when training teachers not to confuse research instruments with those that should be used in the classroom.

For the moment, let's clearly separate measuring achievement from measuring ability. In measuring abilities, it is surprising to me that there hasn't been more work in a special ability called "reading." For many years we have had tests of special abilities in music, special abilities in mathematics, and other fields, yet we almost seem to ignore the possibility that reading ability might be special or might follow the normal distribution curve. We seem to accept the

school boards' assumption that everybody should read at the same level or at least that the normal distribution curve shouldn't apply below the mean.

Achievement Tests

Thorndike and Hagen (1969) state that future historians will probably call the current period, the last twenty years, the period of the increased utilization of the achievement battery. Some type of reading achievement test is used at least several times in most student school careers in most school districts. Even with the millions of tests being given, we are almost at a primitive stage in content, quality. By simply looking at several major reading achievement batteries, it is readily apparent that there is little agreement on content or types of skills tested. Davis (1971) attempted to statistically analyze reading comprehension items and had great difficulty separating many factors; vocabulary stood out from all the rest, but even the terminology used to describe the factors is somewhat ambiguous. At most, he had eight categories, but some of those were barely independent.

The current increased use of test management systems that incorporate criterion referenced tests into a prescriptive system for classrooms should tend to heighten our concern for factors that might or might not be relevant to measuring reading success, and to factors that should be included into the teaching of reading. I recently had an interesting argument with an undergraduate class that insisted that the reading field knew what factors should be tested and taught. I couldn't convince them that we were so dumb.

The strength of the criterion test movement is that they measure skills to an absolute mastery criterion, independent of a norm reference group. Their weakness is that there is not strong research basis for determining what skills to measure.

An interesting blend of achievement and criterion referenced tests can be seen in the Basic Word Vocabulary Test completed by Dupuy for the National Center for Health Statistics (1975). Dupuy is a trained statistician, but being unacquainted with the reading field, he boldly set out to measure the vocabulary size of the non-institutional population of the United States. After overcoming initial hurdles, he succeeded quite well. The first hurdle was to find out, "What is a 'word'?" If, for example, we look at "run, runs, running, and ran," are they really one word, two words, or four words? Carroll in the American Heritage study would call them four words. A second hurdle is the so-called semantic count; is "run" to the store the same "run" as a "run" on the stock market?

Dupuy solved the problem by going to the unabridged dictionary and taking a 1% sample of main entries, omitting derivations, foreign, technical and compound words, then testing only the first meaning. Despite the fact that Carroll found 87,000 words and unabridged dictionaries claim over 200,000 entries, Dupuy found only 12,300 basic words. He made a test of 1% or 123 items and administered it to over 3,000 Virginia students who also had taken other normed vocabulary tests. He was thus able to anchor his test to U.S. norms and have a criterion score which was the total number of basic words known by multiplying raw score by 100. One interesting interpretation of his findings is that fifth graders can read and know the meaning of about one-quarter of the basic

words, tenth graders know about half the words, and Ph.D.'s know about three-quarters.

An interesting implication of this is the rather solid evidence that half the vocabulary growth takes place between grades 5 and 10. Some of us have been so nearly brainwashed about the importance of language development in early childhood, that this information might give us some heart about what is being done, and what needs to be done, in the middle school years. Last, but not least, Dupuy's test is interesting in its extreme range; few other instruments can measure growth in any reading skill from third grade to beyond Ph.D. with both norms and a clearly defined criterion. This tool has implications for both longitudinal school achievement and language development in a neglected range. It also gives a Vocabulary Development Quotient which is similar to an Intelligence Quotient.

Before we leave the achievement tests, let me suggest that I suspect that there is some valuable data hidden in the computer memories of most major test companies. A majority of schools now pay for computer scoring, and they program pages of output on individual, class, school and district score reports. They have data on every item, age of children, sex, school, sub-test scores in other subjects, IQ, time of year, and often repeated testing in later years. Studies could investigate many possibilities. To cite a few of the many possible examples:

1. maybe a few math skills can predict vocabulary growth years later;
2. maybe certain reading items indicate future failure or underachievement;
3. maybe there are regional differences in skill development;
4. maybe some reading skills don't make a bit of difference to mature comprehension;
5. maybe some types of items favor boys or girls.

Research Design

Finally, I would like to suggest that research design and techniques are very definitely tools. Controlling for *training time* or *order of presentation* can be every bit as important in trying to gain knowledge as using the right instrument.

Subjective judgment is not dead. Oftentimes, it is the only way to get at attitude or reading ease. Singer (1975) successfully had college students judge readability by comparing unknown passages with passages of known difficulty (the SEER Technique). Carver (1976) utilized carefully selected and trained judges to compare test passages against standard passages in the Rauding Scale, and at Rutgers, we have been attempting to verify the Kernel Distance Theory by having students judge sentence pairs of equal vocabulary and length, but differing syntax.

Our investigations should vary in size. Much information about brain functioning has and will continue to be gained by studying individual aphasia victims, or looking at the longitudinal academic growth of a few children. Clinical studies, especially, make good use of controlled or structured observations which are partly exemplified by the use of a checklist.

Large scale studies can also give us information that is not obtainable by any other means. The National Assessment of Educational Progress is valuable and should be expanded. We need more data on the overall reading ability of the nation.

The USOE First Grade Studies in the mid-1960's were the largest scale reading experiment ever. They helped us to see what was going on in different instruction methods, and they slowed down the claims of zealots who had some sure-cure methods. Something like the First Grade studies should be repeated about every decade. Right now we need large-scale comparisons of test management systems like the Wisconsin Design with book controlled systems like the basal readers, and studies of individualization versus traditional grouping.

The First Grade Studies also raised an interesting question, yet unanswered. Why did a reading method like the Language Experience Approach, which has almost no specific skills emphasis, succeed in teaching reading as well as other highly skills-oriented programs like the so-called linguistic or phonics approach?

Large-scale studies might not be important for model builders, but they are certainly important for superintendents, curriculum committees, and classroom teachers trying to decide what to do next year.

They also have important implications for society as a whole. For example, one important finding of the National Assessment that confirmed smaller studies, was the high correlation between reading achievement and socio-economic status. Our newspapers which castigate reading methods usually fail to point out the solid correlation between poor reading achievement and such SES factors as parent education, number of parents in the home, or family income.

Miscellany

The computer is too big to be called a tool; it is a whole garage full of tools. It has a myriad of uses besides its standard uses for large-scale data handling and high level statistical analyses. Computer simulation has been an influence on the way we analyze tasks and try to reproduce them. MIT is now closing in on a computer that can read orally from any printed page. Computer control of the teaching process as in CAI offer not just a way to supplement instruction which is good, but a way of controlling teaching variables so that one variable can be changed while all others are held constant. On-line computer monitoring of sensitive instrumentation promises to discover things that the human eye or the human mind cannot detect. Its use in information search and retrieval are still in an infancy period.

A list of research tools would not be complete without mentioning research summaries and document collections. Those who don't know history are condemned to repeat it. The *NRC Yearbook* summaries by Bleismer and others and the *Reading Research Quarterly* annual summaries are very useful to current and future researchers. Incidentally, the total *William S. Gray Collection of Reading* documents has just become available on microfiche from the Hofstra University. The ERIC collection is another valuable resource tool, as are Kling's (1971) literature summaries and Corder's (1971) summaries of reading research for the USOE Targeted Research. NRC members might be interested to learn that our first eight yearbooks, most of which are out-of-print, are now available on microfiche from ERIC. Our old papers have received a kind of technological posterity.

We face an exciting future in reading research. Undoubtedly that research will uncover new tools. In the meantime, we already possess a number of useful tools — tools that properly refined and sensibly used will serve us well in uncovering

new secrets, verifying old intuitions, and following good hunches about this remarkable process called reading.

Finale

And now the time has come for you to see what research tools have wrought. This year's program at the National Reading Conference is one of the very best. Jaap Tuinman, your President for next year, has labored long and hard to put it together, and most of you have spent countless hours at the typewriter, at the computer center, and chasing down subjects and references to put together your presentations. Most of the research reported here is not funded, or rather, it is funded by the sweat and overtime of the individual investigator. Even when it is funded, there is usually a measure of extra effort which can only be ruled by professional interest, not dollars. So, for all those long hours you have put in to improve the reading of children and adults in America, I for one, would like to say, "Thank you very much."

REFERENCES

Bio-Feedback Systems (catalogue). Boulder, Colorado, 1975.

Brown, B. *New mind, new body: Biofeedback.* New York: Harper Row, 1972.

Carver, R. Measuring prose difficulty using the rauding scale. *Reading Research Quarterly,* 1975-76, *11,* 660-685.

Corder, R. *The information base for reading: A critical review of the information base for current assumptions regarding the status of instruction and achievement in reading in the United States.* Educational Testing Service, Berkeley, California, 1971. (ERIC Document Reproduction Service No. ED 054 922)

Davis, F. B. Psychometric research on comprehension in reading. In M. King (Principal Investigator), *The literature of research in reading with emphasis on models.* New Brunswick, N.J.: Rutgers University, 1971.

Dupuy, H. J. *Basic word vocabulary test.* Highland Park, N.J.: Dreier Educational Systems, 1975.

Geyer, J. J. Perceptual systems in reading: The prediction of a temporal eye-voice span constant. In H. K. Smith (Ed.), *Perception and reading.* Newark, Del.: International Reading Association, 1968.

Gibson, E. J. & Levin, H. *The psychology of reading.* Cambridge, Mass.: The MIT Press, 1975.

Hardyck, C. D. & Petrinovich, L. F. Treatment of sub-vocal speech during reading, *Journal of Reading.* 1969, *12,* 361-368.

Hardyck, C. D. Petrinovich, L. F., & Ellsworth, D. W. Feedback of speech muscle activity during silent reading: Rapid extinction. *Science,* 1966, *159,* 1467-1468.

Huey, E. B. *The psychology and pedagogy of reading.* Cambridge, Mass.: MIT Press, 1968. (Originally published, 1908.)

Klein, G. A. & Klein, H. A. Word identification as a function of contextual information. *American Journal of Psychology,* 1973, *82,* 399-406.

Kling, M. (Principal Investigator), Davis, F. (Ed.). *The literature of research on reading with emphasis on models.* New Brunswick, N.J.: Rutgers University, 1971. (Available from ERIC on microfiche or in soft cover book from IRIS Corp., Box 372, East Brunswick, N.J. 08816.)

McGuigan, F. J., Keller, B., & Stanton, E. Covert language responses during silent reading. *Journal of Educational Psychology,* 1964, *55,* 339-343.

Prestigiacomo, A. J. Plastic tape sound spectrograph. *Journal of Speech and Hearing,* 1957, *22,* 321-327.

Raynor, K. *Eye movements in reading: Historical trends and current developments.* Unpublished Manuscript, University of Rochester, 1975.

Singer, H. "The SEER Technique: A non-computational procedure for quickly estimating readability level." *Journal of Reading Behavior,* 1975, *7,* 254-267.

Sound Spectrograph, Series 700. Somerville, N.J.: Voice Identification Laboratories.

Thorndike, R. L. & Hagen, E. *Measurement and evaluation in psychology and education.* New York: John Wiley and Sons, 1969.

Tinker, M. The study of eye movements in reading. *Psychological Bulletin,* 1958, *55,* 125-232.
Wark, D. M., Tostenrud, D., & Nelms, G. Heart rate and reading. *Yearbook, National Reading Conference,* Clemson, N.C., 1972.
William S. Gray Research Collection (9,000 documents and 18,000 index cards; 300,000 pages). Microfiche Edition, Alvina Treut Burrows Institute, Hofstra University, Hempstead, New York, 1976.