



Response to Intervention in Reading: A Literature Review and Critical Synthesis

Amanda R. Hurlbut, Jemimah Young, Catherine Boggs and Jamaal Young

Abstract

The purpose of this study was to review the existing literature on the use of Response to Intervention (RTI) in reading to critically analyze the methodologies, instruments, and findings within the context of the surrounding literature. RTI remains a key process in special education research and practice. Hence, studies range from intervention effectiveness, implementation fidelity, and methods for determining responsiveness to intervention. There are numerous RTI related research studies indicating that tiered or scripted intervention programs may help students identified as at-risk make academic progress on pre-and posttest measures. However, many of these same studies also indicate that students identified as at-risk do not receive the instructional support necessary to close opportunity gaps in reading. To address this concern, we conducted a systematic review of the RTI reading literature. The results indicate that a wide variety of screening and progress monitoring tools were utilized in reading research, which may account for the vast variation in efficacy across studies. Moreover, researchers cite validity, reliability, and replicability as main concerns in determining true responsiveness to an intervention when such a plethora of resources are available. We conclude that consensus is needed in the literature to determine the best screening and progress monitoring instruments to identify true responsiveness and distinguish the best methods for designing, studying, and replicating intervention programs that sustain academic performance by at-risk learners through an RTI based tiered intervention model.

Keywords: Response to Intervention, Special Education, Literature Synthesis, Specific Learning Disability (SLD)

Contact: Amanda R. Hurlbut, Texas Woman's University

E-mail: amandahurlbut@twu.edu

Jemimah Young, Texas A&M University

ORCID: <https://orcid.org/0000-0001-6598-9196>

E-mail: Jemimah.young@tamu.edu

Catherine Boggs, Texas A&M University

Jamaal Young, Texas A&M University

ORCID: <https://orcid.org/0000-0001-7277-1072>

E-mail: Jamaal.young@tamu.edu

© 2023 *IJLCLE* and Amanda R. Hurlbut, Jemimah Young, Catherine Boggs, Jamaal Young

Response to Intervention in Reading: A Literature Review and Critical Synthesis

Response to Intervention (RTI) is an established process designed to support struggling learners in general education settings. RTI is an approach in which identifies and supports students who are struggling with reading. The RTI framework involves a multi-tiered system of support, interventions are provided at increasing levels of intensity depending on the student's level of need. The goal of RTI is to identify struggling students early and provide them with targeted interventions that will help them catch up to their peers. In the RTI model, students are assessed regularly to determine their reading ability and progress. If a student is struggling with reading, they are provided with an evidence-based intervention that is designed to address their specific needs. The intervention is monitored closely to determine whether it is effective, and if not, adjustments are made until the student shows improvement. If a student does not respond to the first level of intervention, they may be moved to a higher level of support.

Educators use the RTI process to help students struggling with a lesson or skill. In practice, teachers implement interventions with struggling students to foster success, and it is important to note that RTI is considered a general education strategy. However, the RTI process, in which identifies and supports, is used in general education practice to intervene before a student is referred to special education. Thus, increasing the efficacy of RTI has substantial implications for student success in general and special education settings.

The RTI framework is grounded in the idea that all students can learn, and that early intervention is key to addressing learning difficulties (Fletcher & Vaughn, 2009). By providing targeted interventions at increasing levels of intensity, the RTI model allows schools to meet the needs of a wide range of learners. Research has shown that RTI can be an effective way to improve student outcomes in reading, and many schools have adopted the approach as part of their broader efforts to support struggling students.

Although RTI is not used to identify students with special needs, its connection to the special education identification process is important in supporting students who do not respond to previously enacted interventions. Moreover, suppose a student does not respond to initial interventions. In that case, many RTI teams will provide increasingly intense interventions that subsequently work to identify a Specific Learning Disability (SLD) if the student continues to be unresponsive to intervention (Special Education Guide, 2021). As defined within the Individuals with Disabilities Education Act (IDEA), “a specific learning disability is a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations.” (IDEA, 2004, p. 11-12). It is estimated that approximately 80% of all students qualified as having a SLD struggled with poor reading development (Penesetti, 2018). Therefore, much of the RTI research that exists is situated in the reading content area to support the growing number of struggling readers.

The effects of RTI often vary across implementation settings and contexts in the United States. In 2004, Congress reauthorized the Individuals with Disabilities Education Act (IDEA), formerly known as the Education for all Handicapped Children Act PL 94-142. This act had several new updates in regard to the identification and eligibility of students with SLD. These new federal education mandates gave state and local education agencies flexibility for evaluating children with suspected learning disabilities by no longer requiring states to utilize an intelligence quotient or IQ-achievement discrepancy formula. Under this standard, school districts were required to prove that a

significant gap existed between a student's ability or intelligence level and his/her achievement before the student could be identified as having an SLD. This gap was typically identified by indicating that a student's achievement was well below what was projected through a series of intelligence and norm-referenced achievement assessments. This process usually resulted in delayed SLD identification, which often facilitated an increase in academic differences. Many students had to wait several years before they were eligible for special education services based on predetermined achievement differences (Bradley et al., 2007; Ofeish, 2006). Thus, the IQ-achievement discrepancy model became known as the "wait-to-fail" model.

Research demonstrated that this model was not always accurate in identifying students with learning disabilities. The model tended to over-identify students who were not learning disabled, especially students from diverse cultural and linguistic backgrounds. At the same time, it also failed to identify students early on who truly had an SLD (Restori et al., 2008). The IQ-achievement discrepancy model's inability to efficiently identify students for special education led to the need for an alternative. The emergence of RTI as the primary alternative to the IQ-achievement discrepancy model can help literacy educators mitigate the learning loss effects for struggling readers who were without services during the evaluation period and possibly support struggling learners that do not have an SLD gain the skills to succeed in general education.

Researchers and lawmakers proposed RTI to remediate the over and under-identification of students with potential learning disabilities. In this model, at-risk students are identified through a series of screening efforts and then receive scientific or research-based instruction and interventions in the general education classroom. These interventions typically increase in intensity, frequency, and duration if a student continues to display academic struggles through RTI tiers or intervention levels. After some time, if a student fails to respond to scientifically based instruction, as shown through continuous progress monitoring, the student is considered for a special education evaluation to determine eligibility as a student with a SLD.

Given the prevalence of reading challenges present, RTI studies in this area must reflect the highest degree of reliability, validity, and implementation fidelity. Since the reauthorization of IDEA 2004, researchers have extensively studied RTI instructional strategies, intervention processes, methodologies, and instruments being used to determine SLD status. The increased implementation has resulted in numerous primary studies and systematic reviews of the literature (Barrio et al., 2015; Erchul, 2011). However, there is a large degree of variance in the efficacy of study results amongst these studies. Prior research syntheses focused on implementing RTI as a moderator of student responses to intervention (Cartledge et al., 2016; Fletcher & Vaughn, 2009). One of the few literacy-specific reviews of RTI focuses on the beliefs, assumptions, institutional structures, and divisions of expertise that have emphasized special education identification rather than remediation (Johnston, 2011). In the present study, we focus on the moderating effects of research practices on student RTI outcomes in reading intervention studies. This is a necessary shift because it is important to recognize that "methods matter," as these studies' results have implications for literacy instruction.

For this literature review, we focus on research studies conducted evaluating students' response to specific, tiered interventions in reading that support the RTI framework as a research-proven way to provide intensive intervention to struggling learners. Additionally, this discussion will address methodologies and instruments used to screen, progress monitor, and determine responsiveness or define academic risk status and SLD identification. To address these goals, we will first review the design challenges present in RTI studies conducted in the reading content area. Then, we systematically review the most impactful RTI studies in the area of reading (i.e., the most cited

studies). Finally, we discuss the implications of the current research trends and major themes on the efficacy of RTI in the reading content area.

Design Challenges in Reading Intervention Studies

Much of the research on reading RTI supports using a tiered model with increasing intervention intensity, frequency, or duration as students fail to show progress. Research in reading focuses mainly on early intervention and demonstrates an “inch deep, mile-wide” mentality in the scope of research completed. Additionally, various screening, progress monitoring, and IQ-achievement measurements are used to determine responsiveness and future interventions for students. This process's complexity requires that all of the possible moderators are examined and critiqued to maximize student learning and successful intervention implementation.

Since RTI is not a new topic in the literature, there is a wealth of information available. Moreover, research in early intervention methods is rapidly gaining momentum. Several research studies in RTI focus on a wide range of issues from Curriculum-Based Measurement (CBM), the tiered intervention model, intervention fidelity, intervention implementation, and responsiveness to intervention. Much more information in each of these areas is needed to corroborate emerging data's efficacy in these topics. In the sections that follow, we review the prior research on RTI in reading intervention studies to establish what is currently known in the field and lingering empirical concerns.

Tier Structure in Reading Intervention

To provide appropriate interventions to students, teachers and other educators must accurately identify, monitor, and assess the students' ever-changing needs. Unfortunately, many students fail to respond to interventions, presenting challenges for parents, teachers, and schools. Understanding design and implementation factors that moderate the variance in RTI implementation is essential to reducing the number of non-responders to intervention. Here we review the key studies that have attempted to address the challenges of unresponsiveness in reading intervention studies.

Non-responders to interventions are a major concern. Thus, numerous studies have attempted to pinpoint the mediating factors that inhibit student responses to reading intervention. McMaster et al. (2005) distinguished students who responded to regular classroom instruction in reading from the non-responders to compare the growth of the non-responders to that of the typically developing peers. The researchers specifically examined performance after receiving varying treatments of intensive reading instruction. Additionally, McMaster et al. (2005) sought to identify non-responders through a dual-discrepancy approach and used a variety of instruments to determine responsiveness to intervention, including several subtests of the *PALS*, reading words of the *Woodcock Reading Mastery Test-R or WRMT-R* (Woodcock, 1987), spelling on the *Weschler Individual Achievement Test or WIAT* (Psychological Corporation, 1992), and Nonword Fluency Probes on the Dynamic Indicators of Basic Early Literacy Skills (*DIBELS*) assessment (Good & Kaminiski, 2002). Despite the innovative approaches, the study results revealed no statistically significant differences between the varying treatment types. However, the non-responders did make some gains as a whole, which is of practical significance.

However, McMaster et al. (2005) did find validity in using a dual discrepancy approach to identify and describe non-responders by using both growth and cut scores. This is substantial for future research considerations because it shows a need for a prolonged study period with more extensive research measures in place to compare intervention strategies. The findings of this study support broader research that intervention pullout methods are in many cases more effective than

general classroom instruction for some students (McMaster et al., 2005). Pullout tutoring is typically expensive because it often requires additional school support personnel, yet due to this strategy's effectiveness, it is gaining momentum in research and practice.

Results from a preventative reading study by Gilbert et al. (2013) corroborate the findings of McMaster et al. (2005) and reveal that students who do not respond to standard tier one classroom instruction can benefit from supplemental reading tutoring compared to students who remain in the classroom and receive no additional intervention. However, participation in this tier two level tutoring does not necessarily prevent future reading difficulties and the need for continued or future secondary or tertiary intervention (Gilbert et al., 2013). At the end of first grade, 41% of the students who received the supplemental tier two tutoring failed to score in the normal range on the Rapid Letter Naming task and the fluency probes of high-frequency words; this is consistent with the 45% of non-responders that McMaster et al. (2005) identified in their study, although the screening and progress monitoring instruments used were different. These results highlight the need to further investigate the effects of design and implementation differences on reading intervention effectiveness. Thus, in the next section, we examine the effect of sampling on student responsiveness to reading interventions.

Sampling in Reading Intervention

Population sampling in reading interventions studies use a variety of characteristics and procedures to select sample participants in studies. These are crucial yet often underexamined considerations related to treatment efficacy in intervention studies. VanDerHeyden et al. (2007) examined developmental age when evaluating young learners' early literacy progress by considering development age as a factor in the research. In their study, preschool-aged students participating in a government-sponsored Head Start program demonstrated growth in the areas of rhymes, alliteration, fluency, and letter-naming on probes of the *DIBELS*. However, one possible mediating factor is that rather than a control group of peers at-risk for SLD not receiving the intervention, the experimental group was a cohort of public preschool-aged students. The effects of this sampling decision could afford or constrain the implementation of the results of this study.

The absence of a comparable control group was also a consideration in the outcome of a study conducted by Little et al. (2012), which compared kindergarten student achievement when using a commercial-based early reading intervention versus a school-designed intervention. Students were randomly assigned to either of the two intervention groups with no control group present. Findings from the study indicated that there were no statistically significant differences between the two intervention groups; however, effect sizes in the commercial intervention group differed on the measures of sound matching, nonsense word fluency, oral reading fluency (*DIBELS*), and written spelling with effect sizes ranging from (.34 to .59). It is important to note that this study's findings are divergent from previous studies, where statistically significant differences were found. Many students in both groups performed above the 30th percentile on all performance measures post-treatment. There are two limitations of this study, however, the first is the small sample size, and the second is the geographical region of the participants used for the study, which is different from previous studies in this area.

While in one of the few studies that applied random sampling assignment, Fuchs et al. (2008) assessed the long-term effects of small group secondary intervention tutoring on identified at-risk students, to determine how the instruments used to identify risk status influenced measures of responsiveness offering methods and measures that define responsiveness versus non-responsiveness with consensus needed in the broader literature. This study did not specifically discuss how students performed on each intervention and instrumental measure, but rather focused on the need for

instruments with more precise, “sensitivity, specificity, severity, and prevalence” (p. 433) to accurately define at-risk students and measure responsiveness. This leads us to the next possible moderator of effects—instrumentation.

Instrumentation in Reading Intervention

The effects of RTI tend to vary based on the screening tool administered, how it is administered, and who it is administered to in the intervention. Tools used as screening instruments in reading intervention include the Harcourt Trophies Pre-K Beginning Sound Awareness (CBM) (Harcourt School Publishers, 2002) and the Beginning Sound Awareness subtest of the Phonological Awareness Literacy Screening for Preschool (PALS)-PreK (Invernizzi et al., 2002). Moreover, many of the instruments used in reading RTI interventions include IQ-achievement tests (Woodcock Reading Mastery Test (WRMT-R-R, Woodcock- Johnson III (WJ-III), Wechsler Individual Achievement Test (WIAT), etc.). These assessments often require parental permission before a school can administer the test to students and usually must be given in a one-on-one setting. Other screening and progress monitoring tools such as the DIBELS, Texas Primary Reading Instrument (TPRI), or Curriculum Based Interventions (CBI) are generally adopted as part of the regular, general education curriculum and, therefore, do not require parental permission to administer since they are used as screeners and academic progress monitoring tools. School professionals can typically administer these assessments to small or large groups of students without compromising test validity or reliability. These tests are often affordable ways for a district to collect data on student achievement and academic progress. Unfortunately, there is a consistency in the instrumentation efficacy of these results as the instruments mentioned above were associated with varying levels of success when implemented.

The DIBELS literacy assessment is one of the most popular screening tools, as evidenced by its extensive use in prior studies. A specific purpose of the study by Goffreda, Diperna, and Pedersen (2009) was to determine the predictive validity of a student’s DIBELS scores upon future reading proficiency in district and state standardized scores. The researchers used DIBELS data in first grade and then later compared student results to their performance on the TerraNova (CTB/McGraw-Hill, 2002) in second grade and the Pennsylvania State Assessment in Reading (PSSA) in third grade. Findings from the study indicate that the oral reading fluency subtest of the DIBELS was the only significant predictor of future TerraNova and PSSA at $p < .001$; performance on the other DIBELS subtests did not statistically significantly increase the odds of predicting performance on the other two assessments. This study’s two major limitations were the limited sample size and the lack of representation among diverse ethnic and racial groups, as demonstrated by the campuses where the samples were drawn.

Additionally, screening results also indicated a sustained intervention effect based upon the administration of the DIBELS in kindergarten (Good & Kaminski, 2002). A severe limitation of this study is the fact that very young children develop at varying rates. Thus, the growth could easily be falsely attributed to the tiered interventions provided and not to the individual child’s development. Additionally, the study focused on the response to tier two interventions; it is unknown whether the non-responders continued to receive secondary intervention or considered more strenuous tertiary intervention. Other areas in reading development have been evaluated in the intervention literature and often require different instruments.

A study focusing on tier three reading interventions in fluency, decoding, and comprehension demonstrated that students made meaningful gains in these areas, based on the following measures Test of Word Reading Efficiency or TOWRE (Torgensen, Wagner, & Rashotte, 1999), Gray Oral Reading Test or GORT-4 (Wiederholt & Bryant, 2002), and the WJ-III (Denton et al., 2006).

However, it is important to note that these studies are distinct from other studies. Specifically, certified teachers and not graduate students or research assistants implemented the interventions. Thus, making the program unique in that practitioners rather than researchers delivered the instruments. Hence calling to question the external and internal validity of the results and possibility limiting the study's generalizability to prior and future studies.

Intervention Implementation

One of the most important considerations related to the effectiveness of RTI is the fidelity of the implementation of the interventions. Koutsoftas and colleagues (2009) focused specifically on the nature of the interventions provided. The findings support the existing literature that many students make significant growth after being provided with intensive interventions and continuous progress monitoring in phonemic, phonological, and print awareness. The researchers observed that 71% of children responded to the treatment intervention through small group, scripted tutoring; however, the results were not compared to that of a control group. In contrast, Denton et al. (2011) sought to examine the effects of duration and scheduling on tier two intervention progress.

Unlike previous studies, Denton et al. found that a longer intervention time and modified schedules did not significantly impact intervention effectiveness and growth. However, student progress across the three intervention schedules was compared via a pretest-posttest design rather than against a control group. Approximately 77% to 83% of students demonstrated adequate instructional response on the screening measures' decoding criterion. However, the authors state that many students, both responders, and non-responders, may require more extensive intervention in the future.

Other studies also evaluated the relationship between the growth of non-responders and duration. Vaughn et al. (2009) examined the effects of intervention duration on response to secondary and tertiary intervention. One key element of the Vaughn et al. study was the total number of sessions that researchers completed using intensive reading instruction in second grade for non-responders receiving secondary and tertiary interventions and in the fidelity measures to ensure for validity and reliability of intervention implementation. The findings support the need for increased intervention time versus intervention variety by showing that non-responders demonstrated statistically significant differences in progress over time. The researchers also observed statistically significant gains for non-responders in word identification and passage comprehension but not in oral reading fluency. The authors suggest that more intensive and extensive interventions are needed for students with continued low or non-response to the interventions and conclude that more investigation of providing alternative or specialized interventions is needed.

The data presented from this literature review indicate that RTI can be an effective means to support students at risk for SLD. However, design and methodological considerations must be reconciled to support the efficient implementation of these reading interventions. The purpose of this paper is to provide a critical review of the research findings from high-impact studies in the reading RTI literature. We focus specifically on the methodologies and instruments used in reading RTI studies cited at least 20 times based on the Web of Science database's data.

Method

Two distinct searches were conducted to locate relevant research studies for this literature review using computer database search engines. These databases include Academic Search Complete, Education Research Complete, Educational Resources Information Center (ERIC), PsycINFO, and

the Professional Development Collection. In the first search, the keywords of “response to intervention” and “reading” yielded 3,259 results and were reduced to 1,645 after requiring the articles to be peer-reviewed, published in academic journals, and published after the year 2004. The year 2004 was chosen since the focus of the literature review was targeting RTI related research after the reauthorization of IDEA 2004, as previously discussed. Specifying certain keywords and topic themes further narrowed these results. These keywords included quantitative, reading and reading intervention, education, school-based intervention, special education, and learning disabilities as related terms. This reduced the results to 157 articles for review.

To select studies that best met the purposes of this review, the researchers determined article eligibility by inspecting elements in more depth by reviewing the title and abstract to determine research purpose, full-text availability, age of the participants being studied (elementary vs. secondary) as an initial impression of appropriateness for the present review. If the title and abstract were insufficient, the article was retrieved and reviewed in its entirety. Studies conducted in elementary classroom settings or elementary learners were preferred, as RTI is often defined as an elementary intervention.

Articles that met these minimum criteria were stored in an electronic folder for further investigation. A total of 23 reading studies were initially recovered. Because we wanted to include the most impactful studies in the field, we only included studies with at least 20 citations. To account for the influence of the “vintage effect,” we only included studies that were conducted between 2004 and 2014. We retrieved our data between August 2020 and December 2020, allowing all studies at least five years to acquire citations to account for the vintage effects. The literature search yielded a total of 23 studies in reading intervention, 11 of which were chosen for closer analysis due to the purpose of the study, study design, instruments used to screen and progress monitor students, and results reporting. Many of the studies that were not included either differed substantially in the overall research purpose (determining an instrument for predictive validity) or did not demonstrate consistency with the purposes stated in this literature review. The full scope and description of each study is provided in table 1.

Results

Table 1.1 provides a summary of the reading studies, and an overview of methodological and design approaches present in each of the studies. In the following sections, we examine trends that emerged across the following study elements: participant characteristics, methodology, instruments, and study results.

Table 1. *Methodological Summary of Reading Studies*

Author	Purpose	Participants	Method	Instruments	Validity/ Reliability	Results	Citation (f)
Compton et al. (2012)	To identify necessary data to predict at-risk failure to respond at tier 1 and 2 to bypass interventions to tier 3	N=485 first through fourth-grade students across three years (cohorts) across 15 schools in urban/suburban Nashville	Quantitative Comparative Treatment vs. Control	CTOPP WJ-III WRMT-R TOWRE	Intervention provided by research team; Fidelity measures of intervention not discussed	Effect sizes reported for each instrument subtest and comparison regressions. Tier 2 data may not always be necessary to predict unresponsiveness at tier 2 to accelerate intervention to tier 3; Data distinguished non-responders with 90% sensitivity and 80% specificity	109
Denton et al. (2006)	To study tertiary intervention in oral reading fluency	N=27 K – 3rd-grade students in 4 schools	Quantitative Experimental Treatment vs. Treatment	TOWRE GORT-4 WJ-III	Small, heterogeneous sample size; no control group; Interventions provided by certified teachers	Effect Sizes reported as standard errors in each instrument subtest. Intervention increased test scores; Reading ability remained lower than average. No indication of sustainability	405
Denton et al. (2011)	To study different forms of tier two intervention schedules	N=273 first grade students in 9 schools, 2 districts (1 large urban, 1 small rural in Texas)	Quantitative Experimental 3 treatment conditions	TPRI WJ-III CPMERS TOWRE GRADE	Treatment group results compared against each other; not compared to a control; Findings not consistent with prior research; Intervention provided by research staff	Results reported in Means and <i>SD</i> pretest, after 8 weeks, and after 16 weeks of intervention; Effect sizes for each subtest were reported. Varying intervention schedules did not reveal significant differences in student growth. Pool of at-risk students decreases the actual sample size	72
D. Fuchs et al. (2008)	To determine long-term effects of secondary intervention; To study how the measures used determined student responsiveness to the interventions.	N=252 From 42 first grade classrooms in 16 schools in an urban and suburban school district in Tennessee; Half of the schools designated as Title I (low SES)	Quantitative Experimental Treatment vs. Control	WIF and RLN from the CTOPP PALS DIBELS WRMT-R WIAT	Intervention provided by research staff	Effect sizes reported for each instrument subtest in addition to Means and <i>SDs</i> of pre and posttest batteries; Results revealed significant main effects on time; in other words, performance increased with added time.	247

Gilbert et al. (2013)	To study the effect of a preventative, multi-tiered intervention in the form of tutoring upon first-third grade achievement	N=437 students screened across two years; more than 95% of the sample was considered economically disadvantaged	Quantitative Experimental Treatment vs. Control	Researcher created fluency probes and rapid letter naming tasks	Short intervention duration; lack of program differences between tier 2 and tier 3 intervention; did not account for tier 1 intervention; Intervention provided by the research team	Varying tools have varying determinants of identifying responsiveness; Treatment group improved more than a control group on the raw score in WIF Results reported in pre, and posttest means and SDs; all made gains but some more than others; ES reported in standardized mean difference between treatment and control was .19	119
Goffreda et al. (2009)	To determine predictive validity in student's DIBELS scores upon future reading proficiency on state assessment	N=67 at a rural school in Pennsylvania; 21 student participants were economically disadvantaged	Quantitative Comparative Instrument predictability	DIBELS TerraNova-CAT Pennsylvania State Assessment	Sample was not representative of diverse racial and ethnic groups in the school population; generalizability	At-risk students benefited from a supplemental tier 2 reading program outside of the classroom; tier 2 responsiveness did not necessarily prevent later reading difficulties Effect sizes reported as Beta scores for each DIBELS subtest predictor; ORF category of the DIBELS was the only significant predictor of future TerraNova and PSSA proficiency (B=2.77, p=.001) The ORF portion of the DIBELS is administered later in the year, while other portions are administered early on.	124
Koutsoftas et al. (2009)	To study secondary intervention in phonemic awareness and determine sustainability in kindergarten	N = 34 at-risk preschool students	Quantitative Experimental Pre vs. Post-treatment	Tropics PreK CBM PALS PreK DIBELS	Small sample size; no control group; No control on tier 1 intervention; Intervention provided by research staff	Effect sizes computer for each child on the baseline, pre, and post-intervention Means and SDs used.	155
Little et al. (2012)	To compare kindergarten achievement on a commercial versus school developed	N = 90 students from 8 elementary schools;	Quantitative Comparative Experimental	Florida Assessment for	Small sample size; no control group; students in both groups were controlled for	Intervention increased scores. Results were sustainable on DIBELS in kindergarten ES range (.34 to .59) on students in the commercial ERI tutoring; however, not	21

	intervention program; to determine effects of tier 2 intervention	approximately 63-92% of students designated as economically disadvantaged	Treatment vs. Treatment	Instruction in Reading CTOPP DIBELS WRMT-R Peabody Picture Vocabulary Test Test of Written Spelling-4 PALS DIBELS CBM WIAT WRMT-R	intervention dosage and group size; Intervention provided by schoolteachers with research trained PD sessions	statistically significant; ES reported for each instrument subtest. Students in both groups performed above the 30 th percentile on all performance measures post-treatment.	
McMaster et al. (2005)	To explore the validity of utilizing a dual discrepancy approach for determining responsiveness; to compare the achievement of students in varying treatment intervention programs	N = 496 students at 8 Metropolitan Nashville schools; 4 schools classified as Title I	Quantitative Experimental Treatment vs. Treatment		Low statistical power of the study; effect sizes not statistically significant; small sample size when responders removed from the sample; lack of control group; School teachers provided intervention	Effect sizes reported for each instrument subtest and responder groups in treatments; Effect sizes post-test (between .30 to .50) suggest that tutoring is most helpful for reducing unresponsiveness. There were no statistically significant differences across the treatment approaches on the monitoring measures or other reading-related measures; Found validity in utilizing a dual-discrepancy approach to responsiveness	457
VanDerHeyden et al. (2007)	To identify/compare the at-risk status of preschool children on varying measures and with different forms of intervention	N = 20 Preschool Head Start N = 15 Preschool in a rural setting	Quantitative Experimental/ Comparative Treatment group vs. Treatment group	DIBELS Brigance Preschool Screen First Steps Screening Test for Evaluating Preschoolers	Small sample size; no control group; Intervention provided by research staff	ES Cohen's d = .43, p = .249 that children in Head Start had slightly lower scores, not statistically significant but low sample size A single measurement tool could not predict risk status with 100% accuracy on other instruments; Whole group intervention affected growth more than individual intervention	103
Vaughn et al. (2009)	To examine how first-grade non-responders would continue to respond with intervention at the secondary and tertiary intervention level	N = 275 students in two cohorts; Cohort 1 N = 153, Cohort 2 N = 121; 7 elementary schools	Quantitative Experimental Treatment vs. Control	WRMT-R DIBELS Peabody Picture	Did not use both growth and benchmark measures (dual discrepancy) when identifying responsiveness; small sample size after the non-responders	Effect sizes computed and reported for each instrument subtest.	226

in a small district
near a big city in
southwest

Vocabulary
Test-III

were identified; Intervention
provided by research team

No statistically significant
effects for tertiary
intervention treatment on oral
reading fluency and word
attack; lower responders did
make progress over time, but
many continued to
demonstrate a lack of or low
response to the treatments

Participant Characteristics

Many of the research studies (six out of eleven) included in this analysis relied on participants from different settings to study since the research topic typically required an academic at-risk sample. Many of these studies attempted to locate participants from a variety of regions, school districts, and individual campuses and due to participant age, studies required parental consent in addition to district and campus permission for the study. A common trend that emerged was a narrowing of the sample size. A smaller number of the original sample met the qualifying criteria by struggling to respond to the general curriculum and became labeled as “at-risk.” An original sample of 500 to 600 students was easily reduced to a smaller sample of 200 to 300 when considering the need for tiered intervention as evidenced by the screening and progress monitoring tools. Two studies specific to pre-school intervention had a very small sample size of less than 40 (i.e., Koutsoftas et al., 2009; VanDerHeyden et al., 2007), while two other studies had a moderate participant size of 67 and 90, respectively.

Additionally, six of the eleven studies addressed economic disadvantages in the student population as well as the relationship to academic risk status. In these studies, the researchers attempted to match treatment and control groups (if utilized in the research design) similar to the original sample. However, the reduction in the original sample size did not always make this feasible. The participants being studied were typically young elementary-aged students, pre-school to third graders. This is aligned to the nature of RTI early reading intervention since pre-school through third grade is the prime age where most students acquire the necessary skills to read text. Finally, the research occurred at districts and campuses from various large-city urban, suburban, and rural areas, mainly located in the Southern regions of the United States.

Methodology

All of the research studies used in this literature review were quantitative, with most being experimental. Typically, a study was designed to test an intervention protocol or methodology and compare against either an alternate form of treatment or against a control group of at-risk students to determine intervention effectiveness. When designing an intervention protocol to use in a treatment study, many researchers relied on graduate research assistants and paid trained tutors as the intervention providers. This will be discussed more in-depth in the section on study results but is an important component to mention here as part of the methodology.

The researchers put extensive fidelity measures in place to ensure that similar or identical intervention protocols were being delivered to ensure the study's internal validity. If classroom teachers were delivering the intervention, there was typically an alternate protocol to ensure that teachers strictly adhered to the intervention and measurement timeline put into place by the researchers to improve internal and external validity. Four of the eleven studies utilized a treatment vs. control methodology. One study was comparative. One study focused exclusively on pre and post-intervention test results of a single group following an intervention. The remaining five studies evaluated two or more treatment conditions without a control group.

In an instrument comparison study, Goffreda et al. (2009) sought to determine if the DIBELS screening battery given to first-grade students was a reliable predictor of future performance on the TerraNova and state assessment. This study was distinct from the other studies, as it did not incorporate a treatment protocol. However, the limitations of the other studies discussed how often classroom, tier-one instruction was not controlled. This study set out to look at how one instrument predicted success using another measure with no specific intervention. With tier one general instruction, it was included as part of the literature review.

Instruments

A consistent theme in measuring reading intervention effectiveness is using a wide variety of instruments or multiple screening and progress monitoring tools to define student responsiveness to an intervention. Researchers repeatedly discussed determining intervention effectiveness upon a single screening or progress monitoring tools since validity and reliability issues will arise. When choosing tools to use, researchers must consider the research conducted on the individual tool, the purpose, and the instrument's tendencies to either over-identify students who are not truly at risk or fail to identify students who need additional academic support. This phenomenon is referred to as identifying a “false positive” or “false negative” in RTI literature (D. Fuchs et al., 2008). Subsequently, determining what classifies a student as non-responsive to the intervention is subjective. It varies depending on the intervention being used, screening and assessment measures, and the cut point for identifying non-responsiveness.

As previously discussed in the broader literature, the DIBELS reading assessment is one of the more widely supported instruments based upon its utilization across the included studies. The data from this literature review indicate that the DIBELS can be administered to large groups of students as part of the general education curriculum, does not require parental permission for a school to administer, and is easily or readily available in many schools (Goffreda et al., 2009; McMaster et al., 2005; Vaughn et al., 2009). However, the decision to implement is usually a district or school-based decision. It may not necessarily be available in all educational settings unless the researcher chooses to purchase and implement. This makes gathering data somewhat more difficult than simply using instruments that are already available and used by a school.

Study Results

Reporting practices varied substantially based upon the nature of the study. The majority of the researchers reported their findings in the form of mean scores pre-and post-intervention, standard deviations, and effect sizes comparing the intervention with each subtest of the instrument used, but also in the effect sizes comparing the results from the treatment groups with other treatment or control groups (Compton et al., 2012; Denton et al., 2011; Vaughn et al., 2009). When necessary, researchers reported actual *p* values. They discussed the statistical significance of each finding or the lack thereof and possible reasons for the findings (e.g., small sample size, design study, etc.). Because of the studies' varying nature in question, there was limited ability to quantify each of the effect sizes to compare results from one study to the next. Rather, the effect sizes reported were used as an indication of instrument reliability in predicting responsiveness or determining the effectiveness of an intervention compared against an alternate treatment or control group for the design of that study.

Almost every study addressed in some form reliability and validity issues, which usually corresponded to how the intervention treatments were given. For instance, in the studies that relied on graduate assistants and paid tutors to deliver the intervention, extensive measures were put into place to ensure the intervention protocol's internal validity. In many cases, the interventionists were recorded and/or observed to ensure adherence to the intervention plan or script, and the results were quantified to measure intervention consistency. These measures were replicated in the assessment procedures to ensure that screening and progress monitoring results were also valid. In the few instances where classroom teachers provided the intervention tutoring or scripted instruction (Denton et al., 2006; Little et al., 2012; McMaster et al., 2005), similar procedures were implemented and evaluated to account for validity and reliability of intervention fidelity across multiple classrooms.

While using trained graduate research assistants and paid tutors was certainly easier for researchers to control when considering intervention fidelity, this practice does have limitations. The purpose of many RTI studies is to find intervention effectiveness in the program, instruments, and intervention protocols used to be replicable at other districts and campuses. However, when trained research assistants deliver intervention lessons without considerable fidelity measures in place, it makes RTI implementation more difficult in the broader context of an educational system without tutors and researchers' assistance, thus affecting future reliability. This was the case in six out of the 11 studies. In contrast, four studies had active involvement from the certified teachers in the classroom, while the remaining study did not have an applicable intervention protocol.

Another frequent limitation in intervention protocol research was the lack of researcher control or design on tier one intervention. Many of the researchers in these studies sought out to deliver tier two and three small group and individualized tutoring without first studying what had been done in the classroom show students as unresponsive to the tier one general curriculum. Only two of the studies attempted to address this concern in the present study by ensuring that the screening occurred at a certain point during the regular curriculum before delivering intervention lessons (Compton et al., 2012; Gilbert et al., 2013). Most researchers attempted to discuss the findings within the larger context of the existing research and detailed discrepancies and proposed hypotheses for inconsistencies that arose.

Finally, as discussed in the section on participant characteristics, a common limitation listed by almost all studies was the reduction in the size of the participant sample or a low sample size due to the study's scope. As mentioned, six of the studies had more than 200 participants after being reduced from an original, larger sample. A common theme in reading RTI research is reducing the sample because of the need to study students' responses to intervention protocols. The student participating in the intervention must be identified as a non-responder at one or more levels of the process. Thus, many of the children originally selected to participate in a study will not be eligible simply based upon their responsiveness to instruction to tier 1 instruction. These students are often used as comparison groups to determine if intervention protocols can effectively close the achievement gap between responders and non-responders after some time.

Discussion

Several important considerations and limitations have emerged because of the present literature synthesis. First, the issue of sample size repeatedly surfaced in researcher limitations. The requirement for an at-risk population in a particular subject area only narrows an original district or campus population. Thus, there is a great need for improved participant recruitment procedures and sampling designs to meet the field's needs. One recommendation is that researchers consider studying intervention treatments at multiple districts and campuses from various urban, suburban, and rural areas as this form of participant diversification was absent in the data review in the present study.

Additionally, nearly all of the study designs focused on research interventions delivered by a team of researchers rather than practitioners. While having a research team to deliver interventions certainly helps to maintain the fidelity of intervention delivery by reducing threats to internal validity and reliability. Reading RTI research outcomes are virtually all based upon studies where the research team delivered the treatment. This raises the question about reliability and generalizability in the broader context of a teacher-driven intervention program. Moreover, authentic interventions typically occur in a classroom environment delivered by instructional personnel. Furthermore, given the vulnerable nature of the student population of interests, researchers should consider the efficacy of

student-teacher interactions to reduce student anxiety which could be an underreported consideration in student non-responsiveness.

The researchers repeatedly cited issues related to procedures and variations in how the screening and progress monitoring instruments were used to determine intervention responsiveness. Some researchers cited and defined the dual-discrepancy approach as a statistically valid method for determining the intervention's response. However, there is a lack of consensus on which instrument produces the most desirable outcomes consistently. Unfortunately, determining true response and minimizing both false positives and false negatives in at-risk identification remains a challenge. Rather, multiple tools are recommended to increase outcomes efficacy. Future research should focus on synthesizing the literature on each tool, including instrument means, standard deviations, effect sizes, and design methodologies in which the instrument was used. The large body of quantitative literature related to RTI in reading, warrants and updated meta-analysis.

Finally, RTI studies in reading varied greatly in their purpose and subsequent research approach. Some studies set out to establish validity using screening and progress monitoring measures, while other studies have a dual purpose of identifying intervention and treatment effectiveness. A consensus is needed in the literature to determine the best screening and progress monitoring instruments to identify true responsiveness and distinguish the best methods for designing, studying and replicating intervention programs that sustain academic performance by at-risk learners through an RTI based tiered intervention model.

References

- Barrio, B. L., Lindo, E. J., Combes, B. H., & Hovey, K. A. (2015). Ten years of response to intervention: Implications for general education teacher preparation programs. *Action in Teacher Education*, 37(2), 190-204. <https://doi.org/10.1080/01626620.2015.1004603>
- Bradley, R., Danielson, L., & Doolittle, J. (2007). Responsiveness to intervention: 1997 to 2007. *Council for Exceptional Children*, 8-12. <https://doi.org/10.1177/004005990703900502>
- Cartledge, G., Kea, C. D., Watson, M., & Oif, A. (2016). Special education disproportionality: A review of response to intervention and culturally relevant pedagogy. *Multiple Voices for Ethnically Diverse Exceptional Learners*, 16(1), 29-49. <https://doi.org/10.5555/2158-396X.16.1.29>
- *Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., Barquero, L. A., & Bouton, B. (2012). Accelerating chronically unresponsive children to tier 3 instruction: What level of data is necessary to ensure selection accuracy? *Journal of Learning Disabilities*, 45(3), 204-216. <https://doi.org/10.1177/0022219412442151>
- CTB/McGraw-Hill. (2002). TerraNova, the second edition. Monterey, CA.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52(3), 219-232. <https://doi.org/10.1177/001440298505200303>
- *Denton, C. A., Cirino, P. T., Barth, A. E., Romain, M., Vaughn, S., Wexler, J., Francis, D. J., & Fletcher, J.M. (2011). An experimental study of scheduling and duration of "tier 2" first-grade reading intervention. *Journal of Research on Educational Effectiveness*, 4, 208-230. <https://doi.org/10.1080/19345747.2010.530127>
- *Denton, C. A., Fletcher, J. M., Anthony, J. L., & Frances, D. J. (2006). An evaluation of intensive intervention for students with persistent reading difficulties. *Journal of Learning Disabilities*, 39(5), 447-466. <https://doi.org/10.1177/00222194060390050601>
- Erchul, W. P. (2011). School consultation and response to intervention: A tale of two literatures. *Journal of Educational and Psychological Consultation*, 21(3), 191-208. <https://doi.org/10.1080/10474412.2011.595198>

- Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child development perspectives*, 3(1), 30-37. <https://doi.org/10.1111/j.1750-8606.2008.00072.x>
- *Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J. V., & Davis, G. (2008). Making "secondary intervention" work in a three-tier responsiveness-to-intervention model: findings from the first-grade longitudinal reading study of the National Research Center on Learning Disabilities. *Reading & Writing*, 21(4), 413-436. <https://doi.org/10.1007/s11145-007-9083-9>
- *Gilbert, J. K., Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Barquero, L. A., & Cho, E. (2013). Efficacy of a first-grade responsiveness-to-intervention prevention model for struggling readers. *Reading Research Quarterly*, 48(2), 135-154. <https://doi.org/10.1002/rrq.45>
- *Goffreda, C. T., Diperna, J. C., & Pedersen, J. A. (2009). Preventive screening for early readers: Predictive validity of the dynamic indicators of basic early literacy skills (DIBELS). *Psychology in the Schools*, 46(6), 539-552. <https://doi.org/10.1002/pits.20396>
- Good, R. H. & Kaminiski, R. A (2002). *Dynamic Indicators of Basic Early Literacy Skills*. Austin, TX.
- Harcourt School Publishers (2002). *Scientific research base, Harcourt tropics*. Orlando, FL.
- Individuals with Disabilities Improvement Act, 20 U.S.C. § 1414 in et seq. (2004).
- Individuals with Disabilities Improvement Act Regulations Commentary, 71 Fed. Reg. 46540-46753 (2006).
- Individuals With Disabilities Education Act, 20 U.S.C. § 1400 (2004).
- Invernizzi, M., Sullivan, A., & Meier, J. (2001). *Phonological Awareness Literacy Screening for Preschool*. Charlottesville, VA: University Press.
- Johnston, P. H. (2011). Response to intervention in literacy: Problems and possibilities. *The Elementary School Journal*, 111(4), 511-534. <https://doi.org/10.1086/659030>
- *Koutsoftas, A. D., Harmon, M. T., & Gray, S. (2009). The Effect of tier 2 intervention for phonemic awareness in a response-to-intervention model in low-income preschool classrooms. *Language, Speech, & Hearing Services in Schools*, 40(2), 116-130. [https://doi.org/10.1044/0161-1461\(2008/07-0101\)](https://doi.org/10.1044/0161-1461(2008/07-0101))
- *Little, M. E., Rawlinson, D., Simmons, D. C., Kim, M., Kwok, O., Hagan-Burke, S., Simmons, L. E., Fogarty, M., Oslund, E., & Coyne, M. D. (2012). A comparison of responsive interventions on kindergartener's early reading achievement. *Learning Disabilities Research & Practice*, 27(4), 189-202. <https://doi.org/10.1111/j.1540-5826.2012.00366.x>
- *McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*, 71(4), 445-463. <https://doi.org/10.1177/001440290507100404>
- Ofeish, N. (2006). Response to intervention and the identification of specific learning disabilities: Why we need comprehensive evaluations as part of the process. *Psychology in the Schools*, 43(8), 883-888. <https://doi.org/10.1002/pits.20195>
- Psychological Corporation. (1992). *Wechsler Individual Achievement Test*. San Antonio, TX: Harcourt Brace.
- Penesetti, D. (2018). What is specific learning disorder? *American Psychiatric Association*. <https://www.psychiatry.org/patients-families/specific-learning-disorder/what-is-specific-learning-disorder>
- Restori, A. F., Gresham, F. M., & Cook, C. R. (2008). 'Old habits die hard': Past and current issues pertaining to response-to-intervention. *California School Psychologist*, 13, 67-78. <https://doi.org/10.1007/BF03340943>

- Special Education Guide. (2021). *The special education process explained*. Special Education Guide. <https://www.specialeducationguide.com/>
- Torgesen, J.K., Wagner, R.K. and Rashotte, C.A. (1999). Test of word reading efficiency (TOWRE). PRO-ED, Austin, Texas.
- *VanDerHeyden, A. M., Snyder, P. A., Broussard, C., & Ramsdell, K. . (2007). Measuring response to early literacy intervention with preschoolers at risk. *Topics in Early Childhood Special Education*, 27(4), 232-249. <https://doi.org/10.1177/0271121407311240>
- *Vaughn, S., Wanzek, J., Murray, C. S., Scammacca, N., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention: Examining higher and lower responders. *Exceptional Children*, 75(2), 165-183. <https://doi.org/10.1177/001440290907500203>
- Wiederholt, J. L. & Bryant, D. R. (2002). *Gray oral reading test* (4th ed.). Austin, TX.
- Woodcock, R. W. (1987). *Woodcock Reading Mastery Test-Revised*. Circle Pines, MN: American Guidance Service.
- Woodcock, R. W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.