

2001

## **A Treatment-Based Reading Assessment.**

Gary Joseph Duhon Jr  
*Louisiana State University and Agricultural & Mechanical College*

Follow this and additional works at: [https://digitalcommons.lsu.edu/gradschool\\_disstheses](https://digitalcommons.lsu.edu/gradschool_disstheses)

---

### **Recommended Citation**

Duhon, Gary Joseph Jr, "A Treatment-Based Reading Assessment." (2001). *LSU Historical Dissertations and Theses*. 378.  
[https://digitalcommons.lsu.edu/gradschool\\_disstheses/378](https://digitalcommons.lsu.edu/gradschool_disstheses/378)

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact [gradetd@lsu.edu](mailto:gradetd@lsu.edu).

## **INFORMATION TO USERS**

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**ProQuest Information and Learning  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
800-521-0600**

**UMI<sup>®</sup>**



# **A TREATMENT BASED READING ASSESSMENT**

**A Dissertation**

**Submitted to the Graduate Faculty of the  
Louisiana State University and  
Agricultural and Mechanical College  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy**

**in**

**The Department of Psychology**

**by**

**Gary J. Duhon, Jr.  
B.A., Nicholls State University, 1992  
M.A., Louisiana State University, 1999  
December 2001**

UMI Number: 3031760

UMI<sup>®</sup>

---

UMI Microform 3031760

Copyright 2002 by Bell & Howell Information and Learning Company.

All rights reserved. This microform edition is protected against  
unauthorized copying under Title 17, United States Code.

---

Bell & Howell Information and Learning Company  
300 North Zeeb Road  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

## TABLE OF CONTENTS

<b>ABSTRACT .....</b>	<b>iii</b>
<b>CHAPTER</b>	
<b>1 INTRODUCTION.....</b>	<b>1</b>
<b>2 REVIEW OF LITERATURE.....</b>	<b>5</b>
The Treatment Validity of Reading Assessment.....	5
Intervention Development and CBM.....	15
Developing Interventions for Academic Skill Deficits..	17
Low Inference Assessment Linked to the IH.....	27
Evaluating the Treatment Validity of Assessments.....	29
<b>3 METHOD.....</b>	<b>32</b>
Subject Selection and Setting.....	32
Grouping Criteria.....	32
Materials.....	34
Response Measurement and Observer Agreement.....	35
Treatment Packages.....	36
<b>4 PROCEDURE.....</b>	<b>42</b>
Overview.....	42
Group Placement.....	42
Assessment-Response to Treatment.....	43
Assessment Validation.....	47
Experimental Design: Extended Analysis.....	48
Procedural Integrity.....	51
Comparison to an Alternative Assessment Procedure...	51
<b>5 RESULTS .....</b>	<b>54</b>
Assessment and Extended Analysis.....	55
Extended Analysis with Alternative Treatments.....	60
Comparison to Instructional Placement Standards.....	62
<b>6 DISCUSSION.....</b>	<b>64</b>
Summary of Findings.....	64
Implications.....	65
Limitations.....	68
Future Research.....	69
<b>REFERENCES .....</b>	<b>72</b>
<b>APPENDIXES .....</b>	<b>78</b>
<b>VITA .....</b>	<b>83</b>

## **Abstract**

Each year a large number of students are identified as having problems in reading. The methodology for the identification of reading problems has been established and in practice for several years, however, the tools necessary for the identification of remediation procedures appropriate for the individual student lack this level of refinement. This study is an attempt at refining current assessment practices for the purpose of identifying remediation procedures that are effective on the level of the individual student. A conceptual model of learning (instructional hierarchy) was combined with a treatment based functional assessment methodology, resulting in a behavior analytic approach to reading assessment designed to identify remediation procedures that have a known level of effectiveness for an individual student. Eighteen first, second and third graders were assessed across a range of fluencies resulting in treatment recommendations based on individual responses to treatment components. These recommendations were then evaluated in an extended validation, which was designed to ascertain the level of effectiveness of these recommendations over baseline and against each other. Results indicate that this behavior analytic approach to the assessment and treatment along with the incorporation of the instructional hierarchy produced valid treatment recommendations. When these results were compared to alternative forms of treatment recommendations the behavior analytic approach produced superior results.

## INTRODUCTION

Difficulty with reading is one of the most frequently cited academic problems in general education (Shapiro, 1996). Data from the National Assessment of Educational Progress revealed that 44 percent of fourth graders nationwide are performing below basic reading levels (Daonahue, Finnegan, Lutkus, Allen, & Campbell, 2001). Reading instruction in the U.S, as currently practiced, for many students has not proven to be an effective way to establish adequate reading skills. Due to the importance of reading across all aspects of education and the growing demand for higher literacy rates in an increasingly technical society, the impact of reading difficulties on individuals and society as a whole becomes greater every year.

For the students reading at or above basic reading levels, the instruction that they are receiving seems adequate, but for the 44 percent mentioned above the instruction they are being exposed to is not effectively producing meaningful reading achievement. This does not mean that the reading instruction these students are receiving is faulty or ineffective as a whole, but that for those students who are not reading, instruction is inappropriate or misplaced. Valid treatments are those that produce a favorable outcome for the person to whom they are applied (Hayes, Nelson & Jarrett, 1987). If we consider reading instruction to be a treatment designed to improve reading skills then for some students the treatment is not valid.

Why current instructional procedures are effective for some students and not others is an issue that is and will continue to be debated for many years and is beyond the scope of this paper. The issues to be addressed here center around providing effective services for those children who fail to read under normal circumstances in



~~general education. What can be done for those students if the reading instruction they are receiving is ineffective? Due to the importance of this problem there is an obvious need for the development of procedures, which can not only assess reading skill, but do so in a manner that leads to the development and delivery of effective interventions for remediating reading deficits.~~

For alleviating reading deficits a plethora of effective interventions are available. The problem for school-based practitioners is not identifying effective treatments, but identifying which treatment will be effective for the individual reader of concern. This can be accomplished by employing assessment techniques that can discriminate between treatments based on their effectiveness (i.e., their level of treatment validity) at an individual level. However, current assessment practices do not base assessment decisions on how the subject may respond to treatment. Because they are not treatment based, the assessment techniques tell us very little about what type of reaction to a given treatment a student might have. However, techniques exist that can be used to evaluate how subjects may respond to specific types of treatment or even classes of treatments.

Hypothesis testing is an assessment procedure that uses brief exposures to treatments that are designed to predict response to extended forms of similar treatments.

Hypothesis testing is generally based on a particular conceptual model used to guide treatment selection and to make it possible to generate alternative forms of treatment should the need arise. The idea of using brief exposures to hypothesis based treatments as an assessment model builds on similar work with severe behavior problems (Repp, Felce, & Barton, 1988; Sprague & Horner, 1992; Neef & Iwata,

1994) and other social behaviors (Karsh, Repp, Dahlquist, Carol & Munk, 1995)

where a person's initial response to a brief treatment was predictive of the response to extended treatment. With hypothesis testing, hypotheses about the function of the behavior of concern are evaluated. Based on the results of the assessment a hypothesis regarding the function of aberrant is evaluated by applying a treatment based upon the hypothesized function. If the treatment is effective then the hypothesis is accepted and the treatment continued. If, however, the treatment is ineffective the hypothesis is not supported and another hypothesis is recommended.

### Intervention and Reading Deficits

The remediation of reading deficits usually begins with an assessment of student performance designed to evaluate the problem and guide selection of alternative and hopefully effective instructional procedures. The alternative instructional procedure is an attempt to identify a procedure or procedures, which will provide the student with instruction that would significantly improve reading performance and eliminate the need for further remediation. In other words, based on an assessment that identified poor performance or poor growth, a new and presumably effective instructional procedure is forwarded which is designed to improve reading skill. This instructional procedure is employed and evaluated for effectiveness. If the problem remains, another new instructional procedure is used until some form of satisfactory outcome is achieved.

The purpose of assessment in this context is to verify and specify student skills as well as make decisions about student problems (Salvia and Ysseldyke, 1995) for the purpose of intervention. However, not all assessment techniques are capable of

providing adequate information for attaining the goal of valid treatment development. Several assessment techniques exist that have the capacity to classify student performance and evaluate response to an intervention, but lack the capacity to evaluate this classification with reference to multiple treatments (Daly & Martens, 1994). However, it should be noted that these assessment procedures were never designed to provide specific treatment recommendations, just the recommendation that an alternative approach should be attempted. The result of this type of assessment is a pick and hope procedure. Pick an intervention and hope it works and if it doesn't work another is selected. This leads many practitioners with little help in determining *which treatment* to use for low performing readers.

To assist in identifying which treatment to use, the type of assessment employed should be linked to the particular outcome the assessment is designed to produce (Gresham, 1992). In other words, if the purpose of the assessment is to make treatment recommendations, the assessment should be valid for that purpose. With reference to the problem of identifying treatments for low achieving readers an assessment that has treatment validity would not only classify and evaluate progress of student reading, but also provide recommendations with regard to which interventions would be most appropriate. Assessment procedures could be employed that are based on empirically validated procedures designed specifically to provide treatment recommendations derived from a student's responses to treatment. An assessment procedure, like this, that leads to better treatment outcomes is said to have treatment validity (Gresham, 1992).

## **REVIEW OF LITERATURE**

### **The Treatment Validity of Reading Assessment**

In the next section, the treatment validity of various types of assessments (eg. norm-referenced and criterion-referenced assessments) will be briefly reviewed. The purpose of this section is to highlight particular outcome each type of assessment is designed and has been validated to generate and the relevance of these instruments for treatment development.

**Norm-referenced assessments.** Norm-referenced assessments, or more specifically, nationally standardized norm-referenced assessments are one of the most commonly used methods for evaluating academic performance. These assessments contain items that are designed to sample specific academic skills within a content area. Student performance is then compared to that of a norm group and a relative standing of student performance, as compared to the norm, is advanced. Results from norm-referenced assessments are often used to determine eligibility for programs such as special education.

Several norm-referenced assessments exist and range from intellectual assessments to standard achievement test. One of the most common NRA is the Wechsler Intelligence Scale for Children - III. It is widely used for the purpose of classification and determining eligibility for services in school or hospital settings. Other examples of NRA are the Peabody Individual Achievement Teat-Revised (PIAT-R) and the Kaufman Assessment Battery for Children. Both of which are achievement tests that provide information regarding relative standing of a child compared to a normative group.

~~Norm-referenced assessments provide information about student performance;~~

however, the results are only relevant to the degree that the skill assessed actually matches the skills required by the student's curriculum. Given that a norm referenced assessment is one test that samples a finite number of skills and is intended for use across many areas with many different curricula, the degree of overlap between the skills assessed and a curriculum could range from very high to very low. The degree to which a test and a curriculum do overlap is considered "curriculum-test overlap". Poor test overlap between the curriculum and the test make it difficult to determine the relation of the assessment to actual curriculum achievement. So norm-referenced assessments that have very little or low "curriculum-test overlap" do not provide much information about whether the student has the skills needed to perform what the curriculum is requiring.

The outcomes NRA are designed to generate are best used to classify students with respect to the norm group (Witt, Elliott, Daly, Gresham & Kramer, 1998). This ultimately is only useful for determining relative standing in reference to the norm group. From this relative standing it can be determined, based on a pre-established criterion, if the student is in need of remediation or is functioning appropriately, but decisions about what to do if remediation is needed is not evident. Given the possibility that the test is not relevant to the curriculum, it is difficult to see how a NRA with low "curriculum-test overlap" test would contribute to treatment development. Even with regard to NRA with high "curriculum-test overlap", outcomes about relative standing, lack the information necessary to predict student

~~responses to treatment and therefore have little validity with reference to treatment development.~~

Methods for identifying treatments based on norm-referenced assessments have been evaluated by researchers in the form of aptitude-treatment interactions (Arter & Jenkins, 1979). Aptitude-treatment interactions (ATI) are based on inferred mental constructions called “aptitudes” (Witt, et al, 1998). These aptitudes are hypothetical constructs that are believed to represent a particular student’s “learning style”. By identifying a student’s aptitude or learning style, instruction can be modified to take advantage of the aptitude and improved academic behavior should result (Shapiro, 1996). The purpose of assessment in this case is not only to evaluate student performance against the norm, but to identify a student’s aptitude for use in intervention development.

For example, in evaluating a student with a reading deficit it may be determined that a preference for visual over auditory learning style exist. A remedial reading program, for this student, would then consist of instruction based more on visual presentation of instruction rather than auditory. Despite the logical ATI approach there is not widespread support for aptitude by treatment interactions (Arter & Jenkins, 1979; Gresham & Witt, 1997; McMacmann, & Barnett, 1997). In fact several studies that examined the effectiveness of interventions based on ATIs discovered that in many cases the treatment based on the ATI was least effective in improving academic skills (Ayres & Cooley, 1986; Good, Vollmer, Creek, Katz and Chowdri, 1993)

Criterion-referenced assessments. Similarly, criterion referenced assessments are useful for comparisons. However, the comparison for the CRA is not a norm group, but an absolute standard or benchmark for a specific skill (Hambleton, 1982). CRA consist of a comparison of student performance in a particular skill area against an absolute standard that represents acquisition of that skill (Shapiro, 1996). The results of this type of assessment are evaluated as either meeting or not meeting the established criterion (Witt, et al, 1998). For example, a school district may establish a benchmark that all second grade students, by the end of the year, will be able to calculate 30 addition problems, sums to 18, with 90% accuracy. Students would be assessed on this skill and performance is compared to this standard to determine if the student assessed has met criterion or did not. Results from criterion-referenced assessments can be used to evaluate individual student performance or educational programs for groups of students by determining what areas or specific skills are below the criterion (Hambleton, 1982). By evaluating student performance against a criterion, strengths and weaknesses can be identified. In other words, CRA identify what a student can and can't do.

Probably the most familiar and widely used type of criterion-referenced assessment is the series of Brigance inventories. Brigance has developed a series of widely used CRA for preschoolers (Brigance Inventory for Early Development), elementary aged students (Brigance Diagnostic Inventory of Basic Skills) and for secondary age students (Brigance Diagnostic Inventory of Essential Skills). These assessments cover a wide range of sub-skills associated with specific behavioral objectives for each sub-skill (Shapiro, 1996).

~~Criterion-referenced assessments can assess specific skill at the level of the~~  
individual, but just like with NRT above, the possibility exists for poor “curriculum-test overlap”. With CRT designed for national use, the extent to which the curriculum the student is being assessed is represented by the samples on the criterion-reference assessment may be high or low. Low “curriculum-test overlap” makes it difficult for criterion-referenced assessments to detect student changes within the curriculum. These assessments may be able to provide information about what a student can and can’t do, but it may have little relation to what they are being asked to do. However, increasing the amount of “curriculum-test overlap” would increase the relevance of the assessment to the student in a particular curriculum and increase the usefulness of the assessment results for evaluating student performance in relation to a criterion and/or evaluating the effectiveness of an instructional program. However, the primary problem with criterion-referenced assessments are establishing suitable mastery criterion (Witt, et al, 1998). In other words, what skills should be assessed and what level of accuracy constitutes meeting the criterion. This problem is compounded in that often times mastery criterion is established through the use of logical and not empirical means (Shapiro, 1996).

Criterion-referenced assessments are designed to be useful in providing information about what a student can and cannot do. Where there is high overlap between the curriculum and the assessment, this information can be used as a starting point for helping a professional make a decision about the presence and severity of student difficulties. It can also help identify where to begin future assessments and what skills need intervention (Shapiro, 1996). However, this assessment provides no



information for determining which treatments to use for remediating academic areas of concern (Shapiro, 1996). They were designed and validated for providing information about current student performance in reference to the curriculum, but it cannot provide information on what the curriculum expects of the student or what to do if the student needs remediation.

Direct assessment procedures. Direct assessment procedures are based on the assumption that “one should test what one teaches” (Shapiro, 1996 p. 16). Although several different models of direct assessment exist one component that is consistent for each model is the use of the instructional curriculum as the content of the assessment. As compared to other less direct forms of assessment this reduces concern about poor “curriculum-test overlap” because what is tested is what has been taught. This also avoids any inferences about what the student can and can’t do in reference to performance in the curriculum.

Various models exist for assessing student performance based on the curriculum. Each has been designed for specific purposes with regard to the evaluation of academic performance, but they all consist of direct observation and recording of student performance in the curriculum as a basis for gathering information used to make instructional decisions (Deno, 1987).

Blankenship (1985) developed a model called “curriculum-based assessment” where student performance is evaluated based on curricular objectives. Assessments of one or more of these curricular objectives are repeated over several days to establish a baseline of student performance. Regular assessments using these curricular objectives are conducted to determine if progress toward the objectives is

being made. Progress within CBA is based on a pre-established criterion. The primary purpose for administering assessments with CBA is to provide the teacher with information for instructional planning (Shinn, 1989). By providing frequent information regarding the current performance of a student or class, the teacher has the opportunity to either continue or alter current instructional procedure. This ability to quickly and frequently evaluate the effectiveness of instructional procedures makes it possible for more timely modification to be made in unproductive instructional procedures. Within this model intervention consist of altering the current instructional procedures based on inadequate advancement in the curriculum as evaluated by an established criterion.

Another model of CBA was developed by Gickling and colleagues (Gickling & Havertape, 1981; Gickling & Rosenfield, 1995), which evaluates student skills in terms of “knowns” and “unknowns”. Designed primarily for math and reading responses are evaluated based on an accuracy model and a ratio can be calculated for a given level of material. The proportion of “knowns” and “unknowns” within the instructional content is then manipulated to keep students within a range that is presumed to best promote progress. It is believed that once appropriate instructional placement has been determined through the evaluation of “knowns” and “unknowns” student progress is assured (Shinn, 1989). In other words this model attempts to control the difficulty level of instructional delivery in order to maximize the potential for success. Interventions derived from this model result in an alteration in student placement in the curriculum, but not a modification in instructional procedure.

**“Curriculum-based evaluation” is a third model, which was developed by Howell, Fox and Moorhead (1993). This model is wider in scope than the previous models of curricular assessment. It is not only designed to evaluate basic skills within a curriculum, but it is also designed to address various subskills such as reading comprehension, decoding, mathematics, written communication, and social skills (Shinn, 1989). This model employs skill probes along with task analysis, direct observation as well as other evaluation tools to provide detailed information about student performance within a skill level and suggestions for intervention programming (Howell et al. 1993). In addition, decision rules are provided for making determinations about when to modify or alter interventions. Within CBE suggested intervention programming is based on presumed functions of performance deficits derived from assessment outcomes.**

**The final model discussed is the most extensively researched model. Various referred to as curriculum-based assessment or curriculum-based measurement, this model was originally developed at the University of Minnesota and based on research published in “data-based program modifications” (Deno & Mirkin, 1977). This model is designed mainly as a progress monitoring system, which uses frequent and repeated administration of curricular probes designed to assess the student’s progress within the curriculum (Deno, 1987). The skills assessed within this model are not necessarily the skills being instructed, but are looked upon as “vital signs” measuring improvement and acquisition of the curriculum content (Shapiro, 1996). Included within the decision making model of CBM are instructional placement standards that are designed to help appropriately place students based on**

skill level. These standards are based on the premise that there is a range of fluency characterized by a level of difficulty that best promotes instructional growth. In other words students should achieve the maximum amount of academic growth when instruction is delivered using curricular materials from this instructional level. Care is taken not to provide materials that are too difficult (frustrational level materials) or too easy (mastery level materials) because although growth can occur at these levels the rate of growth will be significantly less than that of the instructional level. Based on this premise students are evaluated and placed in the curricular level that corresponds to the instructional level. Because this model is primarily designed for progress monitoring and not intervention development (Shapiro, 1996), ensuring that the curricular materials are of appropriate difficulty is the initial intervening step. However, if that does not produce effective change the next step is to alter the instructional procedure and monitor progress.

These procedures have good psychometric properties (Shinn, 1989) and have been demonstrated as good predictor of other reading behaviors such as comprehension (Shapiro, 1996). However, Curriculum based assessment procedures were developed to identify where to place a child instructionally within a curriculum and then to assist in evaluating interventions (Shinn & Hubbard, 1992). Aside from CBE these procedures do not lead directly to recommendations for how to instruct the student, only where to instruct and when to make changes in instructional procedures. CBM assesses fluent responding, but is not tied to any instructional model (Daly et al., 1996).

CBE is, however, tied to an instructional model. Based on subject responses to probes a “functionally assumed cause” (Howell, Fox & Morehead, 1993, p 228) is advanced. Based on this assumed cause treatment recommendations are identified and implemented. Although CBE does provide treatment recommendations they are not based on a response to treatment, but the identification of these “functionally assumed causes” of deficient behavior. In other words, based on observed response characteristics a theory about why this deficit exists is raised and a treatment based on the cause is implemented. This is an advance over most other forms of assessment in that the results do provide treatment recommendations, however these recommendations are based on assumptions about cause and therefore only represent a guess as to how treatment will affect behavior.

Although they are not designed to identify interventions, CBM procedures do provide information about student performance relevant to instructional placement and intervention evaluation. Interventions are evaluated using CBM through the use of continuous progress monitoring. Based on established benchmarks or goals (generally a goal for year end performance) student progress within a given instructional treatment is evaluated. If a student is consistent with attaining specified goals, then whatever intervention is currently being employed, is continued. However, if a student is not achieving these goals the instructional treatment is evaluated and either modified or changed (Shinn 1989). If the modified treatment is then effective the program is continued until the next evaluation.

## **Intervention Development and CBM**

Intervention development using the CBM model is derived from instructional placement standards. Instructional placement standards are employed to ensure a student is being instructed at a level, which produces optimal growth (Mirkin, Fuchs & Deno, 1982). Fluency rates in reading are divided into instructional, mastery and frustrational levels. These fluency rates are designed to correspond to difficulty levels of reading materials for students across different grade levels. Mirkin, Fuchs and Deno (1982) have suggested that passages with mid-range difficulty (instructional level) produce the steepest slopes for performance growth. Frustrational level materials are said to be too difficult to produce meaningful growth, while the material at the mastery level will produce less than optimal growth because these materials are too easy for the student and result in a ceiling effect (Mirkin et al., 1982). For example, a particular student's reading ability may establish his instructional level within the 4<sup>th</sup> grade curriculum. Frustrational level would then be at and above the 5<sup>th</sup> grade curricular materials and mastery at or below the 3<sup>rd</sup> grade.

To evaluate a student's "reading level" fluency measures of oral reading would be compared to the instructional placement standards. For any particular student this is accomplished by measuring his or her fluency level using content controlled grade level reading passages and evaluating the child's performance against the instructional placement criteria which have been established in the literature (e.g. Deno & Mirkin, 1977; Fuchs & Deno, 1982; Lovitt & Hansen, 1976; Shinn, 1989; Starlin, 1982).

Although instructional placement standards are widely utilized, several issues remain unresolved concerning their use. The main source of concern is the lack of

empirical data to support their origins. For example, the earliest and possibly the most cited source for the instructional placement standards is Deno & Mirkin (1977). This referenced text is a manual entitled "Data-Based Program Modification: A Manual". The manual is a description of the implementation of CBM procedures and not an empirical study. In fact, Deno and Mirkin cite Starlin & Starlin (1974) as the actual source for the standards. It should also be noted that this citation is inaccurate. In attempting to locate this publication contact with the first author (Clay Starlin) revealed that the actual citation is Starlin & Starlin (1973). The Starlin and Starlin (1973) reference a one-page pamphlet describing the instructional placement standards that appear in Deno & Mirkin (1977). Starlin and Starlin provide no information as to where these numbers actually came from in this pamphlet. Further contact with the first author (personal communication with Clay Starlin 2001) revealed that the origins of the standards can be traced back further to Starlin (1970), an unpublished dissertation, which indicated that the instructional placement standards were arrived at by dividing the performance of a fluent adult reader by half. There was no evidence that these standards were empirically validated for children.

Other instructional placement standards are also suspect because of a lack of empirical support for their selection. The "standards" advanced Lovitt & Hansen (1976) were arrived at based upon what the authors "decided" (Lovitt & Hansen, 1976 p. 351) the standards should be. From this it appears that instructional placement standards derive from what amounts to clinical judgements and not from studies, which show that children who are placed at a particular curricular level will perform better than children who are placed at other curricular levels. This calls into question

~~the validity of these widely used instructional placement strategies for curricular~~  
selection. Given the manner in which they are used, establishing their validity for  
placement in the curriculum is a necessary but not sufficient prerequisite in efforts to  
establish treatment validity.

The assessment procedures described above including curriculum-based  
measurement were not designed to provide treatment recommendations. In fact, they  
may be of questionable value in making instructional placement decisions. However,  
because curriculum based assessments have proven to be valid and reliable measures  
of academic performance (Shapiro, 1996) they may be useful as part of assessment  
procedures designed to provide treatment recommendations. In fact many advocates  
of CBM would not suggest that it has ever been claimed that CBM should be used to  
derive treatment recommendations (Shapiro, 1996). Generally, practitioners have  
been urged to use CBM to establish an instructional level and then to use some  
validated intervention combined with formative evaluation. If one intervention is not  
effective, then it is modified or replaced until an effective one has been implemented.  
This amounts to trial and error treatment development, which can be a time consuming  
and costly process.

#### Developing Interventions for Academic Skill Deficits

Attempts to identify alternative instructional procedures or treatments to  
remediate reading problems can be addressed along a continuum from indirect to  
direct procedures (Shapiro, 1996). Indirect procedures attempt to increase academic  
performance by improving the underlying academic learning processes or inferred  
mental constructs. An excellent example of this type of indirect procedure is the



**“aptitude-treatment interaction” (ATI). With ATIs the underlying construct is assumed to dictate the treatment. With a reading deficit, an underlying strength such as auditory processing or visual discrimination would be matched with a treatment that takes advantage of this strength. The assumption is that if the identified strength is matched to a treatment designed specifically for that strength the student’s skill will improve. For example, if a student is identified as having an auditory processing strength, an intervention to remediate reading may be to provide more auditory cues tied with the visual cues already present in reading. Conversely, if a student has a visual processing deficit, any interventions developed would take that into consideration and avoid excessive reliance on visual cues. Although this process has a great deal of logical appeal there is little empirical evidence in support of ATIs. Several comprehensive reviews of ATI literature have been conducted (Arter & Jenkins, 1979; Ysseldyke & Mirkin, 1982) and they consistently fail to show significant support for the validity of these indirect procedures for the purpose of treatment development.**

**Direct procedures on the other hand, attempt to improve performance by intervening on the academic behavior of interest. For example, with a reading deficit, word production is the behavior of interest. Unlike indirect procedures they do not assume that underlying constructs are the cause of academic deficits. One direct approach to remediation of reading problems consists of increasing time on task or opportunities to respond to improve reading skills (Shapiro 1996). Several procedures have been developed specifically to increase time on task or opportunities to respond. These procedures consist of peer tutoring and cooperative learning strategies (e.g.,**

Delquadri, Greenwood, Stretton, & Hall, 1983; Phillips, Fuchs, & Fuchs, 1994) and are designed specifically to increase the target response of reading. Several direct approaches designed to intervene on the behavior of interest were employed in this study. LPP, error correction, repeated readings. Last phrase in incomp sentence

A second direct approach to instruction is more germane to the present study and is based on a behavior analytic approach to treatment of academic problems. These procedures consist of modifying the academic environment, through the use of contingencies, to improve academic responding. These modifications have been employed to improve reading comprehension (e.g., Lahey & Drabman, 1973) oral reading rates (Lovitt, Eaton, Krikwood, & Perlander, 1971) as well as other academic task. Both time on task procedures and behavior analytic procedures have been validated experimentally as effective for improving reading ability in students, but more importantly changing academic behaviors were accomplished without concern for underlying cognitive skills (Shapiro, 1996). In other words the internal condition of the subject of concern was not taken into account when employing these interventions. Change in reading skill was accomplished with regard only to environmental factors and subject responses.

Both of these areas of research have contributed to a rich pool of resources and interventions, which are available to intervene with problem readers. A crucial issue for the service provider, however, is which of the many interventions to choose. All of the interventions are not equally effective with all types of problems (Daly, et al., 1996). Hence, methods are needed for assisting professionals to determine which interventions are most effective with particular problems.

## Designing Assessment and Intervention Procedures

The brief review of the assessment and intervention techniques above represents only a small portion of the literature devoted to the research and refinement of these procedures. However, "in spite of our well developed knowledge base, one reason for poor literacy rates is that the assessment and intervention tools available to educators are (still) not sufficiently refined to provide the kind of information necessary for making good educational decisions about children on an individual basis" (Daly et al., 1996). Current methods for assessment and treatment identification lack two main elements.

First, many assessment practices are not linked to conceptual models of learning that have been empirically evaluated. Incorporating a conceptual model of academic responding into the assessment and intervention development process will tie the results of the assessment to a treatment recommendation based on the conceptual model. This will provide specific recommendations for treatment based on the results of the assessment. Traditional methods of assessment are either not tied to a conceptual framework of learning (i.e. curriculum based measurement) or are rooted in models that have little or no empirical support (i.e. ATIs).

Curriculum Based Measurement procedures have good psychometric properties (Shinn, 1989), and provide useful data about the academic performance, but CBM is an assessment process which is not tied theoretically or conceptually to an instructional model and therefore does not indicate which instructional techniques are likely to be most effective in remediating academic deficits (Daly et. al., 1996). Generally the approach to intervention using CBM is formative evaluation. This

amounts to a “try something, see if it works, and change it if it doesn’t”. This type of intervention framework is nothing more than an advanced mode of trial and error. Given the difficulty, expense, and consequences of interventions to improve reading performance, reducing any delay in identifying an appropriate intervention would be of great importance.

Aptitude by treatment interaction (ATI) procedures represent another type of procedure that are limited in utility for identifying treatments. ATI are grounded in a conceptual model of learning, however the model has not stood the rigors of empirical evaluation (e.g. Arter & Jenkins, 1979; Ysseldyke & Mirkin, 1982). Therefore, interventions based on an identified aptitude have not proven to be an effective method for remediating academic deficits.

A primary example of an assessment procedure that is linked to a conceptual model and has stood the rigors of empirical evaluation is “performance and skill assessment”. Initially discussed with reference to social skills (Gresham 1981), performance and skill assessments are rooted in a very simple model, which evaluates behavioral deficits in reference to the presence of or absence of the ability necessary to perform the required skill. With performance and skill assessments the results of the assessment indicate the presence or absence of the skill of interest and result in what are called skill deficits or performance deficits. Treatments can then be recommended based on these outcomes. If a subject possesses the skill necessary to engage in the behavior he or she would be said to have a performance deficit. If a subject does not possess the necessary skills to engage in the behavior then the assessment would indicate a skill deficit (Gresham 1981). Based on these results treatments are

recommended. These treatments either attempt to elicit the existing behavior or teach the skill necessary to perform the behavior. An example of a treatment for a performance deficit may be the addition of reward for engaging in the target behavior designed to produce a reinforcement effect. Treatments for skill deficits can be dealt with by providing instruction for the skill of interest in order to make it possible for the subject to perform the behavior.

Second, the level of inference with even the best technology does not provide more than a probabilistic chance at an effective intervention. A primary intent of assessment is to classify for the purpose of planning, executing and evaluating treatments (Korchin & Schuldberg, 1981). The assumption within an assessment is that based on its results, predictions about the type of intervention that would be beneficial can be made. In other words, the assessment is designed to predict a treatment's effect on behavior. However, like all predictions this one entails a level of inference. Because the effect of a treatment on a behavior cannot actually be known until that treatment is implemented, the use of something other than the actual treatment to make statements about the potential effects of a treatment requires an inference. The greater the level of inference between the assessment results and the treatment based on those results, the more likely an error may exist between the prediction made by the assessment and the actual outcome of the treatment. The smaller the level of inference the more likely the assessment result will accurately predict a treatment's effect on behavior. Along these lines, Martens (1992) has argued that the lower the inference between assessment and treatment the greater the possibility for the assessment to produce valid treatments. If our goal in assessment is

to produce valid treatments, the level of inference from assessment to treatment recommendation should be as low as possible. However, methods for identifying instructional procedures remain only hypotheses with regard to effective treatments and therefore will be effective in some but not all cases.

Linking assessment to intervention using a conceptual model. The Instructional Hierarchy is a heuristic framework designed for generating instructional treatments based on the level of skill development (Haring, Lovitt, Eaton, and Hansen, 1978). This heuristic was derived from behavior analytic methods and principles, and focuses on the variables that change behavior (Haring et al., 1978). Within the Instructional Hierarchy exist four levels of skill development. These levels of skill development refer to stages of learning that apply to a learner's performance of a given target behavior. These stages of learning are acquisition, fluency, generalization and finally adaptation (Daly et al. 1996). With the acquisition stage of learning the learner acquires the ability to perform the new skill with accuracy. At the fluency skill level the newly acquired skill is developed to a more fluent level. With the generalization stage of learning, the skill is generalized to a novel context. Finally, in the adaptation stage of learning the learner is able to modify the response according to novel demands (Daly et al. 1996).

Each of the stages of learning in the Instructional Hierarchy is associated with specific instructional procedures designed to initiate mastery of each stage of learning (Daly & Martens, 1994), and are derived from a behavior analytical approach to learning. The acquisition stage of learning is associated with modeling and prompting. These procedures are designed to build the skill to an accurate level.

Fluency stage of learning is associated with instructional procedures of drill and practice. The purpose of these procedures is to improve the rate of responding once a behavior is acquired. The generalization stage of the Instructional Hierarchy relies on training in the natural context to enable the application of a skill so performance can progress to novel contexts. In the final stage, adaptation, the learner is asked to apply the skill to several different novel tasks. The purpose of treatments derived from the adaptation skill level is to promote problem-solving or modification of the skill in face of novel environmental demands (Haring et al., 1978).

A basis for applying a conceptual model within an assessment was demonstrated by Noell, Gansle, Witt, Whitmarsh, Freeland, LeFleur, Gilbertson & Northup (1998). Noell et al. (1998) used treatments based on the conceptual model of performance and skill deficits, originally proposed by Gresham (1981), to develop an assessment for differentiating between students with reading deficits. This study employed the application of contingent reward and instructional procedures as assessments designed to make distinctions between students who's reading problems were due to performance deficits or skill deficits. By observing a student's response to either performance treatments (contingent reward) or skill treatments (instruction) a determination was made about the cause of the reading problem. This study was not only successful in differentiating between students with performance or skill deficits, it also identified treatments that when implemented resulted in a substantial increase in reading fluency. This study successfully tied an assessment to a conceptual model and reduced the level of inference from assessment to treatment by using treatments as assessment components. This paper is similar in that it attempts to use a conceptual

model to drive assessment targets and it focuses on the use of treatments within the assessment to produce valid treatments. However, it is an extension of this work in that it employs a different conceptual model (Instructional Hierarchy) and focuses strictly on the assessment of student skill level as opposed to other potential causes of poor achievement such as a performance deficit.

Daly & Martens (1994) have applied the instructional hierarchy as a conceptual model for assessment development in an attempt to remediate reading deficits. The authors used the instructional hierarchy to compare the effects of three instructional interventions on subject's oral reading fluency. Three instructional procedures were employed that contained treatment components associated with the IH. The first treatment consisted of listening passage preview (LPP), a combination acquisition, fluency building and generalization treatment. Second, subject passage preview (SPP), was implemented as a fluency building only treatment. Finally, taped words (TW) a strictly acquisition component was also part of the study. Four students with learning disabilities participated in the study. After a brief baseline the three treatments were applied in a multielement design using frustrational level materials. The purpose of the study was to examine the effect of each treatment on reading of instructed passages and on instructed word list. The authors found that LPP (the treatment package with the most treatment components) produced the greatest gains in reading and that TW (the treatment package with the fewest treatment components) produced the least amount of growth over baseline. The results of this study reveal that, for the four subjects involved, when instructional interventions contained more active components prescribed by the Instructional Hierarchy these interventions were



more likely to be effective. Because this research employed the IH as a basis for intervention identification, these results lend empirical support for the use of the Instructional Hierarchy for the development of interventions designed to increase reading fluency.

In a second study, Daly, Martens, Dool and Hintze (1998) combined a functional analysis methodology with the conceptual model of the instructional hierarchy in an attempt to identify the most effective intervention for increasing reading fluency. Four treatments were employed as components of a functional assessment of reading fluency. Contingent reinforcement (CR), repeated readings (RR), listening passage preview (LPP) with phrase drill (PD) and error correction were used in isolation and in combination throughout the study. CR consisted of the delivery of a preferred item contingent upon reading at or above a predetermined standard. This condition was an attempt to rule out the possibility that poor performance was the result of a performance deficit. RR was an instructional procedure based on the fluency skill level of the instructional hierarchy and consisted of 4 readings of the same readings by the participant. This procedure was designed to increase the participant's opportunities to respond in order to increase reading fluency. RR was considered the simplest instructional procedure to implement and was the first instructional procedure implemented after CR. Following RR the procedure implemented was LPP/PD. This procedure corresponded with the acquisition skill level of the instructional hierarchy and was designed to provide modeling and error correction to improve reading fluency. Subjects listened as the passage was read allowed, repeated the passage and with experimenter assistance reread all phrases with

~~error words in it. This final procedure was implemented last~~ due to its more complex and effortful procedure. After a single baseline data point, CR, RR and LPP/PD were each implemented for one session in the current sequence. Bases on subject responses to each of these treatments combinations of treatments were evaluated. The simplest most effective treatment was then considered the identified treatment and a replication of its treatment effect was attempted. For the three subjects assessed a treatment recommendation was identified based on the subject's response to treatment and the replication was successful in producing similar results. Similar to the previous Daly & Martens (1994) study this study also support the use of the conceptual model of the IH for development of interventions to increase reading fluency. It also extends this application of the IH to the functional analysis literature and by doing so provides a framework for developing other forms of function based reading assessments for use with the IH.

#### Low Inference Assessment Linked to the IH

The functional analysis literature provides a useful model in which very direct assessments can be linked logically and functionally to intervention. With a functional analysis, an assessment is conducted by directly observing the effects of mini-treatments on the behavior of concern. The subject's response to the brief treatments leads to a hypothesis about the function of the behavior. The hypothesized functional relationship is then used to develop a treatment for the behavior of concern (Repp, 1994). With this method of assessment there are two levels of inference. First, assumptions are made about the accuracy of the assessment results as a true sample of the behavior of interest and the degree to which these are measure of criterion

responses (Goldfried & Kent, 1972). The direct nature of the observation makes this an inference of the lowest level possible without obtaining a more extended sample of behavior. However, by extending the duration of the sampling procedure feasibility for assessment using this method is reduced. The second assumption made is that the treatment derived from the hypothesized functional relationship will be effective when generalized to the treatment context. The level of inference here is small because treatments derived from functional assessments are based on hypothesized function, and these treatments have been demonstrated to have a higher probability of effectiveness than arbitrarily chosen treatments (e.g., Carr & Durand, 1985; Iwata, Pace, Cowdery, & Miltenberger, 1994; Repp, Felce & Barton, 1988).

Because of their low level of inference, foundation in a conceptual model and the success of these procedures in surviving scientific rigor, these technologies are some of the most sound currently in use. However, the fact still remains that no assessment is perfect and all assessments no matter how low their inference provide only hypothesis with reference to treatment outcome. Nonetheless, it may be possible to reduce this inference to an even greater extent by reducing even further the inference made by the assumption that a treatment derived from the hypothesized function will be effective.

By incorporating treatments based on the conceptual model of interest into the assessment process, and making classification decisions based on the obtained response to the treatments, the inference would be removed. This would be similar to hypothesis testing (eg. Repp, 1994; Repp, Felce, & Barton, 1988; Repp & Karsh, 1994) where multiple hypotheses about the function of a behavior are tested by

observing response to these function-based treatments. Response to one function-based treatment and not others would indicate a subject's placement within a conceptual model. It also provides information on what treatment or treatments are effective in altering the behavior of interest because a treatment has already been demonstrated effective. For example, if aberrant behavior were hypothesized to be due to positive reinforcement and a treatment designed to reduce aberrant behavior based on this positive reinforcement hypothesis was applied the results of the treatment would serve as the assessment of the hypothesized function. Positive response to the treatment would not only identify a function, but also illustrate a treatment with a known effectiveness.

In applying this same logic to the Instructional Hierarchy, the acquisition and the fluency building levels of skill development refer to two different placement levels. Each placement level is associated with a separate class of treatments designed to promote mastery of that skill level. Response to a treatment designed specifically for one skill level, and a lack of response to the other should indicate that the student is performing within that skill level. Results such as these would not only indicate current performance, but provide recommendations for interventions that have been demonstrated effective.

#### Evaluating the Treatment Validity of Assessment

In developing assessment procedures, the validity of those procedures is always important. As described earlier, valid assessments are ones that contribute to beneficial treatment outcome (Hayes et al. 1987). However, in evaluating an assessment similar to the one described above treatment utility consist of more than

the display of a treatment effect. Because the assessment produces more than one outcome (acquisition or fluency skill level) and more than one treatment recommendation (acquisition or fluency treatment) mere improvement over baseline as a result of either treatment is not sufficient to establish treatment validity. In a case such as this where either treatment could potentially produce positive effects on behavior a more elaborate form of evaluation is necessary. For an assessment to have treatment utility it must contribute to a beneficial treatment outcome (Hayes et al. 1987). In a situation where two treatment options are present the potential that both treatments can produce some beneficial effects exist. Because of this both treatment and therefore both assessment outcomes could be viewed as valid, and if both assessment outcomes are valid there is no purpose for conducting the assessment. So the question that must be answered is not if the assessment produces a valid treatment, but if the assessment produces the most valid treatment of the options present. Hayes, Nelson & Jarrett (1987) outlined a procedure for evaluating treatment utility of this type of assessment. Described as an "obtained differences-two or more treatments study" this procedure directly compares the two treatment outcomes to evaluate the extent to which the assessment identified the treatment that had the greatest impact on behavior. After a baseline is established both treatments are applied in an alternating treatment design, which allows differential treatment effects to be determined within an individual (Hayes, Nelson & Jarrett, 1987). Using this methodology an assessment would be considered to have treatment utility if it resulted in an assessment outcome that produced larger treatment effects than the alternative outcome. In other words did the assessment predict the most effective outcome?

~~This paper has three aims. The primary purpose of this study was to evaluate~~  
the treatment validity of a reading assessment. This was attempted by first evaluating  
the ability of a treatment based assessment to discriminate between subjects based on  
their responses to intervention and second determining the extent to which a brief  
treatment conducted within an assessment protocol produces the most valid treatments  
when compared to an extended application. A secondary goal was to compare the  
treatment validity of the response to treatment assessment with the treatment validity  
of instructional placement standard recommendations. A third goal was to examine  
the generalizability of treatment recommendations derived from the treatment based  
assessment to multiple type of treatments based on the IH. In other words, will the  
results of the assessment not only predict responses to the treatments employed, but  
will they also predict responses to other treatments in the same class based on the  
instructional hierarchy?

## **METHOD**

### **Subject Selection and Setting**

Participants in the study were first, second and third grade students selected from the general education population of two public schools. Six subjects from each grade were selected based upon a match between reading ability and the pre-established grouping criterion described below. Potential subjects were initially identified for participation based upon a screening of reading fluency. This screening consisted of a one minute individually administered grade level reading probe scored for words read correctly per-minute (Shinn, 1989). The first six students in each grade who met the initial criterion for the study and returned parent permission forms were included in the study. First grade students that met the above criteria were Karen, Monty, Rhoda, Dana, Gerald and Ronnel. Karen, Monty, Rhoda and Dana were 7 years of age while Gerald and Ronnel were both 6-years-of-age. Second grade students included in the study were Mark, Don, Bob, Kelly, Tony and Clarence. All of whom were 8-years-old at the time of the study with the exception of Clarence who was 9 years-of-age. Finally, the third graders included in the study were Ken, Tina, Conner, Chris, Jan and Didra, all of whom were 9-years-old at the time of this study. All study procedures were conducted in the school's library or the computer lab depending upon availability. Doctoral students in school psychology conducted all assessment and intervention procedures.

### **Grouping Criteria**

A pre-established grouping criterion was created in order to insure a heterogeneous sample across grades and across a range of reading fluencies. This

grouping criterion was derived from instructional placement standards established by Fuchs and Deno (1982). These instructional placement standards consist of fluency ranges, which assist in determining the correct placement of a student within a hierarchically graded reading curriculum. For example, a particular student, when presented with a second grade reading passage, would be classified as functioning in the mastery, instructional or frustrational range based upon oral reading fluency. Because the focus of this study was on instruction, only children in the frustrational and instructional ranges were included. In addition to providing information to place a child within a reading curriculum, Shapiro (1986) has suggested instructional placement criteria can be linked to appropriate instructional practices. That is, the placement standards help to identify performance that is associated with a level of reading ability where fluency building would be appropriate (instructional range) or a level of reading where accuracy training would be appropriate (frustrational range) (Shapiro, 1986).

Subjects were selected from throughout the instructional and frustrational ranges specified by Fuchs and Deno (1982) so as to be able to more precisely define meaningful instructional placement standards should small differences in fluency make a difference in treatment outcome. The six groups within each grade level ranged from the bottom of the frustrational range through the top of the instructional range, divided as evenly as possible between the six fluency groups. As shown in Table 1, for first and second grade, the six groups consisted of a range of ten words correct per minute, beginning with 0 to 9 and ending with 50 to 60. Third grade consisted of a range of approximately 17 words correct per minute, beginning with 0



to 16 and ending with 86 to 100. Table 1 shows the stratification of the groups across grades and fluency level. Based on these divisions, for each grade level, 4 groups were considered frustrational and 2 groups were considered instructional for grade level materials according to Fuchs and Deno (1982). Placement into groups occurred during the initial steps of the assessment.

Table 1. Grouping divisions.

Grade	Fluency Level/Group					
1-2	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 - 60
3	0 - 16	17 - 34	35 - 51	52 - 69	70 - 85	86 - 100

### Materials

Reading Probes. Reading passages used throughout the study consisted of content controlled grade level passages of 100 words or more. Because the curriculum in the school district where the study was conducted used a literature-based series, which was not content controlled, that reading series was not be used in the study. Instead, the reading passages used in this study came from grade level probes of the “Great Leaps” reading series (Campbell, K., 1998).

Reading passages were screened for reading level in an attempt to increase the consistency of the reading difficulty within each grade level. A computerized version of the Flesh-Kincaid readability index was used to test for readability. This index was chosen because of its ease of use and convenience as well as previous findings of validity supporting its use (Clariana, 1993). Readability was assessed on all passages contained within first, second and third grade levels of the “Great Leaps” reading

series and any passages that were not within the assigned grade level were dropped.

The mean readability index of passages used in this study based on Flesch-Kincaid were first grade, 1.11 (range 0.1 to 1.89); second grade, 2.43 (range 2.1 to 2.87); and third grade, 3.44 (range 3.15 to 3.93).

Rewards. Students were surveyed about reward preferences. Students were given a list of 30 different stimuli and asked to select the items they would like to earn. These items were considered preferred and the list of the student selected preferred items was presented to each student's teacher. The teacher had the opportunity to eliminate any items, which were unacceptable from the list. Using the list of acceptable preferred stimuli, a "treasure chest" was created containing tangible items identified as preferred by the students. The treasure chest was used with the treatment packages for reward delivery. Reward delivery consisted of allowing the student access to the treasure chest for the purpose of selecting a tangible reward. Typical items listed on the reinforcer survey are pencils, pens, erasers, colors, candy, stickers and small toys.

#### Response Measurement and Observer Agreement

The dependent variable for this study was words read correctly. Words read correctly (WRC) were defined as words pronounced correctly with no more than a 3 second delay after the completion of the previous word (Shinn, 1989). Both words read correctly and errors were scored by an experimenter. This was accomplished by marking errors on a separate copy of the reading passage as the participants read aloud. An error was defined as any word containing mispronunciations, substitutions,

omissions, or any word not completed correctly within 3 seconds of the completion of the last word (Shinn, 1989)

During 37% of the sessions a second experimenter independently scored words read correctly for participants in order to evaluate the reliability of scoring. Each word was scored as either an agreement or a disagreement based on whether the two observers scored the word the same in reference to correct or incorrect. Interobserver agreement was calculated by dividing the number of agreements by total number of agreements plus disagreements. The total agreement across all reliability sessions was 93% with a range of 84 – 100% agreement.

#### Treatment Packages

Two treatment packages were assembled, based on empirically tested procedures, which were designed to be effective treatments for the initial two levels of the instructional hierarchy (IH): the acquisition level and the fluency level. The first of the two treatment packages, termed the acquisition treatment package, was designed to establish new reading skills and contained the active treatment components of modeling and error correction, both of which have been demonstrated to improve correct responding (Daly et al. 1996). These specific treatments were selected because of their correspondence to the acquisition skill level of the IH. The second treatment package, termed the fluency building treatment package, was intended to increase fluent responding and was designed to be an appropriate intervention for students who were at the fluency level of the IH. This package contained the active treatment component of repeated reading (practice), which has been shown to improve reading fluency and was selected because of its correspondence to the fluency building (Daly

**& Martens, 1994). level of the IH. Throughout the study these two treatment packages were components of both the brief assessment and the extended validation. For example, the acquisition treatment package was implemented three times as part of the skill level assessment, and over several sessions as part of the extended validation.**

**Each of the two treatment packages contained three elements. These elements consisted of an initial fluency evaluation, an active treatment component (described above) and a final fluency evaluation, delivered in that order. The initial fluency evaluation and the final fluency evaluation were identical to each other and were identical in both treatment packages. The only element that varied between the treatment packages was the active treatment implemented.**

**The initial fluency evaluation and the final fluency evaluation each consisted of a one-minute timed reading of the grade level probe assigned for that session. The probe was scored for words read correctly (WRC) as described above and error words were recorded. In an attempt to promote discrimination of the active treatment components and to reduce the effects of practice within the acquisition treatment package the initial fluency evaluation preceded the active treatment component by a minimum of one hour and a maximum of two hours. However, the final fluency evaluation occurred immediately after the active treatment.**

**Both treatment packages included the same reward component. A performance goal was established for each session based on the initial fluency evaluation. Prior to the initiation of the active treatment component the student was informed of the performance goal and instructed that if he or she surpassed the performance goal in the final fluency evaluation they would have an opportunity to**

select a reward from the "treasure chest". The reward component was employed in an effort to control variability in student responding that may be attributable to student attention and effort. The specific procedure employed for each treatment package is described below.

The acquisition treatment package. According to Haring et al (1978), students at the acquisition level of the IH will benefit most from instructional procedures designed to establish a new skill. For this study, modeling and error correction were selected because of their efficacy (Daly & Martens, 1994). The active treatment within the acquisition treatment package was a combination of two procedures for improving reading accuracy. The first component was listening passage preview (LPP), a modeling procedure, where the experimenter modeled the reading of the passage for the student prior to the student reading independently. The second component in this treatment package was error correction. Error correction consisted of the experimenter modeling correct responding whenever the student made a reading error and prompting correct responding from the student. Both procedures independent of each other have been empirically tested (LPP with Rose & Sherry 1984; Salend & Nowak 1988 and error correction with Rose, McEntire & Dowdy, 1982; Singh & Singh, 1986; Singh, Singh & Winton, 1984) and combined (Daly & Murdoch, 2000) are intended to serve as a powerful treatment for improving reading skills for students at the acquisition skill level.

In the acquisition treatment package an experimenter presented the student with a grade level reading passage and asked the student to read the passage for a one minute timing (initial fluency level evaluation). The words read correctly (WRC) as

defined above and the errors were recorded. All error words were recorded on a separate sheet of paper for later use in the error correction portion of this procedure. The WRC from this reading also became the performance goal used with the reward component.

A minimum of one hour lapsed between the initial evaluation and the beginning the active treatment component of the treatment package. After this delay, but prior to a two-hour laps the experimenter informed the student of the performance goal from the initial fluency evaluation and told the student that if he exceeded the performance goal he would be allowed to select one item form the “treasure chest”. The experimenter then modeled the same reading passage (LPP procedure) by reading orally 1.5 times the number of words read correctly from the student’s one minute timing. For example, if the student’s read correctly 30 words in the one minute timing, then the experimenter would read (modeling component) a total of 45 words from the reading passage. After listening passage preview was complete the experimenter presented the error words, which has been recorded from the initial reading. The words were presented one at a time to the student and the student was asked to read them. If the student correctly read the word the experimenter moved on to the next word. However, if the student incorrectly said the word when presented in isolation or did not respond in 3 seconds the experimenter began the error correction procedure for that word. The error correction procedure, in this experiment, consisted of three steps. First the experimenter pronounced each syllable of the incorrect word orally and the participant was then asked to repeat the syllables orally. For the second step the experimenter blended the syllables together to form the correct word. The

participant was then asked to repeat the blended form of the error word and finally say the word immediately after without a prompt. For example, if the incorrect word is “three” the experimenter broke the word down into its syllables (th+r+ee) and pronounced each syllable independently. Following that demonstration, the participant was asked to say each syllable. The experimenter then blended the syllables by saying the complete word (three) at a normal speaking rate and also had the participant say the error word. Finally the participant was asked to say the word independently without experimenter prompts. This was repeated for the entire set of error words from the one-minute timing (initial fluency evaluation).

Upon completion of this package the student completed the final fluency evaluation while the experimenter scored for words read correctly. If the student exceeded the performance goal, the student was given an opportunity to select a preferred item from the “treasure chest”. The words read correctly from the final fluency evaluation was used as the dependent measure for this condition.

The fluency treatment package. Based on the IH, students at the fluency building level have established some initial competency with the skill and merely need to practice in order to increase fluency, hence, such students would benefit from repeated practice (Daly & Martens, 1994) to increase reading rate. The second treatment package consisted of a component that is designed to improve reading fluency. This component was repeated readings (RR), a practice procedure, in which a student repeatedly reads a passage for a one-minute timing prior to evaluating reading performance. This procedure has also been empirically tested (Stoddard, Valcante, Sindelar, O’Shea & Algozzine, 1993; Rashotte & Torgensen, 1985) and is

intended to serve as a powerful treatment package for improving reading skills for students performing at the fluency skill level.

With the fluency treatment package the experimenter presented the student with a reading passage and asked the student to read the passage for a one-minute timing (initial fluency evaluation). The words read correctly (WRC) as defined above and the errors were recorded. The WRC from this reading also became the performance goal used with the reward component.

The experimenter allowed a minimum of one hour, but not more than two hours to pass before beginning the active treatment component of the treatment package. Then the student was informed of the performance goal, based on the initial fluency evaluation, and told that if the performance goal was exceeded he would be allowed to select one item from the "treasure chest". The experimenter then presented the student with the same grade level reading passage and asked the student to read the passage orally for a one-minute timing. The experimenter scored on a separate copy of the reading probe the words read correctly for each reading. This process was repeated two more times with the same reading passage for a total of three repeated readings for the treatment package. During these repeated readings the experimenter did not provide correction of any kind, except a prompt to "move on" if the student remains fixed on a word for more than 3 seconds.

The same reading was presented again for the final fluency evaluation and served as the dependent measure for this condition. If, in the final fluency evaluation the student exceeded the performance goal he or she was given an opportunity to select a preferred item from the "treasure chest".



## **PROCEDURE**

### **Overview**

The procedure for this study consisted of three steps. The first step consisted of participant placement into groups based on the pre-established grouping criteria outlined in Table 1. Once a participant was assigned a group based on his or her grade level fluency a response to interventions assessment was conducted. With this assessment participant responses to the acquisition and the fluency treatment package were evaluated. As part of the assessment student responses to intervention were evaluated and recommendations for treatment were made. Finally, these recommendations were evaluated in an extended analysis to determine if the assessment produced recommendations that resulted in valid treatment outcomes.

### **Group Placement**

The first step in the procedure was to assign participants to groups based on the grouping criteria presented in Table 1. Initially, a fluency level for current grade level materials was identified for each participant. Fluency levels for the purpose of this study were derived for each student following the procedure described by Shinn (1989). Three one-minute grade level reading probes were administered to each child and scored for words read correctly (Shinn, 1989). The median score of the three reading probes was the score used to identify the participants for placement into one of the pre-established groups shown in table 1.

## **Assessment – Response to Treatment**

Following subject selection for the experiment, each subject was administered a response to intervention assessment, designed to determine whether students responded better to the acquisition treatment package or the fluency building treatment package. The response to treatment identification assessment consisted of an assessment of the student's initial fluency level followed by the application of the acquisition assessment and the fluency building assessment described below. Results from the response to intervention assessment were then evaluated using the assessment interpretation methodology described below.

**Initial fluency level.** The student's initial fluency level was identified by administering three one-minute grade level reading probes (Shinn, 1989), and calculating the words read correctly for all three passages. The median words read correctly was considered the student's current performance level. This median score was used as a baseline for evaluating assessment results.

**Acquisition assessment.** The acquisition assessment consisted of three applications of the acquisition package described above. This portion of the assessment was designed to determine the effect of modeling and error correction on words read correctly. The median words read correctly from the three applications of the acquisition package was then compared to the student's initial fluency level in the assessment interpretation phase below.

**Fluency building assessment.** The fluency building assessment consisted of three applications of the fluency-building package described above. This portion of the assessment was designed to determine the effect of repeated readings on words

read correctly. The median words read correctly from the three applications of the fluency-building package was then compared to the student's initial fluency level in the assessment interpretation phase below.

Sequence of assessment. The assessment consisted of 3 one-minute probes designed to identify initial fluency followed by 6 treatment-based probes. In order to reduce the possibility of a sequence effect the order in which the treatment-based probes were delivered was a pre-established semi-random repeating pattern. Students were then randomly assigned a number from 1 to 6 for assignment into two different sequence groups. For each grade level, students with assigned numbers of 1, 3 and 5 were designated the pattern ABBABA and for students 2,4 and 6 the pattern was BAABAB. Therefore, the assessment was considered complete when at least 9 sessions were completed.

Assessment interpretation. Because the assessment consisted of treatments designed to increase the rate of reading for participants, the assessment interpretation consisted of an evaluation of the treatment effects in comparison to both the initial fluency evaluation (baseline) and relative to the other treatment. To evaluate differences over initial fluency levels a standard of 20% was adopted. In other words a difference was considered to exist if a difference in WRC of 20% or more was observed in either treatment over the initial fluency level. An increase of 20% was derived from Carnine, Silbert, and Kameenui's (1990) recommendation of 40% increase in reading fluency as an appropriate weekly goal for reading instruction based on repeated readings of the same passage. Twenty percent was adopted with the rational that if brief exposures to treatments could produce, in three sessions, half of

~~what is considered appropriate for one week that growth would be considered promising.~~

The median words read correctly from the two brief treatments were compared to the student's current performance level. If either or both of the treatment packages produced an increase over the initial fluency level of 20% or more, a difference between baseline and treatment was considered to be present and a treatment effect was said to exist. If neither treatment packages showed increases of 20% or more, the student repeated the assessment phase of the study until a difference was observed by the above criteria. If, after repeating the assessment three times no difference was observed, the study for that participant was discontinued.

The second phase of interpretation was a comparison of the two treatment packages. This stage of the assessment interpretation occurred only if a difference was present in the first evaluation. Given that at least one of the treatments had shown improvement over the initial fluency assessment of 20% or more this evaluation was designed to determine if differences between the two treatment packages could be observed. To accomplish, this visual inspection of the graphed assessment results was conducted. A difference was considered to exist in the two data paths if there were no overlapping data points during the response to intervention assessment. However, if an overlap in data paths did occur the student repeated the acquisition and fluency assessment phases of the study until a difference was observed by the above criteria. If, after repeating the assessment three times no difference was observed the study for that participant was discontinued.

The results of the assessment could produce one of three outcomes. No treatment effect, which was the result of growth less than 20% for both treatment packages. Treatment effect for one package, which was the result of at least a 20% improvement over the initial fluency assessment for one or both treatment packages and no overlap in the data paths. In this case the treatment package with the greatest level of improvement over baseline was considered the effective treatment. Treatment effect for both treatment packages, which was the result of a 20% or more improvement over the initial fluency assessment for both treatment package and an overlap in the treatment packages data path.

Intervention Identification. Based on the results of the assessment, treatment recommendations were made. For any given participant, if one treatment produced an increase that was considered effective by the above standards (i.e. there was a treatment effect for the acquisition or fluency package) that treatment was considered to be the more appropriate intervention for the participant. If treatment effects were observed for both treatment packages, both treatments were considered appropriate for the participant. However, if neither treatment demonstrated effectiveness, neither treatment was considered to be an appropriate match for the participant. For example, if both treatment packages produced more than 20% growth over the initial fluency assessment with no overlap in the data paths and the acquisition package showed the greatest level of increase, it was considered the most appropriate treatment according to the assessment. If, however both the fluency assessment package and the acquisition assessment package produced an increase over the initial fluency

assessment greater than 20%, and there was overlap in the data paths, both the fluency and acquisition treatments would be considered appropriate treatments.

### Assessment Validation

The purpose of the assessment validation was to determine if the recommended treatment in the assessment phase of this study was accurate in predicting the most effective treatment. As described in the introduction of this paper an “obtained difference-two treatment study”, (Hayes, Nelson & Jarrett, 1987) was implemented in an attempt to determine if the assessment could identify the most appropriate treatment. Treatment validity is the extent to which an assessment produces a favorable treatment outcome (Hayes, Nelson & Jarrett, 1987). However, because in the present study two potentially effective treatment recommendations can result from the assessment, simple improvement over baseline does not provide information about the utility of the assessment. In other words, both treatment recommendations have been validated as effective and have the potential for improving reading fluency. Improvement over baseline in this case is not evidence that the assessment contributed to the outcome. To show treatment validity with an assessment of this type it is necessary to show that the assessment makes needed distinctions for treatment identification (Hayes, Nelson & Jarrett, 1987).

Based on this information, this assessment would only have treatment validity if the treatment recommendation produced a more favorable outcome over the alternative treatment recommendation. The extent to which the treatment recommendation from the skill level identification assessment predicts or matches the

more effective intervention in the extended validation would be its level of treatment validity.

#### **Experimental Design: Extended Analysis**

To evaluate treatment validity an alternating treatments design was used to determine whether the treatment recommendation from the assessment matched the extended analysis of treatment effectiveness. For all subjects receiving treatment recommendations both treatments were implemented in an alternating treatments design. The extent to which the assessment predicted the most effective treatment was evaluated. Both treatments were alternated in a semi-random repeating pattern. The pre-established pattern for the design was ABABABBABAABBABA for participants 1,3 and 5 and BABABAABABBAABAB for participants 2,4 and 6 for each grade level.

**Baseline.** The baseline for comparison was the initial fluency assessment of student readings collected in the assessment phase of the study.

**ATD Phase.** The acquisition treatment applied in the alternating treatments design consisted of one application of the acquisition package for each session and was scored for words read correctly during the one-minute timing. The fluency building treatment applied in the alternating treatments design consisted of one application of the fluency-building package for each session and was scored for words read correctly during the final evaluation assessment. The procedures for implementation of both the acquisition treatment and the fluency building treatment were identical to the procedure described above.

Alternative treatment packages. This phase was designed to examine the robustness of the IH. The primary question addressed by this study centered around the extent to which brief treatments derived from the IH could identify effective interventions. Evidence supporting the treatment validity of the assessment does not necessarily provide evidence for the assessment's relationship to the IH. In other words, even if the assessment can identify valid treatments (valid for individual treatment recommendations), this result is not convincing support for the correspondence between assessment outcomes and student skill level, based on the instructional hierarchy. However, if one could show interchangeability between a class of interventions linked to specific levels of the IH, then it would lend support to the conceptual underpinnings of the IH. That is, if members of the class of interventions designed to establish new skills were most appropriate for children at the accuracy level and were shown to be interchangeable then this would be the type of evidence needed to support the generality of the model. In this phase participants that responded to one type of treatment (either acquisition or fluency building) were evaluated to determine if they also responded to other members of the class of that type of treatment and not others. For example, if participants who responded to acquisition treatments and not to fluency treatments also responded to other forms of acquisition treatments and not to other forms of fluency treatments, then this would be additional evidence supporting the IH.

Four participants were selected to continue the extended validation using alternative forms of the acquisition and fluency building treatment packages. After the completion of the initial assessment and the extended validation the alternating



treatments design was continued using alternative treatment packages. All aspects of the treatment packages remained the same except the active treatment components. Initial and final fluency assessments were still conducted as described above.

For this phase of the study an alternative intervention was substituted for both the fluency package and the acquisition package. The alternative acquisition package consisted of modeling and error correction, similar to the original acquisition package. However, the modeling component consisted of the student listening to a taped recording of the entire reading passage. The use of audio taped passage preview has been empirically demonstrated effective for improving reading fluency (Medcalf, 1989). The alternative acquisition package also utilized error correction in the form of word drill. Word drill has also been empirically demonstrated to be effective at improving reading fluency (Fleisher & Jenkins, 1983; and O'Shae, Munson, & O'Shae, 1984). Word drill in this study consisted of presenting flashcards with error words on them to the participant and asking the participant to say the word. If the participant said the word in less than three second the experimenter put the card in a correct pile and went on to the next word. If the word was said incorrectly or three seconds lapsed the experimenter said the word and asked the participant to repeat the word. The card with the error word was then placed back in the stack to be presented again. This was continued until all flash cards were in the correct pile. The participants then continued to the final fluency assessment component of the acquisition package.

In the alternative fluency package all aspects of the treatment package were also identical to the initial fluency-building package including reward, except the

active treatment component, which in this phase consisted of phrase drill. With phrase drill, individual phrases are repeatedly reread by the student to increase fluency. Phrase drill has been empirically demonstrated to be effective for improving reading fluency (Daly, Martens, Dool, & Hintze, 1998 and O'Shae, Munson, & O'Shae, 1984). Following the initial fluency assessment the experimenter presented the participant with the sentences from the reading passage separated, one phrase per page. The participant was instructed to read each sentence three times orally. The experimenter gave no further prompts or instructions except to move on when the participant remained fixed on a word for more than 3 seconds. Following the completion of the fluency package, final fluency assessment component was completed.

#### Procedural Integrity

An observer assessed the integrity during 39% of the intervention. Sessions evaluated were selected randomly across each subject and session and integrity was evaluated by completing a procedural integrity checklist. Procedural integrity was calculated by dividing the number of steps completed in each session by the number of steps required in each session and multiplying by 100. Procedural integrity across all observed sessions was 100%.

#### Comparison to an Alternative Assessment Procedure

Using the results of the extended analysis as a standard for comparison the match between the extended analysis and the treatment recommendations deriving from the treatment based assessment as well as the treatment recommendations deriving from instructional placement standards as defined by Fuchs and Deno (1982)

~~were assessed. The purpose of this process was to determine whether the~~  
recommendations derived from the instructional placement standards defined by Fuchs and Deno (1982) or the treatment based assessment better match the results of the treatment outcomes from the extended validation. This was determined by calculating the percent of match between each assessment and the extended validation.

The percent of match with each assessment was calculated by dividing the number of assessment matches by the total number of cases completing the extended validation. An assessment match for the Instructional Hierarchy existed when the assessment outcome predicted accurately the results of the extended validation. For example, if the assessment outcome reveals that the student is at the Acquisition level of the Instructional Hierarchy and the accuracy treatment was the most effective treatment, the assessment would match the extended validation. Assessment match for the instructional placement standards as defined by Fuchs & Deno (1982) was also calculated. A match with the instructional placement standards exists when treatment recommendations based on instructional placement standards (using data from the initial assessment baseline) matches with the extended validation outcome. For example, the median from the assessment baseline was above 40 words correct per minute, the student was functioning in the instructional range according to instructional placement standards for first or second graders. According to Shapiro (1996) children in the instructional range would benefit from fluency building procedures while those in the frustrational ranges (i.e., first and second graders whose baseline WCM was below 40) would benefit most from acquisition type interventions in order to establish new skills. Hence it would be predicted that the extended

validation outcome for a student in the instructional range would be of a higher level for the fluency based treatment than for a student in the frustrational range who should benefit most from a treatment designed to enhance skill acquisition.

## RESULTS

Table 2 contains each case, the extended validation outcome and the results of the evaluation of match based on the treatment based assessment and the IPS.

Table 2. Outcome results.

Name	Grade	Treatment based assessment outcome	Extended validation outcome	Instructional placement standards recommendation
<b>Acquisition treatment recommendation</b>				
Dana	1 <sup>st</sup>	Acquisition	Acquisition	Fluency
Gerald	1 <sup>st</sup>	Acquisition	Acquisition	Acquisition
Ronnel	1 <sup>st</sup>	Acquisition	Acquisition	Acquisition
Don	2 <sup>nd</sup>	Acquisition	Acquisition	Fluency
Bob	2 <sup>nd</sup>	Acquisition	Acquisition	Acquisition
Kelly	2 <sup>nd</sup>	Acquisition	Acquisition	Acquisition
Clarence	2 <sup>nd</sup>	Acquisition	Acquisition	Acquisition
<b>Fluency treatment recommendation</b>				
Monty	1 <sup>st</sup>	Fluency	Fluency	Fluency
Ken	3 <sup>rd</sup>	Fluency	Fluency	Fluency
Tina	3 <sup>rd</sup>	Fluency	Fluency	Fluency
Jan	3 <sup>rd</sup>	Fluency	Fluency	Acquisition
Didre	3 <sup>rd</sup>	Fluency	Fluency	Acquisition
<b>Both treatments effective</b>				
Rhoda	1 <sup>st</sup>	Acquisition and Fluency	Acquisition and Fluency	Acquisition
Mark	2 <sup>nd</sup>	Acquisition and Fluency	Acquisition and Fluency	Fluency
<b>Two assessments completed with acquisition treatment recommendation</b>				
Karen	1 <sup>st</sup>	Acquisition	Acquisition	Acquisition
Tony	2 <sup>nd</sup>	Acquisition	Acquisition	Acquisition
Connor	3 <sup>rd</sup>	Acquisition	Acquisition	Acquisition
Chris	3 <sup>rd</sup>	Acquisition	Acquisition	Acquisition

### **Assessment and Extended Analysis**

The results of the initial intervention based assessment process indicated that for most children, there were marked gains in reading fluency and clear differentiation between the acquisition and fluency based treatments. For some children the response was greater for the acquisition based treatment and for some the response was greater to the fluency based treatment. For a few children, there was no clear differentiation. Because of the commonalities across children who responded to one treatment versus the other, the results will be discussed separately depending on the response to the initial assessment.

Dana, Gerald, Ronnel's (Figure 1) response to intervention assessment indicated that the acquisition package produced a better response than the fluency package. Don, Bob, and Kelly (Figure 2) and Clarence (Figures 3) also indicated the same. For each participant both the acquisition and the fluency treatment packages produced gains of more than 20%, however, because the greatest increase occurred in response to the acquisition package, this package was considered the appropriate treatment. In the extended validation the acquisition treatment package continued to produce improvements of greater than 20% over the initial fluency assessment while continuing to produce greater improvement than the fluency-based treatment.

Monty, Ken, Tina, Jan and Didre's response to the intervention assessment (see Figure 4 and Figure 5) indicated that there was a better response to the fluency treatment. Both the acquisition and the fluency treatment packages produced gains of

more than 20%, however, because the greatest increase occurred in response to the fluency package this package was considered to the indicated treatment. In the

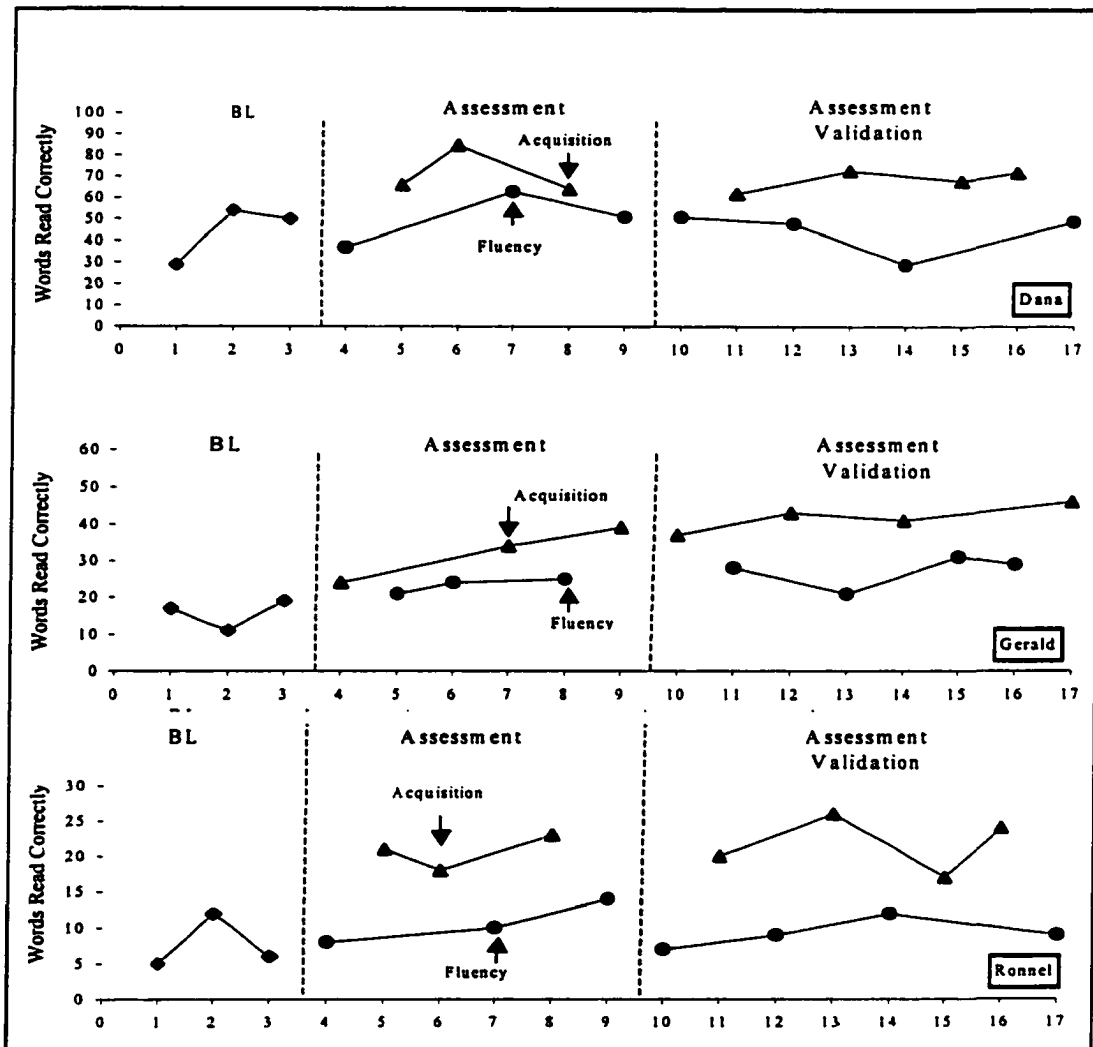


Figure 1. Words read correctly graphed by session and assessment condition for Dana, Gerald and Ronnel.

extended validation the fluency treatment package continued to produce improvements of greater than 20% over the initial fluency assessment while maintaining its advantage over the acquisition based treatment package.

For four participants, Karen, Tony, Connor, and Chris, the first assessment (see Figure 6 and Figure 7) produced gains of more than 20% over the initial fluency

assessment, however, the assessment resulted in overlapping data points between the two treatment packages. Because of this a second assessment was conducted to determine if the assessment could differentiate between the two treatments. The second assessment for all students indicated a potential treatment effect for the

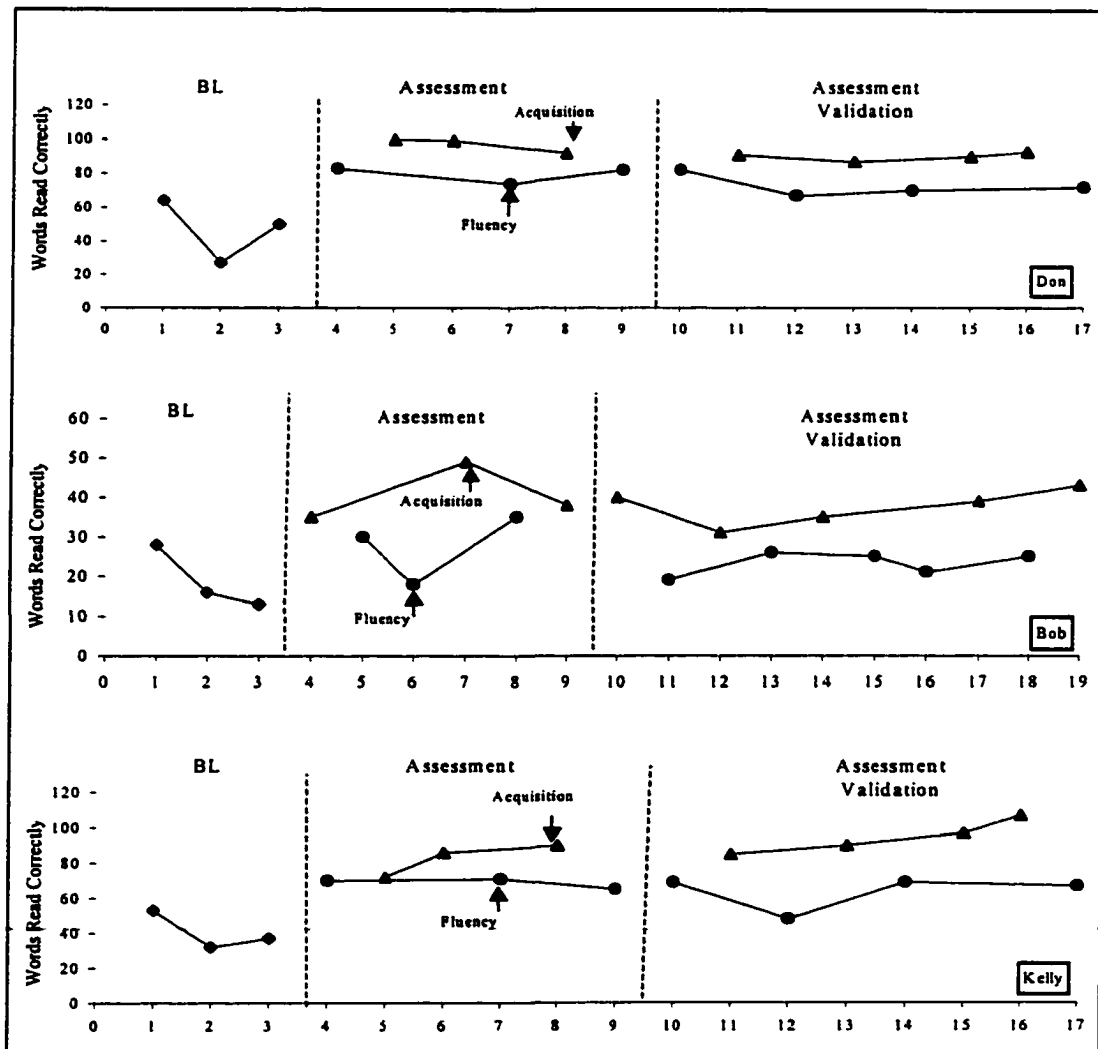


Figure 2. Words read correctly graphed by session and assessment condition for Don, Bob, and Kelly.

acquisition treatment. Both the acquisition and fluency treatment package produced gains of more than 20%, however, because the highest increase occurred in response to the acquisition package this package was considered the appropriate treatment. In



the extended validation, the acquisition treatment package continued to produce improvements of greater than 20% over the initial fluency assessment while continuing to produce greater improvements in reading fluency than the fluency based treatment.

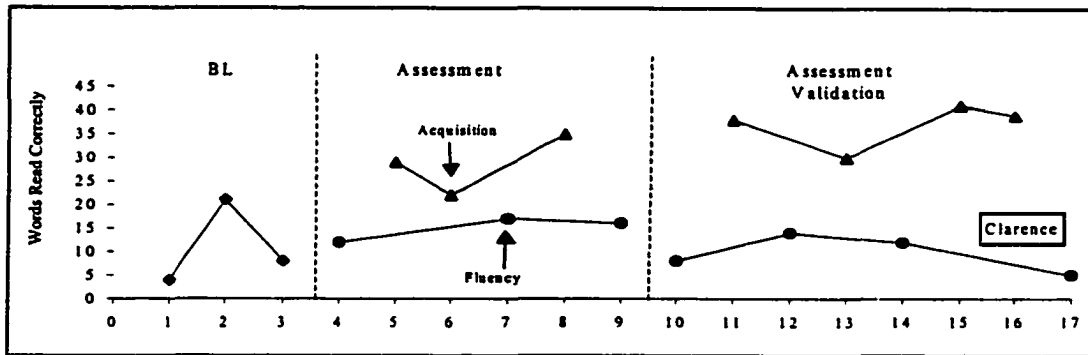


Figure 3. Words read correctly graphed by session and assessment condition for Clarence.

Of the 18 subjects evaluated in this study two subjects revealed a potential treatment effects for both treatments. Rhoda and Mark (see Figure 8), showed improvement of at least 20% over the initial fluency assessment for both treatment packages; however, the data paths for the packages from the first assessment were overlapping. When additional assessment was conducted, this overlap continued during a second and third assessment. Because both packages showed potential treatment effects the acquisition and fluency treatments were both considered to be appropriate treatments for these students and assessment were discontinued. Within these cases because results remained consistent throughout the second and third assessment these were viewed as the extended validation of the initial assessment results.

The Purpose of this phase of the study was to determine the extent to which the assessment procedure had treatment validity. During this phase of the study, the assessment for 16 of the 18 participants resulted in a differentiation between the two

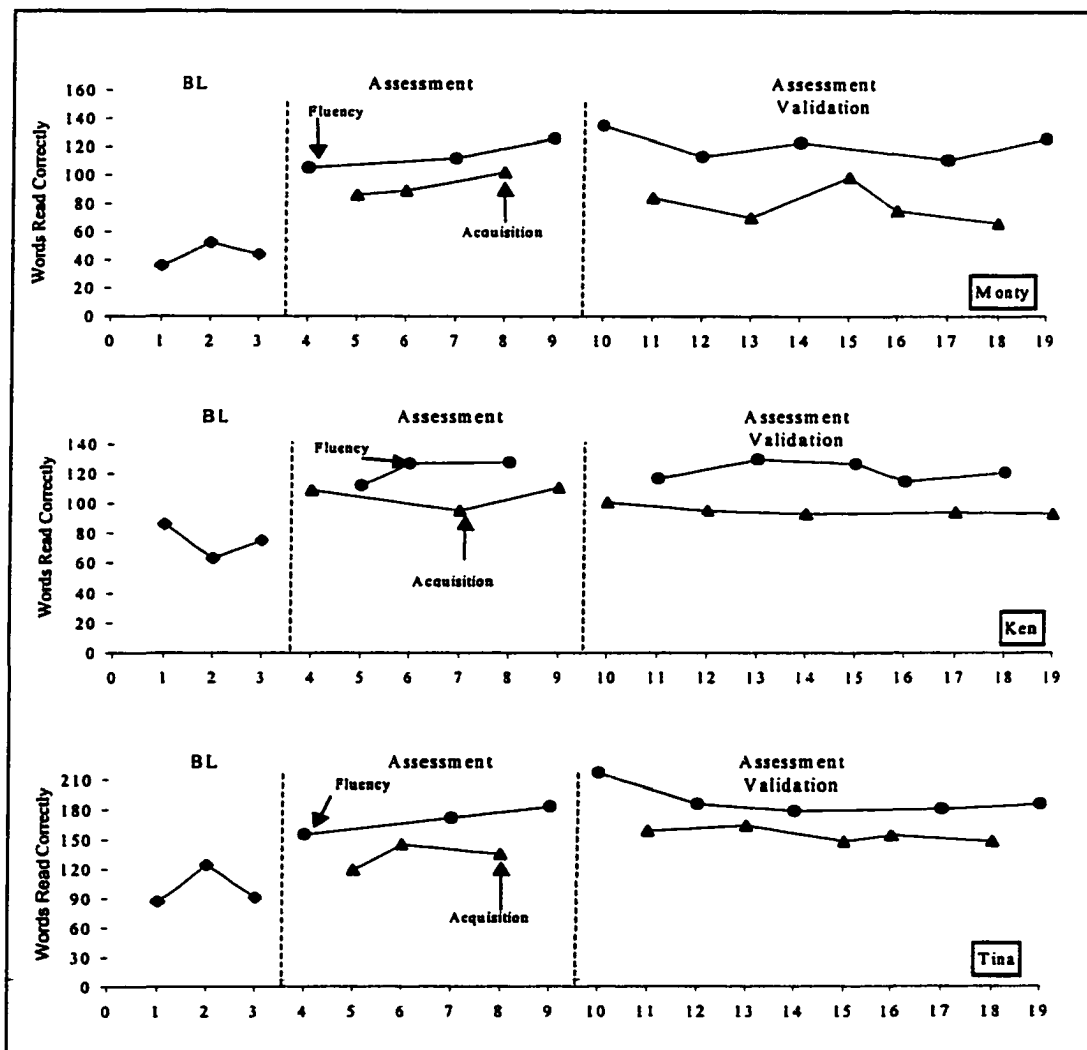


Figure 4. Words read correctly graphed by session and assessment condition for Monty, Ken and Tina.

potential treatments, based on the responses of the children to intervention. For these 16 participants all treatment recommendations matched with the extended validation.

### Extended Analysis with Alternative Treatments

To test the robustness of the IH, Monty, Rhoda and Bob also completed a second phase of the extended analysis in which an alternative treatment was implemented within the treatment packages. The alternative treatment packages substituted an alternative intervention, which incorporated that same active treatment

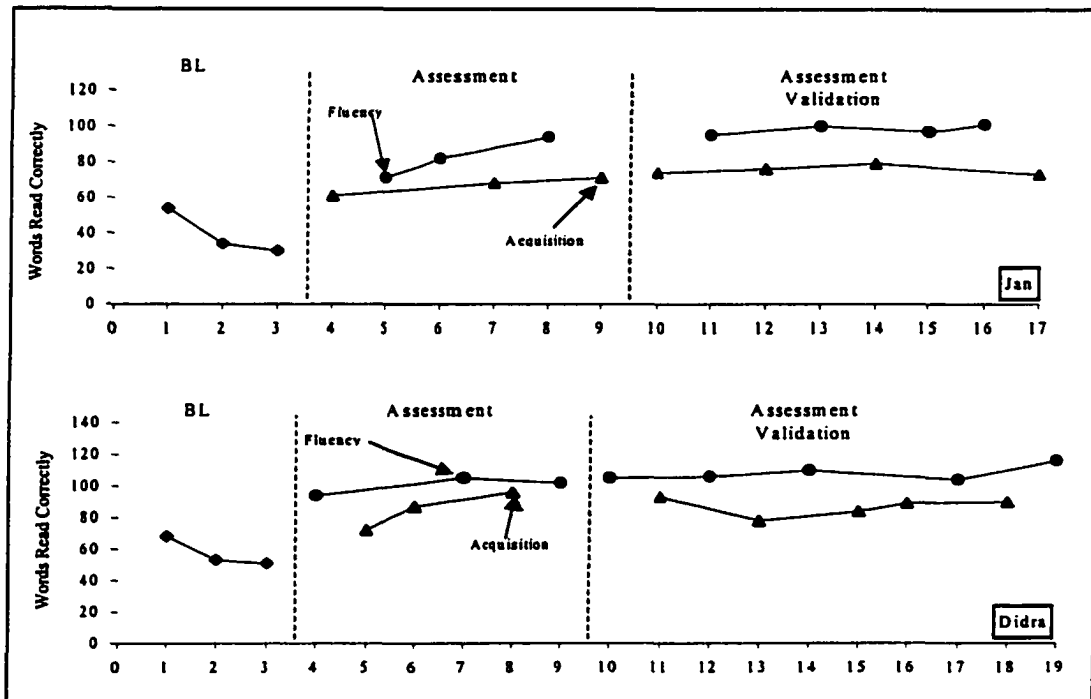


Figure 5. Words read correctly graphed by session and assessment condition for Jan and Didra.

component. In this phase (see Figure 9) the effects of the treatments remained the same relative to each other and the initial fluency evaluation. Bob, who responded more favorably to the acquisition treatment, continued to respond in a similar fashion when the treatment packages were altered. Monty, a better responder to the fluency building treatment, also continued to respond better to the fluency treatment after the treatment packages were altered. For Monty, the magnitude of the effects were slightly less with the alternative treatment packages, but the level and trend were

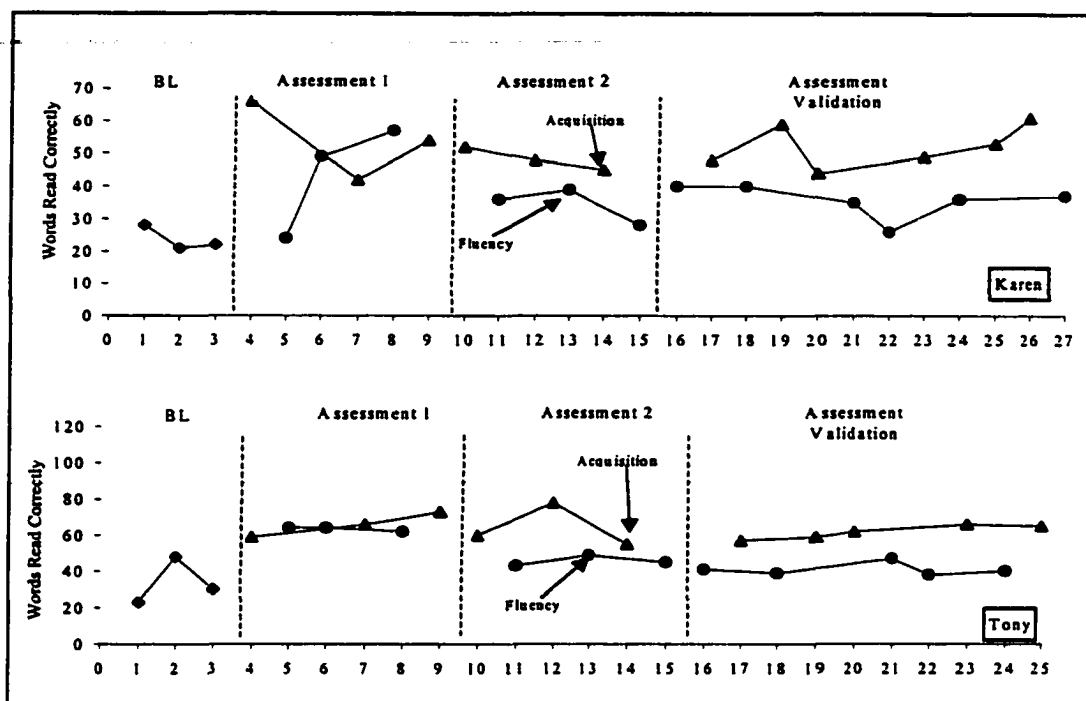


Figure 6. Words read correctly graphed by session and assessment condition for Karen and Tony.

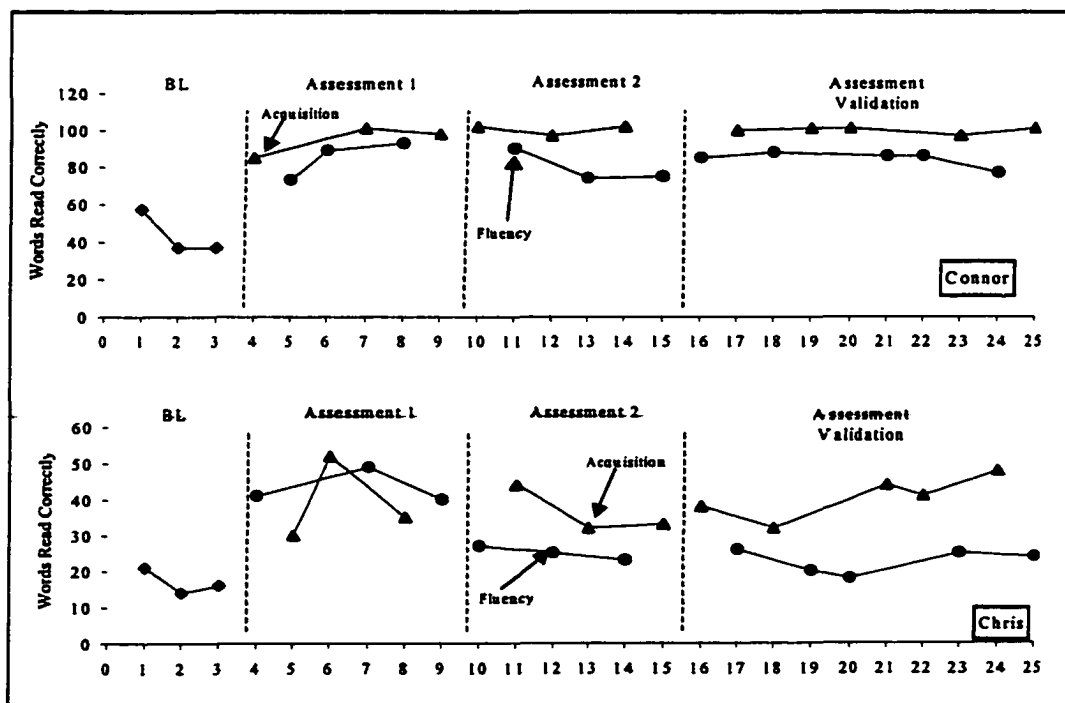


Figure 7. Words read correctly graphed by session and assessment condition for Connor and Chris.

identical. Finally, for Rhoda, who responded to both treatments without clear differentiation between the two, continued to respond similarly to the alternative forms of the treatment.

### Comparison to Instructional Placement Standards

The question of whether the treatment based assessment or the use of instructional placement standards were more accurate in identifying an effective treatment was addressed by calculating the degree of match between each type of assessment with the results of the extended validation. The results of treatment based assessment matched the extended validation for 16 out of 16 cases. This

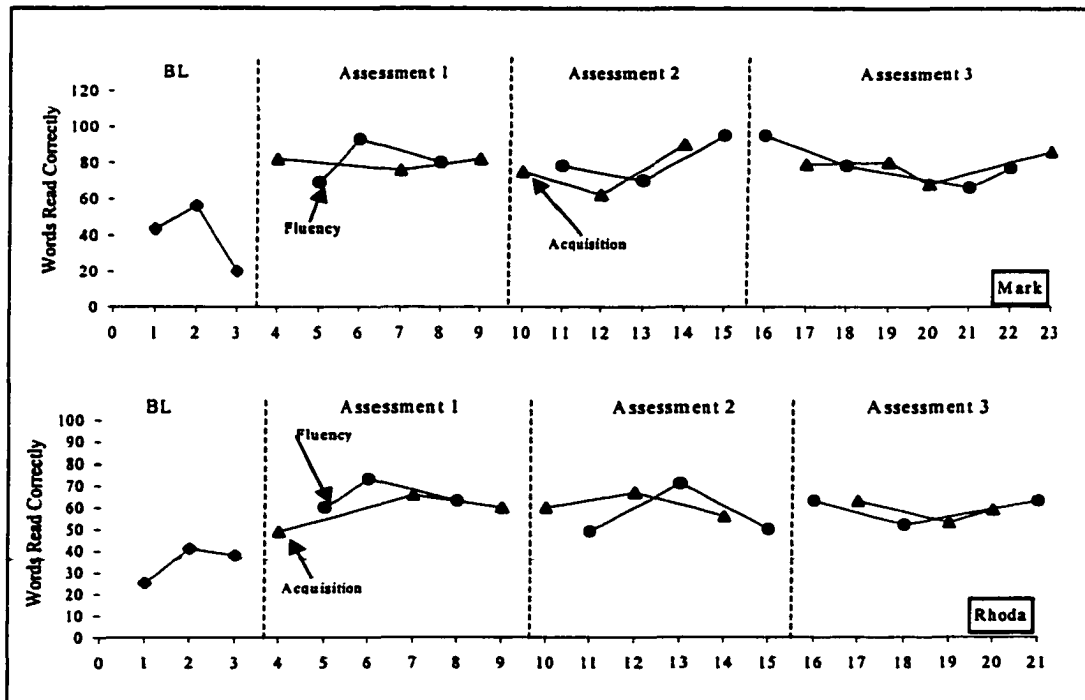
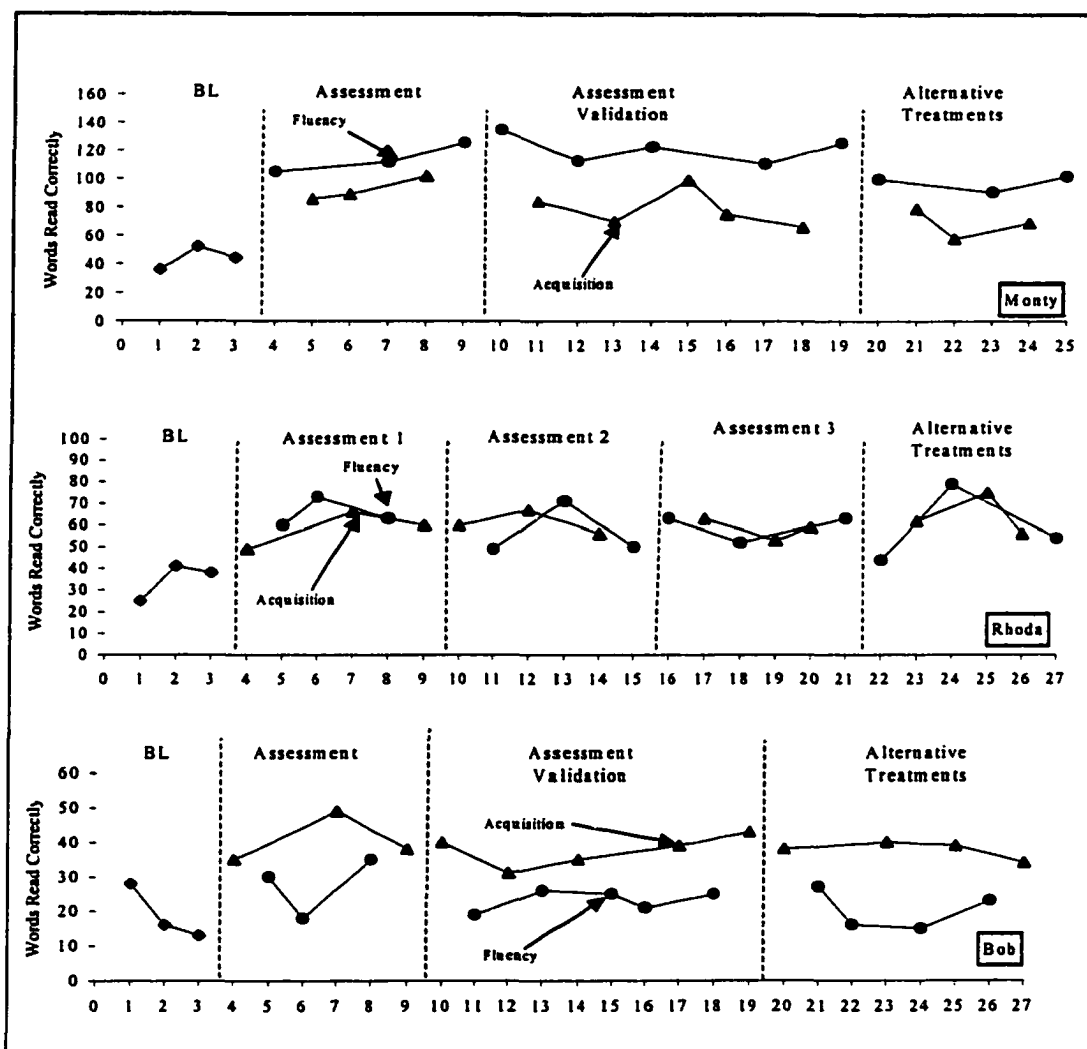


Figure 8. Words read correctly graphed by session and assessment condition for Mark and Rhoda.

resulted in an accuracy of 100%. For the instructional placement standards 11 of the 16 recommendations matched with the extended validation. This resulted in a degree of match of 69%. The two cases where no discrimination occurred were excluded

from this evaluation because there was no extended validation to serve as a standard for comparison.



## DISCUSSION

The primary purpose of this study was to evaluate the treatment validity of a reading fluency assessment, which employed a child's response to treatments as the method for evaluating reading fluency. A secondary purpose was to simultaneously evaluate instructional placement standards described by Fuchs and Deno (1982) as a basis for developing intervention recommendations for reading. Finally this research sought to examine the generalizability of treatment recommendations derived from the treatment based assessment by incorporating the same active treatment component into alternative interventions. This was done to determine if the assessment not only predicted response to the same treatments employed during the assessment, but also predicted responses to similar treatments derived from the instructional hierarchy.

### Summary of Findings

To address the primary research question, treatment validity was evaluated by a comparison of the assessment recommendation derived from the treatment based assessment to the extended analysis. Bases on this comparison, all 16 participants completing the extended validation received a recommendation that proved to be the most valid in the extended analysis. In other words the assessment results predicted the most effective treatment from the extended analysis 100% of the time.

The intervention "recommendations" from the instructional placement standards derived from Shapiro's (1996) suggest that students performing at the frustrational level need accuracy training and students performing at the instructional level benefit from fluency training. That is, intervention development using traditional curriculum based measurement is based upon reading fluency with fluency based

treatments associated with for students in the instructional range and acquisition level instruction for students in the frustrational range (Shapiro, 1996). In evaluating these recommendations, the extended analysis produced a match for 11 of the 16 participants completing the extended validation. This was markedly less accurate in identifying a treatment that matched with the extended validation than the treatment based assessment.

The application of alternative treatments within the extended validation for 3 participants provided information regarding the generalizability of the assessment to other treatments based on the same active treatment component within the instructional hierarchy. For all three participants similar results were obtained for treatments that were related (i.e., they were either acquisition or fluency treatments) based on the instructional hierarchy. The results provide preliminary support for the interchangeability of treatments within the instructional hierarchy.

### Implications

The results support and extend the use of brief functional assessment or treatment based assessments for identifying effective treatments. This extends work by Broussard & Northup (1995), Carr & Durand (1985), Iwata, et al. (1994), Repp (1994), Repp, Flece, & Barton (1988) and others who used brief assessments to identify effective treatments by broadening this methodology to academic behaviors such as reading fluency. Similar to Noell et al. (1998) where reading deficits were identified as related to either deficits in performance or skill, this work focused on the identification of appropriate treatments where the question was which treatment would best meet the needs of a child with a skill deficit. This extends Noell et al. (1998) by



showing that this type of brief assessment can be used to discriminate not only between skill and performance problems but also between different types of skill problems.

The study also has implications for conducting CBM. In current practice the most common method of deriving treatments from curriculum based measurements is thru the use of instructional placement standards. Shapiro (1996), for example, has stated that students in the instructional range are best instructed with interventions designed to build fluency and students in the frustrational range will benefit more from accuracy based treatments which focus on establishing new skills. Based on the results from this study, the use of instructional placement standards for developing treatment was effective in 11 of the 16 cases for participants completing the extended analysis. For these participants the indicated treatment, using only instructional placement standards, was accurate only 68% of the time. Using a treatment based assessment with these same participants proved effective in 16 of 16 cases or 100% of the time.

These findings extends the work of Daly et al. (1999) and Daly et al. (1998) who utilized brief exposures to different instructional procedures to identify procedures that subsequently resulted in marked gains in reading fluency. Because this study incorporated the IH as a conceptual model on which treatments can be based, it allows for the potential of multiple treatment recommendations to derive from a single assessment. That is, if treatments with the same active treatment component are substitutable, then school-based teams have a large menu of interventions from which to choose and can base treatment selection on other important factors such as ease of

implementation. It also goes a step further by basing all treatments on student current grade level. In the work by Daly et al. (1998) curricular level was one variable that was manipulated in the analysis to improve reading. However, this curricular change may be a problem in that it does not answer the question of what procedures are most effective for the student based on where he or she is expected to perform. What I was trying to say is that if I change the students curriculum to make instruction work I can't tell you how to teach him in his regular curriculum.

Finally the results contribute to the growing body of literature pertaining to the instructional hierarchy (Daly et al. 1999; Daly et al. 1998; Daly & Martens 1994; Haring et al. 1978). Given that alternative treatment methods proved similarly effective when derived from the instructional hierarchy, this lends support to the use of the IH as a model for intervention development in reading and potentially other academic areas. These findings extend the work by Daly & Martens (1994) by showing that knowing a child's functional level within the IH may allow other treatments appropriate for that level of the IH to be substituted. It should be noted, however, that the small number of subjects completing this aspect of the study limit the confidence of any statement made with regard to the ability of the assessment to identify skill levels based on the IH.

Results pertaining to the two participants whose results during the assessment did not differentiate may also lead to other research, which could contribute to theory and practice regarding the IH and its relation to reading acquisition. For both subjects there was no differentiation across several sessions, however, both subjects showed marked gains in reading fluency as a result of each treatment. In examining these

results it may be that, for reading, the IH should not be viewed as a dichotomy, but a continuum. In other words, perhaps students move gradually from the acquisition to the fluency level within the IH, it is possible that at any given time, their skills may be acquisition for some words and fluency for other. For example, with each word we are asked to read, it is possible that we are at an acquisition, fluency, generalization or adaptation skill level. Because of this in any given passage, a complex combination of skills may exist. However, if the vast majority of words in a passage are at the acquisition skill level for us it could be assumed that an acquisition treatment would be most effective for improving reading. However, if a near equal distribution of acquisition and fluency words is present both acquisition and fluency building treatments may be effective. Although this question has not been posed empirically it is potentially an area of future research. As a practical matter, if the assessment indicated that both treatments were equally effective, then that is also valuable info for school based decision making.

### Limitations

While the results of the current study are encouraging there are at least 4 limitations should be noted. First, as mentioned in the previous section, the small number of students with whom the alternative treatments was conducted make it impossible to make any definitive statements about the degree to which treatments with the same active treatment component are interchangeable. However, the data so far are encouraging. Second, this study was conducted outside of the classroom setting during which reading instruction continued in the regular classroom. It is possible that some instructional procedures within the student's regular classroom

could have affected the results. However, due to the brief nature of this procedure and the differentiation between the two treatments for most of the participants, this threat to internal validity is unlikely to have been present. If classroom instruction did have an effect on fluency levels obtained in this study, these effects would most likely have occurred across both treatments. Third, because there were no generalization data taken for reading fluency, it is not possible to determine if either treatment had any effect on reading outside of the study. Relatedly, the study focused only on oral reading fluency. While fluency was selected as an important and reliable component of reading (Fuchs & Fuchs, 1986; Shinn, 1989), it is not known the extent to which the results would generalize to other important components of reading such as reading comprehension. Finally, the brief nature of the extended analysis may have contributed to false conclusions about the most effective treatment. Many of the extended analyses consisted of 8 to 10 sessions. Although the extended validation produced stable data, a more long term evaluation of the effects of each treatment could reveal that, over time, the two treatments may not prove as stable.

### Future Research

There are many directions in which future research in this area can be taken. This study and its results could be extended in several ways. First, refining the assessment procedure to minimize the time and effort required to identify an effective treatment will be important. . With this assessment a total of 9 data points were collected prior to a recommendation. Each data point in the assessment and validation phase consisted of approximately 5 minutes of one on one contact with the student, therefore, the total time for assessment was about 45 minutes. This may not be a large

amount of time compared to other forms of reading assessment, but it may yet be possible to improve upon these procedures so recommendations could be made in a much shorter time frame. Second, the validation phase could be extended even further to determine the long term effectiveness of these treatments as well as the long term stability of the treatments derived from the assessment. Third, alternative treatments could be applied to a larger number of participants to evaluate the assessment's ability to identify student skill with respect to the instructional hierarchy. What I was trying to say here was that by attempting to replicate the alternative treatments with more subjects a more definitive statement about this assessments ability to identify student skill level based on the IH could be made.

The methodology could also be used with other academic or behavioral skill areas. Once basic procedures are established, other behaviors and treatments could be substituted to evaluate whether the procedures used here are more broadly applicable. The development of such procedures could potentially improve the efficiency of assessment procedures and provide treatments that are effective sooner.

In summary, the procedures utilized in this study were successful for identifying effective treatments directed toward the improvement of reading fluency by evaluating participant responses to various treatments. The present results will need to be replicated, but these data provide preliminary support for the use of assessment procedures that are derived from treatments for academic problems. Perhaps there are more questions raised than answered by these results and certainly this study provides areas of future exploration. Although additional research is needed to refine the procedures used in this study, these data suggest that treatment based

assessments can potentially serve as a powerful assessment tool for treatment identification.

## REFERENCES

Arter, J. A., & Jenkins, J. R. (1979). Differential diagnosis-prescriptive teaching: A critical appraisal, Review of Educational Research, 49, 517-555.

Ayres, R. R., & Cooley, E. J. (1986). Sequential versus simultaneous processing on the K-ABC: Validity in predicting learning success. Journal of Psychoeducational Assessment, 4, 211-220.

Blankenship, C. S. (1985). Using curriculum-based assessment data to make instructional management decisions. Exceptional Children, 42, 233-238.

Campbell, K. U. (1998). Great leaps elementary stories collection (k-5). Gainesville, FL: Diarmunid.

Carnine, D., Silbert, J., & Kameenui, E. J. (1990). Direct instruction reading (2<sup>nd</sup> ed.). Columbus, OH: Merrill.

Carr, E. G. (1977). The motivation of self-injurious behavior: A review of some hypotheses. Psychological Bulletin, 84, 800-816.

Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. Journal of Applied Behavior Analysis, 18, 111-126.

Clariana, R. B. (1993). Using readability formulas to establish the grade level difficulty of software. Journal of Computing in Childhood Education, 4, 115-161.

Daly, E. J., III, Martens, B. K., Hamler, K. R., Dool, E. J., & Eckert, T. L. (1999). A brief experimental analysis for identifying instructional components needed to improve oral reading fluency. Journal of Applied Behavior Analysis, 32, 83-94.

Daly, E. J. III, Martens, B. K., Dool, E. J., & Hintze, J. M. (1998). Using brief functional analysis to select interventions for oral reading. Journal of Behavioral Education, 8 (2), 203-218.

Daly, E. J. III, & Martens, B. K. (1994). A comparison of three interventions for increasing oral reading performance: Application of the instructional hierarchy. Journal of Applied Behavior Analysis, 27, 459-469.

Daly, E. J., III, & Murdoch, A. (2000). Direct observation in the assessment of academic skill problems. In E. S. Shapiro & T. R. Kratochwill (Eds.), Behavioral assessment in schools: Theory, research, and clinical foundations (2<sup>nd</sup> ed.). New York: Guilford.

Daly, E. J. III, Lentz, F. E. Jr., & Boyer, I. (1996). The instructional hierarchy: A conceptual model for understanding the effective components of reading interventions. School Psychology Quarterly, 11 (4), 369-386.

Daonahue, P. L., Finnegan, R. J., Lutkus, A. D., Allen, N. L., & Campbell, J. R. (2001). NAEP 2000 reading report card for the nation and the states. Washington, DC: National Center for Education Statistics.

Delquadri, J., Greenwood, C. R., Whorton, D., Carta, J., & Hall, R. V. (1986). Classwide peer tutoring. Exceptional Children, 52, 535-542.

Deno, S. L. (1987). Curriculum-based measurement. Teaching Exceptional Children, 20, 41.

Deno, S. L., & Mirkin, P. K. (1977). Data-based program modification: A manual. Reston, VA: Council for Exceptional Children.

Fleisher, L. S. & Jenkins, J. R. (1983). The effect of word and comprehension emphasis instruction on reading performance. Learning Disability Quarterly, 6 (2), 146-154.

Fuchs, L. S., & Deno, S. L. (1982). Developing goals and objectives for educational programs. Washington: American Association of Colleges for Teacher Education.

Gickling, E., & Havertape, J. (1981). Curriculum-based assessment. Minneapolis, MN: National School Psychology Inservice Training Network.

Gickling, E. E., & Rosenfield, S. (1995). Best practices in curriculum-based assessment. In A. Thomas & J. Grimes (Eds.), Best practices in school psychology-III (pp.587-595). Washington, DC: National Association of School Psychologists.

Good, R. H., III, Vollmer, M., Creek, R. J., Katz L., and Chowdri, J. (1993). Treatment utility of the Kaufman Assessment Battery for Children: Effects of matching instruction and student processing strength. School Psychology Review, 22, 8-26.

Gresham, F. M. (1992). Misguided assumptions of DSM-III: Implications for school psychological practice. School Psychology Quarterly, 7, 79-95.

Gresham, F. M. (1981). Assessment of children's social skills. Journal of School Psychology, 19, 120-133.

Gresham, F. M., & Witt, J. C. (1997). Utility of intelligence test for treatment planning, classification, and placement decisions: Recent empirical findings and future directions. School Psychology Quarterly, 12, 249-267.



Goldfried, M. R., & Kent, R. N. (1972). Traditional versus behavioral personality assessment: A comparison of methodological and theoretical assumptions. Psychological Bulletin, 77(6), 409-420.

Hambleton, R. K. (1982). Advances in criterion-referenced testing technology. In C. R. Reynolds & T. B. Gutkin (Eds.), The handbook of school psychology. New York, NY: John Wiley & Sons.

Haring, N. G., Lovitt, T. C., Eaton, M. D., & Hansen, C. L. (1978). The fourth R: Research in the classroom. Columbus, OH: Merrill.

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. American Psychologist, 45, 999-1017.

Howell, K. W., Fox, S. L., & Moorhead, M. K. (1993). Curriculum-based evaluation: Teaching and decision making (2<sup>nd</sup> ed.). Pacific Grove, CA: Brooks/Cole.

Iwata, B., Dorsey, M., Slifer, K., Bauman, K., & Richman, G. (1982). Toward a functional analysis of self-injury. Analysis and Intervention in Developmental Disabilities, 2, 3-20.

Iwata, B. A., Pace G. M., Cowdery, G. E., & Miltenberger, R. G. (1994). What makes extinction work: An analysis of procedural form and function. Journal of Applied Behavior Analysis, 27, 131-144.

Karsh, K. G., Repp, A. C., Dahlquist, C. M., & Munk, D. (1995). In vivo functional assessment and multi-element interventions for problem behaviors of students with disabilities in classroom settings. Journal of Behavioral Education, 5(2), 189-210.

Korchin, S. J., & Schulberg, D. (1981). The future of clinical assessment. American Psychologist, 36, 1147-1158.

Lalli, E. P., and Shapiro, E. S. (1990). The effects of self-monitoring and contingent reward on sight word acquisition. Education and Treatment of Children, 13, 129-141.

Lahey, B. B., & Drabman, R. S. (1973). Facilitation of the acquisition and retention of sight word vocabulary through token reinforcement. Journal of Applied Behavior Analysis, 6, 101-104.

Lovitt, T. C., Eaton, M., Kirkwood, M. E., & Perlander, A. (1971). Effects of various reinforcement contingencies on oral reading rate. In E. Ramp & B. L.

Hopkins (Eds.). A new direction for education: Behavior analysis (pp. 54-71). Lawrence: University of Kansas Press.

Lovitt, T., & Hansen, C. (1976b). Round one--placing the child in the right reader. Journal of Learning Disabilities, 9, 347-353.

Martens, B. K. (1992). The difference between a good theory and a good treatment is a matter of degree. School Psychology Quarterly, 7(2), 104-107.

McMacmann, B. M., & Barnett, D. W. (1997). Reliability interpretations for Kaufman's "intelligent testing" approach to the WISC-III. School Psychology Quarterly, 12, 197-234.

Medcalf, J. (1989). Comparison of peer tutored remedial reading using the pause, prompt and praise procedure with an individualized tape-assisted reading programme. Educational Psychology, 9 (3), 253-262.

Mirkin, P., Fuchs, L., & Deno, S. (1982). Considerations for designing a continuous evaluation system: An integrative review (Monograph No. 20). Mineapolis: University of Minnesota, Institute for Research on Learning Disabilities.

Neef, N. A., & Iwata, B. A. (1994). Current research on functional analysis methodologies: An introduction. Journal of Applied Behavior Analysis, 27, 211-214.

Noell, G. H., Gansle, K. A., Witt, J. C., Whitmarsh, E. L., Freeland, J. T., LaFleur, L. H., Gilbertson, D. A. & Northup, J. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. Journal of Applied Behavior Analysis, 31, 659-664.

O'Shae, L. J., Munson, S. M. & O'Shae, D. J. (1984). Error correction in oral reading: Evaluating the effectiveness of three procedures. Education and Treatment of Children, 7 (3), 203-214.

Phillips, N. B., Fuchs, L. s., & Fuchs, D. (1994). Effects of classwide curriculum-based measurement and peer tutoring: A collaborative research-practitioner interview study. Journal of Learning Disabilities, 27, 420-434.

Rashotte, C. A., and Torgesen, J. K. (1985). Repeated reading and reading fluency in learning disabled children. Reading Research Quarterly, 20, 180-188.

Repp, A. C. (1994). Comments on functional analysis procedures for school-based behavior problems. Journal of Applied Behavior Analysis, 27, 409-411.

Repp, A. C., Felce, D., & Barton, L. E. (1988). Basing the treatment of stereotypic and self-abusive behaviors on hypotheses of their causes. Journal of Applied Behavior Analysis, 21, 281-290.

Repp, A. C., & Karsh, K. G. (1994). Hypothesis-based interventions for tantrum behaviors of persons with developmental disabilities in school settings. Journal of Applied Behavior Analysis, 27, 21-31.

Rose, T. L., McEntire, E., & Dowdy, C. (1982). Effects of Two error correction procedures on oral reading. Learning Disability Quarterly, 5, 100-105.

Rose, T. L., and Sherry, L. (1984). Relative effects of two previewing procedures on LD adolescents' oral reading performance. Learning Disability Quarterly, 7, 39-44.

Salend, S. J., Nowak, M. R. (1988). Effects of peer-previewing on LD students' oral reading skills. Learning Disability Quarterly, 11(1), 47-53.

Salvia, J. A., & Ysseldyke, J. E. (1995). Assessment in special and remedial education (6<sup>th</sup> ed.). Boston: Houghton Mifflin.

Shapiro, E. S. (1996). Academic skills problems: Direct assessment and intervention (2<sup>nd</sup> ed.). New York: Guilford.

Shinn, M. R. (Ed.). (1989). Curriculum-based Measurement: Assessing Special Children. New York: Guilford Press.

Singh, N. N., Singh, J. (1986a). Increasing oral reading proficiency: A comparative analysis of drill and positive practice overcorrection procedures. Behavior Modification, 10(1), 115-130.

Singh, N. N., Singh, J. & Winton, A. S. (1984). Positive practice overcorrection of oral reading errors. Behavioral Modification, 8(1), 23-37.

Sprague, J. R., & Horner, R. H. (1992). Covariations within functional response classes: Implications for treatment of severe problem behavior. Journal of Applied Behavior Analysis, 25, 735-745.

Starlin, C. M. (1982). On reading and writing. Iowa Monograph Series. Des Moines, Iowa: Department of Public Instruction.

Starlin, C. M. (1970). The use of daily direct recording as an aid in teaching oral reading. Unpublished doctoral dissertation, University of Oregon.

Starlin, C., & Starlin, A. (1973). Guides for continuous decision making. Bemidji, MN: Unique Curriculum Unlimited.

Stoddard, K., Valcante, G., Sindelar, P., O'Shea, L., & Algozzine, B. (1993). Increasing reading rate and comprehension: The effects of repeated readings, sentence

~~segmentation, and intonation training. Reading Research and Instruction, 32(4), 53-65.~~

Witt, J. C., Elliott, S. N., Daly, E. J. III, Gresham, F. M., Kramer, J. J. (1998). Assessment of at-risk and special needs children (2<sup>nd</sup> ed.). Boston, MA: McGraw-Hill.

Ysseldyke, J. E., & Mirkin, P. K. (1982). The use of assessment information to plan instructional interventions: a review of the research. In C. R. Reynolds & T. B. Gutkins (Eds.), The handbook of school psychology (pp. 395-409). New York: Wiley.

## **APPENDIXES**

### **Appendix 1**

#### **Sample first grade reading passage**

##### **The Wind Slows Cold**

**The wind blows cold**

**Blows oh so cold**

**Freezing me to the bone**

**as cold as marble stone Blowing on and on. (26)**

**The damp chills me and I shiver**

**Freezing me to the bone**

**as cold as marble stone Going on and on. (47)**

**My toe bones ache my ears burn**

**I'm frozen to the bone**

**as cold as marble stone Aching on and on. (68)**

**The winter goes on I am so tired**

**and frozen to the bone**

**as cold as marble stone Going on and on. (go)**

**The wind blows cold**

**Blows oh so cold**

**Freezing me to the bone as cold as marble stone blowing on and on. (112)**

## Appendix 2

### Sample second grade reading passage

#### Excuses

I would make all A's if my teacher wasn't mean. If I only had the time, my room would be clean.

If you were nice to me, I wouldn't make a scene.

I would make all A's if my teacher wasn't mean. (42)

Seems you've got an excuse for everything. So you don't ever have to do anything. It all brings to mind an old, old saying

About a scared old frog on a stump just praying, (76)

"if I only had wings

I would take that jump. If I only had wings

My butt wouldn't have to bump.

If I had my wings

I'd up and leave this stump. If I had my wings

Then, I'd take that jump. (118)

I would take that jump

and be not afraid

I would leave this stump," The old frog prayed.

He remained afraid, while he prayed and prayed, He wouldn't jump, so he stayed and stayed.

Frogs don't have wings, it's the nature of things - and when they jumps, their butts do bumps. (169)

### **Appendix 3**

#### **Sample third grade reading passage**

##### **Preventing a Cold**

Do you catch colds every winter? If you do, then there are things you can do to help prevent them; things certainly worth the try. (28)

During the cold season it is wise to wash your hands several times a day. The cold "germs" on your hands will then be washed down the sink.

Better there than in you. (61)

You need to get out of the habit of rubbing your eyes. Many doctors think that is the main way you catch a cold. The cold enters your body through the eyes. Soon, you become sick with a cold. (100)

You can eat fresh fruit and take Vitamin C to prevent colds. Stay in shape with exercise and get plenty of sleep each night. If you are strong, you may be able to fight off infection. (136)

By doing these things, you may avoid a cold next year. In fact, you may never catch another cold. Now wouldn't that be nice? (160)

## Appendix 4

### **Protocol for Acquisition based treatment.**

Check all that the experimenter completed.

1. \_\_\_\_ Select the appropriate passage for the session.
2. \_\_\_\_ Have the student read a grade level passage for a one-minute timing. Be sure to mark any words as errors that are said incorrectly or take longer than 3 seconds to pronounce.
3. \_\_\_\_ Score the reading passage and record any error words on the "Error Correction worksheet".
4. \_\_\_\_ Allow for a one-hour time lapse.
5. \_\_\_\_ Identify a point in the same passage that is 1.5 times that of the student's score from the reading in #2. Read the passage back to the student to that point having them follow along while it is read.
6. \_\_\_\_ From the "Error Correction Worksheet" present each word from the error list and ask the student to pronounce the word. If the student reads it correctly move on to the next word. If the student takes more than 3 seconds to say the word or pronounces the word incorrectly begin the error correction procedure.  
Error Correction Procedure - pronounces each syllable of the incorrect word orally then blend the phoneme segments together to form the correct word. Have the student repeat the phonemic naming and the phonemic blending. Repeat this for all words that are not said correctly.
7. \_\_\_\_ Present the performance goal to the student (words read correctly from #3).
8. \_\_\_\_ Have the student read the same passage for a one-minute timing. Score words read and errors in the student's scoring sheet.

Number of steps completed \_\_\_\_\_

Integrity \_\_\_\_\_/8 = \_\_\_\_\_%



## Appendix 5

### Integrity Checklist for Fluency Building based treatment.

Check all that the experimenter completed.

1. \_\_\_\_ Select the appropriate passage for the session.
2. \_\_\_\_ Have the student read a grade level passage for a one-minute timing. Be sure to mark any words as errors that are said incorrectly or take longer than 3 seconds to pronounce.
3. \_\_\_\_ Score the reading passage for words read correctly and errors.
4. \_\_\_\_ Allow for a one-hour time lapse.
5. \_\_\_\_ Have the student read the same passage for a one-minute timing. Score words read and errors on the student's scoring sheet.
6. \_\_\_\_ Have the student read the same passage for another one-minute timing. Score words read and errors on the student's scoring sheet.
7. \_\_\_\_ Have the student read the same passage for another one-minute timing. Score words read and errors on the student's scoring sheet.
8. \_\_\_\_ Present the performance goal to the student (words read correctly from #3).
9. \_\_\_\_ Record the words read correctly and errors from the 4th reading on the student's record sheet.

Number of steps completed \_\_\_\_\_

Integrity \_\_\_\_\_/9 = \_\_\_\_\_%

## **VFFA**

Gary Joseph Duhon, Jr. began his work in psychology at Nichols State University of Thibodaux, Louisiana, and obtained his bachelor of arts in psychology in May of 1992. His work within the graduate program in psychology at Louisiana State University began in 1997 where a specialty of school psychology was pursued. Within the program a master of arts was awarded in December of 1999 followed by the degree of doctor of philosophy in December of 2001. Gary Duhon is currently a faculty member in the school psychology program at Oklahoma State University in Stillwater, Oklahoma.


# DOCTORAL EXAMINATION AND DISSERTATION REPORT

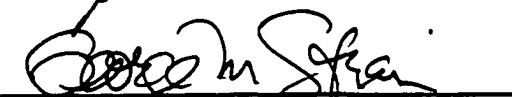
**Candidate:** Gary J. Duhon Jr.

**Major Field:** Psychology


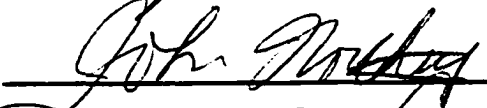


**Title of Dissertation:** A Treatment Based Reading Assessment

**Approved:**

  
Major Professor and Chairman

  
Dean of the Graduate School

## EXAMINING COMMITTEE:

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

**Date of Examination:**

26 June 2001