REMEDIAL OUTCOMES WITH DIFFERENT READING PROGRAMS

Remedial Outcomes with Different Reading Programs

Stephen Truch

The Reading Foundation

Poster session presented at the International Dyslexia Conference, San Diego, 2004

Abstract

In this paper, pre and post-test results are first presented on a relatively new research-based remedial reading program called *Discover Reading*. This program was developed at The Reading Foundation, a private clinic in Calgary, Alberta, that offers remedial reading services to students of all ages. Pre and post-test data for 146 students given 80 hours of one-on-one intervention using the *Discover Reading Program* is presented. Strong and significant gains ($p<.001$) were noted for phonemic processing; letter and sound knowledge; visual memory; word attack; word recognition; spelling and contextual fluency.

Outcomes from the *Discover Reading Program* are then compared to the outcomes from two other programs previously used at the clinic – the *Lindamood Phoneme Sequencing Program* (LiPS) (Lindamood and Lindamood, 1998) and *Phono-Graphix* (McGuinness and McGuinness, 1998). Since all three programs contain many of the features required for effective remediation, the author hypothesized that remedial outcomes among the three programs would not be significantly different from each other, which indeed was supported by the statistical analysis. All three programs produced strong gains on all variables with large effect sizes also noted for each program.

This author also presents a comprehensive framework of features that can be used to analyze any remedial program. Remedial programs with strong outcomes still differ

from each other on important features and the framework can be used to analyze similarities and differences among programs.

Remedial Outcomes with Different Reading Programs

Background

Reading research since the mid 1980's has consistently shown that remedial reading programs that develop phonemic awareness bring positive outcomes for students with reading disabilities (Torgesen, Wagner & Rashotte, 1997). Indeed, a compelling case for a *causal* connection between weak phonemic awareness and reading and spelling was first postulated by Bradley & Bryant (1983).

Research in phonemic processing and its relationship to both reading and spelling is prolific (Vellutino, Fletcher, Snowling and Scanlon, 2004); has continued for several decades and has also been conducted in languages other than English (e.g. Olofsson, A., & Lundberg, I., 1985). A causal hypothesis has been supported by much of that research (Wagner, 1988).

Developing weak phonemic awareness is now seen as one of the essential components in teaching children to read and a cornerstone for any remedial program or intervention (see for example, the *Report of the National Reading Panel*, 2000).

However, while stimulating phonemic awareness is necessary for effective remediation, *by itself* such stimulation does nothing for reading unless the link between phonemes and graphemes is *also* directly taught to the student (Truch, 1991; 1993; 2005).

It is therefore critically important for remedial reading, that as many of the "known ingredients" for effective remediation be incorporated into a cohesive program. Creating such programs is beyond the reach of many special educators due to lack of time, expertise, or both. Most teachers appreciate the value of lessons that are already prepared. Therein lies the value of a comprehensive and commercially-produced remedial reading program. Reading programs that include many or most of the research-based "ingredients" are very useful to classroom teachers. However, the number of such commercial programs is still limited and of the ones that do exist, pre and post-test results are often lacking. As a consequence, teachers who want research reports about a specific commercial remedial reading program usually cannot find much evidence for its effectiveness in the scientific literature.

The information presented in this study deals with the outcomes from three commercial remedial reading programs. The information should therefore be of some practical value to special educators, school psychologists and school districts that have the responsibility of providing remedial services to students. Educators often purchase commercial programs but the paucity of research on outcomes using them can make the decision of which program to use much more difficult. The data in this paper can therefore be of some use to educators facing such decisions.

While the research on phonemic awareness and reading is prolific, there are still many areas to explore and questions to answer, including methodology (Troia, 1999). Of concern to most special educators though, are questions like: "Is Program x better than

Program y?" "Is there truly any one program that produces the best results for students?" "Is Program x useful for special populations other than the reading disabled?" "Does Program x contain spelling activities?" And so on. This paper provides a framework to respond to most of those questions.

While research on commercial remedial reading programs is scarce, *some* research data is available. For example, the author was trained (in 1988) in a commonly used and phonemically-based intervention program called (at that time) the *Auditory Discrimination in Depth Program*. He then started The Reading Foundation clinic in Calgary in 1990. The clinic used the *Auditory Discrimination in Depth Program* (now called the *Lindamood Phoneme Sequencing Program* or *LiPS* – Lindamood and Lindamood, 1998) for several years in the "intensive immersion" fashion pioneered at the Lindamood-Bell Clinic. As at the Lindamood-Bell clinics, students attend The Reading Foundation clinic for four hours of one-on-one remediation each day for an average of four weeks or 80 hours of remedial immersion in any given program. (The Reading Foundation uses four different programs to meet various learning needs in reading, writing and math.) Students are first assessed for their needs and the clinical interventions are designed to strengthen their areas of weakness so not all clients receive any phonemically-based remediation.

Truch compiled the results of using the *LiPS Program* in this intensive immersion fashion with Reading Foundation clients and in 1994, the pre and post-test outcomes on

280 clients were published in *Annals of Dyslexia* (Truch, 1994). Significant gains ($p<.001$) on all variables were noted.

In 1997, the author was trained in another phonemically-based program called *Phono-Graphix* or *PG*. (McGuinness C., & McGuiness, G., 1998). Outcomes from the *PG* program were first published by the developers of the program in *Annals of Dyslexia* in 1996 (McGuinness, C., McGuinness D., & McGuinness, G., 1996). The authors of the *PG Program* claimed to achieve outcomes much faster than other remedial programs such as *LiPS.* The Reading Foundation in Calgary made the transition from using the *LiPS Program* to the *PG Program* partly because of that claim and again compiled data on the outcomes for 195 students who were "immersed" in the *PG Program*. Highly significant pre and post-test treatment gains ($p<.001$) were also noted for the *PG Program* (Truch, 2003).

However, the clinical data on the 195 clients at The Reading Foundation did not support the claim for much faster outcomes for the *PG Program* compared to the *LIPS Program.* The *PG Program* did produce some significant changes in students on important reading variables that were noticeable in as little as 12 hours of instruction as claimed, but in terms of overall outcomes after the same amount of one-on-one instructional time (80 hours) in the same clinic (The Reading Foundation), statistical comparison between the two programs showed that, with minor differences, the outcomes between them were about the same (Truch, 2003). After seeing these results the author concluded that no one program was likely to produce better outcomes than another so

long as the program contained the major essential ingredients needed for effective remediation and sufficient time is provided for students to respond to it.

We do know that some of the "essential instructional ingredients" for any program include activities that develop phonemic awareness; provide direct instruction in letter and sound connections and give practice in helping students use these to "sound out" words and spell them. However, there are still many questions to answer about the classroom implementation of such skills, particularly in terms of how much time needs to be spent in each of these areas (*Report of the National Reading Panel*, 2000). Other features of a remedial reading program such as contextual applications, fluency practice, attention to visual memory for letters, a sequential approach to teaching decoding of simple, complex and multisyllable words, morphological and comprehension activities are also necessary but again, very little is know about how much instructional time needs to be devoted to such activities. And reading programs including commercial ones, do vary greatly in how comprehensive such teaching activities are. For example, the *PG Program* has activities dealing with multisyllable words, but those teaching activities are very limited in scope compared to those used in both the *LiPS* and *Discover Reading* programs. And neither the *LiPS* or *PG* programs contain *any* activities specifically directed to comprehension. Spelling activities are given more attention in the *LiPS* and *Discover Reading* programs compared to the *PG* program. The *PG Program* provides no activities for morphological processing, and so on. Educators often have no way to compare such features (or even to know what the features should be) from the literature provided by the publisher or authors of the programs or from the scientific literature.

The Appendix in this article contains a list of many such features and compares the three programs discussed in this study to each other.

Since effective phonemically-based remedial programs produce similar outcomes, the question of what program a teacher should use with a student comes down to several factors. One of them is seeing if there is *any* research on outcomes with that specific program and then examining the program as a whole to try to determine if the features a teacher finds appealing in the program are a match for his or her classroom and/or the school district's needs. The features outlined in the Appendix should therefore be helpful to educators seeking to make such comparisons. Those features can be used to compare any other remedial programs with which readers may be familiar. The three programs in the Appendix were chosen just because of the extensive clinical experience that The Reading Foundation has had with all of them but it could also be used to compare other phonemically-based programs to each other such as the *Wilson Reading Program, Slingerland Program, Open Court, Orton-Gillingham, Project Read, Literacy Links* and the *Spalding Method*. The list of features in the Appendix, while comprehensive, is not exhaustive. Other features that educators consider important can easily be added so that readers can create their own unique comparisons.

The Reading Foundation created its own program – *Discover Reading* (Truch, 2003) – which contains all the known essential instructional ingredients required for effective remedial reading and a number of other features that make the program very

teacher and student friendly as well as extremely comprehensive. Indeed, *Discover Reading (DR)* might be the most comprehensive commercial program yet developed.

While outcomes from both the *LiPS Program* and the *PG Program* have been previously published, the outcomes from The Reading Foundation's *DR Program* are presented here for the first time.

The outcomes from *DR* are then compared to those from the *LiPS* and *PG* programs to determine whether any differences in outcomes are present. Based on the author's prior experiences with *LiPS* and *PG*, it was hypothesized that such differences in fact would *not* be found.

Method

Subjects

Clients who attend The Reading Foundation (which opened its doors in 1990) usually come by word-of-mouth referrals from other parents or by professional referral. The results from a grand total of 155 such clients (collected over a two year time period starting in 2001) were used in this analysis. Of these, 36 students were ages 6-7 years; 59 students were ages 8-9; 43 students were ages 10-16 and 9 students were ages 17 and up. The ratio of males to females was 1.5 to 1. The average oral vocabulary score of this entire group was quite high at 108.97 (minimum of 66 and maximum of 146).

No attempt was made in this study to formally classify or diagnose the students as "dyslexic" or "learning disabled" nor to determine the severity of their presenting reading difficulty. (While such differences are important, collecting and analyzing data in a clinic such as The Reading Foundation is not an easy thing to do since the prime purpose of the clinic is to provide a service and not conduct research.) On average though, the majority of students would meet traditional criteria as "learning disabled" since most showed a discrepancy of about one standard deviation between reading potential, as measured by oral vocabulary scores and actual reading performance (the average oral vocabulary score, as mentioned, was 108.97 and the average pre-test Reading score on the WRAT 3 was 86.06). However, some of the students had very "severe reading disabilities" where discrepancies of two or more standard deviations were present. (Indeed, as some of the data later suggested, The Reading Foundation over the years may have attracted a population of more "severe" clients than in earlier years, when it first opened.) Some of the students were also "mentally challenged." Many students also had overlapping symptoms of "attention deficit disorder." Other students were previously diagnosed with Asperger's Syndrome, with Fragile x Syndrome, with Autism, with Fetal Alcohol Syndrome and with various symptoms of language impairment. No attempt was made to analyze results in these sub-categories but merely to provide overall outcome data. What was common to every student was a reading weakness that had a clearly identified basis in weak phonemic awareness and attendant difficulties in decoding, spelling and reading fluency. Students who attended the clinic for other issues such as

reading comprehension difficulties in the absence of decoding difficulties or for math remediation were not included in this analysis.

Measures

All clients are screened on a standard battery of tests that included:

Oral Vocabulary

Vocabulary was measured in most cases using standard scores from the Comprehensive Receptive and Expressive Vocabulary Test (CREVT) (Wallace and Hammill, 1994) or the Comprehensive Receptive and Expressive Vocabulary Test – Second Edition (CREVT 2) (Wallace and Hammill, 2002). In some cases, students had previously been administered the Wechsler Intelligence Test for Children – Third Edition (WISC-III). For this analysis, only the Expressive scores were used, either from the CREVT or the WISC-III Verbal IQ or Vocabulary subtest.

Phonemic Awareness and Processing

Students were assessed using three measures of phonemic awareness and processing:

(1) Auditory segmenting was measured by an informal test developed by the author. The segmenting test consisted of 13 real words ranging from three to five phonemes in length and 10 nonsense words ranging from three to six phonemes in length. Students were presented each word orally and asked to provide the sounds (not the letters) in that particular word. For example, if the word presented were “chomps,” the student would have to respond /ch/, /o/, /m/, /p/ and /s/ to receive a full 5 points. The ceiling score on this test is 100 (one point for each phoneme segmented correctly from each of the words).

(2) For the auditory blending test, an informal measure was developed by the author. Students are presented sounds from a total of 10 real words and asked to decide what the actual word is. For example, if the sounds presented by the examiner were /w/, /e/, /n/ and /t/, the student should say “went.” Words ranged from three to six phonemes in length and the maximum score was 10 (one point for each word blended correctly).

(3) Students were also assessed on the Lindamoood Auditory Conceptualization Test (LAC) (Lindamood and Lindamood, 1979) as an independent measure of phonemic manipulation. On this test, students use colored wooden blocks to represent sounds. They are then asked to match the patterns of sounds presented orally by the examiner using those blocks. The ceiling score on this test is 100.

Letter to Sound Knowledge

Students are assessed on two different informal measures developed by the author:

(1) One test is called “Sound Links to Sounds.” There is one version of this test for younger students (Basic Level which covers up to age 12) and one for older students (Advanced Level). For this test, the student sees a particular letter or letter combination (these are called “Sound Links” in the *DR Program*) and must say the sound for that Sound Link. So, if the student sees “ay,’ she should say /ae/ as her answer. A total of 50 Sound Links are sampled at each level and the score converted to a percentage correct out of 100.

(2) The other is called “Sounds to Sound Links.” For this measure, the examiner says a sound such as /b/ and the student prints the letter or letter combination needed to visually represent that sound. In this case, the student should print or write the letter “b.” A total of 25 sounds (15 consonants and 10 vowels) comprise this informal measure.

Visual Memory for Letters

This is another informal measure developed by the author. Clinical experience has demonstrated that while phonological processing is an essential ingredient in learning to read, it is not the whole story by any means. In fact, many students whose phonological skills are obviously improving still have great difficulty retaining the graphemes (Sound Links) used to represent sounds. This process must involve the internal retention of the Sound Links the student is exposed to as she goes through the program. But some students retain the Sound Links easily while others require a substantial degree of reinforcement and repetition. Individual differences in this process form a key component, in the author's view, to the subsequent development of fluency. However, the process is difficult to define and to measure, so the informal measure used in this program should be seen only as a crude attempt to determine the student's ability to form a mental representation for letters only. For this test, the student is presented with a series of letters printed on a card. The student is told not to try to create words from the letters, since they are just random letters. Her job is to just remember the letters and repeat them back to the examiner once the card is placed face down on the table. For example, the letters H C appear on the first card, which is flashed to the student for just one second. Once the student sees the card, the examiner places it on the table and asks the student to name the letters she saw. A point is given for each letter stated in its correct sequence. The test has a total of 10 cards

ranging from two letters up to six letters per card so the ceiling score on this test is 40.

A limitation of this test is the fact that the student may simply rehearse the letters in his head without actually visualizing them and thus, it may not be measuring “visual memory” at all. However, the procedures outlined in the Discover Reading Manual do allow the examiner to question the student to see what strategy the student is using – if she is mentally rehearsing the letters without visualizing them, just visualizing the letters or doing both. Most students reply they do both so this is a confounding element. An occasional student uses a creative strategy such as reading part of the letters as a word or a word they are reminded of.

Decoding

The Woodcock Word Attack subtest (Woodcock, 1987) was used to measure decoding ability. Students must read a total of up to 50 nonsense patterns on this test, containing simple, complex and multisyllable patterns. The student’s raw score was converted to a grade-equivalent score. While a newer version of this subtest is available, the author tried to maintain consistency from the first study (Truch, 1994) and the second one (Truch, 2003), where the 1987 version was used along with grade-equivalent scores.

Word Identification

The Reading subtest of the Wide Range Achievement Test – Third Edition (WRAT 3) (Wilkinson, 1993) was used. Students read a graded word list ranging from easy to very difficult until they make 10 consecutive errors. Their raw scores are converted to standard scores for their age.

Spelling

The Spelling subtest of the WRAT 3 (Wilkinson, 1993) was used for this measure. Students spell words orally presented to them ranging in degree of difficulty until they spell 10 words in a row incorrectly. Their raw score is converted to a standard score for their age.

In addition, each spelling word on the WRAT 3 was scored for its approximation to a real word using a developmental spelling scale presented by Gregg and Mather (2002). Standard scores by themselves do not necessarily capture spelling progression since students receive a score of “0” on the WRAT 3 if the word is not exactly correct. However, students could still be making spelling gains if their errors are more phonetically accurate so a scale to capture such progress provides an additional source of information to educators on spelling growth. For that reason, each word on the WRAT 3 was informally

scored on a scale of 0 to 7 as per Gregg and Mather's (2002) scoring system. A total numerical score is calculated for each student pre and post.

Reading in Context

The old Gray Oral Reading Test (GORT) was used as a measure of reading in context. Again, a newer version of this test is available (Wiederholt and Bryant, 2001) but was not used, partly to maintain consistency with data collection from the previous two programs. On the GORT, students read a series of graded passages out loud. The examiner times the student and counts the errors for each passage. Once a student makes too many mistakes on a passage, testing is discontinued. The examiner converts errors by time to a grade-equivalent score using tables provided in the Manual. The grade-equivalent score is interpreted as a "fluency index." While there are comprehension questions on the GORT, the student responses to recall questions were not analyzed for this study.

Procedure

Most of the students in this analysis attended The Reading Foundation clinic on an intensive-one-on-one basis. Students attended for four consecutive hours either in the morning or the afternoon and missed school for half-days in order to do so. Older students typically attended after-school for two hours twice a week and then a Saturday

for four hours to give them the recommended minimum of 8 hours per week to a total of 80 hours for the program.

Some students required less than 80 hours to complete the program while others required many more. For this analysis, the number of hours completed was averaged for the group as a whole.

This instructional variable (intensity of service) was also used in prior years at the clinic with both the *LiPS Program* and the *PG Program.*

Clinicians providing the service to students were all first trained in the *DR Program* and diligently followed procedures provided in the *Discover Reading Manual* (Truch, 2003). Each hour, students were given instruction in phonemic segmenting, blending and phonemic manipulation using the "Sound Links" (letter and letter combinations) for each word. Sound Links are taught from context using specially created Books. For the *Discover Reading Program,* Basic Level – covering the age span from 5 to 12 years – the author analyzed the 100 most common words in children's literature (Bodrova E., et. al, 1998) and first categorized those words as either simple, complex or multisyllable in nature. He then analyzed the *sounds* in each of the words and determined their frequency. He then created a tightly structured sequence of letters and sounds that could be used for instructional purposes and wrote contextual material (Books) representing those letters and sounds. Unlike both the *LiPS* and *PG* programs, the phonemes and graphemes in the *DR Program* are taught through contextual stories

(the Books). At the Basic Level, there are a total of three such Books. At the Advanced Level (used with older students) there is just one Book and it was based on the 300 most common words as a starting point and extended to the 5500 most common words (based on *The Educator's Word Frequency Guide* – Zeno, S.M., et. al, 1995). For each Book at both Levels, a number of Activity Booklets were written which contain the sequenced series of instructional activities taught in the program using the sounds and Sound Links to that point only. The teaching activities include ones which stimulate phonemic processing in all its aspects, visual memory for letters, word recognition, spelling, writing, reading in context and development of fluency. Comprehension activities are included with each contextual activity, thus some meta-cognitive elements are also part of the *DR Program.* (It should be noted however, that the prime purpose of the program is the development of word recognition skills; students who have a separate issue with the comprehension process require remedial work in that process and so a separate program for that purpose is used at the clinic.)

An important feature of the *DR Program* is that students are taught from *sounds to letters*, rather than *letters to sounds* (as used in most phonics programs). This makes the student's task of learning the complexities of English spellings much easier. The student is taught the basic principle that *every word consists of sounds and each sound has to have a Sound Link.* The student discovers both the sounds and the Sound Links as she progresses through the program and places her discoveries on Sound Link Charts created for the program. There is a chart for consonants and another for vowels. In the end, the student learns to sort the Sound Links into their various phonemic categories and

is able to see the various ways in which the same sound can be represented. Since the task is simply one of sorting, there is no need for any phonics kind of "rules." This instructional approach was used successfully in the *PG Program* and to a lesser extent by the *LiPS Program.* The approach of going from sounds to letters has a compelling rationale (McGuinness, D., 1997) since English is, after all an alphabet system. The approach makes a great deal of sense for students who often marvel at the relationships once they are presented. The *DR Program* has the added feature of *discovery* by the student. Sound Links are discovered, rather than presented at once as in the *PG Program.* Students progress their way through the series of Activity Booklets and learn all the skills required to decode simple, complex and multisyllable English words. The initial assessment with the student provides the information the teacher needs to determine which of three possible program sequences a teacher should use with any given student. Since students vary in age and prior knowledge, it is important to start remediation at a place that is not too easy nor too difficult for that student; hence the need for different starting and end points. Keeping track of student progress is accomplished with the use of Student Progress Charts developed for each sequence.

Students who attended The Reading Foundation were tested after 12 hours of instruction (three clinical days) on some of the pre-test measures. This procedure allowed comparison to the 12 hour results first presented by McGuinness et al in their study of the *PG Program* (1996).

Students were again tested after 24 hours and finally post-tested, usually near the end of 80 hours but sometimes sooner, if the student was able to complete the program in less than 80 hours. The purpose of the 24 hour testing was to provide a further measure of comparison between the *DR Program* and the *PG Program*, since 12 and 24 hour measures were also collected for the latter.

Results

Outcomes for all students using the Discover Reading Program

The means for the pre-test, the 12 hour testing, the 24 hour testing, and the post-test scores for all students are presented in Table 1.

Pairwise comparison using t-tests for paired samples and two-tailed significance were used for the analysis. Since the Woodcock and the GORT use grade-equivalent scores, the use of parametric measures is not entirely appropriate. Those two variables were therefore subjected to further analysis using the Wilcoxon Matched-Pairs Signed-Ranks Test (Siegel, 1956). The outcomes were still highly significant with a two-tailed p value of $p<.0001$ on both variables.

Table 1

Outcomes For All Students (N= 146) After 80 Hours of One-On-One Intervention Using the Discover Reading Program

Variable	**N**	**Pre**	**12 hours**	**24 hours**	**Post**	**Significance (Pre to Post)**
LAC	144	56.90	NA	NA	85.61	p<.0001
S.D.		22.59			17.41	
Segmenting	145	39.74	79.65	82.81	92.28	p<.0001
S.D.		26.53	21.55	20.01	13.78	
Blending	145	6.58	8.46	8.92	9.50	p<.0001
S.D.		2.51	1.90	1.61	1.24	
Sound Links To Sound	145	51.68	65.00	69.91	82.22	p<.0001
S.D		18.36	16.97	16.47	16.72	
Sounds to Sound Links	140	18.87	NA	NA	23.72	p<.0001
S.D.		4.96			2.60	

Visual						
Memory	135	23.64	NA	NA	27.72	p<.0001
S.D.		7.98			7.83	
Word						
Attack	141	2.75	4.71	5.16	7.13	p<.0001
S.D.		1.38	3.03	3.37	3.85	
Word						
Identification	146	86.27	92.79	93.20	100.68	p<.0001
S.D.		12.11	12.08	11.75	12.49	
Spelling	146	87.61	90.90	91.07	94.95	p<.0001
S.D.		11.77	11.68	9.93	11.04	
Developmental						
Spelling	140	100.70	NA	NA	137.83	p<.0001
S.D.		48.77			52.91	
Fluency	136	2.89	NA	NA	3.77	p<.0001
S.D.		2.17			2.52	

Discussion of DR Results

The results indicate highly significant gains from pre to post-test for all subjects and all variables. This suggests that the *DR Program* accomplishes what it was intended to accomplish – help develop the overall decoding process for students of all ages by strengthening all the known components of that process.

Further analysis (data not presented here in tabular form) shows that the gains per hour are cumulative and that the scores after 12 hours and 24 hours are significantly stronger than the pre-test scores and from each other (p. <.001). In other words, students make progressive gains hour by hour in the program. This sort of hourly gain was also obtained from the *PG Program.* It is entirely possible that similar quick gains were also present when students did the *LiPS Program.* Unfortunately, only pre and post-test gains after 80 hours were collected for that study (Truch, 1994).

The various analyses also showed that when the data was broken down into the three age-groups noted earlier, that the gains were very strong. Each age group made cumulative gains on each of the variables after every testing interval with the overall result of highly significant gains for students of every age group on every variable (p.<.001 in all cases).

Overall then, it can be stated with confidence that the clinical usage of the *DR Program* using an intensive immersion approach brings strong gains in the decoding and

spelling process for students of all ages. Strong gains for fluency are also apparent but 80 hours is too short a time period in which to “normalize” the process of fluency. Indeed, normalizing this process remains very challenging, as noted in some of the more recent research on fluency. For example, Torgesen, et. al (2003) report that most of the successful phonemically-based programs improve reading accuracy and comprehension, but usually not fluency. Lyon and Moats (1997), after a review of research, report the “persistent finding” that “improvements in decoding and word-reading accuracy are far easier to obtain than improvements in reading fluency and automaticity.” Torgesen and his group are currently studying this issue in depth.

At The Reading Foundation clinic, the author has observed that many students who have completed the intensive immersion process do not immediately show large gains in fluency, but re-assessments done months and in some cases years later on the same student, usually show much larger fluency gains. It’s as if time is needed for some of the process to “percolate” and then to finally surface as fluent reading. However, this issue merits much more systematic and long-term data collection and analysis than is possible from a study of this nature.

Comparing Discover Reading Outcomes To Those From The LiPS and Phono-Graphix Programs

The Reading Foundation now has clinical experience of over 14 years using three different phonemically-based programs, all delivered in the same intensive one-on-one fashion and with “due diligence” as to the procedures intended by the authors of those programs. The question arises as to whether or not any one of the three programs delivers stronger outcomes.

To answer that question, all data collected over the 14 year time span of the clinic from three different cohorts of students were re-analyzed.

The post-test scores for all relevant variables were compared for all three programs and subjected to an analysis of variance. However, there were some challenges just because of the long period of time over which the data was collected so it was not a case of comparing “apples to apples” in every instance.

For example, one of the test instruments had changed over the years. The WRAT 2 was the test used for the *LiPS* group, but the WRAT 3 was used for both the *PG* and *DR* groups. The segmenting and blending scores from the *PG* and *DR* groups were based on different informal tests. Visual memory was introduced as a variable for the first time in the data collection for the *DR* group.

That left the LAC test, the Woodcock, the WRAT (with the above caveat) and the GORT as intact over the years. The comparison analysis was therefore done for only those variables.

Table 2 shows the pre and post test mean scores on those variables among the three different programs. Again, scores on the Woodcock and the GORT were subjected to further analysis using the Wilcoxon Matched-Pairs Signed-Ranks Test because they were calculated as grade-equivalent scores in the first place. All results remained highly significant ($p < .0001$).

The LAC score has a ceiling of 100. The Woodcock scores are expressed as grade-equivalents. The WRAT scores are standard scores and the GORT scores are grade-equivalent fluency scores.

Table 2

Reading Outcomes From Three Different Programs

	LiPS (N=280) (1994)		PG (N=195) (2003)		DR N=122) (2004)		
	Pre	Post	Pre	Post	Pre	Post	Sig.
Variable							
LAC	64.45	92.37	63.47	90.39	55.56	85.07	p<.01
S.D.	19.76	12.70	21.95	11.63	22.88	18.21	
Word Attack	3.00	7.31	3.28	8.01	2.70	6.87	n.s
S.D.	1.71	3.74	2.24	3.91	1.42	3.87	
WRAT Reading	75.88	93.17	83.51	98.64	86.42	101.07	n.s

S.D.	13.15	14.55	13.29	12.19	11.99	12.82	
WRAT	76.11	86.94	83.99	95.62	87.84	96.75	p<.03
Spelling							
S.D.	12.50	14.05	12.88	13.14	12.25	11.86	
GORT	3.45	4.51	3.80	4.82	2.84	3.62	n.s
S.D.	2.70	3.16	2.87	3.20	2.15	2.51	

Examination of Table 2 shows that the mean LAC score, the mean Woodcock Word Attack score and the mean fluency score on the GORT are all lower to begin with for the *DR* group (however, their WRAT mean scores are higher). Overall though, this suggests that the *DR* students may have had more significant reading issues to begin with and were a more "severe" group from the other two. That would make some sense since The Reading Foundation clinic was by then, established for a much longer period of time and was perhaps attracting students with more severe reading issues by way of word of mouth reputation. To account for pre-test score differences, analysis of covariance was conducted, resulting in non-significant differences on all variables but the LAC Test.

However, examination of the raw data showed that 14 students from the DR group had raw scores on the LAC Test lower than 30 as did 14 from the PG group and just 5 from the LiPS group. When raw scores lower than 30 were eliminated from all three groups, and the analysis of covariance done with the remaining students, the significant difference between groups was no longer present ($p<.09$).

The significant difference for Spelling on the WRAT favored both the *PG* and *DR* groups over *LiPS*. However, as noted earlier the reason may be due in part to different norms on the WRAT 3 over the WRAT 2.

McKenna & Violato (2003), in their study of the effects of the *Literacy Links* program, another relatively new remedial program, found a very small effect size on spelling outcomes. This may be due to the teaching methods in the program or it may be

due to the limited time students spent in the program – an average of 30 to 35 hours versus the 80 hours in this study. Teaching methodology may be the issue however, since spelling gains after just 12 hours using both the *PG* and the *DR* programs were stronger than those from 30-35 hours of *Literacy Links.* It should also be noted however, that the spelling procedures outlined in the *PG* manual are very limited. When The Reading Foundation made the switch to using the *PG Program* clinically, much re-writing of the program was done, including more emphasis on spelling. Thus, the spelling outcomes for *PG* may not be as positive if teachers using the program follow only the procedures in the *PG Manual.*

McKenna & Violato (2003) also quote Lovett & Steinbach (1997) and Oakland et al. (1998) who found only limited spelling gains on programs they evaluated. Many studies, they state, do not include developmental measures of spelling which could reflect spelling changes not found when standardized measures alone are used. This study included both types of measures and perusal of Table 1 shows significant gains on the WRAT 3 standard scores as well as on the developmental scale employed on the same WRAT 3 spelling items.

Another measure of gains being used in reading research studies is “effect size.” An effect size measures the magnitude of a treatment effect, as opposed to a significance test, which merely determines if there is a significant difference between two samples. There are a number of measures of effect size. The one used in this study is “Cohen’s d” (Cohen, 1988). Using this measurement, an effect size of 0.2 or less is considered

"small." An effect size of 0.3 to 0.5 is considered "medium" and an effect size of 0.6 to 0.8 is considered "large." The scale does however, go up to 2.0 but effect sizes of 1.0 or more are rare.

Table 3 shows the effect size using Cohen's d for the three programs used at The Reading Foundation over the years. The variables included for this analysis were phonemic processing using the Lindamood Auditory Conceptualization Test (LAC); word recognition using the Wide Range Achievement Test (WRAT); spelling using the Wide Range Achievement Test (WRAT) and sound and letter connections using different informal measures. Because the Woodcock and the GORT used grade-equivalent scores, calculating an effect size on such scores is not appropriate.

Table 3

Effect Sizes for LiPS, PG and DR

Variable	LiPS (1994)	PG (2003)	DR (2004)
LAC	1.729	1.525	1.775
WRAT Reading	1.732	1.769	1.609
WRAT Spelling	1.269	1.300	1.138
Sound/letters	NA	1.118	1.290

Again, these effect sizes are generally comparable with one program showing no major advantage over the other on any particular variable.

Discussion

The results of this analysis give support to the effectiveness of the *Discover Reading Program.* The data also show that differences between and among phonemically-based programs are not going to be easy to find. Good programs produce good outcomes and no one program can claim superiority over another overall; however, some differences on selected variables may exist. Only a rigorous study with random assignment of students to programs done over the same time period could answer such questions but such studies, as mentioned earlier, are generally not found in the scientific literature, particularly those involving commercial programs.

Many educators are also very concerned as to whether or not there is transfer from phonemically-based reading interventions to “real reading.” The whole point of reading is comprehension of text and there is still the fear from many educators that too much emphasis on the “bottom-up” skills of phonemic processing and “phonics” will detract the young reader from the main point of reading, which is to understand the information or the story. Many of the phonemically-based intervention studies do not include measures of reading comprehension; indeed, many of them contain no comprehension activities at all, so do students who receive such interventions actually learn to read and comprehend text? The study by McKenna & Violato (2003) showed in fact, strong gains

on a measure of reading comprehension using a phonemically-based program. (However, that program does contain some comprehension activities.) The McKenna and Violato study is valuable in that regard since the test batteries used in many studies, including this one, do not include measures of comprehension. Certainly, the clinical experience at The Reading Foundation has been that comprehension actually improves as a result of improved decoding ability, but measuring such change is a different matter, so it is encouraging to see some data on that issue.

Because outcomes from good remedial programs are similar, it is equally important to determine which programs are teacher and student friendly. The number of phonemically-based remedial programs is still small at this point, but that number is bound to grow over time, particularly since there are initiatives in place, such as the *No Reader Left Behind* one in the United States that will encourage the development of more reading programs which are "science-based" in nature. There will be plenty of "research-based" programs that will make an appearance in the next few years, but it is not necessarily the case that any research will actually have been done on the program. It is therefore important for educators and those involved in the special education field to have some way of evaluating those programs. A list of features for such comparison is found in the Appendix.

The limitations of this study include the fact that it is quasi-experimental in nature; no control groups were used and the data collection occurred over a long period of time. As a result, some changes in the measurement instruments themselves occurred

over the long time span. As well, the clinical population and "severity" of reading disability may also have changed over the years. Nevertheless, there is still some confidence that can be placed in the major conclusions of this study which are:

(1) To be effective for remedial purposes, a reading program needs to include phonemic awareness activities and other research-based activities including explicit teaching of letter and sound connections into the program.

(2) Any program that does include such activities is likely to produce positive outcomes for students, provided instructors are trained in the program and consistently follow the procedures.

(3) The outcomes from this study shows that all students of all ages, regardless of initial diagnostic categories such as "mentally handicapped," "attention deficit disorder," Asperger's Syndrome," etc. benefit from intensive remedial reading work so long as the student's reading problem involves weak phonemic awareness in the first place.

(4) No one program can claim, at least at this point in time, superiority of outcomes over any other program following about 80 hours of remedial time. It may well be that some programs may produce superior outcomes in some subtle ways over another, but the well-controlled scientific studies required to truly demonstrate superiority of outcomes have not yet been done.

These conclusions should therefore also be treated as hypotheses and be subjected to future research.

While this study provided outcomes with students in a clinical setting with one-on-one conditions, such conditions do not generally exist in school settings. When it is possible to do so however, it appears the outcomes in schools are similar to those found in a clinical setting (e.g. Torgesen, et al, 2001).

Not many school districts can provide a one-on-one service however. Data for outcomes using phonemically-based programs with students in small groups and classroom-based situations is scarce in the scientific literature but it has been the experience of many teachers who are using these commercial programs (and others) that they can be successfully adapted for small group and classroom use. More research in this area is certainly very much in need.

References

Bell, N. (1997). *Seeing stars.* San Luis Obispo, CA: Gander Publishing.

Bodrova, E., Leong, D.J., Semenov, D. *100 most frequent words in books for beginning readers.* The list is posted on the website at www.bestpraceduc.org/TOTM/HundredsWords.html.

Bradley, L., & Bryant, P.E. (1983). Categorizing sounds and learning to read – A causal connection. *Nature, 30,* 419-421.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, N.J. Lawrence Erlbaum Associates.

Gray, W.S. (1963). *The Gray Oral Reading Test.* Indianapolis, IN: The Bobbs-Merrill Company, Inc.

Gregg, N., & Mather, N. (2002). School is fun at recess: Informal analyses of written language for students with learning disabilities. *Journal of Learning Disabilities, 35, (1),* 7-22.

Lindamood, C.H., & Lindamood, P.C. (1979). *Lindamood auditory conceptualization test.* Austin, TX: Pro-Ed., Inc.

Lindamood, P., & Lindamood, P., (1998). *The Lindamood phoneme sequencing program (Third Edition).* Austin, TX: Pro-Ed., Inc.

Lovett, M.W., & Steinbach, K.A. (1997). The effectiveness of remedial programs for reading disabled children of different ages: Does the benefit decrease for older children? *Learning Disability Quarterly, 20,* 189-220.

Lyon, G.R., & Moats, L.C. (1997). Critical conceptual and methodological considerations in reading intervention research. *Journal of Learning Disabilities, 30,* 578-588.

McGuinness, C., & McGuinness, G. (1998). *Reading Reflex: The Fool-Proof Phono-Graphix Method For Teaching Your Child To Read.* New York: The Free Press.

McGuinness, C., McGuinness, D., & McGuinness, G. (1996). Phono-Graphix: A new method for remediating reading difficulties. *Annals of Dyslexia, 46,* 73-96.

Oakland, T., Black, J.L., Stanford, G., Nussbaum, N.L., & Balise, R.R. (1998). An evaluation of the dyslexia training program: A double-deficit perspective. *Journal of Learning Disabilities, 31*(2), 140-147.

Oloffson, A., Lundberg, I. (1985). Evaluation of long-term effect of phonemic awareness training in kindergarten: Illustrations of some methodological problems in evaluation research. *Scandinavian Journal of Psychology, 26,* 21-34.

Paivio, A. (1990). *Mental representations: A dual coding approach.* New York: Oxford University Press.

Report of the National Reading Panel: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction. Reports of the Subgroups. (2000). Washington, D.C.: National Institute of Child Health and Human Development.

Report of the National Reading Panel (Executive Summary). (2000). Washington, D.C.: National Institute of Child Health and Human Development.

Sadoski, M., & Paivio, A. (2001). *Imagery and text: A dual coding theory of reading and writing.* Mahwah, NJ: Lawrence Erlbaum Associates.

Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

Swanson, H.L. (1999). Reading research for students with LD: A meta-analysis of intervention outcomes. *Journal of Learning Disabilities, 32*(6), 504-532.

Thaler, V., Ebner, E.M., Wimmer, H., & Landerl, K. (2004). Training reading fluency in dysfluent readers with high reading accuracy: Word specific effects but low transfer to untrained words. *Annals of Dyslexia, 54, (1),* 89-111.

Torgesen, J.K., Wagner, R.K., & Rashotte, C.A. (1997). Prevention and remediation of severe reading disabilities: Keeping the end in mind. *Scientific Studies of Reading, 1(3),* 217-234.

Torgesen, J.K., Alexander, A.W., Wagner, R.K., Rashotte, C.A., Kytja, K., Voeller, S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34, (1),* 33-58.

Troia, G. (1999). Phonological awareness intervention research: A critical review of the experimental methodology. *Reading Research Quarterly, 34,* 28-52.

Truch, S. (1991). *The missing parts of whole language.* Calgary, AB: Foothills Educational Materials.

Truch, S. (1993). *The WISC-III companion.* Austin, TX: Pro-Ed. Inc.

Truch, S. (1994). Stimulating basic reading processes using Auditory Discrimination in Depth. *Annals of Dyslexia, 44,* 60-80.

Truch, S. (2003). *The Discover Reading Program.* Calgary, AB: Weigl Publishing.

Truch, S. (2003b). Comparing remedial outcomes using LiPS and Phono-Graphix: An in-depth look from a clinical perspective. Poster session presented at the International Dyslexia Conference, San Diego. November, 2003.

Truch, S. (2005). *The WISC-IV companion.* Austin, TX: Pro-Ed. Inc.

Vellutino, F.V., Fletcher, J.M., Snowling, J.M., & Scanlon, D.M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry, 86,* 24-53.

Wagner, R. (1988). Causal relations between the development of phonological processing abilities and the acquisition of reading skills: A meta-analysis. *Merrill-Palmer Quarterly, 32,* 261-279.

Wallace., G., & Hammill, D.D. (1994). *Comprehensive receptive and expressive vocabulary test.* Austin, TX: Pro-Ed., Inc.

Wallace, G., & Hammill, D.D. (2002). *Comprehensive receptive and expressive vocabulary test – Second edition.* Austin: TX: Pro-Ed., Inc.

Wechsler, D. (1991). *Wechsler intelligence scale for children – Third edition.* San Antonio, TX: Psychological Corp.

Weiderholt, J.L. & Bryant, B.R. (2001). *Gray oral reading tests – Fourth edition.* Austin, TX: Pro-Ed., Inc.

Wilkinson, G.S. (1993). *The wide range achievement test 1993 edition.* Wilmington, DE: Wide Range, Inc.

Woodcock, R.W. (1973). *Woodcock reading mastery tests.* Circle Pines; MN: American Guidance Service.

Zeno, S.M., Ivens, S.H., Millard, R.T. & Duvvuri, R. (1995). *The Educator's Word Frequency Guide.* Touchstone Applied Science Associates, Inc.

Appendix

Features of an Effective Remedial Reading Program

Feature	LiPS	Phono-Graphix	Discover Reading
Detailed Scope and Sequence			
Charts	None	None	Yes
Clear Lesson Plans	None	None	Yes
Age Levels	All	All	All
Classroom or Small Group?	No (with difficulty)	Yes	Yes
Homework?	No	Yes	No
Flexible Delivery?	Somewhat	Somewhat	Very
Phonemic Awareness	Yes	Yes	Yes
Articulatory feedback	Yes	No	No
Auditory segmenting	Yes	Yes	Yes
Auditory blending	Yes	Yes	Yes
Phonemic manipulation (all 5 possible moves)	Yes	Partial	Yes
Sound/Letter Connections	Yes	Yes	Yes
Explicit and structured	Yes	Yes	Yes
Sound to letter	Partial	Yes	Yes

Use of phonics “rules”	Partial	No	No
Visual Memory	No	No	Yes
For letters	No*	No	Yes
For words	No*	No	Yes
Word Attack	Yes	Yes	Yes
Word Identification	Yes	Yes	Yes
Spelling	Yes	Yes	Yes
Reading in Context	Must be added	Some	Yes**
Written Activities	No	No	Yes
Fluency Development			
At the sound/letter level	No	No	Yes
At the word level	No	No	Yes
At the sentence level	No	No	Yes
At the contextual level	No	No	Yes
Comprehension Activities	No	No	Yes
Multisyllable processing	Extensive	Adequate	Very extensive
Error-handling Procedures			
Responding to the Response**	Yes	No	No
Directed error-handling	No	Yes	Yes
Informal Assessments	LAC only	Yes	Yes
Training Times	Two weeks	One week	Four days

*Can be added using “Seeing Stars” from Nanci Bell

**The Discover Reading Program also uses much more contextual material from the start than do the other two programs. All lessons and activities are taught from context.

*** “Responding to the response” as an error-handling procedure is excellent, though difficult to master for some. The rationale is that this procedure leads to greater self-correcting on the part of the student. However, self-correcting is also very evident from the more direct methods used in both the Phono-Graphix and Discover Reading Programs.

Author Notes

The author would like to thank the staff of The Reading Foundation for their collective efforts over the years.

Correspondence concerning this article should be addressed to Steve Truch, The Reading Foundation, Suite 207, 2310-2nd St. S.W., Calgary, Alberta, Canada T2S 3C4 or email to thereadf@telus.net.