THE ROLE OF CONTEXT IN THE ASSESSMENT OF “UNRESPONSIVENESS” WITHIN RESPONSIVENESS-TO-INTERVENTION: THE “RELATIVE SLOPE-DIFFERENCE DISCREPANCY MODEL” (RSDDM)

Georgios D. Sideridis, Susana Padeliadu and Faye Antoniou

ABSTRACT

The purpose of the present study was to evaluate the role of context in the identification of learning disabilities (LD) within the responsiveness-to-intervention (RTI) model. In Study 1, using a sample of students with and without LD (N = 167) and data from a reading assessment, we tested whether the decision making regarding literacy disabilities is significantly different if we take into account variability within the schools and school characteristics. Initially a logistic multilevel model was fit to the data to assess prevalence rates of LD identification. The validity of these
estimates was substantiated by bootstrapping the sample’s parameters using 1,000 replications and by evidencing negligible bias parameters. Subsequently, the relationship between reading ability and LD identification was established by means of a multilevel model including random effects. The significant slopes linking reading to LD identification (i.e., fluency and overall reading ability ratings by teachers) were predicted by cross-level interactions involving schools’ location (rural, urban, and suburban). The results of Study 1 demonstrated the moderating role of school context, as the slopes linking fluency and reading achievement to LD placement were moderated by the area in which a school was located. Study 2 was designed to present a relative discrepancy identification model by taking into account information from the school (i.e., district). Using 29 students from one district, whose writing ability was evaluated three times within the semester, comparisons were made between a specific low-ability student and the rest of his/her class. Through fitting a multilevel model in which within-student and between-student variance was assessed, Study 2 demonstrated that the specific pattern of responsiveness of a target student can be tested against the norm of his/her school district in order to have a more sensitive relative criterion of what constitutes both responsiveness and the norm. Thus, by utilizing a multilevel framework that involves school characteristics into our assessment we demonstrated that decision making is much more informative and likely more “accurate” under the RTI model. Certainly more research is needed to verify the usefulness and applicability of the proposed “relative slope-difference discrepancy model.”

The most important attribute one can possess at school is achievement. Achievement has the most important implications regarding one’s academic and social standings, his/her psychological functioning (acceptance, rejection), etc., particularly so for students with learning disabilities (LD; Elias, 2004; Elksnin & Elksnin, 2004; Greenway & Mihe, 1999; Grolnick & Ryan, 1990). Thus, decision making with regard to the presence or absence of LD has implications regarding one’s functioning and well-being (Filippatou, Dimitropoulou, & Sideridis, 2009; Heath & Ross, 2000). The thesis of the present paper is to contribute information that may be useful in the valid identification of students with learning and other disabilities, particularly within the responsiveness-to-intervention (RTI) model (Ehren & Nelson, 2005; Fuchs, Mock, Morgan, & Young, 2003). Specifically the present paper
will deal with the "unresponsiveness" aspect of the RTI model by presenting issues that affect responsiveness and a proposal to accommodate those issues.

IDENTIFICATION OF LEARNING DISABILITIES: THE RESPONSIVENESS-TO-INTERVENTION (RTI) MODEL

Historically, the IQ-achievement discrepancy posits that learning disabled students have a significant academic deficit, which is demonstrated by a significant discrepancy between a student’s expected ability (for his/her age) and his/her actual ability, as demonstrated using normative criteria. This model has been with the field of LD for more than 30 years (since P.L. 94-142) and has been criticized on several grounds and for some has outlived its usefulness (for analytical discussions and critiques see Fletcher et al., 1998; Francis et al., 2005; Scruggs & Mastropieri, 2002).

Recently, the field has moved toward a noncategorical classification system that defines LD as an inadequate response to documented effective instruction (called the responsiveness-to-intervention (RTI) model; e.g., Hollenbeck, 2007). Specifically, the IDEA reauthorization (2004) suggested that “a local educational agency may use a process that determines if a child responds to scientific research-based intervention as a part of the evaluation procedures.” (20 USC §§1400). The essence of this model is to rule out the hypothesis that poor instruction is accountable for low achievement (Fuchs, Fuchs, & Hollenbeck, 2007; Sampson, Faggella-Luby, & Fritschmann, 2005). Fuchs et al. described that the RTI has two goals: (a) to classify students as having LD based on their inability to respond to effective instruction and (b) to provide effective instruction to students who need it, as early as possible. Based on the RTI model, a student is initially selected based on “unresponsiveness,” that is, his/her inability to demonstrate adequate growth in academic subjects using Tier-1, traditional (effective) instruction (Mastropieri & Scruggs, 2005; Mellard, Byrd, Johnson, Tollefson, & Boesche, 2004). Subsequently, that student is subjected to a set of educational practices that have proved to be effective and his/her academic behavior is evaluated systematically with the expectation that the intensive instruction will lead to improved academic outcomes (Tier-2; see Marston, 2005). If this practice is yet not effective, then the student moves toward more intensive and systematic instruction in order to improve (Tier-3 instruction; see Vaughn, Linan-Thompson, & Hickman, 2003). Lack of responding following Tier-3 would likely signal the presence of an LD.
However, with regard to the RTI, there is at least one important question posited by Compton (2006), “What type of criteria should be used to identify children who do not respond to validated secondary intervention?” (p. 170). The answer to this important question has implications regarding who, among students, will receive intensive instruction, how long will that instruction last, when it should take place, when it will be discontinued, and what criteria will signal discontinuation of the applied intervention. This concern is discussed in detail below.

How Is “Unresponsiveness” Evaluated?

What is the golden criterion, deviation from which constitutes a significant deficit? Should it be (a) the deviation between a child’s score from his/her previous performance (Berninger & Abbott, 1994), (b) the deviation between a student’s performance and the mean or slope of his/her class, (c) the deviation between a student’s performance and the mean or slope of his/her school (See Fig. 1), (d) the deviation of a student’s performance from the norms (grand mean), (e) the deviation of a student’s performance from an average based on an elaborate weighted model that involves personal and situational parameters, (f) the deviation between a student’s performance and a specific learning profile, or (g) the meeting of critical benchmarks (Good et al., 2001)? We believe that all of the above concerns should be placed within an ecological framework that takes into account information from the school itself.

For example, what would happen to a student whose performance is at the lowest 10th percentile among students in his/her school (School B, Fig. 2) and how would his/her performance be evaluated and judged if the same student were educated in a low-achievement school (School A)? Obviously, there would be a significantly lower probability for a low-achieving student in a high-ability school to be identified as having LD compared to the (high) probability of identification in the high-achieving school. Similarly, an average-ability student in a low-achieving school would have a high probability of being identified as having LD in a high-ability school. This concern is not addressed currently within the RTI, as information about the school or setting in which a student is educated are not weighted upon. Unfortunately, schools within or across districts may be very different not only at their mean level but also in their rate of growth in a given academic subject (see slopes of two different-ability schools in Fig. 3).
Currently, several important models have been proposed, but none of them involves (or weighs) information from the school or school district regarding students with LD identification or classification. For example, Fuchs and Fuchs (1998) have proposed the “dual discrepancy” model in which evaluations are based on both growth (slope of improvement) and mean-level improvement at the end of a treatment (see also Burns & Senesac, 2005). Torgesen et al. (2001) used standard scores (e.g., 10th percentiles) as a means of differentiation using normative assessments. Fuchs, Fuchs, and Compton (2004) proposed the “slope discrepancy” model in which academic achievement is measured periodically and then evaluated against a predefined standard. Good et al. (2001) suggested the need to meet specific learning milestones (benchmarks) that need to be defined both empirically and theoretically. Last, Vellutino et al. (1996) proposed a median split model in comparing the growth of a target student to the mean slope of his/her class. Although the above models did not explicitly state inclusion of information from the school or district, several of the above researchers have commented

Fig. 1. The effects of type of centering on the prediction of a school outcome. Schools have equal slopes but differ in their mean levels.
Fig. 2. Two examples of schools, a low-achieving (School A) and a high-achieving (School B) from the population of schools within a state.

Fig. 3. Two examples of schools, a low-achieving and a high-achieving from the perspective of growth differences in reading.
on the need to do so. For example, Fuchs and Deshler (2007) suggested that the point of reference could potentially be a mean that would be computed from a data pool (norm) defined by same age peers from the same school or district (or nation) using multiple curriculum-based measurements (CBMs). Mastropieri and Scruggs (2005) raised concerns about the consistency of decision making across schools, districts, and states, and similar concerns were raised by Kavale, Holdnack and Mostert (2005) as an explanation of the different identification rates across states (see also Johnson, Mellard, & Byrd, 2005). The above concerns could be addressed using an ecological approach to LD identification.

The Importance of Context: An Ecological Perspective to the Identification of Learning Disabilities

Dean, Burns, Grialou, and Varro (2006) stated that “When considering a classification of LD and the need for special education services it would seem that more than the specific achievement standards within a given school or district should be taken into account to warrant the descriptor of ecological” (p. 161). Thus, an ecological approach may be particularly valuable in order to make the most informal decisions. According to the ecological perspective (Bronfenbrenner, 1986), in order to understand a child’s behavior one needs to take into account all parameters of the child’s environment (home, school, community, culture, etc.).

Within the school, for example, it is important to understand that a student’s achievement is a function of the activities taking place in that school. For example, the provision of ample practice and reinforcement by teachers, the implementation of effective practices, adherence to the curriculum, the ability of teachers to manage challenging behaviors or keep a fast-paced lesson are only some factors that explain the achievement levels of the students. This is one explanation why classrooms and schools vary in their achievement level. That is, specific differences in context are likely accountable for the observed differences in mean achievement. Does it then make sense to evaluate students in the absence of context?

The Relative Achievement Discrepancy Model (Peterson & Shinn, 2002)

The purpose of the relative achievement discrepancy (RAD) model is to take into account information from the district or the school in order to make
informed decisions about placement and identification. Peterson and Shinn (2002) described the (RAD) model as a means of drawing inferences about what constitutes deficient achievement by using information from a student’s district or school (see also Deno, 1989; Shinn, 2002). Peterson and Shinn (2002) added that there is no information about the adequacy of the RAD model. However, there is another consideration that may contribute to the need to adapt an RAD model. That consideration pertains to the regression to the mean phenomenon (Labouvie, 1982; Nesselorade, Stigler, & Baltes, 1980). Based on that model, individuals’ repeated assessments of a construct tend to cluster around the mean of their (representative) distribution. In other words, students who lie further away from the mean tend to come closer with repeated testing. Thus, based on that consideration it is likely that students who are educated in a low-achieving school would likely cluster around the mean of their distribution and that effect would contribute to increased differences between schools. This consideration further strengthens the need to take into account the variability between classrooms, schools, or districts and the use of relative criteria in the identification of LD.

Purpose

The purpose of the present studies was twofold. First, to evaluate how context affects decision making regarding identification of LD, and secondly to propose an RAD model that would assist in defining “unresponsiveness,” within the RTI model. Specifically, Study 1 tested the importance of educational context in drawing conclusions regarding deficits in achievement. Study 2 attempted to expand on the RAD (Peterson & Shinn, 2002) by presenting a multilevel model that would use “relative” achievement criteria in order to define “unresponsiveness” in LD.

Specifically, Study 1 attempted to provide answers to the following research questions:

1. What are prevalence rates of LD?
2. Are the prevalence rates of LD predicted by students’ actual reading ability and teacher ratings of that ability?
3. Is the relationship (slope) between reading ability and LD identification moderated by the area in which a school is located?
STUDY 1: METHOD

Participants and Procedures

Participants were 167 students, out of which 149 were typical and 18 had been diagnosed as having LD through a state diagnostic team. There were 75 boys and 92 girls. All students came from state schools and identification of LD was based on the discrepancy between ability and achievement. Students came from general education/inclusive classrooms and were from grades 1 to 9 (elementary = 124, junior high school = 43). Students were Caucasian and represented three different nationalities. There were also 15 students who were bilingual.

Measures

Learning Disabilities Screening Scale for Teachers

Primary and secondary school educators who were well acquainted with students’ skills and achievement were asked to fill in the Learning Disabilities Screening Scale for Teachers (Padeliadu & Sideridis, 2008). The educators rated 116 behaviors based on frequency (9-point Likert-type scale where 1 = always and 9 = never). The behaviors represented the following academic subjects: listening, speaking, reasoning, reading, writing, and mathematics. Each subscale contained 17–20 items representing the multidimensional nature of the problems that students with LD face as defined from the National Joint Committee on Learning Disabilities (see Hammill, 1990). For example, in reading, educators rate behaviors related not only to decoding, fluency, and comprehension, but also to reading comprehension strategy use (cognitive and metacognitive). For the present study’s purposes, however, only the reading scale was used. Its internal consistency estimate was equal to 0.995.

Tests of Reading Achievement

Five constructs, related to decoding, fluency, morphology, syntax, and comprehension, comprised the reading instrument. This reading test was similar in principle and concept to the existing Test of Reading Performance (TORP; Padeliadu & Sideridis, 2000). The difference is that all items corresponded to the lexical, morphological, and syntactical level of each grade (3rd–9th).
Decoding

The ability of decoding is based on the usage of the orthographic and phonological strategy (dual-route hypothesis; Castles & Coltheart, 1993) as well as on semantic knowledge (Perfetti & Hart, 2001). Therefore, three subtests were implemented in order to evaluate the threefold elements of decoding. These subtests were the following.

Word Decoding. A list of 53 words, presented in ascending difficulty, assessed students’ orthographic ability (Griffiths & Snowling, 2002) to correctly decode words with a meaning. These words ranged between 1 and 8 syllables. Students were asked to read aloud the real words. Any decoding errors such as missing or added letters/syllables, word replacements, and/or incorrect stressing were scored with a 0, while phonologically correctly read words were scored with a 1. A discontinuation rule involved five consecutive errors. The internal consistency (Cronbach’s alpha) estimate of the scale was equal to 0.831.

Pseudoword Decoding. A list of 24 pseudowords, presented in ascending difficulty, assessed students’ phonological strategy (Share & Stanovich, 1995) while decoding words without meaning. Syllables in pseudowords varied from 1 to 6. Students were asked to read the pseudowords accurately and to stress them appropriately. When students read items incorrectly (i.e., made mistakes in stressing, missed or added letters/syllables, and/or replaced part of the word that reminded them of an existing word) they received the score of 0; correct reading was linked to a score of 1. The discontinuation rule involved also five consecutive errors. Internal consistency (Cronbach’s alpha) of the scale was 0.799.

Word and Pseudoword Decoding. In order to find out the amount that vocabulary knowledge supported decoding, students were asked to read aloud only the words with meaning derived from a pool of mixed words and pseudowords. Students had to correctly choose and decode aloud only the 18 words with meaning included within 32 items, which were presented in rows including 3 to 5 words and pseudowords. For each correctly chosen word, students were given 1 point, whereas a 0 signaled a failed item. Alpha reliability was 0.898.

Reading Fluency

Students were presented with an unfamiliar expository text of 279 words and were asked to read the words as accurately and fast as they could for
one minute. After the lapse of time, students' score was evaluated by adding
the words read correctly. For this construct, alpha reliability was 0.928.

**Morphology and Syntax**

Two grammar and two syntax subtests were included in the reading measure
(Carlisle, 2003). Students were asked to transform words included in
brackets into the correct grammatical form (the first two subtests) or to put
words in the right order by the help of a picture (the last two subtests).
Correct responses were given 1 point, and incorrect 0 point. The internal
consistency of the construct was 0.754.

**Comprehension**

Reading comprehension was evaluated by two subtests. The first subtest
included four groups of sentences and students were asked to find the two
sentences in the group with the same meaning. The passages for the second
reading comprehension exercise were one narrative and two expository texts
ranging from 97 to 127 words and of ascending difficulty. Seven questions
corresponded to each text and measured the three types of the reading com-
prehension question taxonomy of Pearson and Johnson (1978): textually
explicit, textually implicit, and scriptually implicit. All right responses
were scored with 1 point while all wrong responses were scored with 0.
The stopping rule involved five errors in responding to questions. The
internal consistency of reading comprehension was equal to 0.957 based on
Cronbach’s alpha.

**Demographics of School**

Schools were identified by the area in which they were located (rural,
suburban, urban) based on data from the national statistics agency. There
were 103 schools located in urban areas, 27 in suburban, and 37 in rural areas.

**Statistical Analyses**

Multilevel random coefficient modeling (MRCM) was implemented to aid
diagnostic decision making (Kreft & de Leew, 1998; Roberts, 2004; Shin,
Espin, Deno, & McConnell, 2004). We evaluated whether identification of
a child as having an LD would relate to his/her achievement levels (e.g., in
reading) or also on the characteristics of his/her school, such as the area in
which the school was located. For that purpose we fitted the Bernoulli
model with the dependent variable (LD identification) being binary
In order for the resulting logistic coefficient to be transformed into percentage points, we employed the following formula:

\[
\text{Probability of an outcome} = \frac{1}{1 + \exp \{-\eta_{ij}\}}
\]

(1)

with \(\eta_{ij}\) being the logistic regression coefficient.

**RESULTS OF STUDY 1**

*Unconditional Model of LD Identification*

Initially an unconditional model (i.e., without predictors) was fit to the data in order to ascertain the prevalence rates of an LD, using a population-based model (Raudenbush & Bryk, 2002):

**Level-1 Model**

\[
\text{Probability of Diagnosis} = \phi_{ij}
\]

\[
\log \left[ \frac{\phi_{ij}}{(1 - \phi_{ij})} \right] = \beta_{0j}
\]

**Level-2 Model**

\[
\beta_{0j} = \beta_{00} + u_{0j}
\]

with the \(\phi\) (phi) dichotomy representing the prevalence (mean) rates. This statistic is transformed into log units and is expressed with the term \(\beta_{0j}\).

Subsequently, prevalence rates are examined as a function of their intercept \(\beta_{00}\), the error \(u_{0j}\), around that intercept, and any other predictor variables at Level-2 of the model. In this unconditional model the Level-1 variance is the reciprocal of the Bernoulli variance \(\phi_{ij} = (1 - \phi_{ij})\).

Results indicated that 10.84% of the students in the population were identified as having LD. However, in order to ascertain whether one type of school was predictive of LD identification, we bootstrapped the population mean (Chernick, 2007; Efron, 1982) to ensure that there were not substantial biases due to participant selection (Diaconis & Efron, 1983). These analyses were conducted using S+ and are described in the following section.
Bootstrapping the Mean Prevalence Rates of LD in the Population

In order to be confident that our sample estimates of the prevalence rates of LD were accurate, we bootstrapped the mean rate using the nonparametric bootstrap and 1,000 replications. The mean bias of the bootstrap distribution (see Fig. 4) was 0.00006707, which was miniscule. Furthermore the standard error of measurement was 0.025. Last, the empirical 95% confidence intervals were between 7.2% and 16.2%. Thus, these findings increase our confidence that our estimates of prevalence were close to population estimates.

Learning Disabilities as a Function of Reading Ability in the Absence of Context

The purpose of the present analysis was to test the hypothesis that students’ reading achievement was predictive of an LD identification/diagnosis, in the
absence of any other predictor. Thus, the following multilevel model was fit to the data:

**Level-1 Model**

\[
\text{Probability}_{\text{Diagnosis}} = \phi_{ij}
\]

\[
\log\left(\frac{\phi_{ij}}{1 - \phi_{ij}}\right) = \beta_{0j} + \beta_{1j}(\text{Decoding}) + \beta_{2j}(\text{Grammar})
\]

\[
+ \beta_{3j}(\text{Syntax}) + \beta_{4j}(\text{Fluency}) + \beta_{5j}(\text{Comprehension})
\]

\[
+ \beta_{6j}(\text{Reading Rating}) + u_{0j}
\]

**Level-2 Model**

\[
\beta_{0j} = \gamma_{00} + u_{0j}
\]

\[
\beta_{1j} = \gamma_{10}
\]

\[
\beta_{2j} = \gamma_{20}
\]

\[
\beta_{3j} = \gamma_{30}
\]

\[
\beta_{4j} = \gamma_{40}
\]

\[
\beta_{5j} = \gamma_{50}
\]

\[
\beta_{6j} = \gamma_{60}
\]

Results highlighted the presence of two significant findings (see Table 1). There was a significant slope linking fluency and teacher’s ratings of reading with LD placement. Let us see what these significant slopes mean.

**Table 1.** Multilevel Random Coefficient Model Predicting Prevalence of Identification Rates Due to Student Level Variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>SE</th>
<th>t-ratio</th>
<th>p</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept $B_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept $\gamma_{00}$</td>
<td>-1.762</td>
<td>0.403</td>
<td>-4.369</td>
<td>0.000**</td>
<td>66</td>
</tr>
<tr>
<td>Slopes of predictors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decoding $\beta_1$</td>
<td>0.006</td>
<td>0.054</td>
<td>0.118</td>
<td>0.907</td>
<td></td>
</tr>
<tr>
<td>Grammar $\beta_2$</td>
<td>0.656</td>
<td>0.498</td>
<td>1.318</td>
<td>0.190</td>
<td>115</td>
</tr>
<tr>
<td>Syntax $\beta_3$</td>
<td>-0.266</td>
<td>0.189</td>
<td>-1.044</td>
<td>0.163</td>
<td>115</td>
</tr>
<tr>
<td>Fluency $\beta_4$</td>
<td>-0.026</td>
<td>0.009</td>
<td>-2.893</td>
<td>0.005*</td>
<td>115</td>
</tr>
<tr>
<td>Comprehension $\beta_5$</td>
<td>-0.010</td>
<td>0.033</td>
<td>-0.301</td>
<td>0.764</td>
<td>115</td>
</tr>
<tr>
<td>Teacher ratings $\beta_6$</td>
<td>-0.189</td>
<td>0.094</td>
<td>-2.014</td>
<td>0.046*</td>
<td>115</td>
</tr>
</tbody>
</table>

*Notes: The coefficients reflect the fixed effects part of the model; SE = standard error of measurement; *$p<0.05$, **$p<0.01$; as a probability level of zero cannot exist.*
The intercept represents the log odds of LD identification for a student with average levels \((z = 0)\) in decoding, grammar, syntax, fluency, comprehension, and teacher ratings of reading. When applying the formula \(1/1\{\exp\}\) the value of \(-1.76\) is transformed into a probability and was 14.7%. This probability represents prevalence rates for students who are average across all independent variables. However, in order to evaluate the contribution of the significant predictors we need to apply predicted values.

Given that all independent variables, except area, were grand mean centered, then all partial regression coefficients become zero as we try to model an average student across all variables (with mean \(z = 0\) across predictors) except fluency for which we model the effect for 1 SD above and below the mean. Thus, before we apply the equation above to evaluate the effects of context we first need to evaluate the two significant effects in the absence of context. By applying predicted values at 1 SD above and below the mean in fluency and ratings of reading we obtain the following:

For Fluency levels at \(-1\) SD

\[ \eta = \gamma_{00} + \gamma_{40} (\text{Fluency}) + u_0 \]

which gives us:

\[ \eta = -1.76 + (-0.026)(-1) = -1.734 \text{ (i.e., 15% prevalence rates)} \]

For Fluency levels at \(+1\) SD

\[ \eta = \gamma_{00} + \gamma_{40} (\text{Fluency}) + u_0 \]

which gives us:

\[ \eta = -1.76 + (-0.026)(1) = -1.786 \text{ (i.e., 14.36% prevalence rates)} \]

In other words, if a student had fluency levels 1 SD above the mean his/her probability to be identified as having LD is significantly less (14.4%) compared to having 1 SD below the mean in fluency (with the probability leveling up to 15%). This may seem like a small difference but it is evaluated as a function of the error of measurement and was nevertheless significant. The effects are more dramatic, however, when evaluating the effects of reading ability through teacher ratings (see below).

For reading ability ratings that place a student at 1 SD below the mean

\[ \eta = \gamma_{00} + \gamma_{60} (\text{Reading Ratings}) + u_0 \]
which gives us:

$$\eta = -1.76 + (-0.189)(-1) = -1.571 \text{ (i.e., 17.2\% prevalence rates)}$$

For reading ability ratings that place a student at 1 SD above the mean

$$\eta = \gamma_0 + \gamma_6 \text{ (Reading Ratings)} + u_0$$

which gives us:

$$\eta = -1.76 + (-0.189)(1) = -1.786 \text{ (i.e., 12.47\% prevalence rates)}$$

Thus, for individuals who are rated as being 1 SD below the mean in reading ability as measured by teacher ratings, chances are that in approximately 17\% of the time they will be identified as having LD. On the contrary, for students who are rated at reading ability levels above 1 SD from the mean chances for identification lie at approximately 12.5\%. This finding substantiates the role of teacher ratings regarding overall reading ability. Subsequent modeling tested whether the relationship between fluency/reading ratings and LD placement is affected by the context in which students are educated (i.e., school area). In other words we attempted to test the stability of the slope coefficient as a function of the areas in which the school was located.

Cross-Level Interactions Between Context and Reading Ability for the Prediction of LD Identification

The above model suggested that students’ actual fluency levels and their teachers ratings of their reading ability were significant predictors of LD identification. The purpose of the present modeling was to test the hypothesis that the above two significant relationships (slopes) would be moderated by school area/location (see Table 2). Thus, the following multilevel model was fit to the data:

Level-1 Model

Probability\text{Diagnosis} = \phi_{ij}

$$\log \left[ \frac{\phi_{ij}}{1 - \phi_{ij}} \right] = \beta_{0j} + \beta_{1j}\text{(Decoding)} + \beta_{2j}\text{(Grammar)} + \beta_{3j}\text{(Syntax)} + \beta_{4j}\text{(Fluency)} + \beta_{5j}\text{(Comprehension)} + \beta_{6j}\text{(Reading Rating)} + u_{0j}$$
In the “absence of context” model tested previously, when all independent variables were treated as predictors of LD prevalence rates, fluency emerged as a significant predictor. Specifically, for individuals 1 SD above the mean in fluency there was a significantly lowered probability of being diagnosed as having LD. However, when the relationship between fluency and LD identification was moderated by the area in which the school was located there was a significant moderating effect (cross-level interaction) of the area in which the school was located. This significant finding suggested that the relationship between fluency and LD identification was, in part, dependent upon the area in which the school was located. Similarly, with regard to teacher ratings, their relationship to LD identification “changed” as a

**Table 2.** Multilevel Random Coefficient Model Predicting Prevalence of Identification Rates Due to the Moderating Role of School Location.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>SE</th>
<th>t-ratio</th>
<th>p</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept $B_0$</td>
<td>$-1.778$</td>
<td>$0.411$</td>
<td>$-4.321$</td>
<td>$0.000^{**}$</td>
<td>$66$</td>
</tr>
<tr>
<td>Intercept $\gamma_{00}$</td>
<td>$0.001$</td>
<td>$0.053$</td>
<td>$0.019$</td>
<td>$0.985$</td>
<td>$113$</td>
</tr>
<tr>
<td>School location $\gamma_{41}$</td>
<td>$-0.021$</td>
<td>$0.009$</td>
<td>$-2.498$</td>
<td>$0.014^*$</td>
<td>$113$</td>
</tr>
<tr>
<td>Teacher ratings $\gamma_{61}$</td>
<td>$-0.304$</td>
<td>$0.234$</td>
<td>$-1.298$</td>
<td>$0.197$</td>
<td>$113$</td>
</tr>
<tr>
<td>Teacher ratings $\gamma_{61}$</td>
<td>$0.046$</td>
<td>$0.089$</td>
<td>$0.517$</td>
<td>$0.605$</td>
<td>$113$</td>
</tr>
</tbody>
</table>

Notes: The coefficients reflect the fixed effects part of the model; SE = standard error of measurement; $^*p<0.05$, $^{**}p<0.01$. 

**Level-2 Model**

\[
\beta_{0j} = \gamma_{00} + u_{0j}
\]

\[
\beta_{1j} = \gamma_{10}
\]

\[
\beta_{2j} = \gamma_{20}
\]

\[
\beta_{3j} = \gamma_{30}
\]

\[
\beta_{4j} = \gamma_{40} + \gamma_{41} \text{ (School Area/Location)}
\]

\[
\beta_{5j} = \gamma_{50}
\]

\[
\beta_{6j} = \gamma_{60} + \gamma_{61} \text{ (School Area/Location)}
\]
function of school location, although the moderator by itself did not exceed conventional levels of significance. Let us see how area$^4$ affected the relationship between fluency and reading ability and LD identification by applying predicted values:

For Urban Area and the Relationship Between Fluency and LD Identification

$$\eta = \gamma_{00} + \gamma_{10} \text{(Decoding)} + \gamma_{20} \text{(Grammar)} + \gamma_{30} \text{(Syntax)}$$

$$+ \gamma_{40} \text{(Fluency)} + \gamma_{41} \text{(Fluency \times Area)} + \gamma_{50} \text{(Comprehension)}$$

$$+ \gamma_{60} \text{(Reading Ratings)} + u_0$$

Fig. 5 (upper and lower panels) displays predicted values for reading achievement ratings (upper panel) and fluency (lower panel). With regard to reading achievement ratings, the relationship between LD identification and reading is apparently moderated by the area in which the school is located. As shown in Fig. 5 (upper panel) the rates of identification are higher in urban schools compared to rural, for the same level of reading ability. In other words, students who possess the same exact levels of reading achievement (low) are more likely to be identified as having LD in an urban school compared to a rural school. These effects, even more pronounced, were observed with regard to fluency levels (see Fig. 5, lower panel). These findings confirmed the moderating role of context in the identification of students with LD.

BRIEF DISCUSSION OF STUDY 1

The purpose of Study 1 was to test the hypothesis that context would moderate the relationship between LD identification and achievement in reading. The first important finding of Study 1 was that students’ reading ability (actual or through teacher ratings) was predictive of their LD. This finding was in the predictive direction and was a prerequisite of subsequent modeling. The second and most important finding was that the above relationship was moderated by the area in which the school was located. In urban schools, for a specific reading ability, the probability of being identified as having LD was significantly higher compared to the same-ability student who was located in a rural school. This finding, as disturbing as it can be seen, suggests that our identification criteria should take into account information from the schools in order for their estimates to be valid. Below we present a modified RAD model in order to aid identification in LD.
Fig. 5. Prevalence rates of LD identification as a function of reading ability level (teacher ratings) in urban versus rural school locations. Predictions are for reading achievement ratings (upper panel) and fluency (lower panel).
Proposed Model of LD Identification

The present authors see merit in the RAD model. We would like to extend this idea to a model that takes into account information from both the student and his/her context. We would, thus, like to propose a modification of the RAD model by taking into account information from the student and his/her variance and compare that to information about the mean of his/her class/district or other grouping. In essence, the present model proposes a comparison of two slopes or two growth parameters: (a) the trajectory of a student’s growth and (b) the trajectory of growth for his/her class/district or other reference point.

Let us assume that the individual slope of a student is positive and at 0.2 (i.e., within person slope). We can compare that slope that we assume reflects the performance of a student over time to the average slope of his/her class/district or other reference point (between-students slope). If the two slopes are significantly different (in favor of the group’s slope), then we can conclude that the growth and, thus, responsiveness of a student to instruction is significantly lower compared to that of his/her class/district (grand slope).

The advantages of this modeling, compared to the absolute model of LD identification, is that information about the within-student variance as well as that of his/her class/district are taken into account in the identification process. Also, differences are estimated at the growth parameter and not at the mean level (which may reflect a “variable” and likely unreliable point estimate of a student’s abilities). Thus, one has a better picture of the “responsiveness” attribute as it is established from multiple measurements (i.e., measurements over time). Study 2 presents a working example of the proposed “relative slope-difference discrepancy model (RSDDM)” that may aid the valid assessment of a student’s responsiveness to an effective treatment.

STUDY 2

The purpose of Study 2 was to present a working example of the RSDDM within the multilevel data analysis framework. Data came from a study by Antoniou et al. (2009).

METHOD OF STUDY 2

Participants and Procedures

Participants were 29 students, whose writing ability was evaluated three times within a semester. All students came from one rural district, and
included 19 boys and 10 girls. Also, two students were bilingual. Of these students 6 were 4th graders, 13 were 5th graders, and 10 were 6th graders. More information about the methodology of the study can be traced in Antoniou et al. (2009).

Measures

A standardized writing composition scale developed by Porpodas (2008) was implemented in the present study which involved measurements of structural parts of the produced composition such as coherence, relevance to the purpose, syntax, etc. The alpha reliability of the scale was 0.792.

Data Analyses

Multilevel modeling was again implemented in order to model the slopes of individual students compared to the mean of their district. Following this modeling, the two slopes were compared with each other using multivariate \( \chi^2 \)-tests.

RESULTS OF STUDY 2

Modeling Students’ Growth in Writing Within a District

In order to evaluate the hypothesis that there was ample variability at the intercept and slope levels, the following multilevel model was fit to the data:

**Level-1 Model**

\[
Y = \beta_0 + \beta_1 \text{(Growth)} + r_0
\]

**Level-2 Model**

\[
\begin{align*}
\beta_0 &= \gamma_{00} + u_0 \\
\beta_1 &= \gamma_{10}
\end{align*}
\]

with \( \beta_0, \gamma_{00}, \) and \( \gamma_{10} \) representing intercepts and \( \beta_1 \) representing the growth parameter. The terms \( r_0 \) and \( u_0 \) represent error terms at levels one and two, respectively.

This model testified that means and slopes were nonzero. In the absence of, at least, a significant slope, further modeling would make no sense.
Results indicated that the grand slope was equal to 0.982, which was significantly different from zero \( t(82) = 3.329, p<0.01 \). Thus, this model acted as a baseline model in order to proceed with further testing. Subsequently, a student’s growth was compared to a sample of students from his/her school district (we call him/her student 28) by creating a dummy term.

**Growth in Writing Between Student 28 and a Sample of Students from His/Her District**

In order to evaluate the hypothesis that student 28 was not responding equally well to effective instruction, compared to his/her peers, the following multilevel model was fit to the data:

**Level-1 Model**

\[
Y = \beta_0 + \beta_1 \text{ (Growth)} + r_0
\]

**Level-2 Model**

\[
\beta_0 = \gamma_{00} + u_0 \\
\beta_1 = \gamma_{10} + \gamma_{11} \text{ (Dummy)}
\]

Results (see Fig. 6), indicated that there were significant differences between the slope of the students who came from District A \( (b = 1.086) \) \( t(81) = 2.718, p<0.01 \) and the slope of student 28 who came from the same district \( (b = -1.83) \) \( t(81) = -2.082, p<0.05 \). This finding suggests that student 28 did not respond adequately to instruction in writing compared to the average growth produced by students of the same educational environment (district). These findings confirm the importance of moving into relative models of ability for the identification of LD as the respective finding would be saliently different if the reference criterion would involve the grand mean (i.e., through standardized assessments).

**BRIEF DISCUSSION OF STUDY 2**

The purpose of Study 2 was to present a working example of the proposed RSDDM within the multilevel data analysis framework. Study 2 demonstrated that it is possible to compare the “responsiveness” of a student to a treatment by comparing his/her growth parameter to that of a “relative” reference point (his/her class, district, random sample of the population
The point to be made here is that the use of a relative, compared to absolute, criterion may prove to be a more appropriate reference point compared to federal norms. Furthermore, the application of the multilevel framework demonstrated the sensitivity of the model to estimate and compare growth parameters across different groups. Nevertheless, it is important to bear in mind that alternative “relative points of reference” exist and it may be worth exploring them. For example, one could select low-achieving students and evaluate hypotheses that a given low-ability student is significantly worst compared to low-achieving peers. Other alternative models could involve comparisons between pairs of students of low/average/high ability or the comparison of profiles for a given area of achievement.

**GENERAL DISCUSSION**

The purpose of the present studies was twofold: (a) to evaluate the role of context in the identification of LD within the RTI model and (b) to propose
a model that would take into account context in the assessment of “unresponsiveness” within the RTI.

In Study 1, the potentially moderating role of school context was evaluated on whether it affects the relationship between achievement and LD identification. It is important to emphasize here that the purpose of the present studies was not to be critical about the RTI model. Nor was it to identify all parameters at the school level that may affect decision making. Instead, the primary purpose was to highlight the role of school context in order to aid decision making, as each decision regarding responsiveness should be (in our opinion) tied to the particular class/school/district in which the decision is made. Thus, the present findings aim to assist valid identification of students with LD, and suggest that the specifics of each school (meaning school achievement, school area) need to be taken into account in such diagnostic decisions.

The most important finding of Study 1 was the fact that the relationship between reading ability/achievement and LD identification was moderated by the area in which a school was located. Specifically in urban areas, compared to rural areas, the probability of being identified as having LD was significantly higher for a given ability student. In other words, in urban schools, low-achieving students had a much higher probability of being identified as having LD compared to rural and suburban schools. This finding highlights the moderating role of a school’s characteristics in LD identification decision making and agrees with earlier concerns of leading researchers in the field on the salient role of context (Dean et al., 2006; Fuchs & Deshler, 2007; Johnson et al., 2005; Kavale et al., 2005; Mastropieri & Scruggs, 2005).

The above finding suggests that for a given student’s low ability, the urban school’s diagnostic teams and criteria of implementation are more likely to lead to a conclusion for identification compared to the decision made in rural schools. What factors contribute to this differentiation? We can only speculate what some of these factors maybe. For example, a schools efforts to increase its mean levels, to achieve excellence, to compete at the state level, maybe some relevant reasons. Nevertheless, we can only speculate why decision making at urban schools is much more strict, compared to the respective decision in rural and suburban schools.

Ecological Validity

The findings from Study 1 supported an ecological approach to decision making regarding identification in LD (Dean et al., 2006; Peterson et al., 2002).
The purpose of Study 2 was to present a working example of a modified RAD model in order to demonstrate how such a model works. The results of Study 2 suggested that multilevel modeling can aid our decision making as we can directly compare the growth parameters of individual students compared to a reference point (his/her group mean, grand mean, or other). Fig. 6 demonstrated that effect. Thus, the example of Study 2 strongly supported the premise and also use of an RAD model. This demonstration agrees with the recommendations of Fletcher, Francis, Morris, and Lyon (2005) who reported several psychometric problems of discrepancy models and concluded with a recommendation to apply hybrid models.

Limitations

It is important to acknowledge certain limitations of the present studies. With regard to Study 1, the findings are correlational in nature. Thus, conclusions about cause and effect relationships should not be drawn. Field studies in which school, teacher, assessment, and other environment characteristics are assessed are important in order to evaluate the validity of the responsiveness (compared to unresponsiveness) ratings (such as the Peterson & Shinn, 2002 study). Furthermore, the sample of the study was relatively small for estimating population parameters in means and relationships although we attempted to substantiate the effects by use of intensive simulations (i.e., the bootstrap).

With regard to Study 2, it is important to establish criteria pertaining to what constitutes a significant difference between a point/slope estimates of a student compared to his/her reference group11 (class/district/norms/or other). The problem stems from the fact that such comparisons would be limited by the population of the class (anywhere between 15–25 students) or district (likely more than 25 students) and the time series representing responsiveness, thus low power will likely be an issue (Cohen, 1992). One potential solution would be the use of the bootstrap to estimate population parameters in growth based on (e.g., student 28) estimates (Chernick, 2007; Efron, 1979, 1982, 1985; Efron & Tibshirani, 1993). Another potential solution would be to use the standard error estimate of the reference population (i.e., class/district/other) as that of the individual low-achieving student. A third solution would be to “borrow” standard errors from the literature of the population of LD that are specific to the specific object under investigation. Following that decision regarding standard errors, one could use custom slope difference tests to assess differences between the two
slopes (student’s vs. his/her reference group). Other recommendations may involve the implementation of clinical means in assessing differentiation in growth (compared to statistical inferences that may likely reflect low power; see also Onwuegbuzie, Levin, & Leach, 2003), particularly if these means are defined by expert multidisciplinary teams (Fuchs & Deshler, 2007).

Conclusions and Recommendations

Fuchs and Deshler (2007) stated:

It is untrue and misleading to claim that we currently have a necessary and sufficient knowledge base to guide the implementation of RTI as a process of early intervention and disability identification across all grades, for all academic skills, in all content areas, and for all children and youth. (p. 134)

Similar concerns have been raised by other leading researchers (e.g., Kavale et al., 2005). The intent of the present studies was to increase our knowledge base about how the RTI would be best implemented in order to deliver instruction to those who needed it the most (as well as diagnostic mean). Specifically, the present studies provided information on how “unresponsiveness” can be conceptualized using an ecological framework and the use of dual discrepancy criteria (evaluation of means and slopes) using a multilevel mathematical framework. We hope that the present work will stimulate more discussion regarding the determination of “responsiveness” for which, today, there is no consensus. Until this consensus is reached, different schools will be identifying different-ability students as those who need additional services and student identification would be tied to the specific school in which assessments are made.

Future Directions

In the future it is important to elaborate on the influential effects of school, family, and other contextual factors that may affect decision making within the RTI model. For example, Sideridis, Padeliadu, and Antoniou (2008) found a biasing factor related to the diagnosis made by male teachers compared to female teachers. This significant bias could be rooted to teacher’s education, experience, personality, and other factors. Nevertheless, it is important to evaluate the contribution of personal and contextual factors that contribute error variance to our decision making regarding
identification of students with LD. This systemic view of LD identification is particularly more important within the RTI model as responsiveness is likely a function of several influential factors involving the teacher, the school psychologist, the use of specific CBMs, and many more. Much more research work is needed in order to reach a valid model of LD identification.

**UNCITED REFERENCES**

Bryk & Raudenbush (1992); Burns, Appleton, & Stehouwer (2005); Choi (2001); Diakidou, Stylianou, Karefillidou, & Papageorgiou (2005); Fuchs & Fuchs (2006); Fuchs et al. (2007); Hammill & Bryant (1998); Nezlek (2001); Nezlek (2003); Padeliadu & Antoniou (2008).

**NOTES**

1. Fig. 1 will be discussed next within the multilevel modeling framework as a means to accommodate the variability due to classrooms/schools/districts and other means of differentiation.
2. As Good, Simmons, and Kame‘enui (2001) stated those benchmarks would have to meet the criteria of empirical and theoretical soundness, and social validation.
3. Estimates are taken from Table 1.
4. “Area” and “school location” were used interchangeably.
5. This is the essence of the “relative” aspect of this model. To be able to compare the performance of a given student with the most appropriate reference point (point estimate) or rate of growth (slope). Whether that reference point is the mean of a student’s class, or his/her district or other needs to be ascertained but nevertheless needs to be built in the model of LD identification. We certainly support a multiple measurement model that provides multiple points over time and provides a more consistent estimate of a student’s ability.
6. Model in which the reference criteria involve the mean of a normative group.
7. These $\chi^2$-tests run with 1 df.
8. The dummy variable represents student 28 with a 0 and the remaining students of the class with a 1.
9. Nevertheless, because the two slopes were evaluated against the absolute reference point of zero a $\chi^2$-custom test was employed to test the hypothesis that the two coefficients were significantly different from each other. Results indicated that the null hypothesis of no differences was rejected [$\chi^2(1) = 7.085, p<0.01]$.
10. Certainly the most important element of “relative” models is definition of the relevant population to act as the reference point.
11. For Berninger and Abbott (1994), the criterion should be “within” the student and should be solely based on an intraindividual model, thus, comparisons between a
student’s performance at a given time from his/her common responding. In our RSDD model that would be equivalent to comparing the slope of the student to the absolute criterion of zero.

REFERENCES


The Role of Context in the Identification of Learning Disabilities


Dear Author,

During the preparation of your manuscript for typesetting, some questions may have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin of the proof or compile them as a separate list.

**Disk use**
Sometimes we are unable to process the electronic file of your article and/or artwork. If this is the case, we have proceeded by:

- Scanning (parts of) your article
- Rekeying (parts of) your article
- Scanning the artwork

**Bibliography**
If discrepancies were noted between the literature list and the text references, the following may apply:

- The references listed below were noted in the text but appear to be missing from your literature list. Please complete the list or remove the references from the text.
- **UNCITED REFERENCES:** This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or delete it. Any reference not dealt with will be retained in this section.

**Queries and/or remarks**

<table>
<thead>
<tr>
<th>Location in Article</th>
<th>Query / remark</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU:1</td>
<td>Please check the suggested running head title.</td>
<td></td>
</tr>
<tr>
<td>AU:2</td>
<td>Please check the levels of all the section headings.</td>
<td></td>
</tr>
<tr>
<td>AU:3</td>
<td>Please check that the following references are not included in the reference list: Fuchs, Fuchs, &amp; Compton (2004); Padieliadu &amp; Sideridis (2000); Perfetti &amp; Hart (2001); Share &amp; Stanovich</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(1995); Pearson &amp; Johnson (1978); Kreft &amp; de Leew (1998); Shin, Espin, Deno, &amp; McConnell (2004); Peterson et al. (2002); Sideridis, Padeliadu, &amp; Antoniou (2008);</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td><strong>AU:4</strong> Please check whether CBM has been correctly expanded.</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td><strong>AU:5</strong> Please check that Roberts (2003) has been changed to Roberts (2004), per reference list.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td><strong>AU:6</strong> Please check that the symbols (plus and minus) have been removed from the terms, <code>+/-1 SD above and below</code>, <code>+1 SD above</code> and <code>-1 SD below</code>. Is this OK?</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td><strong>AU:7</strong> Please check the added initials of the author names in Good et al. (2001).</td>
<td></td>
</tr>
</tbody>
</table>