JSLHR

Research Article

Predicting Response to Treatment in a Tier 2 Supplemental Vocabulary Intervention

Elizabeth Kelley,[a] Emily Leary,[b] and Howard Goldstein[c]

Purpose: To effectively implement a response to intervention approach, there is a need for timely and specific information about student learning in response to treatment to ensure that treatment decisions are appropriate. This exploratory study examined responsivity to a supplemental, Tier 2 vocabulary intervention delivered to preschool children with limited language abilities.
Method: A secondary analysis of a cluster-randomized trial of a supplemental vocabulary intervention was conducted. Responsivity (e.g., adequate learning) to the intervention was examined, and learning in the 1st few weeks of intervention was evaluated as a possible predictor of response to intervention.
Results: Using a criterion of learning of 20% of target vocabulary, nearly one third of participants were identified as poor responders. A 1st unit benchmark was identified that maximized the sensitivity to identification of children who were likely to respond to the intervention.
Conclusions: Even for generally effective interventions, there is likely to be a substantial proportion of children who are not responsive. Learning in the 1st few weeks of intervention may be a useful indicator of appropriate response to treatment and could inform instructional decisions.

Many children begin school with limited language abilities, placing them at high risk for academic failure. The goal of a response to intervention (RtI) model is to efficiently match children to appropriate levels of instruction to reduce preventable disabilities. Efficiency depends on our ability to determine if children are learning from the current instruction and to make adjustments so that children receive a tier of instruction that meets their needs. Although a substantial body of research has examined the implementation of RtI models to improve academic outcomes and prevent academic difficulties (Berkeley, Bender, Gregg Peaster, & Saunders, 2009; Burns, Appleton, & Stehouwer, 2005), the issue of identifying children who are, and are not, responding to intervention, remains a challenge.

[a]Department of Communication Science and Disorders, University of Missouri, Columbia
[b]Department of Biostatistics and Research Design, University of Missouri, Columbia
[c]College of Behavioral and Community Sciences, University of South Florida, Tampa
Correspondence to Elizabeth Kelley: kelleyej@missouri.edu

Editor: Sean Redmond
Associate Editor: Patricia Eadie
Received October 12, 2016
Revision received June 27, 2017
Accepted September 8, 2017
https://doi.org/10.1044/2017_JSLHR-L-16-0399

RtI

In an RtI model, children are matched to tiers of instruction that may vary in intensity, structure, and content to provide instruction that will ensure adequate educational progress for all children (Fuchs, Compton, Fuchs, Bryant, & Davis, 2008; Fuchs & Fuchs, 2006; Gersten et al., 2008). In a typical model, the first tier of high-quality classroom instruction is provided to all children. One presumes that Tier 1 instruction will be effective in promoting adequate progress for the majority of children (e.g., 80% of students; Bender & Shores, 2007). For the subset of children for whom Tier 1 instruction is not effective, more intense and explicit Tier 2 interventions are delivered, often in small group settings, and for a smaller proportion of children, individualized interventions are provided at Tier 3. Thus, a successful RtI approach requires effective educational interventions at three or more tiers of instruction.

RtI in Early Childhood

Much of the research on RtI has been conducted with elementary school-aged children, especially in the domain

Disclosure: *Howard Goldstein is an author of Story Friends and has a financial interest, as he receives royalties from sales of this product through Paul Brookes Publishing. This interest has been reviewed by the University in accordance with its Individual Conflict of Interest policy, for the purpose of maintaining the objectivity and the integrity of research at the University of South Florida.*
Elizabeth Kelley is an author of Story Friends and has a financial interest, as she receives royalties from sales of this produce through Paul Brookes Publishing.

of literacy (see, e.g., Compton et al., 2010; Fuchs et al., 2008; Vaughn, Linan-Thompson, & Hickman, 2003). The small number of studies that have examined the application of RtI in early childhood settings have identified challenges particular to this context. For example, because participation in early childhood education is not universal, many early childhood settings serve a population that is high risk (Ball & Trammell, 2011; Kaiser, Cai, Hancock, & Foster, 2002). This means that the proportion of children who require instruction beyond Tier 1 is likely to be higher than the 20% predicted by school-age models. In one study of a variety of early childhood settings, Carta et al. (2014) found that the proportion of children who were likely to require additional tiers of instruction was quite high, around 30%–35%. In state-funded prekindergarten classrooms, 50% of children qualified for Tier 2 or 3 services on the basis of a measure of vocabulary. Although children in all settings made gains in language and early literacy during the prekindergarten year, children who began the school year with lower levels of skill did not make sufficient progress to catch up to peers with age-appropriate skills (Greenwood et al., 2013). Given the high proportion of children who are likely to require instruction beyond Tier 1, effective, low-cost interventions at Tiers 2 and 3 are necessary. Further, measures that are sensitive to learning are required to determine if these additional tiers of instruction are effective in serving the needs of all children.

Assessment in RtI

An RtI approach requires valid, reliable, and sensitive measurement tools to address several purposes. In a typical model, assessment in RtI consists of universal screening, provided to all children a minimum of once per year and, often, three times per year (fall, winter, and spring). For children identified as at risk, progress monitoring is provided more frequently to examine learning in response to instruction (Gersten et al., 2008; McConnell, McEvoy, & Priest, 2002). Measurement tools must be carefully selected to appropriately match students to tiers of instruction, to evaluate progress within tiers of instruction, and to reassign children to tiers on the basis of their progress (Snyder, Wixson, Talapatra, & Roach, 2008; Ysseldyke, Burns, Scholin, & Parker, 2010).

In the domain of early literacy, several well-established measures of early literacy skills exist to guide decision making in RtI (e.g., Dynamic Indicators of Basic Early Literacy Skills–Sixth Edition; Good & Kaminski, 2002; Individual Growth and Development Indicators of Early Literacy; McConnell, Bradfield, Wackerle-Hollman, & Rodriguez, 2015). These measures have demonstrated utility both as screening measures to identify children who may require additional instruction (Goffreda, Diperna, & Pedersen, 2009) and as progress monitoring tools that are sensitive to learning in response to targeted interventions (e.g., Ziolkowski & Goldstein, 2008). In the domain of oral language, emerging evidence supports the use of measures of vocabulary as screening tools (Bradfield et al., 2014; Marcotte, Clemens, Parker, & Whitcomb, 2016). However, there are fewer research-based assessments for monitoring progress in response to instruction (Gersten et al., 2008). Measurement in the domain of oral language is challenging. Vocabulary learning is dependent upon exposure to or teaching of specific words. Learning proceeds incrementally and is not mastered as an all-or-none phenomenon. For sensitive assessment of learning in response to treatment, curriculum-based measures, measures that are closely aligned with the targets of instruction, are likely the best approach.

To more efficiently match children to the appropriate type of instruction, several research groups have examined a dynamic assessment approach. A dynamic assessment approach combines teaching and testing; performance on an outcome is measured in relation to brief instruction. For example, a child might be assigned a score on the basis of the amount of instruction necessary to master a skill (e.g., Fuchs, Compton, Fuchs, Bouton, & Caffrey, 2011). Alternatively, a child might be tested before and after a brief teaching session to measure learning (e.g., Kapantzoglou, Restrepo, & Thompson, 2012). In comparison to static measures, dynamic assessments can provide a stronger indication of a child's potential response to instruction (Caffrey, Fuchs, & Fuchs, 2008), making these measures a logical fit for RtI models (Grigorenko, 2009). For example, Fuchs et al. (2011) found that a dynamic assessment of decoding ability was a significant predictor of responsiveness to literacy instruction in first grade. Performance on dynamic assessment measures may be useful to quickly assign children to higher tiers of instruction.

Identification of Children Who Are Unresponsive

Even for treatments that are generally effective, a proportion of children do not adequately respond to the instruction. Al Otaiba and Fuchs (2006) identified 25% of children as poor responders to an early literacy intervention. Similarly, O'Connor, Bocian, Beebe-Frankenberger, and Linklater (2010) found that 28% of participants did not respond to an early literacy intervention. In the domain of oral language, most of the studies of language intervention have not specifically reported the numbers of students who did and did not respond to treatments (Loftus, Coyne, McCoach, & Zipoli, 2010; Neuman, Newman, & Dwyer, 2011; Pullen, Tuckwiller, Konold, Maynard, & Coyne, 2010; Zucker, Solari, Landry, & Swank, 2013). However, the large standard deviations that have been reported indicate substantial variability in student learning, suggesting that some children did not respond to intervention.

To effectively implement an RtI model, it is important to efficiently and accurately identify those children who are not responding so that adjustments in treatment can be made, rather than waiting for children to fail. Fuchs (2003) highlights three key issues related to the identification of poor responders: timing of assessment of RtI, the criterion for determining RtI, and the specific characteristics of the intervention. The timing of assessments to evaluate poor responders requires a balance of priorities.

Although it is important to identify poor responders promptly, adequate time must be allowed for participants to demonstrate RtI. The criterion for determining RtI is often determined relative to the performance of other children, either a normative population that includes all children or a subgroup of children who have participated in an intervention. Without agreed-upon criteria for adequate learning, it is difficult to identify children who are not responding to instruction. The nature of the intervention, whether it is a generally effective educational approach (e.g., Tier 1) or a more intensive treatment (e.g., Tiers 2 and 3), plays a role in decisions about poor responders.

Poor responders have been identified by a static measure of performance level (e.g., posttest score), a measure of growth (e.g., gain score), or a combination of both. Approaches that incorporate information about both level and growth in RtI, a form of dynamic assessment, may be most effective in matching children to appropriate tiers of instruction. For example, McMaster, Fuchs, Fuchs, and Compton (2005) identified poor responders on the basis of both level and rate of growth on measures of early literacy. The authors determined that including both level and growth most reliably identified poor responders, in contrast to methods that used criteria based only on performance level or growth.

In a study of a large-group narrative intervention, Spencer, Petersen, Slocum, and Allen (2014) categorized participants into groups on the basis of response to treatment using performance level and growth. A criterion of 0.5 *SD* below the group mean, on the basis of McMaster et al. (2005), was applied to scores on measures of narrative retell and narrative comprehension. "Minimal responders" were those children who were below criterion on both growth and outcome level; 20%–28% of participants were categorized as minimal responders. An additional 5.5% of participants achieved the criterion for growth, but not outcome level. Although this approach was effective in identification of poor responders, these children were not identified until the end of the 3-week intervention.

Dynamic assessment has been applied to the identification of poor responders. Spencer, Petersen, and Adams (2015) examined a dynamic assessment approach to identify those children likely to be poor responders to a narrative intervention. Narrative skills were measured before and after a brief, 3-day narrative treatment. Children who scored above the cutoff at pretest or posttest were determined to not need intervention. Children who scored below the cutoff were identified as poor responders and were placed into the Tier 2 intervention. Children, who had such limited language skills that they could not be tested, were determined to be candidates for a Tier 3 intervention. This study provides an example in which poor responders are identified quickly, and thus, children are more efficiently matched to appropriate tiers of instruction.

To efficiently identify children who are not responding to targeted interventions in the domain of oral language, a dynamic assessment approach might be particularly useful. By making use of existing curriculum-based measures to provide the most sensitive measure of growth in RtI, poor responders can be identified as those who make limited progress in the first days or weeks of an intervention. This practical, inexpensive approach could immediately inform decisions about instruction. In the current study, we examine responsiveness to a targeted vocabulary intervention, the Story Friends program, and explore the use of learning in the first few weeks of instruction as a method of identification of poor responders.

Story Friends Program

Story Friends is a curriculum designed to teach vocabulary words and question-answering skills to preschool children with limited oral language. In the context of an RtI model, Story Friends can serve as a Tier 2 curriculum to supplement existing Tier 1 instruction. The program was developed during a multiyear iterative development process (Kelley & Goldstein, 2015). A series of studies conducted during the development process indicated strong effects on vocabulary learning, with more moderate effects on comprehension (Greenwood et al., 2016; Kelley, Goldstein, Spencer, & Sherman, 2015). Subsequently, a cluster randomized design trial was conducted in 32 preschool classrooms with 195 participants (Goldstein et al., 2016). Children in the treatment group participated in the Story Friends program and listened three times per week to prerecorded storybooks with embedded lessons for challenging vocabulary words and question answering. Children in the comparison classrooms listened to the Story Friends books without embedded lessons. Learning of vocabulary words and question answering was assessed at approximately monthly intervals.

Goldstein et al. (2016) found large, significant group differences in favor of the treatment group for learning target vocabulary words but no significant group differences for the question-answering outcome. Participants in the Story Friends treatment gained an average of 3.4 word points per unit, with large effect sizes of 0.70 (Cohen's f^2) relative to the comparison group. However, substantial variability was observed in the learning of the treatment group (e.g., three children who gained just one word point vs. one child who gained 54 word points), indicating that Story Friends was not equally effective for all children. To incorporate Story Friends into an RtI model, it would be valuable to efficiently and accurately predict which children would be successful in the program.

The automated delivery of instruction meant that overall fidelity of implementation of Story Friends was quite high (95%). The large majority (81%) of participants received the intended dose of instruction. Thus, delivery of treatment did not explain differences in learning. Pretreatment scores on standardized norm-referenced measures of vocabulary and language did not moderate vocabulary learning, meaning that performance on these measures would not be useful in predicting response to the treatment. Further, pretest scores on target vocabulary words were subject to floor effects (i.e., children rarely knew any of the

words prior to treatment). Because pretreatment measures could not accurately predict response to treatment, a next step was to examine whether learning in the first month of the Story Friends curriculum could be a useful predictor.

The current study investigated responsivity to the Story Friends intervention provided to preschool children with limited language ability. This exploratory study seeks to inform the challenge of matching children to tiers of instruction when implementing RtI in the domain of oral language in early childhood settings. This article is a secondary analysis designed to address the following research questions: (a) To what extent does performance of preschool children in response to a supplemental vocabulary and question-answering intervention vary (responsivity analysis)? (b) Can performance in the first few weeks of intervention be used to identify treatment responders and poor responders (sensitivity and specificity analysis)?

Method

The current investigation is a secondary examination of data from the efficacy study of Story Friends conducted by Goldstein et al. (2016). Details of the study, including specifics of the Story Friends program, are reported there. In this section, we provide information about the measures and variables of interest for the current secondary analyses.

Selection and Assignment of Participants

Thirty-two public prekindergarten classrooms in two cities were recruited to participate. These classrooms served primarily students from families with low income and were randomly assigned to either the Story Friends intervention condition or a comparison condition (i.e., Story Friends books without instructional components). An average of five children per classroom participated in 32 classrooms for a total of 193 participants at the onset of the study. In the current study, the 75 participants in the treatment condition who completed outcome measures for all five instructional units are included.

To select participants for the Story Friends program, a multiple gating procedure was followed. The goal was to identify six participants in each classroom. First, two individual growth and development indicators (IGDIs) were administered: Picture Naming and Which One Doesn't Belong (Bradfield et al., 2014; McConnell et al., 2015). The Picture Naming IGDI included 15 photographs, and children were asked to provide a verbal label for each picture. The Which One Doesn't Belong IGDI included 15 cards with three photographs each; children were asked to point to the photograph that did not belong with the others. Children who performed below benchmark expectations on one or both measures moved to the second gate of the selection procedures. These children were given a standardized norm-referenced measure of receptive vocabulary, the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-IV; Dunn & Dunn, 2007). On the PPVT-IV, children are shown a plate of four illustrations and asked to point to the picture that corresponds with a verbal label provided by the examiner. In each classroom, the first six children with standard scores on the PPVT-IV that indicated moderately limited oral language skills (e.g., standard scores between 1.0 *SD* and 1.5 *SD* below the mean) were included as participants. In some classrooms, fewer than six children had scores in this range. In these few cases, we extended the range up and down to include sufficient participants (final standard score range = 71 to 96). Although this extended range included a small number of children with age-appropriate vocabulary scores, the large majority of participants had moderately limited oral language skills (M = 83.90; SD = 5.32). All participants were administered the core language subtests of the Clinical Evaluation of Language Fundamentals Preschool–Second Edition (CELF Preschool-2; Wiig, Secord, & Semel, 2004) as a descriptive measure of oral language abilities.

Primary Outcome Measure

The primary outcome was vocabulary learning of targeted words. Vocabulary learning was assessed using a curriculum-based measure created as part of Story Friends, the Unit Vocabulary Tests (UVTs). Each unit consisted of three instructional books and one review book and lasted approximately one month (one week per book). Each UVT included six items that related to the taught vocabulary words in a unit. UVTs were administered before and after each unit (five units in Kansas, six units in Ohio). Pretests and posttests were spaced approximately one month apart; longer times between pretest and posttest occurred occasionally due to school schedules.

Participants were asked to respond to open-ended definitional questions (e.g., "Tell me, what does enormous mean?"). A standard prompt was used if children did not respond to the initial question (e.g., "Enormous means…" with rising intonation). Participant responses were scored on a 3-point scale: 2 points for a complete definition or appropriate synonym, 1 point for a related but incomplete response, and 0 for an unrelated or no response. The maximum score for targeted vocabulary on each UVT was 12 points. Agreement among scorers of the UVT was high (97%–98%).

Results

Responsivity in the Treatment Group

To analyze responsiveness in the treatment group, we examined growth on the UVT across the five units of instruction. To determine an appropriate criterion for responsivity, we were guided by previous studies and the practical considerations outlined by Fuchs (2003) related to intervention and participant characteristics.

In previous studies, poor responders have been identified based on both performance level and growth in response to treatment. A dual discrepancy approach may be preferable in many RtI applications (Fuchs & Fuchs, 2007). However, in the current study, the vast majority of participants had pretest scores of 0 (i.e., they did not

know any of the target vocabulary words prior to intervention), meaning that posttest performance level and growth in response to treatment were approximately equal. Thus, we considered only growth in response to treatment to identify poor responders.

On the basis of previous work by McMaster et al. (2005), Al Otaiba and Fuchs (2006), and Spencer et al. (2014), we selected a cutoff of 0.5 *SD* below the group mean. The maximum possible cumulative gain for all five units was 60 points (12 points per unit × 5 units). For the group of treatment participants, mean cumulative gain was 17.6 (*SD* = 9.70, range = −2 to 48). Using the 0.5 *SD* criterion, participants with cumulative gains below 13 would be considered poor responders. Of the 75 participants who had complete data for all five units, 22 participants (29.3%) were categorized as poor responders. This proportion was consistent with previous studies in which 25%–30% of participants were identified as poor responders (Al Otaiba & Fuchs, 2006; McMaster et al., 2005).

The 13-point cutoff corresponded to learning of approximately 20% of target vocabulary. When we considered separately the performance of participants above and below the 0.5 *SD* cutoff, the responsive group had a cumulative gain of 22.04 (*SD* = 7.81, range = 13 to 48), representing 37% of words taught, whereas the poor responders had a cumulative gain of 6.95 (*SD* = 3.51, range = −2 to 12), 12% of words taught. From a practical standpoint, it seems reasonable to determine that learning of just 12% of words taught is an indication of poor response to an explicit vocabulary intervention.

First Unit Performance as a Predictor of RtI

This secondary analysis was motivated in part by the findings of Goldstein et al. (2016) that pretreatment scores on standardized, norm-referenced measures of vocabulary and language did not moderate learning in the treatment group. To further explore the relation between pretreatment variables and learning in RtI, we examined correlations between pretest scores on the Unit 1 UVT, standard scores on norm-referenced measures, and cumulative gain at the end of Unit 5. There was no correlation between pretest scores and cumulative gain, but standard scores on both the PPVT-IV and the CELF Preschool-2 were significantly correlated with cumulative gain (PPVT-IV: $r = .38$, $p < .01$; CELF Preschool-2: $r = .32$, $p < .01$). These moderate correlations indicated that there was some relation between pretreatment vocabulary and language abilities and learning in RtI.

However, it would not have been possible to identify poor responders on the basis of pretreatment scores. Although the group of poor responders had significantly lower scores than that of the group of responders on both the CELF Preschool-2 and PPVT-IV at pretest, the group differences were small, and there was substantial overlap between the groups. On the PPVT-IV, the group of poor responders had a mean score of 80.14 with a range of 71 to 91, and the group of responders had a mean score of 85.60 with a range of 74 to 92. Group differences on the CELF Preschool-2 were larger (e.g., *M*s of 74.23 and 86.60), but scores for the two groups overlapped similarly (50–94 for the poor responders, 63–102 for the responders).

The correlation between gain in the first unit and cumulative gain was significant and large ($r = .77$, $p < .01$), suggesting that learning in the first unit might be an effective way to identify good and poor responders. Thus, we examined word learning in the first unit as a potential predictor of overall response to the Story Friends intervention. The first unit of Story Friends was approximately four weeks in duration, and the maximum score was 12 (2 points each for each of the six words). We chose the first unit as our outcome for this exploratory analysis because we knew that many children demonstrated learning in the first unit. As well, this choice meant that poor responders could be identified after one month of intervention, which seemed to be an adequate length of time to make an instructional decision. We hypothesized that children who demonstrated some learning in the first unit would go on to be successful learners throughout the program and that children who demonstrated very little learning might continue to struggle. An alternative was that children who demonstrated very little learning in the first unit might show growth in learning across the duration of the program; this was a pattern that we sometimes observed in previous, small-scale studies (e.g., Spencer et al., 2012).

Our goal was to accurately identify those children who were likely to benefit from Story Friends intervention. We chose to err on the side of leaving children who may not benefit in the intervention, rather than removing children who may benefit. The rationale for this decision was that participation in Story Friends intervention is likely to benefit children, even those who demonstrate relatively little learning. Story Friends is specifically designed to be implemented with high fidelity in classroom settings without placing extensive demands on educational staff. Thus, allowing children to remain in the program was unlikely to have negative consequences, whereas moving children into a Tier 3 intervention prematurely could tax the limited resources of the classroom.

For this exploratory analysis, we examined the potential of a range of first unit gain scores to predict RtI. Gain scores were the differences in the number of word points between the Unit 1 pre- and posttests, a measure of what the child knew prior to the unit compared with after the unit of instruction. We examined sensitivity, specificity, and positive and negative predictive values of gain scores in the first unit. Table 1 includes the definitions for each measure. Because we prioritized identifying those children who were likely to respond to the intervention, we evaluated these measures on the basis of the accuracy of identifying responders.

We chose four different Unit 1 gain scores (2, 3, 4, and 6) to evaluate. These gain scores were chosen to encompass the average performance of the group; the mean gain at Unit 1 was 5.4 (*SD* = 3.5), and the mean gain

Table 1. Definitions for sensitivity, specificity, positive and negative predictive values.

Measure	Mathematical Definition	Definition
Sensitivity	True Positives / True Positives + False Negatives	Of those that are true responders, the percentage that had Unit 1 gain score above criterion
PPV	True Positives / True Positives + False Positives	Of those with Unit 1 gain score above criterion, the percentage that are true responders
Specificity	True Negatives / True Negative + False Positives	Of those that are true poor responders, the percentage that had Unit 1 gain score below criterion
NPV	True Negatives / True Negatives + False Negatives	Of those with Unit 1 gain score below criterion, the percentage that are true poor responders

Note. In this project, True positives are responders (i.e., cumulative gain of 13 points or more) who had Unit 1 gain scores above the criterion. True negatives are poor responders (i.e., cumulative gain of less than 13 points) which had Unit 1 gain scores below the criterion.

across all five units was 3.4 word points per unit. Table 2 includes the sensitivity, specificity, positive predictive values (PPVs), and negative predictive values (NPVs) for each gain score.

The 6-point gain score had 100% specificity; all participants who were poor responders had gain scores less than 6 at the end of the first unit. The PPV for the 6-point gain score was also 100%, meaning that all the children who were true responders scored above 6 points at the end of the first unit. However, the 6-point gain score had relatively poor sensitivity; just 71.70% of responders had gain scores above 6. The NPV was 59.46, indicating that only about 60% of participants identified as poor responders by the 6-point gain score were true poor responders. Thus, if a gain score of 6 was used, about a third of responders would have been misidentified as poor responders. Because our priority was to accurately identify responders, we determined that the 6-point gain score was not useful for our purposes.

The 2-, 3-, and 4-point benchmarks appeared to maximize sensitivity with satisfactory specificity. As shown in Table 2, 98.11% of responders had two or more points at the end of Unit 1. The percentages were slightly smaller for the 3- and 4-point benchmarks, 94.34% and 92.45%, respectively. PPV ranged from 78.79% to 85.96%, meaning that the large majority of participants who were identified as responders by these gain scores were truly responders. Table 3 includes the classification of responders and poor responders for each gain score. Of the 53 responders, one child would have been misidentified as a poor responder using the 2-point gain score, three children with the 3-point gain score, and four with the 4-point gain score. On the basis of sensitivity, it appeared that any of these gain scores would be appropriate, with a slight advantage for the 2-point gain score.

Specificity of the 2-, 3-, and 4-point gain scores were lower but acceptable. Of the poor responders, 36.36% had gained less than 2 points at the end of the first unit, 59.09% had gained less than 3 points, and 63.64% had gained less than 4 points. For the 2-point gain score, NPV was 88.89%, meaning that about 90% of participants who were true poor responders had fewer than 2 points at the end of the first unit. NPV for the 3- and 4-point gain scores were slightly lower but still accurately identified most poor responders. As shown in Table 3, of the 22 poor responders, 14 would have been misidentified as responders using the 2-point gain score, nine would have been misidentified as responders with the 3-point gain score, and eight with the 4-point gain score.

For this exploratory analysis, we prioritized identifying responders, meaning that the 2-point gain score was the best choice. The 2-point gain score would correctly

Table 2. Sensitivity, specificity, and positive and negative predictive values for Unit 1 gain scores evaluated relative to Unit 5 cumulative gain of 13 points.

Unit 1 gain score	Sensitivity	Specificity	PPV	NPV
2	98.11	36.36	78.79	88.89
3	94.34	59.09	84.75	81.25
4	92.45	63.64	85.96	77.78
6	71.70	100	100	59.46

Note. PPV = positive predictive value; NPV = negative predictive value.

Table 3. Participants classified as responders and poor responders for each gain score.

Unit 1 Gain Score		Response: Responder	Response: Poor Responder
2	Responder	52	14
	Poor Responder	1	8
3	Responder	50	9
	Poor Responder	3	13
4	Responder	49	8
	Poor Responder	4	14
6	Responder	38	0
	Poor Responder	15	22

Note. Of the 75 participants who had complete data for all 5 units, 53 had cumulative scores of at least 13 points and were classified as responders. The remaining 22 participants with cumulative scores of less than 13 were classified as poor responders.

identify 52 of 53 responders. However, the 2-point gain score would identify just 8 of 22 poor responders. If the priority was to accurately categorize the largest number of participants, the 3- or 4-point gain scores would be better choices, correctly identifying 63 of the 75 total participants.

Discussion

The purpose of this study was to examine responsivity of preschool children with limited oral language, who participated in a supplemental vocabulary intervention program, and to provide data that could inform decisions for assigning children to tiers of instruction more efficiently. Using a criterion of performance of learning of 20% of vocabulary targets, nearly a third (29.3%) of participants were identified as poor responders. On the basis of estimates from models of RtI (e.g., Bender & Shores, 2007), we would expect approximately 20% of students to be identified for instruction beyond Tier 1. For the majority of these students, Tier 2 intervention would be sufficient. Just 5% of children in the classroom would be expected to require instruction beyond Tier 2. Our proportions are somewhat higher. In classrooms of approximately 20, five participants were selected for Tier 2 intervention (25%). Of those participants, approximately a third did not respond to Tier 2, about 7% of children in the classroom. Because we did not include all children who were eligible for Tier 2 intervention and not all children completed intervention, these proportions are rough estimates. Moreover, the classrooms that participated in this study served children from low-income families, meaning that many children were likely to require instruction beyond Tier 1 to succeed. In similar classrooms, Carta et al. (2014) reported that 50% of children were candidates for Tier 2 or 3 intervention on the basis of a measure of vocabulary but did not distinguish between those tiers. The current analysis suggests that, in classrooms that serve a high-risk population, many children would benefit from tiered intervention in oral language.

In other studies that have examined RtI, the proportion of children who are identified as unresponsive ranges widely (e.g., 8%–80%; Al Otaiba & Fuchs, 2002). This variability reflects the differences in procedures for identification of poor responders. Few studies have examined responsivity for oral language treatments. In a study of a large group (Tier 1) of narrative intervention, Spencer et al. (2014) identified 33.4% of participants as likely to require additional tiers of instruction. A dynamic assessment approach identified a much larger proportion of children (66%) as candidates for Tier 2 and Tier 3 interventions (Spencer et al., 2015). This difference in proportion between these two studies is likely related to the timing of the responsivity analysis (e.g., after a 3-week intervention vs. a 3-day intervention). As others have argued (e.g., Fuchs, 2003; Vanderheyden, 2011), decisions about determining responsivity to treatment must be made carefully.

Identification of Poor Responders Using a Benchmark

Although scores on standardized, norm-referenced measures were helpful in identifying children with limited language who were good candidates for intervention, we could not have used these scores to predict who would not respond to intervention. Pretreatment scores on standardized measures were moderately correlated with cumulative gain, but there was substantial overlap in the groups of responders and poor responders. The lack of learning of the poor responders also could not be explained by fidelity of treatment. Poor responders were from different classrooms and different sites, and all but two received nine of the 10 possible doses of instruction in the first unit.

Gain in the first unit of instruction was strongly correlated with cumulative gain. Thus, we examined learning in the first unit to identify a criterion for predicting RtI. We conducted sensitivity and specificity analyses using several Unit 1 gain scores, referenced against a "success" indicator of learning after five units of instruction. In our choice of gain score, we prioritized the identification of responders. That is, given the choice, rather than misidentifying poor responders as responders, we chose a gain score that would keep potential responders in the intervention. This decision was based on the context of the intervention (i.e., a relatively low cost, low intensity intervention) and the potential consequences of the decision (e.g., placing children in a more costly, individualized Tier 3 intervention).

The sensitivity and specificity analyses provided information to guide the choice of a gain score of learning in response to the first unit. When we selected a 2-point gain score to identify children as poor responders, nine participants would have been identified as poor responders, and eight of the nine went on to gain fewer than 13 points across the five units of instruction; their average cumulative gain was only 4.75 points. Just one participant was misidentified and had a cumulative gain of 15 points. Interestingly, this child demonstrated no learning in the first two units of intervention but went on to gain 15 word points in the last three units. Perhaps, a Tier 3 intervention would have boosted learning earlier for this participant. If the 3-point gain score had been applied, we would have correctly identified an additional five poor responders, while misidentifying two responders as poor responders. The accuracy of first unit gain scores at predicting response to treatment was moderate. The 2-point gain score correctly identified an overall 80% of participants, and the 3- and 4-point gain scores correctly identified an overall 84%.

The 2-point gain score had practical significance as well. On the UVT, a score of 2 points after the first unit indicated that the children had learned to provide the definition for at least one of the six targeted words. (A score of 2 also could indicate that the child had learned a partial definition for two of the words, although this was less common.) It seems logical then to decide that a child who had learned no words (a score of 0 or 1) after the first unit

of intervention was unlikely to demonstrate progress in the Story Friends intervention.

Pretreatment abilities can be useful in predicting who will require additional instruction and are often generally correlated with RtI. Al Otaiba and Fuchs (2002), in a review of early literacy interventions, found that pretreatment abilities, particularly phonological awareness, were predictive of response to treatment. In studies of vocabulary intervention, pretreatment vocabulary scores are often predictive of vocabulary learning in response to instruction (Coyne, McCoach, & Kapp, 2007; Loftus et al., 2010; Pullen et al., 2010). Although pretreatment characteristics may be related to RtI, these characteristics are generally not useful for identification of poor responders. In a study of a reading intervention for children with language impairment, O'Connor et al. (2010) found that, although pretest scores could accurately identify those children who would have reading difficulties, those scores did not identify which children would not respond to treatment. Similar to the findings of the current study, O'Connor et al. (2010) found that children who were identified as poor responders had significantly lower vocabulary scores on average. However, several children with very low vocabulary scores were identified as "fast-responders" who quickly met benchmarks for RtI. Pretreatment scores can help to identify a group of children who are at risk and may require additional instruction. However, pretreatment scores are unlikely to discriminate within that group of at-risk children who will be responsive and who will be unresponsive.

Limitations and Future Directions

A few limitations of the current study are important to note. First, this study was an exploratory secondary analysis. The data analyzed in the current study were not gathered for a responsivity analysis; instead, it was part of a cluster-randomized trial of the Story Friends intervention. Participants were selected based on the priority of examining the efficacy of the intervention, not for examining an RtI model. Further, the identification of participants in the current study did not follow a procedure that would be practical in typical early childhood classrooms, such as administering the PPVT-IV to many children. Without a doubt, a more efficient approach is needed to ease implementation. One option would be to rely on briefer measures like the IGDIs, perhaps in combination with teacher questionnaires. McConnell, Wackerle-Hollman, Roloff, and Rodriguez (2014) describe preliminary research in which use of teacher ratings improve the accuracy of assessment decisions using the IGDIs. In high-risk classrooms, this method would likely identify many children as eligible for Tier 2. However, the procedures of Story Friends allow it to be easily implemented with high fidelity, even for an entire classroom. Next, responsivity in response to a brief trial of intervention could determine which children are better suited for Tier 3 intervention.

A dual discrepancy approach, in which both measures of level and growth are considered, has been recommended for examining responsiveness in an RtI model (Fuchs & Fuchs, 2007). In our study, we determined that the UVT pretest was not useful for describing pretreatment skills due to floor effects (i.e., children did not know words at pretest). On the other hand, individually administered language assessments, such as the PPVT-IV and CELF Preschool-2, are not feasible as screening or progress-monitoring tools. One might expect that pretreatment scores on these measures, considered in combination with gain on the UVT, would improve accuracy in identifying poor responders. For example, the one child who was misidentified as a poor responder by the 2-point gain score had relatively high PPVT-IV and CELF Preschool-2 scores (90 and 92, respectively). The additional two children who were misidentified as poor responders by the 3-point gain score did not have higher scores on the PPVT-IV (standard scores of 85 and 77) but did have relatively high CELF Preschool-2 scores (89 and 92). Of course, it is not possible to draw conclusions from such a small number of children. Instead, we suggest that an examination of a dual discrepancy approach for identifying poor responders warrants consideration in future research.

We applied a criterion of learning of 20% of vocabulary targets (0.5 *SD* below the mean) to identify poor responders. Although other research groups have used a similar cutoff, this is somewhat arbitrary. We do not yet have data on what constitutes adequate progress in response to vocabulary interventions. Without a general outcome measure appropriate for this purpose, we make use of proximal measures likely to be sensitive to growth in RtI. Children rarely learn all the words taught, particularly when participants have limited vocabulary at pretreatment, when targets are challenging vocabulary, and when measures are rigorous (e.g., a decontextualized definitional task). Participants in the Story Friends program learned approximately a third of the words taught (28.33%). Other similar studies have reported a range of learning (e.g., 12.5% of words taught in Loftus et al., 2010; 45.9% in Pullen et al., 2010). The criterion of 0.5 *SD* below the mean corresponded to learning of approximately 20% of words taught, lending some face validity to the choice. Additional research that examines criteria for determining responsivity to oral language interventions will facilitate more accurate decisions.

The sensitivity and specificity analyses were based on learning after the first unit of Story Friends, approximately one month of intervention. To efficiently match children to appropriate tiers of instruction, it is important to make decisions quickly; we determined that one month provided a sufficient opportunity for most children to demonstrate a response.

Implications for Practice

Because there is currently no consensus on what indicates responsiveness to intervention, researchers and practitioners are left to make decisions about identification of poor responders with limited guidelines. As Fuchs (2003)

suggests, these decisions will require consideration of the nature of the intervention and the timing of identification. The current study, in combination with previous research, provides a few guidelines. First, careful selection of measures to determine RtI is necessary. When possible, decision-makers will want to take advantage of research-based measures with well-established benchmarks. This may be possible in the domain of early literacy. However, for decisions about progress in the domain of oral language, curriculum-based measures may be most appropriate, as these measures are likely to be sensitive to incremental progress in RtI. Second, the timing of the responsivity measurement will want to be balanced between waiting too long (e.g., the end of a school year) and assessing before RtI is likely to occur. This timing will likely be informed by the behavior of interest (e.g., learning of discrete phonological awareness skills vs. growth in reading comprehension) but should optimize efficient placement of children into appropriate tiers of instruction. Third, decisions about responsivity will require consideration of the consequences of that decision. In our exploratory analysis, we prioritized the identification of responders. That is, we wanted to err on the side of including all children who may make progress, rather than moving children to Tier 3 who may not require it. One could make the case for prioritizing specificity to ensure that all children not benefiting from Tier 2 are moved quickly to Tier 3, with the intention of moving children demonstrating a boost in learning back into Tier 2 intervention. In an ideal scenario, these decisions would be based on information not only from sensitive measures but also from comprehensive information about an individual child (e.g., risk factors and learning potential).

This discussion highlights the complexities of making data-based decisions about placing children within tiered interventions. Our hope is that analyzing data from larger groups of children who have experienced well-specified, evidence-based tiered instruction will help us establish guidelines that can maximize the utility of screening and progress-monitoring assessment data. Such guidelines have the potential to add to the ease of implementation and the effectiveness of RtI models.

Acknowledgments

This research was partially supported by grants from the United States Department of Education, Institute of Education Sciences, R432A150132 awarded to the University of South Florida and R324C080011 awarded to the University of Kansas.

References

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education, 23*(5), 300–316.

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities, 39*(5), 414–431. https://doi.org/10.1177/00222194060390050401

Ball, C. R., & Trammell, B. A. (2011). Response-to-intervention in high-risk preschools: Critical issues for implementation. *Psychology in the Schools, 48*(5), 502–512.

Bender, W. N., & Shores, C. (2007). *Response to intervention: A practical guide for every teacher*. Thousand Oaks, CA: Corwin Press.

Berkeley, S., Bender, W. N., Gregg Peaster, L., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities, 42*(1), 85–95. https://doi.org/10.1177/0022219408326214

Bradfield, T. A., Besner, A. C., Wackerle-Hollman, A. K., Albano, A. D., Rodriguez, M. C., & McConnell, S. R. (2014). Redefining individual growth and development indicators oral language. *Assessment for Effective Intervention, 39*(4), 233–244.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23*(4), 381–394.

Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment a review. *The Journal of Special Education, 41*(4), 254–270.

Carta, J. J., Greenwood, C. R., Atwater, J., McConnell, S. R., Goldstein, H., & Kaminski, R. A. (2014). Identifying preschool children for higher tiers of language and early literacy instruction within a response to intervention framework. *Journal of Early Intervention, 36*(4), 281–291.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., . . . Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102*(2), 327–340.

Coyne, M., McCoach, D., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly, 30,* 74–88.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test–Fourth Edition*. Minneapolis, MN: NCS Pearson Assessments.

Fuchs, D., Compton, D., Fuchs, L., Bryant, J., & Davis, G. N. (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 and Writing, 21*(4), 413–436.

Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., & Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read: Implications for RTI frameworks. *Journal of Learning Disabilities, 44*(4), 339–347. https://doi.org/10.1177/0022219411407864

Fuchs, D., & Fuchs, L. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*(1), 93–99. https://doi.org/10.1598/rrq.41.1.4

Fuchs, L. (2003). Assessing intervention responsiveness: Conceptual and technical issues. *Learning Disabilities Research & Practice, 18*(3), 172–186.

Fuchs, L., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children, 39*(5), 14–20.

Gersten, R., Compton, D., Connor, C., Dimino, J., Santoro, L. E., Linan-Thompson, S., & Tilly, W. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in primary grades. A practice guide*. Washington, DC: National Center for Education Evaluation

and Regional Assistance, Institute of Education Science, U.S. Department of Education.

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.

Goldstein, H., Kelley, E., Greenwood, C., McCune, L., Carta, J., Atwater, J., . . . Spencer, T. (2016). Embedded instruction improves vocabulary learning during automated storybook reading among high-risk preschoolers. *Journal of Speech, Language, and Hearing Research, 59,* 484–500. https://doi.org/10.1044/2015_JSLHR-L-15-0227

Good, R., & Kaminski, R. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills–Sixth Edition.* Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from http://dibels.uoregon.edu/

Greenwood, C. R., Carta, J. J., Atwater, J., Goldstein, H., Kaminski, R., & McConnell, S. (2013). Is a response to intervention (RTI) approach to preschool language and early literacy instruction needed? *Topics in Early Childhood Special Education, 33*(1), 48–64.

Greenwood, C. R., Carta, J. J., Guerrero, G., Atwater, J., Kelley, E. S., Kong, N. Y., & Goldstein, H. (2016). Systematic replication of the effects of a supplementary, technology-assisted, storybook intervention for preschool children with weak vocabulary and comprehension skills. *The Elementary School Journal, 116*(4), 574–599.

Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities, 42*(2), 111–132. https://doi.org/10.1177/0022219408326207

Kaiser, A. P., Cai, X., Hancock, T. B., & Foster, E. M. (2002). Teacher-reported behavior problems and language delays in boys and girls enrolled in head start. *Behavioral Disorders, 28*(1), 23–39.

Kapantzoglou, M., Restrepo, M. A., & Thompson, M. S. (2012). Dynamic assessment of word learning skills: Identifying language impairment in bilingual children. *Language, Speech, and Hearing Services in Schools, 43*(1), 81–96.

Kelley, E. S., & Goldstein, H. (2015). Building a tier 2 intervention: A glimpse behind the data. *Journal of Early Intervention, 36,* 292–312. https://doi.org/10.1177/1053815115581657

Kelley, E. S., Goldstein, H., Spencer, T. D., & Sherman, A. (2015). Effects of an automated tier 2 storybook intervention on vocabulary and comprehension learning in preschool children with limited oral language skills. *Early Childhood Research Quarterly, 31,* 47–61. https://doi.org/10.1016/j.ecresqu.2014.12.004

Loftus, S., Coyne, M., McCoach, B., & Zipoli, R. (2010). Effects of a supplemental vocabulary intervention on the word knowledge of kindergarten students at risk for language and literacy difficulties. *Learning Disabilities Research & Practice, 25*(3), 124–136.

Marcotte, A. M., Clemens, N. H., Parker, C., & Whitcomb, S. A. (2016). Examining the classification accuracy of a vocabulary screening measure with preschool children. *Assessment for Effective Intervention, 41*(4), 230–242.

McConnell, S., Bradfield, T., Wackerle-Hollman, A., & Rodriguez, M. (2015). *Individual Growth and Development Indicators of Early Literacy* (*Second Edition–Universal Screening Edition*). Saint Paul, MN: Early Learning Labs.

McConnell, S., McEvoy, M., & Priest, J. (2002). "Growing" measures for monitoring progress in early childhood education: A research and development process for individual growth and development indicators. *Assessment for Effective Intervention, 27*(4), 3.

McConnell, S., Wackerle-Hollman, A. K., Roloff, T. A., & Rodriguez, M. (2014). Designing a measurement framework for response to intervention in early childhood programs. *Journal of Early Intervention, 36*(4), 263–280.

McMaster, K., Fuchs, D., Fuchs, L., & Compton, D. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71,* 445–463.

Neuman, S. B., Newman, E. H., & Dwyer, J. (2011). Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly, 46*(3), 103–129.

O'Connor, R. E., Bocian, K., Beebe-Frankenberger, M., & Linklater, D. L. (2010). Responsiveness of students with language difficulties to early intervention in reading. *The Journal of Special Education, 43*(4), 220–235.

Pullen, P. C., Tuckwiller, E. D., Konold, T. R., Maynard, K. L., & Coyne, M. D. (2010). A tiered intervention model for early vocabulary instruction: The effects of tiered instruction for young students at risk for reading disability. *Learning Disabilities Research & Practice, 25*(3), 110–123.

Snyder, P. A., Wixson, C. S., Talapatra, D., & Roach, A. T. (2008). Assessment in early childhood: Instruction-focused strategies to support response-to-intervention frameworks. *Assessment for Effective Intervention, 34*(1), 25–34.

Spencer, E., Goldstein, H., Sherman, A., Noe, S., Tabbah, R., Ziolkowski, R., & Schneider, N. (2012). Effects of an automated vocabulary and comprehension intervention: An early efficacy study. *Journal of Early Intervention, 34,* 195–221. https://doi.org/10.1177/1053815112471990

Spencer, T. D., Petersen, D. B., & Adams, J. L. (2015). Tier 2 language intervention for diverse preschoolers: An early-stage randomized control group study following an analysis of response to intervention. *American Journal of Speech-Language Pathology, 24*(4), 619–636.

Spencer, T. D., Petersen, D. B., Slocum, T. A., & Allen, M. M. (2014). Large group narrative intervention in head start preschools: Implications for response to intervention. *Journal of Early Childhood Research, 13*(2), 196–217. https://doi.org/10.1177/1476718X13515419

Vanderheyden, A. M. (2011). Technical adequacy of response to intervention decisions. *Exceptional Children, 77*(3), 335–350.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69,* 391–409.

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical Evaluation of Language Fundamentals Preschool–Second Edition*. San Antonio, TX: Harcourt Assessment.

Ysseldyke, J., Burns, M. K., Scholin, S. E., & Parker, D. C. (2010). Instructionally valid assessment within response to intervention. *Teaching Exceptional Children, 42*(4), 54–61.

Ziolkowski, R., & Goldstein, H. (2008). Effects of an embedded phonological awareness intervention during repeated book reading on preschool children with language delays. *Journal of Early Intervention, 31,* 67–90.

Zucker, T. A., Solari, E. J., Landry, S. H., & Swank, P. R. (2013). Effects of a brief tiered language intervention for prekindergartners at risk. *Early Education & Development, 24*(3), 366–392.