

**The devil is in the detail regarding the efficacy of Reading Recovery: A rejoinder to
Schwartz, Hobsbaum, Briggs, and Scull**

Meree Reynolds¹, Kevin Wheldall¹, Alison Madelaine¹

¹ *Macquarie University Special Education Centre, Macquarie University, Sydney, New South
Wales, Australia.*

The following is the post-print version of an article that has been peer reviewed and accepted
for publication. It was published by Cambridge University Press on 23 February 2009. Full
reference details are below.

Reynolds, M., Wheldall, K., & Madelaine, A. (2009). The devil in the detail regarding the
efficacy of Reading Recovery: A rejoinder to Schwartz, Hobsbaum, Briggs, and Scull.
International Journal of Disability, Development and Education, 56(1), 17-35.
<https://doi.org/10.1080/10349120802681580>

Abstract

This rejoinder provides comment on issues raised by Schwartz, Hobsbaum, Briggs and Scull (2009) in their article about evidence-based practice and Reading Recovery (RR), written in response to Reynolds and Wheldall (2007). Particular attention is paid to the processes and findings of the What Works Clearinghouse evaluation of RR. The suggestion that this evaluation is flawed casts doubt about some of its findings. The authors maintain their earlier stance that RR is effective for many students but do not accept that there is evidence that initial gains are sustained through the primary grades, that RR is an efficient tier two intervention in a response to intervention approach and that significant cost benefits have been demonstrated in education systems. It is concluded that research into alternative interventions that could be implemented at lower cost is warranted.

The devil is in the detail regarding the efficacy of Reading Recovery: A rejoinder to Schwartz, Hobsbaum, Briggs, and Scull

Introduction

In their response to our earlier article reviewing the research on Reading Recovery (RR) completed over the past 20 years or so (Reynolds & Wheldall, 2007), Schwartz, Hobsbaum, Briggs, and Scull (2009) offer a thoughtful, well-argued and temperate perspective that reaches very different conclusions from our own—albeit based, essentially, on the same dataset of research articles. McNaughton, in his obituary of Marie Clay, stated the case for the success of Clay's RR as follows:

Her applied achievements are recognized most directly in the Reading Recovery programme. This highly effective early intervention programme has near national implementation in New Zealand as well as in parts of education systems in Australia, Barbados, Canada, Denmark, England, the Republic of Ireland, Northern Ireland, the USA and Wales. In 2005, 11,000 NZ children were taught and over a million have been taught worldwide. The U.S. Department of Education recently recognized the quality of the research base to Reading Recovery through its Institute of Education Sciences Clearinghouse, identifying it as an effective programme meeting more standards with higher effectiveness than comparison programmes. (McNaughton, 2007, p. 264)

We would disagree with little in this summary, and in our earlier article we took pains to emphasise the achievements of RR on many dimensions. The point at issue is not, however, and never has been, whether RR is effective or not—we have always acknowledged that it is (see, for example, Center, Wheldall, Freeman, Outhred, & McNaught, 1995); rather it is a question of whether it is effective enough, especially considering the considerable cost

of its implementation, and effective for enough of the population of young, low-progress readers whose reading difficulties it purports to rectify. As in so many things in life, this is not a simple matter: the devil is in the detail.

This article will focus on the four aspects of the RR programme on which Schwartz et al. (2009) based their comments in a response to our earlier article (Reynolds & Wheldall, 2007): evidence of RR's effectiveness, retention of gains, the notion of RR as a component of a Response to Intervention (RTI) model, and cost-effectiveness versus cost-benefit.

Evidence of Effectiveness of RR

Schwartz et al. (2009) took issue with our view that there is limited experimental research into the effectiveness of RR, resulting in a relatively weak research base. We acknowledge that there have been some experimental and quasi-experimental studies into the efficacy of the RR programme. However, these are very small in number in relation to the extent of the programme's implementation and the large number of studies and evaluations that have been conducted. It is surprising that that there have not been many more high-quality studies, given the massive amounts of funding that have been allocated to the implementation of the programme across many education systems over 25 years. It would be reasonable to expect that many systems would have sought rigorous, independent scientific evidence before committing high levels of funding.

Schwartz et al. (2009) used the outcomes of the What Works Clearinghouse (WWC) evaluation of RR (WWC, 2007a) as supporting evidence that RR is effective. This review was published in March 2007 as our previous article was in press, and we therefore were unable to include its findings at the time. As the WWC was established to provide

independent and rigorous reviews of education programmes and approaches, based on scientific evidence from randomised controlled or quasi-experimental studies, we were looking forward to its findings about RR. As anticipated, and in line with the findings of our review, the WWC reported overall positive findings about the effectiveness of the intervention. The evaluation report provides more detail and indicates that RR has positive effects on alphabets and general reading achievement, and potentially positive effects on fluency and comprehension. This is an excellent rating in comparison with the ratings given to other beginning reading programmes.

The finding that RR has a positive effect on alphabets is, however, at odds with previous conclusions drawn in the Reynolds and Wheldall (2007) review, and has prompted a closer look at the constructs and processes used in the WWC evaluation and the studies upon which the conclusions were based. When looking at the details of the evaluation we have found shortcomings in the process that limit the credence of the findings. Recently, other critics (Englemann, 2008; Slavin, 2008; Stockard, 2008) have also voiced serious concerns about the WWC evaluation of RR. Issues pertaining to the selection of studies and reporting of their outcomes, the match between constructs and their measures and the magnitude of effect sizes have become obvious and will be discussed in greater detail.

The WWC Reading Recovery Intervention Analysis and Report

Studies that met the WWC criteria were analysed to determine whether there were significant effects on key components of an early reading programme that have been identified in scientific research (National Reading Panel, 2000). The findings are presented within the domains of alphabets (defined as being made up of phonemic awareness, print

awareness, letter knowledge and phonics), fluency, comprehension (comprising vocabulary and comprehension) and general reading proficiency.

A programme is classified by the WWC as having positive effects if there are two or more studies that demonstrate statistically significant positive effects. At least one of these studies must be a strong design; that is, a randomised control trial. In addition, there should not be any studies that have negative effects that are statistically significant or have a negative effect size greater than 0.25. The rating of potentially positive effects is given when one or more studies have positive effects that are statistically significant or are deemed to be important. At the same time there should be the same number of studies or fewer with indeterminate effects and there should not be any studies with negative effects (WWC, 2006).

Selection of Studies

Seventy-eight studies related to the effectiveness of RR were considered for inclusion, with only five considered to meet the standard for selection. Four of these studies were randomised control trials that met the evidence criteria as being of high standard (Baenen, Bernhole, Dulaney, & Banks, 1997; Pinnell, DeFord, & Lyons, 1988; Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994; Schwartz, 2005). A quasi-experimental study (Iverson & Tunmer, 1993) was considered to meet the evidence standard with reservations. Of the five studies that met the WWC criteria, four were carried out more than 10 years ago. Three of these studies (Pinnell et al., 1988, 1994; Schwartz, 2005) were carried out by RR affiliates. Table 1 presents summary information about the studies that met the WWC evidence standards.

Table 1 about here.

While five studies were stated to have met evidence criteria, the WWC evaluators did not use any of these studies in entirety when they conducted their analysis. The WWC analysed only those parts of the studies that featured randomised control groups and that allowed for a comparison of traditional RR with a very different alternative intervention or no intervention. In the case of the Baenen et al.'s (1997) study, the WWC analysed the results of the first cohort of Year One students, as this was the only one that had a randomised control group. In the remaining studies the WWC evaluators did not look at all comparisons, instead focusing on a traditional RR intervention with a randomised control group that had a completely different intervention such as basic skills instruction, or regular class instruction. In two of the complete studies (Pinnell et al., 1994; Iverson & Tunmer, 1993), RR was compared with RR-like interventions with varied findings.

As the full studies investigated a range of research questions beyond whether RR is effective, the published findings are generally not the same as those noted by the WWC in its evaluation of RR. The study by Baenen et al. (1997) found that RR was not effective, and Iverson and Tunmer (1993) concluded that RR could be more effective if it included more systematic teaching of phonics. The Pinnell et al. (1988) study investigated whether students taught by RR teachers in regular lesson time had better results than students taught by different teachers in regular lesson time. They found that students had a slight advantage when taught by RR teachers in regular lessons, but results were not statistically significant. The Pinnell et al. (1994) study, which compared the achievement of four groups (traditional RR, RR-like individual intervention, RR-like group intervention, and basic skills), found that RR students had superior achievement on general reading achievement measures to students in alternative interventions. Schwartz (2005) compared the outcomes of RR students during

the first and second semesters of the school year, and concluded that RR students in the first half of the year performed better than students in the second half of the year. He found that RR students had significantly higher achievement than control and comparison groups.

Iverson and Tunmer investigated the effectiveness of RR by comparing outcomes in regular RR, modified RR with more emphasis on phonics instruction and a small group intervention.

They concluded that both RR groups performed significantly better than the small group intervention and that modified RR was more efficient than regular RR.

Englemann (2008) and Stockard (2008) have also voiced concerns that the WWC have reported different findings from the Baenen et al. (1997) and Iverson and Tunmer (1993) studies than were published in the source articles. Mathematica, the entity that administers the WWC, has indicated that the reason for the limitation of studies, and the differing outcomes, is that their priority is to investigate the effects of the beginning reading interventions over one year (Dynarski, 2008). They consider that the results of studies that meet evidence standards are more comparable if research data are limited to a 12-month period, rather than being reported on across all complete studies (Dynarski, 2008). Because of this decision, there is limited reporting of retention or non-retention of gains. Englemann suggested that this has resulted in the publication of misleading information about the findings of these two studies. We shall return to this issue of the maintenance of gains later.

One study that was a surprising omission from the review is that of Center et al. (1995), and we note that Schwartz et al. (2009) also disagreed with the WWC's decision not to include it. The reason given by the reviewers is that the study did not meet evidence screens because the groups were "incomparable" (WWC, 2007a, p. 11). In their view, the study was "a quasi-experimental design but did not establish that the comparison group was

comparable to the treatment group prior to the start of the intervention” (WWC, 2007a, p. 11). The Center et al. study, despite being acknowledged as an example of a “more sophisticated” study (Shanahan & Barr, 1995), appears to have been misclassified in the WWC evaluation as it was an experimental study in which subjects were randomly allocated to either the experimental or control groups (Carter & Wheldall, 2008; Center et al.). (It also included a comparison group.) In addition, the published article provides the information that “there were no significant differences between the two groups on any literacy measures at the pretest stage” (Center et al., p. 251). We shall return to this study again later in this article since Schwartz et al. consider its findings in some detail.

Issues Related to Measurement of the Constructs

There are incongruities in the links between the constructs identified by the WWC as the focus of the evaluation and the outcome measures used in the reviewed studies. One notable example is the measurement of effectiveness of RR on student achievement in phonics. The outcome measures that are included for the phonics construct, with one exception, are tests of high-frequency sight words, such as the Ohio Word Reading Test used in the RR assessment battery. This test requires students to read a short list of high-frequency words, a large proportion of which are not phonically regular. We do not accept that this type of test is a valid measure of phonics learning. Young students often begin to read by recognising words they see often in early texts as logographs or by using salient visual features as cues (Adams, 1990). This typically occurs before letter–sound correspondences are used in decoding text. As students become proficient in decoding, they may become so automatic in recognition of phonically regular words that they access them automatically as sight words. It is acknowledged that students rely on fast and automatic phonological and orthographic processing to access words from memory as they gain larger sight vocabularies

and that a word recognition test may provide some information about phonics for more competent readers. However, when assessing young struggling readers, we consider that word recognition tests should not be used to provide information about phonics knowledge.

Table 2 presents details about the assessments that the WWC evaluation relied upon to make judgements about RR's effectiveness on phonics learning. The Iverson and Tunmer (1993) study is the only one that uses a phonics measure that is suitable for beginning readers. As this is a quasi-experimental study and there is a lack of unequivocal evidence from the four experimental studies, we maintain that there is limited scientific evidence for RR's effectiveness in improving knowledge within the phonics construct.

Table 2 about here.

Another major concern with the WWC evaluation of beginning reading programmes is that the alphabetic domain, as defined by the WWC, is very broad and contains a number of crucial constructs; namely, phonemic awareness, print awareness, letter identification and phonics (What Works Clearinghouse, n.d.). In the RR evaluation report, the ratings and improvement indices for each of the specific constructs were amalgamated to make a general finding across the domain. This general finding was presented in summary form under domain headings and is used by RR in promoting the programme. Given the critical importance of each area (and, most specifically, given the importance of the phonemic awareness and phonics constructs for struggling readers), our view is that it would have been more desirable to report separately on each of these crucial constructs in the summary intervention report. We assert that this has the potential to provide misleading information.

Moreover, a review of the links between the measures and the constructs and a consequent re-analysis of outcomes across the alphabetics domain would result in a different finding about RR's effectiveness in this area. Table 3 reflects this re-analysis and indicates that the only construct in the alphabetics domain that demonstrates proven effects is print awareness. An amalgamation of results across all four constructs would not provide a true indication of effectiveness in this domain.

Table 3 about here.

The measures used to determine effectiveness for several of the domains and constructs are those used in RR; that is, they are subtests of An Observation Survey of Early Literacy Achievement (Clay, 1993, 2002). Denton, Ciancio, and Fletcher (2006) cautioned against the use of the letter identification, word identification and dictation subtests to evaluate programme effectiveness, indicating that they “represent a closed set of skills on which mastery is expected and thus have ceiling effects that reduce their sensitivity and usefulness in evaluating end-of-year outcomes in first grade” (p. 32). They suggested that, when these subtests are used, “effect size estimates may also be inflated” (Denton et al., 2006, p. 32). The use of the text reading measures from the Observation Survey of Early Literacy Achievement for programme evaluation is also criticised by Denton et al. They considered that its usefulness is limited because the intervals between levels are not equal and, therefore, this affects statistical analysis and makes comparison of progress difficult. The WWC recognised this issue and only used outcomes of text reading measures from the Schwartz (2005) study as it is the only one where the researcher provided follow-up information to enable WWC reviewers to calculate an effect size and statistical significance (WWC, 2007b).

It has been argued that skewed and biased findings may be caused by the use of the subtests of the Observation Survey of Early Literacy Achievement as they are strongly linked to monitoring and progress throughout the programme (Baker et al., 2002). In the studies reviewed by the WWC, RR teachers carried out most of the assessments that were the subtests of the Observation Survey of Early Literacy Achievement. This practice, in which the assessors have a vested interest in testing outcomes, reduces objectivity and has the potential to produce inflated scores.

The WWC uses a very broad definition of the fluency domain that influences the selection of measures of effectiveness. They define it as “the ability to read text accurately, automatically, and with expression, while still extracting meaning from it” (WWC, 2007a). The Schwartz (2005) study has the only reported positive effect for fluency in the intervention report, and RR is given a potentially positive rating in this construct on the strength of this finding. One measure used for fluency in this study is progress through RR text reading levels. We consider that while this reflects students’ accuracy, it is not a measure of fluency as there is no reporting and analysis of a timed component of text reading to determine whether students have automaticity. The other measure used is the Slosson Oral Reading Test-Revised (Slosson, 1990). This is a test of 200 commonly used words that have been arranged into 10 lists of 20 words in increasing levels of difficulty. Students are scored on their accuracy of reading and there is no timed element. We do not accept this as a measure of fluency for the same reason.

Without the Schwartz (2005) study results, there would be a nil outcome for this domain. In evaluations of other beginning reading programmes, it is not uncommon to have

measures with timed components. The Center et al. (1995) study, for example, incorporated a timed curriculum-based measure that, had the study been included, could have provided evidence of the effects of RR on reading fluency.

In our opinion, the text reading level measures used in RR and curriculum-based measures assess different constructs. In essence, the WWC definition includes four different constructs (accuracy, speed of reading, expression, and comprehension) that could have been analysed separately, or we would suggest that the speed of reading (i.e. automaticity) should be the central or key construct. Shadish, Cook, and Campbell (2002) discussed issues involved in construct validity and suggested that there are four key factors in fostering this in studies:

(1) starting with a clear explication of the person, setting, treatment, and outcome constructs of interest; (2) carefully selecting instances that match these constructs; (3) assessing the match between instances and constructs to see if any slippage between the two occurred; and (4) revising construct descriptions accordingly. (Shadish et al., 2002, p. 66)

In the WWC evaluation of RR there seem to be poor matches between some of the constructs and the tasks that WWC suggest measure these, and this affects the validity of the findings.

Magnitude of Effect Sizes

The WWC has selected an effect size magnitude of 0.25 as a minimum threshold for a study to be considered to have sufficient educational effect to be “important”. This is lower

than the 0.3 (of one standard deviation) that is generally accepted in special education research, and could lead to difficulties in interpreting the significance of studies (Carter & Wheldall, 2008).

Summing Up the Evidence from the WWC Evaluation

The WWC evaluation confirms that RR has positive effects for many students, particularly in relation to gains in general reading achievement. However, when reconsidered, the evidence that RR has strong effects across all of the key areas required for all students to get off to a good start in reading should be treated with caution. While the WWC process appears to provide an objective review of the scientific evidence about programmes, the information can be misleading when the process allows the use of outcome measures that do not adequately reflect the stated constructs. Slavin cautioned that, as the findings of research syntheses such as that of the WWC are relatively “high stakes”, it is “essential that the conclusions be correct, but also that the process by which they are arrived at be open, consistent, impartial, and in accordance with both science and common sense” (2008, p. 7).

More about Evidence of Effectiveness

Schwartz et al. (2009) also used the findings of the meta-analysis carried out by D’Agostino and Murphy (2004) as evidence that RR is effective. These were based on data from 36 evaluations carried out in the United States from 1984 to 1996. Although D’Agostino and Murphy determined that RR was effective for RR students on programme measures and standardised measures, we are somewhat dubious about the quality of this study because it is based on internal evaluations, half of which relied solely on the use of the subtests of the Observation Survey of Early Literacy Achievement. The study carried out by Denton et al. suggested that the use of data from the Observation Survey of Early Literacy

Achievement “may lead to serious problems in evaluating outcomes” (2006, p. 32). We were also concerned that almost all of the studies analysed had not been subjected to peer-review.

Schwartz et al. (2009) are critical of the findings of the Elbaum, Vaughn, Hughes and Moody (2000) meta-analysis of research about one-to-one tutoring programmes for young struggling readers. In this article, the authors discussed in detail issues related to the RR studies, including zero effect sizes for students who were not successfully discontinued and difficulties in using the RR non-standardised measures such as the Text Reading Level measures that may “overestimate intervention effects” (Elbaum et al., 2000, p. 617). Our article included a summary statement about RR from Elbaum et al., but we did not provide Elbaum et al.’s reasons in detail. The full text of the relevant paragraph follows:

Overall, the findings of this meta-analysis do not provide support for the superiority of Reading Recovery over other one-to-one reading interventions. Typically, about 30% of students who begin Reading Recovery do not complete the program and do not perform significantly better than control students (Pinnell, 1988). As indicated in this meta-analysis, results reported for students who do not complete the program may be inflated due to the selective attrition of students from some treatment groups and the use of measures that may bias results in favour of Reading Recovery students. Thus, it is particularly disturbing that sweeping endorsements of Reading Recovery still appear in the literature. For example, in a recent issue of *Educational Leadership*, the official publication of the Association for Supervision and Curriculum Development, (Daniels, Zemelman, & Bizar, 1999) stated that “Reading Recovery has been phenomenally successful” (p. 35). (Elbaum, et al., 2000, p. 617)

One criticism that Schwartz and his colleagues (2009) have of the Elbaum et al. meta-analysis is that it includes a study by Chapman, Tunmer, and Prochow (2001) that uses a retrospective matching design. We accept that the Chapman et al. (2001) study has limitations and that the results need to be treated with caution. However, in their article, Elbaum et al. (2000) provided details of the criteria for inclusion of studies in their analysis and, as the Chapman et al. (2001) study appears to meet these criteria, there seems to be no reason to exclude the study. If the authors made a decision to omit studies with design faults from the meta-analysis, it is highly likely that many others would fit this description and that there would be few studies remaining for analysis. In Elbaum et al.'s article, the outcomes of the Chapman et al. study are briefly summarised and the data used in the meta-analysis, but it appears that these data are not likely to have affected the overall finding that RR students achieve at a higher level than similar students who do not participate in the programme.

We feel that claims of RR's effectiveness are often exaggerated. Certainly the intervention is successful for a proportion of students, but we are aware that percentages that are reported are usually in terms of those students who received a full programme. Because of the way RR successes are reported, it is often difficult to determine the number of students who are removed or referred on from the programme early because they are not making progress. In our earlier review we suggested that many studies that use data from all students in the programme find that 60–70% of students are deemed to be “successfully discontinued”. The more recent reviews and studies have provided no reason to change this percentage and no reason to suggest that the programme is “phenomenally successful”.

Sustained Gains

We agree with Schwartz et al.'s (2009) view that it is difficult to conduct and interpret research on whether RR gains are retained. Longitudinal studies appear to have varying outcomes, with a number of studies showing that many successful ex-RR students achieve at grade level one, two and three years after completing the programmes. However, as we pointed out in our earlier article, a number of studies show that the initial gains of these successful students tend to be reduced over time. This has implications for cost-benefit analysis.

Center et al. (1995) found that RR made significant short-term gains on a number of reading measures. Schwartz et al. (2009) have carried out an additional analysis of the Center et al. data from the RR group and the comparison group at a testing point in mid-second grade, one year after the first post-test. Their conclusions from this quasi-experimental comparison are that the RR group retained gains on all assessments, five of which were not part of the RR assessment battery. If we take this analysis and compare the effect sizes for the RR group with controls at post-test, short-term maintenance (after a further 15 weeks) and the RR group with the comparison group at medium-term maintenance (one year after post-test) we find that there is a gradual decrease in effect size over time. Table 4 presents details of these effect sizes. This supports our view that gains tend to lessen over time.

Table 4 about here.

An evaluation of RR in London schools that was carried out in 2005–06 (Burroughs-Lange & Douetil, 2007) showed strong, initial gains with a slight weakening of these at the 12-month follow-up (Burroughs-Lange, 2008). When tested at the end of the first year, the students who had received RR had an average reading age gain of 20 months on the British

Ability Scale Word Reading Test 11 (Elliot, 1996) and 16 months on the Word Recognition and Phonics Skills Test (WRAPS) (Moseley, 2003). Comparison group students in schools with no RR had an average gain of seven months on the British Ability Scale and 10 months on the WRAPS. Students eligible for RR, but without intervention, in RR schools gained an average of 12 months in reading age on the WRAPS and 11 months on the British Ability Scale throughout the year.

At the end of year two, ex-RR students were an average of 12 months ahead of students in the comparison group on the WRAPS assessment and their average achievement had been maintained at age and grade levels. As students who did not access the programme in the RR schools made reasonable gains, it is possible that not all students identified as low-achievers prior to the programme, needed a programme with such high intensity and high cost. The comparison group was still performing below age and grade expectations, but the gap between them and the ex-RR group had lessened by one month. As these students had a range of different experiences (some had an alternative intervention, some were “taught” by teaching assistants, and many did not have any intervention at all), it is difficult to make comparisons. The results of this evaluation show that RR does make a difference for many students in the short term. The long-term effects of RR within these London schools are yet to be shown, and we hope that further research will investigate whether the gains are sustained over a longer period.

A study by Hurry and Sylva (2007) that investigated the long-term outcomes of two different models of early reading intervention found little effect of RR at the end of primary schooling. Hurry and Sylva compared RR with phonological training based on Bradley and Bryant’s (1985) work. In a quasi-experimental condition, RR students were compared with

their within-school controls and between school controls. Students were pre-tested and post-tested after nine months using standardised tests, subtests of the Diagnostic Survey and phonemic awareness measures. Post-tests using standardised tests were also carried out 12 months after the first post-test and then after another two years. Short-term and medium-term results showed that, while both interventions significantly improved aspects of student's reading, RR had a broader and more powerful effect. The long-term results three and a half years after the interventions showed that there were no significant effects on reading overall. Hurry and Sylva found that RR had a significant effect for a subgroup of children who were non-readers at six years old. These students made more progress than between school controls, but their reading achievement at long-term testing was more than two years below age expectations.

Longitudinal data from the Baenen et al. (1997) study do not support the view that RR gains are retained. The full study was a system evaluation carried out in the Wake County in North Carolina in the United States. After considering achievement data of three cohorts of Year One students in 10 schools, Baenen et al. concluded that RR is not cost-effective. They found that 50% of RR students achieved the grade level in the first year. However, two years later, only one third maintained the grade level, while one-third of comparison students achieved the grade level in the third grade. There were no differences in referral rates of the groups.

The D'Agostino and Murphy (2004) study also investigated the question of whether gains are maintained. Although there were little quality data (see our earlier comments and concerns about the studies used in the meta-analysis), they concluded that successfully

discontinued RR students maintain gains in general reading skills until the end of second grade.

We accept that there are many influences on learning in school settings and that many contextual factors may result in gains not being maintained. We stress that in our earlier article we did not say that RR is not effective, we said that most students made gains but, in some studies, it has been shown that the gains are not maintained.

Programme Evaluation Data from a Response to Intervention Perspective

Schwartz et al. stated that RR has operated as a “Response to Intervention (RTI) approach to the identification of students with learning disabilities related to literacy” (2009, p. 10) since its development in the 1970s. At the time, it was viewed as a preventative intervention (Clay, 1984) and it was ground-breaking in that it introduced an alternative model to the remedial or diagnostic-descriptive models of support for students with literacy difficulties. Clay, in her description of the programme as a “safety net” for students (Clay, 1997) or a “second wave” of teaching, was formulating a model that is aligned with RTI.

Since then, many others have supported this type of approach to preventing literacy failure and the notion of RTI has developed, become widespread and is now mandated in schools in the United States. While there are many variations, it is generally accepted that RTI has the following features: interventions with increasing levels of intensity are available for struggling readers; the interventions are evidence based; they include regular progress monitoring; and they have a problem-solving approach to determine what to teach and how to teach.

A typical RTI model is multi-tiered and is the basis for a whole school approach to dealing with students with learning difficulties. There are many variants, with the number of tiers suggested in the literature ranging from two to four (Fuchs & Fuchs, 2006). Typically, there are three tiers and instruction increases in intensity as students move from tier to tier. In the model that is most widely accepted, the first tier features quality whole class instruction by the classroom teacher with regular curriculum-based progress monitoring. For those students who are not making adequate progress, the classroom teacher will typically implement an intervention that is delivered to students in small groups. Tier-three interventions are generally more intensive individualised programmes provided for those students who are still struggling following tier-two support programmes. Instruction for students in tier three is likely to be provided by a special education teacher.

While proponents of RR suggest that education systems and schools incorporate it as a tier-two intervention in a RTI approach, we consider that it does not fit neatly within the descriptors and that it is not an efficient use of resources. There are two main features of RR that are consonant with RTI. Firstly, it is implemented when students have not responded to initial literacy instruction and, secondly, it includes a procedure for ongoing progress monitoring. What does not fit neatly is that, when RR is used as tier-two intervention, it provides a far more intensive intervention than would be available in most education systems at tier three. This is counter to the concept of RTI as having tiers in which the intensity increases. In typical RTI models, tier-two interventions do not involve individual tutoring. In systems where RR is used as a second-tier intervention, the high cost of delivery is likely to limit the extent of funding available to students who are referred for services in the third tier.

Schwartz et al. (2009) discussed the issue of selection efficiency. As the intervention is implemented so early in schooling, it is likely that there will be a number of false positives among the students selected for the programme. Center et al. (1995) suggested that approximately one-third of the students would have recovered without intervention. This finding from an experimental study places a question mark over findings from studies and evaluations that do not feature a randomised control group. The issue is how many of these students were falsely identified and provided with a costly intervention that they did not need? If we consider the figure of 30% of students that Center et al. suggested would have recovered by themselves, then it seems a very costly exercise to implement such an intensive intervention as a first resort.

The Center et al. (2005) study provides an example that demonstrates the efficiency of a typical three-tier RTI model. This model was not in operation in New South Wales schools in 1991–92 when the study was conducted. At the time, many schools were using a whole language approach to the teaching of reading. It is possible that students in the study who performed poorly on the pre-tests may not have done as poorly if they had had explicit and systematic teaching of the crucial components of reading during the first year of schooling. Whole grade screening at the end of the kindergarten year would have then facilitated the identification of students with difficulties. Tier-two small group interventions could then have been implemented by classroom teachers with the assistance of support teachers allocated to the school, if necessary. Regular progress monitoring would have enabled decisions to be made about the effectiveness of the intervention so changes could have been made or referral made for more intensive support. We believe that this process would have resulted in far fewer students having significant difficulties in their third year of schooling than were evident at follow-up testing in the Center et al. study. The need for individualised assistance

would be much lower and, ideally, those students who have been reliably identified as having significant literacy difficulties would receive appropriate support in the third tier. This process has the capacity to be far more efficient than using a costly individualised intervention in the second tier.

Cost-effectiveness versus Cost-benefits

Schwartz et al. (2009) rightly pointed out that the notion of cost-effectiveness requires a comparison of the cost of RR with alternative approaches. It is for exactly this reason that we suggested in our initial article there should be well-designed experimental studies that compare RR with one or more alternative interventions that may involve small group instruction and synthetic phonics instruction. The latter is suggested because there is clear research evidence indicating that a synthetic approach, in which students are explicitly and systematically taught letter–sound relationships and how to blend or synthesise these to read words, is more effective for struggling readers than an analytic approach in which students begin with whole words and analyse them to learn letter–sound relationships (Ehri, Nunes, Stahl, & Willows, 2001).

There are some costs associated with the ongoing implementation of RR in a system that are not evident yet impact on system implementation and long-term costs. One of these relates to the loss of trained teachers to the programme. An RR teacher is encouraged to go back to a role as a classroom teacher after several years, with a different teacher from the school beginning the training and taking over the role. This facilitates collegial support and knowledge within the school, but the practice is expensive to education systems and schools. We note that, while 2735 RR teachers trained in New South Wales from 1996 to 2004, there were only 931 RR teachers working in the role within the state in 2003–04 (Australian

Reading Recovery Trainer Group, 2005). This loss may also be related to schools making the decision not to continue with programme implementation because of funding issues or because they do not consider it to be effective or cost-effective. Nevertheless, there is a heavy loss of teachers who have been trained at significant expense to education systems and their schools. Meanwhile, the state and other education providers continue to allocate funds on an annual basis to train new RR teachers.

We do not take issue with the high costs of literacy difficulties to society and the need to prevent reading failure as espoused in the Every Child A Reader initiative in England. We reiterate that we are not saying RR would not benefit many students in this initiative, but that other programmes may provide similar benefits at less cost. The problem is that at this time there are no alternatives that have strong scientific evidence to support their implementation. Our suggestion is that other options are fully explored and that they are given fair opportunities to prove whether they can deliver similar or better results to RR at a lower cost. In our experience, there appears to be reluctance on the part of RR personnel to enable independent researchers to compare RR with other programmes.

There are some promising beginning reading interventions such as that used in a synthetic phonics teaching study conducted in Clackmannanshire in Scotland (Johnston & Watson, 2005). This study compared the literacy progress of students who had been involved in one of three group training conditions delivered in early Grade One. One group were taught using synthetic phonics, one group had analytic phonics training, and the third group were taught using analytic phonics combined with some phonological awareness training. Each group had training sessions of 20 minutes per day for a period of 16 weeks. Post-test results after the intervention showed that the synthetic phonics group was superior in reading

words, decoding and spelling. Students from the other two groups were then involved in the synthetic phonics training. After six years students were tested again, and it was found that students who had had synthetic phonics training had continued to make gains throughout the grades. Where their post-test reading achievement showed them to be seven months ahead of their chronological age, by the end of their primary grades they were, on average, three years six months ahead of their chronological age on a standardised reading assessment of word reading (Johnston & Watson, 2005). While comprehension scores showed some washout effect over the years, students' scores were still significantly higher than expected.

In terms of cost-benefits, we agree that we do not know what might occur if the programme was not implemented. We accept that RR has cost-benefits, but we cannot say that there may not be similar interventions that could be implemented with greater cost-benefits. If we take, for example, the study by O'Connor and Simic (2002), we find that those students who accessed the RR programme had significantly lower rates for referral for further assessment (5% lower) and for placement in special education (3% lower) than a comparison group of students with learning difficulties who were initially higher achievers than the RR group. RR, then, at a relatively high cost, has reduced ongoing special education costs in the particular system by approximately 3%. While a study may find reductions such as these to be statistically significant, there may be little real effect on the cost within an education system, and any possible savings may be absorbed into other services or special education programmes.

If RR has the effect of reducing ongoing special education costs it would be reflected in lower total demand for services. We made the point in our earlier article that there is no strong evidence that this has occurred within systems. If we take the New South Wales state

school system as an example, despite RR being implemented for over 15 years (five years of which were as a pilot programme), the number of teachers employed to support students with learning difficulties in primary schools has remained relatively unchanged. While there has been some adjustment in the model of service delivery for students with learning (and reading) difficulties, the major change has been to provide a system so that students throughout the state have access to more specialised services for students with severe learning difficulties. In this time, public schools are experiencing lower student numbers due to growth in private school enrolments. During this time the number of RR teachers working in schools (in all education systems) within the state continued to grow. For example, in the period 1996–2004, more than 2700 RR teachers were trained (Australian Reading Recovery Trainer Group, 2005).

Another way of determining whether a literacy programme has a cost-benefit across an education system would be to investigate whether, after the programme has been in place for a period of time, funding allocations for literacy programmes have been reduced. This appears not to be the case in New South Wales, where a recently released Auditor-General's (2008) report indicated that, from 1998–99 to 2006–07, the funding allocation for literacy and numeracy programmes increased threefold, yet the report indicated that there was little improvement in outcomes for students in this time. RR has been relatively well funded during this period. For example, in 2005 it was allocated approximately 20% of the total funding for literacy and numeracy programmes, including equity programmes (Auditor-General 2008). There are undoubtedly many factors influencing the literacy and numeracy outcomes of students in such a large and diverse education system with many different programmes, but it appears that RR has had little effect on the system during a time in which it has had a marked expansion.

We are also aware that there are some education systems that have implemented RR for some years, but have made the decision not to continue funding for the programme. One of these is the Queensland public school system, which has announced that it will not be involved with RR in 2009 (Queensland Department of Education Training and the Arts, n.d.). Decisions such as this are indications that administrators do not believe that the benefits outweigh the costs.

Conclusion

We acknowledge again that RR has many worthy features and that it has been successful in raising the achievement of many students. The critical factors in its success may relate to just some of RR's elements and may be simple to replicate in alternative interventions. It is highly likely that one key factor in RR's success is the fact that it is implemented early—that is, as students start to struggle—and therefore the students experience what Stanovich (1986) has termed “Matthew effects”. Marie Clay was one of the first literacy experts to recognise the importance of intervening early, and this is one of the most widely accepted concepts in literacy teaching today.

No other intervention has been implemented on the same scale as RR, nor has been as costly. For us the issue is not whether RR has benefits—of course it does. Many students are successful and a proportion of these maintain gains. The issue is whether other interventions could have similar or better outcomes at a lower cost.

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: The MIT Press.
- Auditor-General. (2008). *Improving literacy and numeracy in NSW public schools*. Sydney: Department of Education and Training, Audit Office of New South Wales.
- Australian Reading Recovery Trainer Group. (2005). Submission no. 379 to the National Inquiry into the Teaching of Literacy. Retrieved from <http://www.dest.gov.au/nitl/submissions/361-380.htm>
- Baenen, N., Bernhole, A., Dulaney, C., & Banks, K. (1997). Reading Recovery: Long-term progress after three cohorts. *Journal of Education for Students Placed at Risk*, 2, 161–181.
- Baker, S., Berninger, V., Bruck, M., Chapman, J. W., Eden, G., Elbaum, B., et al. (2002). Evidence-based research on Reading Recovery. Letter to members of US Congress signed by 32 reading researchers. Retrieved July 26, 2008 from <http://ednews.org/articles/2484/1/EVIDENCE--BASED-RESEARCH-ON-READING-RECOVERY/Page1.htm>
- Bryant, P., & Bradley, L. (1985). *Children's reading problems*. Oxford: Basil Blackwell.
- Burroughs-Lange, S. (2008). Comparison of literacy progress: A Reading Recovery follow-up study. Retrieved July 10, 2008 from <http://www.ioe.ac.uk/schools/ecpe/readingrecovery/London%20eval%20follow%20up%20May%2008.pdf>
- Burroughs-Lange, S., & Douetil, J. (2007). Literacy progress of young children from poor urban settings: A Reading Recovery comparison study. *Literacy Teaching and Learning*, 12, 19–46.

- Carter, M., & Wheldall, K. (2008). Why can't a teacher be more like a scientist? Science, pseudo-science and the art of teaching. *Australasian Journal of Special Education*, 32, 5–21.
- Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of Reading Recovery. *Reading Research Quarterly*, 30, 240–263.
- Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (2001). Does success in the Reading Recovery program depend on developing proficiency in phonological processing skills? A longitudinal study in a whole language instructional context. *Scientific Studies of Reading*, 5, 141–176.
- Clay, M. M. (1984). Reading Recovery: A follow-up study. In M. M. Clay (Ed.), *Observing young readers: Selected papers* (2nd ed., pp. 173–191). Exeter, NH: Heinemann Educational Books.
- Clay, M. M. (1993). *An observation survey of early literacy achievement*. Auckland: Heinemann Education.
- Clay, M. M. (1997). The challenge of literacy development. Paper presented at the Meeting the Challenge: NSW State Literacy Strategy Conference.
- Clay, M. M. (2002). *An observation survey of early literacy achievement* (2nd ed.). Portsmouth, NH: Heinemann.
- D'Agostino, J. V., & Murphy, J. A. (2004). A meta-analysis of Reading Recovery in United States schools. *Educational Evaluation and Policy Analysis*, 26, 23–38.
- Daniels, H., Zemelman, S., & Bizar, M. (1999). Whole language works: Sixty years of research. *Educational Leadership*, 57, 32–37.
- Denton, C. A., Ciancio, D. J., & Fletcher, J. M. (2006). Validity, reliability, and utility of the Observation Survey of Early Literacy Achievement. *Reading Research Quarterly*, 41, 8–34.

- Dynarski, M. (2008, September 8). Responses to concerns in 6/25/08 letter. Letter to Dr Jean Stockard. Retrieved October 16, 2008, from http://www.ednews.org/attachments/1/rmresponse_wwc.pdf
- Ehri, L. C., Nunes, S. R., Stahl, S. S., & Willows, D. W. (2001). Systematic phonics instruction helps students learn to read: Evidence from the national reading panels's meta-analysis. *Review of Educational Research*, 71, 393–447.
- Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs for reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, 92, 605–619.
- Elliot, C. D. (1996). *British Ability Scales Word Reading Test II*. London: Nelson.
- Englemann, S. (2008). *Machinations of What Works Clearinghouse*. Retrieved October 17, 2008, from [http://ednews.org/articles/29692/1/Machinations-of-What-Works-Clearinghouse/Page 1.html](http://ednews.org/articles/29692/1/Machinations-of-What-Works-Clearinghouse/Page%201.html)
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 4, 92–99.
- Hurry, J., & Sylva, K. (2007). Long-term outcomes of early reading intervention. *Journal of Research in Reading*, 30, 227–248.
- Iverson, S., & Tunmer, W. E. (1993). Phonological processing skills and the Reading Recovery program. *Journal of Educational Psychology*, 85, 112–125.
- Johnston, R. S., & Watson, J. E. (2005). A seven year study of the effects of synthetic phonics teaching on reading and spelling attainment. Retrieved October 15, 2008, from <http://www.scotland.gov.uk/Resource/Doc/933/0044071.pdf>
- McNaughton, S. (2007). Dame Professor Emerita Marie Clay, DBE, MA (Hons), Dip Ed, PhD, FRSNZ, FNZPsS, FNZEI, 1926–2007. *Journal of Early Childhood Literacy*, 7,

263–266. Moseley, D. (2003). *Word recognition and phonic skills* (2nd ed.). Abingdon, UK: Hodder & Stoughton.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Retrieved July 15, 2008 from

http://www.nichd.nih.gov/publications/nrp/upload/smallbook_pdf.pdf

O'Connor, E. A., & Simic, O. (2002). The effect of Reading Recovery on special education referrals and placements. *Psychology in the Schools*, 39, 635–646.

Pinnell, G. S., DeFord, D. E., & Lyons, C. A. (1988). *Reading Recovery: Early intervention for at-risk first graders*. Arlington, VA: Educational Research Service.

Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly*, 29, 9–38.

Queensland Department of Education Training and the Arts. (n.d.). *Reading Recovery*. Retrieved August 15, 2008 from

<http://education.qld.gov.au/students/procedure/program/reading-recovery/>

Reynolds, M., & Wheldall, K. (2007). Reading Recovery twenty years down the track: Looking forward, looking back. *International Journal of Disability, Development and Education*, 54, 199–223.

Schwartz, R. M. (2005). Literacy learning of at-risk first grade students in the Reading Recovery early intervention. *Journal of Educational Psychology*, 97, 257–267.

Schwartz, R. M., Hobsbaum, A., Briggs, C., & Scull, J. (2009). Reading Recovery and evidence-based practice: A response to Reynolds and Wheldall (2007). *International Journal of Disability, Development and Education*, 56(1): 5–15.

- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Shanahan, T., & Barr, R. (1995). Reading Recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly*, 30, 958–996.
- Slavin, R. E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37, 5–14.
- Slosson, R. L. (1990). *Slosson oral reading test: Revised*. East Aurora, NY: Slosson Educational Publications.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360–407.
- Stockard, J. (2008). *The What Works Clearinghouse beginning reading reports and rating of reading mastery: An evaluation and comment*. Eugene, OR: National Institution for Direct Instruction.
- What Works Clearinghouse. (2006). Study design classification. Retrieved October 20, 2007, from <http://www.whatworks.ed.gov>
- What Works Clearinghouse. (2007a). *Beginning reading intervention report: Reading Recovery*. Retrieved August 15, 2007, from <http://www.whatworks.ed.gov>
- What Works Clearinghouse. (2007b). *Intervention: Reading Recovery appendices*. Retrieved March 23, 2008, from http://ies.ed.gov/ncee/wwc/reports/beginning_reading/reading_recovery/appendix.as
- What Works Clearinghouse. (n.d.). *WWC evidence review protocol for beginning reading interventions*. Retrieved June 23, 2008, from http://ies.ed.gov/ncee/wwc/PDF/BR_protocol

Table 1
Overview of studies used in the WWC evaluation of RR.

Study	Aim(s) of study	Findings	WWC component and type of study	Subjects	Measures used by WWC	WWC findings
Baenen et al. (1997). Evaluation of RR with three cohorts of Year One students in 10 schools in Wake County	To determine:- the extent to which RR assists struggling students to achieve grade levels- the need for support services after Year One- whether discontinued students remain at grade level- whether RR is cost-effective	1/2 of all RR students achieved grade levels in the first year 1/3 of the successful students maintained grade level two years later 1/3 of comparison group were at grade level in third grade No differences in referral rates RR is not cost-effective	Part of larger study:- one cohort from the first year of implementation as this was the only one that had a control group. Randomised control trial	72 RR students and 84 control students	Grade retention at the end of first and second grades The North Carolina End-of-Grade Reading Test administered two years later	General reading achievement component - no significant effects on retention No reported outcomes related to other WWC domains or constructs
Pinnell et al. (1988). Evaluation of RR in 14 schools in the first year of implementation in Ohio	To investigate whether students who were taught by a RR teacher during their regular classroom lessons had better achievement results to students in RR who had different class teachers	RR students taught by the RR teacher in classroom time have a slight advantage, but this was not statistically significant	Part of larger study:- comparison between students in traditional RR (RR teacher not the classroom teacher at other times) and the control group (In small-group skills-based instruction). Randomised control trial	38 in RR; 53 in comparison group	Five subtests of the Observation Survey: writing assessment, CTBS Reading vocabulary subtest, CTBS Reading comprehension subtest	RR has statistically significant positive effects on print awareness, reading comprehension, vocabulary, dictation and writing vocabulary RR has significantly important effects on word recognition No significant effects on letter identification
Pinnell et al. (1994). In schools in 10 districts in Ohio	To investigate the achievement of four treatment groups: 1. traditional RR 2. RR-like individual group 3. RR-like group intervention 4. basic skills group	RR students were superior to the comparison group on four measures	Part of larger study:- comparison between the traditional RR group and the comparison group students who were in the same schools with regular reading instruction only. Randomised control trial	31 in RR; 48 in comparison group	Dictation subtest of the Observation Survey; Woodcock-Gates-MacGinitie Reading Test	Statistically significant positive effects on general reading achievement The WWC found no reported outcomes under other domains or constructs
Schwartz (2005). In 47 elementary schools in 14 states in the USA	To compare the outcomes of RR students and control students during the first and second half of the school year	The RR groups had significantly higher achievement than the control group and a comparison group of average students. Students in RR in the first half of the year performed better than students in RR in the second half	Part of larger study:- comparison of results of students in RR in the first half of the year with students on a wait list. Randomised control trial	37 RR students; 37 in comparison group	Six subtests of the Observation Survey: Yopp-Singer PA test; phoneme deletion; Slosson Oral Reading Test; Degrees of Reading Power Test	RR has positive effects on print awareness, phonics and fluency and general reading achievement (dictation and writing) Positive and important effects on phonemic awareness (potentially positive) Effects on letter identification not statistically significant or important No significant effects on comprehension
Iverson and Tunner (1993). In 30 schools in Rhode Island	To investigate the effectiveness of RR by comparing outcomes in. regular RR, a modified RR programme with more emphasis on phonics instruction, and a small-group intervention	Both RR groups performed significantly better than the small group intervention The modified RR condition was more efficient than regular RR No difference between two RR groups after 12 months	Part of larger study:- comparison of the outcomes of students in regular RR and the small group intervention. Quasi-experimental	32 in RR; 32 in comparison group	Five subtests of the Observation Survey; Dolch Word Recognition; Yopp-Singer PA Test; Phoneme deletion; Pseudoword reading	Statistically significant positive effects on phonemic awareness, print awareness, letter identification, phonics and general reading achievement (dictation and writing vocabulary)

Table 2

Measures of the WWC evaluators used to determine effectiveness within the phonics construct.

Measure	Study
Word Recognition subtest of RR assessment battery	Iverson and Tunmer (1993)Pinnell et al. (1988)Schwartz (2005)
Dolch Word Recognition Test	Iverson and Tunmer (1993)
Pseudo-word decoding task	Iverson and Tunmer (1993)

Table 3

Re-analysis of findings related to the alphabetics domain.

Research study	Phonemic awareness	Print awareness	Letter identification	Phonics
Baenen et al. (1997) (WWC strong design)	No reported measures	No reported measures	No reported measures	No reported measures
Pinnell et al. (1988) (WWC strong design)	No reported measures	$ES = 0.89$	$ES = -0.24$ (not statistically significant)	Questionable measures
Pinnell et al. (1994) (WWC strong design)	No reported measures	No reported measures	No reported measures	No reported measures
Schwartz (2005) (WWC strong design)	Two measures, $ES = 0.41$ and 0.46 (not statistically significant)	$ES = 1.04$	$ES = 0.23$ (no statistically significant)	Questionable measures
Iverson and Tunmer (1993) (WWC-met design standards with reservations)	Two measures, $ES = 1.73$ and 2.33	$ES = 1.58$	$ES = 0.94$	One measure $ES = 1.43$ and two questionable measures
Suggested revised rating	Potentially positive effects	Positive effects	Potentially positive effects	Not able to be determined
Note: ES , effect size.				

Table 4

Effect sizes from the Center et al. (1995) study on the post-test and short-term maintenance test for RR and control students (from Center et al., 1995) and on the medium-term test for RR and comparison students (from Schwartz et al., 2009).

Measure	Post-test ^a	Short-term maintenance test ^a	Medium-term maintenance test ^b
Clay's Book Level Test	3.05	1.55	1.55
Neale Analyses of Reading	2.03	1.50	1.12
Burt Word Reading Test	1.55	1.35	1.07
Passage Reading Test	1.57	1.45	1.23
Waddington Diagnostic Spelling Test	1.48	1.17	N/A
Phonemic Awareness Test	0.79	0.76	0.64
Cloze Test	0.42	0.69	N/A
Word Attack Skills Test	1.88	0.88	0.77
Woodcock-Passage Comprehension Subtest	N/A	N/A	0.98
Note: ^a RR = 22, control = 30; ^b RR = 23, comparison = 32.			