DOCUMENT RESUME

ED 026 208 RE 001 306

By-Baker, Eva L.

Methodological Considerations for Future Readiness Research.

Pub Date Apr 68

Note-12p.; Paper presented at International Reading Association conference, Boston, Mass., April 24-27, 1968.

EDRS Price MF-$0.25 HC-$0.70

Descriptors-Behavioral Objectives, Predictive Measurement, Prereading Experience, Reading Achievement, Reading Programs, *Reading Readiness, *Reading Readiness Tests, Reading Skills

The following criticism of current practices for determining reading readiness are discussed: (1) the variables measured by readiness tests are in many cases unrelated to instruction, (2) while the variables measured are correlated with reading achievement, this association does not attribute causality, and (3) the criterion selected for measuring reading achievement is often an achievement test written to maximize individual differences rather than to measure success in a given program. The subcomponents or prerequisites of a particular reading program need to be developed through research. A criterion-referenced test could be developed which would measure the skills that should be mastered before formal instruction is begun. (RJ)

SYMPOSIOM I
EVALUATING READINESS FOR DEVELOPMENTAL LANGUAGE LEARNING:

Implications for Further Research and Practice

Methodological Considerations for Future Readiness Research

Eva L. Baker

Southwest Regional Laboratory for Educational

Research and Development

Paper presented at the Thirteenth Annual Convention of the International Reading Association. Boston, Mass. April, 1968

I should like first to agree generally with Dr. MacGinitie's initial comments. He has identified many areas of apparent deficiency and has done so from the vantage point of an esteemed expert in the field. There is, of course, a measure of presumptiveness in my coming to you today to comment on the area of readiness, for I am not an expert in that domain, nor particularly esteemed.

However, I would like very much to raise some questions, in indisputable innocence, about the readiness concept which seems to spawn so much research and attract so much attention. Perhaps, like Blake's lost little boy, I'll hit on a question to which there is no answer but ought to be. As one who has come late into the field of reading from instruction in general, I find the entire concept of readiness perplexing. Perhaps I am disturbed because I don't understand what the readiness people are after. Perhaps I am disturbed because I do understand. In either case I feel queasy about

anything which commands the research time and interest of a great body of qualified people. What does one get from readiness and what can one do with it?

Readiness testing, interpreted simply, seems to mean that a test measuring certain factors can be given to children before they learn to read. The factors measured on this test are presumably clear enough that a teacher can decide who is ready to learn what. The selection of variables to be measured on readiness tests seems to be primarily analytic rather than experimental-- intuitive rather than empirical. From what I can observe, the validation of the readiness tests typically involves giving children readiness instruments, measuring the children after they have completed a certain period of reading instruction and finally, working out regression equations for using their performance on the first set of tests to predict the criterion of reading achievement. It has been pointed out that it is inappropriate to measure the criterion variable first, then to work backwards to abilities which the good achiever possesses. I certainly agree.

However, I also feel that other currently popular methods of investigating readiness are not without some problems. The question of readiness predictors is troublesome because I am not entirely sure what we will have when we have identified good predictors. Among the sources of my confusion, the following is paramount: because a child scores well on a test which later correlates with his reading achievement, can we say that high scores of other children at a given moment in time mean that those children are then <u>ready</u> for reading? The question can't be adequately answered and seems

to be directly related to the fact that there are vastly different classes of predictors. These different categories should lead to different and perhaps mutally exclusive decisions by the teacher.

The first category seems conceptually related to the idea of "readiness". Predictors of this type can be largely maturational in nature. Can a child profit from instruction regarding how to hold a pencil, or is his coordination such that this is impossible? The difficulty with this type of predictor is the manner in which judgments about it are promulgated. Uninstructed children are observed and that which they do at a given age is taken as the norm. No attention is given to the fact that these behaviors may be amenable to instruction. These are status predictors which are not easily varied in a conventional educational context. Within reading and not limited to readiness, other types of status variables are the favored factors to investigate. I suppose this is so because the data from them is often "hard" rather than soft: convenient for the investigator to obtain in a form amenable to statistical analysis. For example, research has been conducted which attempts to show relationships between the child's ability to read and laterality, birth order, or parental occupation. It is easy to classify a child with respect to any of these variables, he is the first born child or he is not.

There are other readiness factors which don't appear to be so patently unrelated to instruction as birth order but still can be labeled as status variables. Factors like perceptual skills and knowledge of common concepts, seem more relevant to the reading criterion. Unfortunately, their appearance may be misleading. My suspicions regarding this type of predictor are under-

scored by something in readiness literature, often readiness factors are discussed as traits. A trait can be interpreted as a rather stable attribute. Perhaps I am reading more into these remarks than what is there, but identifying a trait which predicts reading achievement doesn't help the classroom teacher very much, unless the teacher can exclude children who do or do not have particular traits. In addition, any correlation between reading achievement and presumed prerequisites, such as some perception tests, which persists regardless of the instructional strategy used to teach reading becomes very suspect. If these factors are impervious to instruction, then their relationship to reading achievement is almost beside the point. As with an intelligence test, if one can't do much to alter the child's general ability, why worry about it in the first place? It will develop and operate with or without us.

Aside from the sense of unity it gives one to understand many of the variables which contribute to reading, there is very little practical benefit to be derived from such knowledge. However, assuming that among these factors there are some reasonably related to the skills of reading, the procedure of constructing readiness tests themselves, so there will be high reliability and concomitant variation substantial enough to detect correlations, may be defeating the practical use of these variables. This procedure is completely consistent with measurement of traits, but the point is we are not measuring, I think, qualities which we should like to be extremely stable regardless of treatment. We should want to be able to make a child "ready", and this desire is demonstrated by the host of readiness activities provided in most reading programs. The teachers, because they have difficulty

in interpreting the tests, fall back on a single measure of "readiness" and think of it as a rather immutable entity, like strong ankles or good looks. This is of course incompatible with the idea of readiness learning experiences. If helpful readiness tests are to be devised, they must measure adequately and representatively certain skills found to be prerequisite to reading, prerequisites which the teacher can clearly identify and deal with instructionally.

The terms adequate and representative were deliberately chosen rather than reliable and valid. Tests which measure prerequisite skills should measure the attainment of these skills. The child's score should be interpreted in terms of the skill itself, like 70 percent correct rather than the 70th percentile. If a child can perform adequately on reasonable tests of auditory discrimination, it seems beside the point to lengthen the test and include more difficult items in order to increase the variation between children. If the test is easy for the child, this simply indicates that he has mastered the skill and instruction on this particular skill is not required. The test should be totally criterion-referenced. This type of test will not be likely to provide the teacher with enormous variation among students as a typical norm-referenced test might. But even here, more variation will exist on a criterion-referenced test than teachers can take account of instructionally. The criterion-referenced framework which I am proposing for the readiness tests themselves can be applied as well to the criterion test in most readiness validation studies. The criterion is most frequently scores obtained on a respected standardized test. More than ever, people involved in the development of instruction are recognizing that

standardized tests should not function to evaluate the acquisition of a set of desired skills. Standardized tests come from the great and thoroughly unimpeachible tradition of psychometry. Standardized tests allow us to do things, which, in a behavioral field where any data are shaky, enable us to feel scientific. We can construct these tests in a way to maximize the variation which we will get between respondents. This enables us to pick up reliability coefficients that are in the respectable range. We can talk about comparing the scores of sub-groups of children. All of these things are fine when they star in their journal report world and refrain from contaminating the real world. When standardized tests in any field, and particularly in reading, are used as dependent variables in a research study, questions should be asked.

Reading research so far has suffered from most of the flaws common to other areas of research in education, for example, teacher evaluation. We see almost monthly published studies where two teachers use Method A, two teachers use Method B and their children are measured on some standardized test. It is clear that a study of this type probably ought not to be done at all, or if conducted, designed in a markedly different fashion. The teachers undoubtedly provide a source of variation so great that it is easy to understand the rash of non-significant differences obtained. And if we find differences, they can most reliably be attributed to the teachers rather than the methods. Aside from the analysis problems which usually seem to accompany these studies, for instance, the use of the number of children rather than the number of classrooms to contribute to the analysis, the study is still almost hopeless and the hopelessness resides in the choice

of the dependent variable -- a standardized test.

Because standardized tests are not constructed to assess a unique program, but skills of widely divergent children, their use to evaluate any program tends to be inadequate. As it has been pointed out, if the test is constructed with a heavy emphasis on irregularly spelled, esoteric words, then sight approaches will show up better than phonics or other approaches. An added difficulty, of course, is that the teacher in the classroom doesn't know which particular approach influenced the construction of the standardized test which he must use. This is not important when the teacher's role is custodial. But, when he is committed to bringing about observable changes in the skills of his students, he ought to be aware of the inner workings of the standardized test. What I am suggesting should be used to evaluate a reading program is again the criterion-referenced test. This is a test in which all items are representive measures of the stated objectives of a program. If two programs with disparate objectives are used and a comparison is desired, then two criterion-referenced tests should be used. There are additional benefits to be accrued from forcing oneself into a criterion-referenced framework. First, this requires that a priori judgments be made regarding the "goodness" of the objectives to be measured. Should comprehension be measured by pictures? Should it be measured by multiple-choice questions? The statement of objectives permits the teacher full knowledge of what is being measured. She does not become awed in the presence of a test whose title seems to bear no relation to what the children are doing. Secondly, the teacher can obtain valuable instructional cues and might seek to specifically bring about particular changes in the learner's

behavior rather than rely too heavily on the author of the reading series for help. This brings us along the way to making a reading program definable and replicable in terms of getting good results in different classrooms, and emphasizes the conscious choices to be made by the teacher. Shifting into a criterion-referenced model is not easy, and perhaps, therefore, it won't be quickly done, but it is the real way to measure what is happening in reading.

Among other things, what must be occuring to most of you is that what I am suggesting won't work in the present readiness research model. If we have criterion-referenced tests of reading prerequisites and criterion-referenced tests for dependent variables, then neither test will have ranges within it to enable us to carry on every meaningful prediction studies. This is the intended implication. Much of the difficulty in deciding if high correlations indicate that the child is ready to read or to engage in readiness activities at a particular time has to do with the attribution of causality made from associational data. I do not say this is the purpose of the writers of readiness tests. They are skilled experimenters and have clear understandings of the limitations of predictive research. However, the results of these tests and validations attempts are often misinterpreted to the general detriment of many children. It is misinterpretations of gross readiness data which generates the all-or-nothing readiness fallacy.

Obviously, I have some problems with readiness research. I am suggesting that there are areas of it which are inappropriate. It studies relationships which are correlated, not caused; many of the predictors used are those least amenable to modification by instruction and potentially least useful

to know about; the dependent variable is often an achievement test written to maximize individual differences rather than to study the effects of reading program. There do seem to be alternatives without completely abandoning the idea of "readiness" assessment of young children. The most important typ of research to be done is the determination of what the sub-components of reading actually are. This would seem to be best accomplished in an experimental model. While experts have often mentioned this, I think it must be emphasized. An inspection of the literature reveals far too few studies which have proceeded experimentally. The way to go about this is direct but laborious. First, we should establish specific objectives for reading. We should make these explicit, operational and understandable. They should be stated in terms of student performance. Then we should write test items which are designed to measure these objectives representatively rather than worrying about item difficulties, esoteric validities, and other psychometric verities. We will, even at this preliminary point, encounter difficulty since most people won't agree about which are the particular skill which adequately define reading. The entire population of potential skills ideally should be described specifically and items devised for them. In this way members of different reading ideologies could attent to the tests measuring <u>their</u> conceptions of the objectives of reading.

Once the potential dependent variables are established, the second phase of research can be conducted. Experimental studies investigating manipulable variables should be pursued. It would appear reasonable that research in reading should be concerned with the effects of things which can be included or withheld from the curriculum. If auditory discrimination

practice yields high reading scores, we can do something with that information, that is, include auditory discrimination practice in the classroom. If, on the other hand, blond children are found to read better than brunettes, we really can't do anything about it, Lady Clairol not withstanding.

Another point regarding the nature of the manipulable variable must not be overlooked. We must ascertain if we in fact have a treatment. An experiment of this type, for example, should not be done: Give Prerequisite A Treatment to one group of children, withhold Prerequisite Treatment A from another group of children, and measure the children on a appropriate criterion-referenced test of reading. The flaw in the study is that no one bothered to determine if the treatment in Prerequisite A resulted in acquisition of Skill A. This is equivalent to letting children in the experimental group work through materials which probably won't be effective in the first place. Appropriate research, dependent upon the development of effective treatments, is a mind-bending task. No one really wants to go through the agony of developing effective treatments which turn out to bear no tangible relationship to the criterion of reading. But it is research of this type which will finally help us decide what to do in the classroom.

There is the added difficulty of determining what potential reading prerequisites ought to be considered as variables in an experimental approach. Directions might be provided through the use of task analysis procedures as explicated by Gagné. As he has suggested, an obvious pattern of successful responses should emerge if learning a particular skill is truly dependent upon the aqusition of a prerequisite. By analyzing the success patterns of children learning behavioral skills, one can see which skills seem to be

crucial to the reading act. If, for example, the child only succeeds at blending when he has previously learned separate letter sounds, one might select an effective program which teaches separate letter sounds as an appropriate treatment variable. Of course, this would not be a comparison of children who had the sound and letter skill and children who had not, but a random assignment of children to receive instruction designed to promote the skill -- real manipulation of the variable. Besides developing adequate criterion-referenced dependent variables and potent treatments, experimenters will have to become willing to use the expertise they verbalize when conducting actual experiments in the classroom. For one thing, classroom unit sampling will have to be used, with sufficient numbers to account for the variability of individual teachers. This means that school district administrators will have to do something which may be disruptive to their on-going programs but which will demonstrate their affective commitment to improving reading in the schools. A little cooperation of this type will do more to elevate children's reading proficiency than punitive teachers' meetings scheduled following publication of the results of the last standardized testing session. After programmatic research of this type, where the actual effects of prerequisites are determined on specific skills, meaningful and useful readiness tests could then be constructed. Many experts are discouraged because teachers don't want to administer and interpret long, tedious tests. Perhaps one solution to this problem might be a modification of the use of item sampling or unmatched testing as it is sometimes called. Not every child responds to every item, but a general estimate of the group's performance is obtainable. Since the ultimate utility of readiness test depends upon

teachers' using them, then the designing of arduous but admirable tests may not result in their widespread use. And again, since the tests are to be a means for the teacher to decide what kinds of instruction are appropriate for children rather than ends in themselves, the tests should function to the specifications for the development of adequate readiness programs for children. These programs should demonstrably increase the child's performance on skills previously shown to be critical to reading.

This paper may sound negative to you. The correlational model is no good, standardized tests are questionable, the predictors don't seem very usable. I have purposely taken extreme, somewhat pontifical positions, in the hopes that you might listen critically to what I am saying.

All of us feel that instruction in reading can be improved. If we are concerned with upgrading the educational opportunities for all our children, then we can no longer wait around for the mood to strike us to do a little hard experimentation. If money is a problem, we had better not even wait around for that. In lieu of a funded, programmatic project to define the prerequisites of reading and how to teach them, individuals directly involved in the teaching of reading as teachers or supervisors should insist on conducting rough-and-dirty studies of the problems they are having with readiness or reading, even if the results generalize only to their own classroom. If any one wished to do such a study, I am sure there are many people in the educational establishment who would be glad to advise them. Tangible findings could be immediately translated into practice. Instructional decisions could be based on evidence as well as analysis. Then we might become really ready to teach reading.