Article

Educational and Psychological Measurement
1–24

DOI: 10.1177/0013164415604440
epm.sagepub.com
SAGE

Assessing Validity of Measurement in Learning Disabilities Using Hierarchical Generalized Linear Modeling: The Roles of Anxiety and Motivation

Georgios D. Sideridis[1]

Abstract

The purpose of the present studies was to test the hypothesis that the psychometric characteristics of ability scales may be significantly distorted if one accounts for emotional factors during test taking. Specifically, the present studies evaluate the effects of anxiety and motivation on the item difficulties of the Rasch model. In Study 1, the validity of a reading comprehension scale was evaluated using the Rasch model with 60 students with learning disabilities (LD). Item parameters were retested for the presence of anxiety and results indicated that the scale was substantially more difficult in its presence. Study 2 replicated the findings of Study 1 using maladaptive motivation and extended with inclusion of adaptive motivational variables in order to reverse the effect. Results using students with and without LD indicated that the difficulty levels of the scale was lower for students with LD, in the presence of positive motivation, compared with a typical student group. Study 3 extended the dichotomous hierarchical generalized linear model with polytomous data. The measures of an ability test were adjusted for the presence of anxiety and results indicated that differential item functioning was observed at both the global level and the most difficult ability item. It is concluded that the difficulty levels of a scale are heavily influenced by situational factors during testing, such as students' entry levels of motivation and affect.

[1]Boston Children's Hospital, Harvard Medical School, Boston, MA, USA

Corresponding Author:
Georgios D. Sideridis, Boston Children's Hospital, Clinical Research Center, Harvard Medical School, 300 Longwood Avenue, Boston, MA 02125, USA.
Email: georgios.sideridis@gmail.com

Keywords

Rasch model, anxiety, motivation, hierarchical generalized linear modeling, assessment, identification, polytomous data

More than 20 years ago Adelman, Lauber, Nelson, and Smith (1989) stated, ''Assessment procedures that can significantly enhance motivation to perform are needed so that poor performance due to low or avoidance motivation is not misdiagnosed as indicating ability deficits or as symptomatic of learning disabilities (LD).'' (p. 234). In their study, they demonstrated that when given proper motivation, students diagnosed with LD could potentially learn and thus, overcome their learning disability. They concluded that motivation should be accounted for in the diagnostic testing of learning disabilities. Thus, early on these authors emphasized the importance of valid assessments in learning disabilities, having the suspicion that motivation could potentially be accountable for false positive or false negative results. Unfortunately, in the past 23 years, motivation was not systematically controlled for, in the identification process of learning disabilities. One purpose of the present studies is to provide a technology to control for the effects of motivation on test administration.

Valid Assessment in Learning Disabilities

Recently, the National Joint Committee on Learning Disabilities (NJCLD, 2010) discussed issues on the assessment and identification of students with disabilities. In the paper, the Committee specifically stated that

> Factors such as poor self-regulatory behaviors (e.g., lack of motivation . . .) . . . are not in themselves considered learning disabilities, but they may be concomitant with learning disabilities. A comprehensive assessment must address all areas of suspected disability, so if these conditions exist it is presumed that they would have been considered and addressed during the assessment. If this is the case, such information would be integrated into the comprehensive assessment report and may affect the program, curriculum, and/or instructional recommendations for the individual student. (p. 9)

It is the first time that a Committee has acknowledged motivation as a factor that needs to be considered for an accurate differential diagnosis of learning disabilities. Specifically, the NJCLD states that motivation, among other factors, should be *considered and addressed during the assessment process*. Although the Committee does not describe *how* motivational issues can be addressed, it acknowledged the potential deleterious role of poor motivation in the assessment/identification process.

Along with motivation, anxiety is a major obstacle for individuals with learning disabilities (Bailey & Andrews, 2003; Nelson & Harwood, 2011) and the problem is greatly exacerbated by the fact that comorbid psychiatric disorder to anxiety are in the range of 75% to 80% (Cooper, 1997; Masi, Favilla, & Mucci, 2000; Raghavan, 1997). At ideal levels anxiety provides the adaptive force needed to cope with task demands and adversities (Cooray & Bakala, 2005), at high levels, however, it interrupts cognitive processing and leads to apprehension, helplessness and task withdrawal (Greulich et al. 2014). In Eysenck, Derakshan, Santos, and Calvo's (2007) terms, anxiety swifts attention to focusing on distractors, rather than the important stimuli, it consumes energy required for storing information in memory and leads to information processing deficits. For students with learning disabilities this trajectory of anxiety to task failure is thought to develop naturally following everyday struggles and embarrassments from failing simple educational tasks. The effects of anxiety may particularly more prominent in the presence of examinations for which the consequences from failure are far more devastating for students with LD compared to the general population. For the former, the assessment situation is yet another instance of potential failure with the consequences of being placed onto undesirable environments (segregated) the development of stigma and the repeat of the vicious cycle of failure and negative emotions.

The topic of valid assessment and identification in learning disabilities has been of major controversy throughout the history of the field (e.g., Fletcher, Coulter, Reschly, & Vaughn, 2004; Fuchs & Fuchs, 2006; Mellard, Deshler, & Barth, 2004). Particularly more so, given that MacMillan and Siperstein (2002) reported that schools differ substantially on how they implement identification procedures. In the present study we will focus on the deleterious effects of a false identification (false positive) due to motivational factors. Mellard et al. (2004) graphically described false identification as errors in the dimensions of our mousetraps. When considering false positive cases, these students take on a diagnosis they do not possess, which is associated with life changing conditions at school. For example, one such effect involves special placement in the form of a special class or an integrated/inclusive setting. Beyond placement, other expectations for a diagnosed child with LD involves, based on the empirical literature, the presence of high levels of depression (Sideridis, 2005), anxiety (Leplow, Murphy, & Nutzinger, 2002), behavioral adjustment problems (Wiener, 2004). As if this is not enough, the student enters a circle of low expectations (Woodcock, & Vialle, 2011), stigma (Zhao, Zhang, & Yu, 2008), and poor socialization (Galway & Metsala, 2011). The above represent extremely serious consequences for a student to experience, especially if this hardship is due to a misdiagnosis.

Importance of the Present Studies

The present studies target at quantifying the effects of adaptive and maladaptive motivation and anxiety on the assessment of students with LD, achievement on specific content areas or ability (e.g., IQ). At present, no attempt has been made to account for social–emotional factors in the assessment process leading to the diagnosis of students with LD. For example, anxiety during testing and motivation potentially contribute significant effects over students' performance. For example, if the average difficulty of an achievement test for a group of students with LD is +1 standard deviation (SD) above the mean, an LD student would need ability levels equal to or above +1 SD from the mean of his/her group in order to be successful on that test. If the estimated difficulty of the test becomes +2 SDs in the presence of high levels of anxiety, an LD student with +1 SD ability who is anxious would fail the test (that now requires +2 SDs ability levels. Thus, in the presence of anxiety, way fewer students with LD would be successful (another 13.6%) compared with not having anxiety as an independent/confounding variable in the environment during testing. The immediate implication and concern that arises from the above description is summarized in the question: ''What if the lowered performance due to anxiety may bring an otherwise 'typical student's performance' below a critical performance level that *defines* a learning disability?'' That would be a devastating decision with tremendous implications for the person's placement, socialization experiences and well-being, knowing that the actual performance level of the student was higher than the observed one. One objective of the present studies (Study 1) was to evaluate and correct for the effects of maladaptive motivation on students' measurement of ability. A secondary objective of the present study (Study 2) was to evaluate the effects of adaptive motivational processes (e.g., hopefulness and approach goals) on students' ability. If the effects are positive for the population of LD, then the heightened ability in the presence of adaptive motivation should be taken as reflective of true ability (based on classical test theory). Thus, adaptive motivation should be practiced and implemented, specifically for the LD population, if those cognitions have positive effects on ability and achievement. Provided that the true score should reflect the true state of affairs regarding ability, person competencies should be corrected for the presence of maladaptive motivation and remain uncorrected in the presence of adaptive motivation. The main objectives of the present study were twofold: (a) to evaluate the effects of anxiety, adaptive and maladaptive motivation on students' with LD ability estimates and (b) to explore the statistical means to account for those influences (i.e., hierarchical generalized linear modeling).

Description of Study 1

The purpose of Study 1 was to test the hypothesis that anxiety exerts deleterious effects on the reading performance of students with LD. Initially, the Rasch model was fit to the data of the reading comprehension scale in order to estimate item difficulties and full scale parameters (information function, measurement error, etc.). Subsequently, a special type of multilevel modeling, the hierarchical generalized linear model (HGLM), was implemented in order to provide not only equivalent estimates to the Rasch model (Beretvas, Meyers, & Rodriguez, 2005; Kamata, 2002; Roberts & Herrington, 2005) but also corrections for the presence of anxiety during testing.

Method of Study 1

Participants of Study 1

Data were collected from 62 elementary and secondary students with LD from Grades 5 through 9 ($N_{5\text{th grade}} = 29$, $N_{6\text{th grade}} = 11$, $N_{7\text{th grade}} = 10$, $N_{8\text{th grade}} = 10$,

Table 1. Characteristics of the Participating Students.

Grade	n	Boys	Girls	Age, years, *M* (SD)	No. of bilingual students
5	29	20	9	10.41 (0.37)	5
6	11	6	5	11.54 (0.44)	2
7	10	8	2	12.48 (0.75)	2
8	10	7	3	14.08 (1.07)	0
9	2	2	0	13.85 (0.88)	0
Total	62	43	19	11.67 (1.53)	9

$N_{\text{9th grade}} = 2$). Students were selected from 16 public schools in the area of southern Greece. There were 43 boys and 19 girls and their mean age was 11.67 years (*SD* = 1.53; minimum = 9.96, maximum = 16.89). Fifty-three children were Greek monolingual learners, while 9 were bilinguals with Albanian being their mother tongue. Thirty-six of the participants had been classified as students with LD from state diagnostic agencies, while the remaining 26 had to meet the criteria based on the discrepancy model (Fletcher, Morris, & Lyon, 2003) in order to be classified as students with LD. The criteria used were (a) adequate intelligence (>85), (b) discrepancy between ability (as reflected in IQ scores) and achievement in the subscales of decoding and spelling of the Software for Screening Learning Skills and Difficulties (LAMDA; Protopapas & Skaloumbakas, 2008), and (c) absence of physical handicaps. Furthermore, all students were identified as having LD using a normative rating scale for the screening of LD (Learning Disabilities Screening for Teachers questionnaire; Padeliadu & Sideridis, 2008). All elementary school students attended the resource rooms of their schools. Table 1 provides an overview of the participant's characteristics.

Procedures of Study 1

Students' participation in the study required completion of an informed consent form provided to both students and their guardians. Their participation was voluntary and researchers emphasized the anonymity of their responses. Assessments took place in students' classrooms during regular school hours and in the special computer room of each school. All students were tested individually in reading comprehension by trained psychology students. The reading comprehension test lasted for 30 minutes. The LAMDA software for screening for LD was administered in a whole-group administration on personal PCs and with the use of headphones. The duration of administration was approximately 40 minutes. All students were aware that they could withdraw their participation at any time during the tests. Students were told that the purpose of these assessments was to obtain a picture of the current level of students' performance in order to evaluate whether new teaching procedures would be effective.

Measures of Study 1

Reading Comprehension. Students' passage understanding was assessed through a text of the reading comprehension subscale of the Test of Reading (Padeliadu & Antoniou, 2008). The reading comprehension subscale includes 9 passages, 5 narratives, and 4 expository with an average number of 120 words. Narrative texts involve the questions who, where, when, what happened and what was the solution of the story. Expository texts are merely informative. Each text is accompanied by 6 multiple choice questions with 4 options for all texts. Students read the passage at their own pace, silently or aloud and then heard the questions (which were also in writing in front of them). Three of the questions in each passage referred to facts, easily identified in the text; the remaining four questions required activation of students' prior knowledge and the making of inferences. There was a discontinuity rule which involved erroneous responding on five consecutive questions. The internal consistency estimate of the reading comprehension subscale was .86.

Social Concerns/Concentration. This subscale comes from the Revised Manifest Anxiety Scale for Children (RCMAS; Reynolds & Richmond, 1978). Cole, Hoffman, Tram, and Maxwell (2000) termed the subscale *social alienation*. The subscale involves student's concerns about not being liked by other students, feeling lonely, being tired, and so forth, and thus is an assessment of social difficulties and adaptation to school. Sample items are ''My child gets mad easily'' and ''My child is nervous.'' Cronbach's alpha of the subscale was .79.

Data Analysis of Study 1

Rasch Model. The Rasch model (Rasch, 1980) was applied as a means of evaluating person abilities theta (θ) given a measure's difficulty levels delta (δ). The comparison between measurements (with and without the contribution of anxiety/motivation will involve plotting test response functions and test information functions. The comparison of the simple structures will be based on a principal components analysis of the residuals using the methods outlined by Guadagnoli and Velicer (1991) by use of the congruence coefficient (RC) (using software developed by Hebbler, 1989). The RC coefficient expresses the similarity of factor solutions and ranges between -1 and +1. Jensen (1998) provided conventions for evaluating the magnitude of the coefficient with values of .90 or greater reflecting a high degree of similarity. More detailed conventions have been provided by MacCallum, Widaman, Zhang, and Hong (1999): ''.98-1.00 = excellent, .92-.98 = good, .82-.92 = borderline, .68-.82 = poor, and below .68 = terrible.''

Estimation of Rasch Parameters Using Hierarchical Generalized Linear Modeling. The approach adopted herein comes from Kamata (1998, 2001, 2002) and others (e.g., Adams, Wilson, & Wang, 1997; Muckle & Karabatsos, 2009; Pastor, & Beretvas, 2006; Raudenbush, Johnson, & Sampson, 2003; Roberts & Herrington, 2005) who demonstrated the estimation of Rasch model parameters by use of HGLM, a variant of multilevel modelling techniques (Raykov & Marcoulides, 2008). Initially, the multilevel model is set up as a two-level (or a three-level) model using the Bernoulli function, with Level 1 containing the items of the scale (with each item representing one level of a repeated factor) and Level 2 the person estimates (Pastor, 2003; Raudenbush & Bryk, 2002; Bryk & Raudenbush, 1992) by use of the logit link function which implies that the mean μ of the distribution is predicted by the logistic regression model. Using this function, $\eta_{ij} = \log(\varphi_{\mathrm{ij}}/1 - \varphi_{ij})$, that is the log odds of success equals the log of the probability of success φ (i.e., 50%) divided by $1 - \varphi$. If the actual probability of success is 50%, then the odds of success are 1 and the logit will be 0 as predicted by the Rasch model. This function is responsible for scaling the data from minus to plus infinity to between 0 and 1. Thus, item responses are treated as hierarchical data with the items being nested within persons. This dichotomous HGLM is equivalent to the Rasch model (Cheong, 2006 and Kamata, 2001) and the log odds of person j endorsing item i is modeled at Level 1 as shown below.

Level 1 (Bernoulli) model:

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_{0j} + \beta_{1\mathrm{j}}X_{1ij} + \beta_{2\mathrm{j}}X_{2ij} + \cdots + \beta_{(k-1)j}X_{(k-1)ij} \tag{1}$$

Level-2 model expressing person estimates:

$$\begin{cases} \beta_{0j} = \gamma_{00} + u_{0j} \\ \beta_{1j} = \gamma_{10} \\ \vdots \\ \beta_{(k-1)j} = \gamma_{(k-1)0} \end{cases} \tag{2}$$

With p_{ij} being the probability that person j will answer item i correctly and X_{ij} being the ith dummy variable for participant j. The parameters of the model must be estimated using the either the full maximum likelihood procedure with numerical integration of the likelihood function or the quasi-likelihood procedure which involves the Taylor series expansion approach (see Hox, 2010). The term β_{0j} reflects the intercept of the model and β_{1j} the slope of variable X_1. Last, the term u_{0j} reflects the random effect of the intercept, which is expected to be normally distributed (with a mean of 0 and variance equal to τ). Since the Rasch model is the 1-parameter model (from the family of item response theory [IRT] models), item difficulties are constrained to be invariant across persons (compared with Birnbaum's [1968] 2-parameter model or the 3- and 4-parameter models). This is why the error terms of the remaining effects were deleted from the subsequent Level-2 equations. As Kamata (2002) and Williams and Beretvas (2006) described, the probability of participant j responding correctly to item i is expressed by the following equation:

$$p_{ij} = \frac{1}{1 + \exp\left\{-\left[u_{0j} - (-\gamma_{i0} - \gamma_{00})\right]\right\}} \tag{3}$$

The final step in the estimation procedure involves the subtraction of each item's slope from the intercept: $(-\gamma_{i0}-\gamma_{00})$ in order to estimate item difficulties. In the present study, for the estimation of Rasch model parameters of the Reading Comprehension subscale the relevant HGLM model was as follows (using a mixed equation for the two levels):

$$\eta_{ij}=\beta_{0j}+\beta_{1j}*\mathrm{RC1}_{ij}+\beta_{2j}*\mathrm{RC2}_{ij}+\beta_{3j}*\mathrm{RC3}_{ij}+\beta_{4j}*\mathrm{RC4}_{ij}+\beta_{5j}*\mathrm{RC5}_{ij} \\ +\beta_{6j}*\mathrm{RC6}_{ij}+u_{0j} \tag{4}$$

with

$$\eta_{ij}=\log\left(\frac{p_{ij}}{1-p_{ij}}\right) \tag{5}$$

and that model was modified, as shown below, in order to account for individual differences in social concerns during the testing situation:

$$\eta_{\mathrm{ij}}=\beta_{00}+\beta_{01}*\text{Social Concerns}_j+\beta_{10}*\mathrm{RC1}_{ij}+\beta_{20}*\mathrm{RC2}_{ij}+\beta_{30}*\mathrm{RC3}_{ij} \\ +\beta_{40}*\mathrm{RC4}_{ij}+\beta_{50}*\mathrm{RC5}_{ij}+\beta_{60}*\mathrm{RC6}_{ij}+u_{0j} \tag{6}$$

Thus, the term "* Social Concerns$_j$" reflects the effects of anxiety on the intercept (and subsequently the estimation of item difficulties). In other words, the above term reflects differential item functioning (DIF) that is common to all items but not specific to each and every item (Williams & Beretvas, 2006). To estimate DIF that is specific to each item, each item must be regressed on the social concerns construct. The evaluation of the two measurements (with and without anxiety/motivation) involved the test characteristic curves (TCCs) and the signed area index (SAI, Wolfe & Smith, 2007):

$$Z_{\mathrm{SAI}}=\frac{\delta_{\mathrm{Reference}}-\delta_{\mathrm{Focal}}}{\sqrt{SE^2_{\delta-\mathrm{Reference}}+SE^2_{\delta-\mathrm{Focal}}}} \tag{7}$$

Results and Discussion of Study 1

Figure 1 displays item characteristics curves (ICCs) based on the Rasch model (upper panel) and controlling for the presence of anxiety (in the form of social concerns/concentration). As shown in Figure 1, lower panel, all item curves were to the right of the respective ones based on the Rasch model (upper panel) suggesting that the items required higher levels of ability in the presence of anxiety. In other words, the scale required below average levels of ability for an average student to be successful but became progressively more difficult when social concerns/concentration was factored in the model. The consistency across all items is remarkable suggesting systematic effects of the presence of social concerns on the ability required by

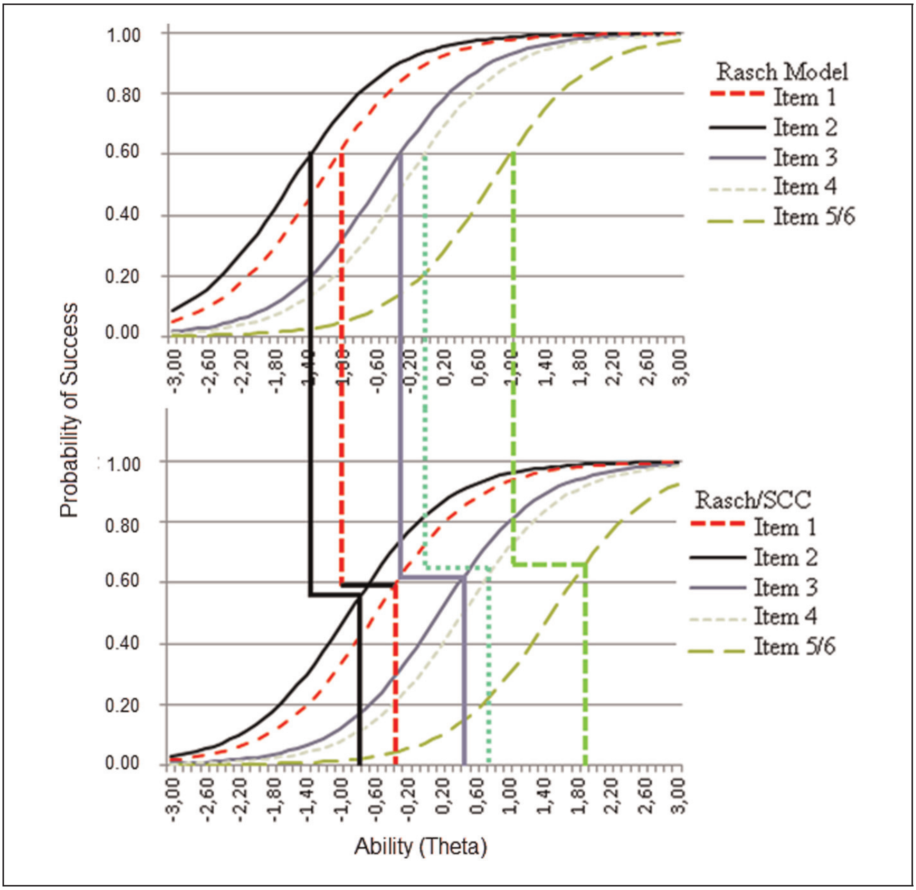

Figure 1. Unadjusted item characteristic curves (upper panel) and adjusted for the presence of social concerns concentration (lower panel) using hierarchical generalized linear modeling in Study 1.

students with LD to be successful. Using the SAI index, results indicated significant differences between the two forms of the reading comprehension subscale. Specifically, the Rasch-based scale was significantly easier for the group of students with LD compared with the difficulty levels of the scale in the presence of anxiety. The Z estimate of the difference between the two TCC curves was equally to 4.35 units, which was significant at $p < .001$ for a two-tailed test.

Description of Study 2

The purpose of Study 2 was to extend the findings of Study 1 by including forms of adaptive motivation in order to evaluate their effects on person ability estimates. A secondary goal of Study 2 was to make comparisons between students with and without LD. Maladaptive motivation involved being engaged out of obligation, termed *ought avoidance* (Sideridis, 2008) a construct that originated in several motivational theories (Carver, Lawrence, & Scheier, 1999; Carver & Scheier, 1998; Higgins, 1997). Its negative propensities are described by the fact that this motive describes engagement out of obligation rather than intrinsic reasons with the goal to avoid failure. In Carver and Scheier's (1998) words: ''regulation regarding an ought involves trying to move toward the ought, as a way of moving away from an unwanted comparison point'' (p. 54). Thus, ought avoidance goals are negatively valenced by obligations and their foci is avoiding negative outcomes (i.e., punishment in Carver and Scheier's words). The negative propensities of ought avoidance goals have included elevated physiological arousal and lower achievement, compared to approach motives that were based on obligations (Sideridis, 2008). The construct was selected because it combines an extrinsic source of motivation (i.e., others) with its drive been based on negative emotions (obligations rather than willful engagement) and with the focus being on avoiding negative outcomes. It is expected that this motive will exert deleterious effects on the achievement of students with LD.

On the opposite side of avoidance motivation, we selected ''hardiness'' (Kobasa, 1979) as reflecting a positive motivational process. Hardiness has been described as a dispositional factor in enhancing performance despite any hardship and stress (Maddi et al., 2006). Maddi and Khoshaba (1994, 2001) described hardiness as expressing psychological vitality and elevated performance. It involves three adaptive motivational processes, namely, challenge, commitment and control. Its functioning as a promoter of motivation has been demonstrated across various outcomes such as problem solving coping, and academic achievement (Maddi & Hightower, 1999). We hypothesized that that hardiness would exert positive influences over the performance of students with LD.

Method of Study 2

Participants of Study 2

Participants were 139 elementary school students. Fifty-two students with LD and 97 typical peers all from an urban area in Greece who were diagnosed using State identification criteria that evaluate discrepancies between potential and achievement. There were 65 boys and 75 girls. There were 70 fifth graders and 70 sixth grader from public elementary schools. Students were assessed individually on the emotional Stroop task (see below a description of the empirical procedure) after consenting to their participation. The measures were carried on by trained research assistants and by use of laptops.

Procedures of Study 2

Emotional Stroop Task. It is a modification of the Stroop paradigm with the purpose of assessing cognitive overload whenever the information presented possess attributes that are emotionally significant to the person (MacLeod, 1991). The primary goal of the task is to complete a matching between the color of the presented stimuli and the respective color that is presented on a palette. The prediction is that emotional stimuli involve cognitive–emotional processes, in addition to cognitive processes alone, thus, they would require more time to produce the correct matching compared with neutral stimuli (that only engage the cognitive process). The primary goal of the person is to complete the matching task as quickly as possible without making errors. We used the computerized Stroop task (Genov, Shay, & Boone, 2002) in which words are presented in the center of a 20-color palette. On word presentation, the respondent selects the correct color from the palette using a mouse. By clicking on a small circle in the center of the screen the chronometer starts and thus, time estimates are also recorded. Every time the software is run, the words appear in a random order. Correct-versus-incorrect word matchings comprised the dependent variable.

Measures of Study 2

Achievement. The amount of errors emitted on the matching of emotionally elicited words with the correct color comprised the dependent variable in Study 2. Thus, this achievement scale comprised 9 dichotomously scored words.

Hardiness. It was assessed using the Personal Views Questionnaire (Maddi & Khoshaba, 1994). The scale included 23 items, modified to be age specific for young students. The scores from all three subscales (challenge, commitment, and control) were aggregated to comprise a total score. Alpha of the total score was .75.

Ought Avoidance. It was assessed using the avoidance subscale of the ought scale (Sideridis, 2008). The purpose of the construct is to assess ones feelings of obligation to engage with a task in order to avoid failure. In other words it assesses whether a student is motivated out of an obligation to avoid the negative outcomes of failure. The subscale was composed of three items: (a) ''Do you feel that you *ought* to avoid, by any means you have, failing in this task?''; (b) ''Do you feel that you *ought* to insist so that you won't fail in this task?''; and (c) ''Do you feel that you *ought* to do well in this task in order to avoid any negative consequences from doing poorly?'' Alpha of the subscale was .83.

Data Analysis of Study 2

The comparisons of models involved an application of HGLM as in Study 1. That is, initially a Rasch model was applied to the data (by use of HGLM see Equation 5) followed by the same model controlling for the presence of positive or negative motivation.

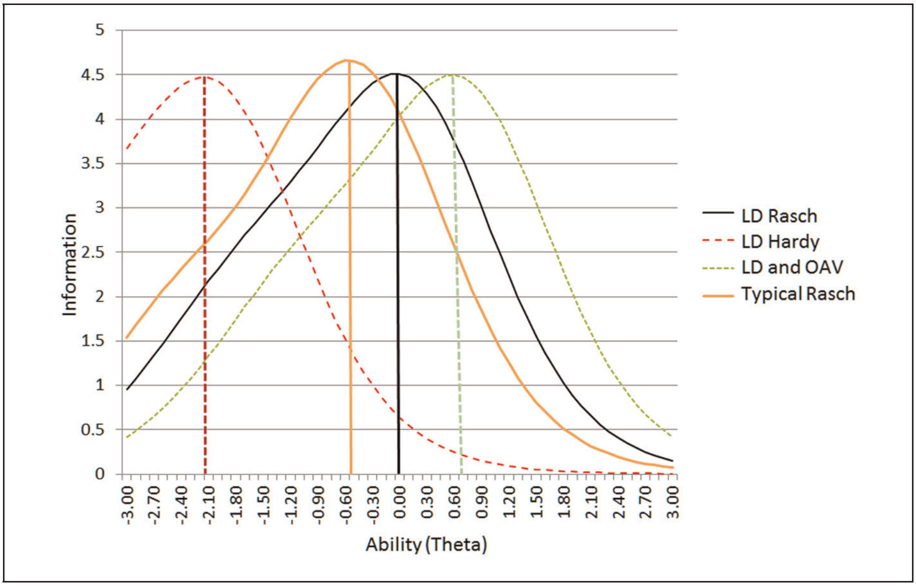

Figure 2. Comparison of test information functions (TIFs) between typical and learning disabilities (LD) student groups in the presence of hardiness and ought avoidance in Study 2.

Results and Discussion of Study 2

When looking at the difficult levels between the two groups, results indicated that the test was significantly more difficult for the students with LD compared to the typical student group[1] (Test DIF = 0.44), as expected. The difference became more severe in the presence of a negative motivational discourse, that is, ought avoidance for the LD group (see Figure 2 test information functions [TIFs] on where the sensitivity of measurement is for each condition). The difference in ability between the two conditions was equal to 1.05 logits. In other words, in the presence of ought avoidance, the test was 1.05 logits more difficult for the LD group compared with typical peers (compared with an earlier difference based on just the Rasch model of 0.44 logits). However, in the presence of positive motivation in the LD, that is hardiness, the effect was reversed significantly. The mean ability levels required by the LD group was −2.61 logits compared with −0.92 logits for the typical group. The difference between the two groups was 1.69 logits, favoring the LD group, which represents a large effect size (Steinberg & Thissen, 2006).

When comparing the TCCs between the typical and students with LD results indicated that the test was significantly more difficult for students with LD compared with their typical peers ($Z = 2.621$, $p < .05$). However, this difference was reversed, in favor of the LD group, in the presence of positive motivation by the latter group ($Z = 10.069$, $p < .001$). Thus, in the presence of hardiness the test was significantly easier for the LD group compared with their typical peers. Within the LD group, the measure was particularly easy in the presence of hardiness (Z = 10860, $p < .001$) and significantly harder in the presence of ought avoidance (Z = 3.110, $p < .01$), with the reference being the baseline Rasch model (in the absence of hardiness and ought avoidance). These findings further substantiate the differential role of motivational attributes and characteristics that affect person ability estimates (see also Swerdzewski, Harmes, & Finney, 2011) suggesting that positive motivational processes exert salient effects over student ability estimates, particularly for students who struggle the most. Study 3 was designed to extend the present methodology (HGLM) with polytomous data, using a measure of anxiety.

Description of Study 3

The purpose of Study 3 was to apply the HGLM methodology with polytomous data in order to control for the detrimental effects of anxiety on a measure of nonverbal ability. The present model, termed Polytomous HGLM (PHGLM; Williams & Beretvas, 2006) makes the HGLM methodology applicable to situations in which more than two response options are available such as partial credit models (e.g., Lord, 1980; Muraki, 1990), and attitude types models (that employ the Likert-type scaling scheme).

Method of Study 3

Participants of Study 3

Participants were 107 elementary school students classified to be at risk for learning disabilities by their teachers (54 boys and 53 girls). The students did not meet State diagnostic criteria, thus, they were at risk for LD. Fifty of them were in Grade 4 and 57 in Grade 5. Students were selected from a low–socioeconomic status school in a rural area of Greece. They were informed of the confidentiality of their participation and took on the ability measure using individual administration (students were called out of their class during regular teaching hours) from trained research assistants.

Measures of Study 2

Anxiety. It was measured using the Revised Children's Manifest Anxiety Scale (RCMAS; Reynolds & Richmond, 1978). Evidence of reliability and various forms of validity have been reported elsewhere (e.g., Chorpita, Moffitt, & Gray, 2005; Stark & Laurent, 2001). Alpha of the general anxiety index was .798.

Ability. It was assessed using Raven's colored progressive matrices, a standardized and normative measure of nonverbal intelligence. It is composed of three 12-item subscales, namely A, AB, and B, which are of increased difficulty. Only Form A was applied in the present report because of its unidimensionality. The first three items for all scales were of minimal difficulty and served as an indicator of the validity of subsequent responding. Alpha of the subscale was .990.

Data Analysis

For the purpose of the present study, a PHGLM model was developed to account for the polytomous nature of the data (ordinal HGLM model; Feldman, Masyn, & Conger, 2009). As Kamata (1998) stated, within the multilevel model this analysis can be carried out using a binomial link function with a k number of trials compared with the Bernulli model, which employs dichotomous data. Below there is a brief conceptualization of the PHGLM model (Williams & Beretvas, 2006). For the current model in which nonverbal intelligence was measured via a partial credit model (ordinal) with 3 options, the HGLM model was conceptualized as follows:

$$\begin{aligned}\eta_{1ij} &= \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{(k-1)j}X_{(k-1)ij}\\ \eta_{2ij} &= \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{(k-1)j}X_{(k-1)ij} + \delta_j\end{aligned} \tag{8}$$

with η_{1ij} being the log odds of responding to Category 1, and η_{2ij} the log odds of responding to Category 2, compared with some reference category (e.g., probability of responding to Category 1 vs. Categories 2 and 3). The Level-2 model is parameterized as follows:

$$\begin{cases}\beta_{0j} = \gamma_{00} + u_{0j}\\ \beta_{1j} = \gamma_{10}\\ \vdots\\ \beta_{(k-1)j} = \gamma_{(k-1)0}\\ \delta_j = \delta\end{cases} \tag{9}$$

with the threshold difference parameter δ being modeled as fixed across both measured items and individuals along with the item level effects of β_{ij} to $\beta_{(k-\ 1)j}$. Based on the above model, item difficulties are estimated using (γ_{00} - γ_{i0} + δ). When attempting to adjust item difficulties for a covariate, the following model can be estimated, which includes a generalized DIF influence of the covariate (on the intercept) and specific DIF effects of the covariate on each of item's difficulty levels:

$$\begin{cases}\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Anxiety}) + u_{0j}\\ \beta_{1j} = \gamma_{10} + \gamma_{11}(\text{Anxiety})j\\ \vdots\\ \beta_{(k-1)j} = \gamma_{(k-1)0}(\text{Anxiety})j\\ \delta_j = \delta\end{cases} \tag{10}$$

The direction and magnitude of the DIF is estimated through evaluating the sign of the coefficients and their significance with a *t*-test statistic (Beretvas & Williams, 2004).

Results and Discussion of Study 3

Results using the PHGLM model suggested that the effects of anxiety were present at a global level as the slope of anxiety after being regressed on γ_{00} was significant ($\gamma_{01} = -0.349$, $p < .05$). Furthermore, an item specific effect was observed for Item A12, the most difficult item ($\gamma_{16} = -0.233$, $p < .05$). This finding is rather unexpected as ability (nonverbal IQ) per se is not malleable to external influences but rather a stable trait. For example, with positive covariates such as motivation, one should not expect that ability would be heightened as it is rather a fixed entity. An effect would only be present if the minimum levels of motivation required to complete the task were not available (as in the presence of amotivation; Ryan & Deci, 2000). Similarly, with anxiety, if there were necessary adjustments those should be on the negative side as anxiety is expected to hinder the cognitive resources required to successfully complete the task at hand. As shown in Figure 3, upper panel, differential ability is not observed at the test level. However, the measure appears to be more sensitive to evaluating individuals of higher ability in the presence of anxiety, as shown through inspecting the TIF.

Results from the PHGLM model also suggested important adjustments of the item difficulties for the presence of anxiety. Interestingly, as mentioned above there were both generalized DIF effects (on the intercept) and also specific ones in Item A12, which was the most difficult item. Thus, it appears that the effects of anxiety are more pronounced when individuals are challenged cognitively and academically. Furthermore, Study 3 extended the dichotomous HGLM model to the case of polytomous item instruments through applying the ordinal model. The application of the PHGLM model suggests that it can be useful in estimating various influences at the person, context, or other levels (e.g., by use of a 3-level model, see Kamata, 1999).

General Discussion

The purpose of the present studies was twofold: (a) to evaluate the effects of anxiety, adaptive and maladaptive motivation on students' with LD ability estimates and (b) to explore the statistical means to account for those influences. Results indicated that anxiety and motivation exerted salient effects over the estimation of item and person abilities of achievement and ability measures and those effects were specialized for typical and LD groups of students.

The present studies have significant implications for the valid assessment of aptitude and abilities in special education, particularly because for special populations performance is associated with a host of decisions, such as placement options, that have significant implications for the well-being of those individuals. For example,

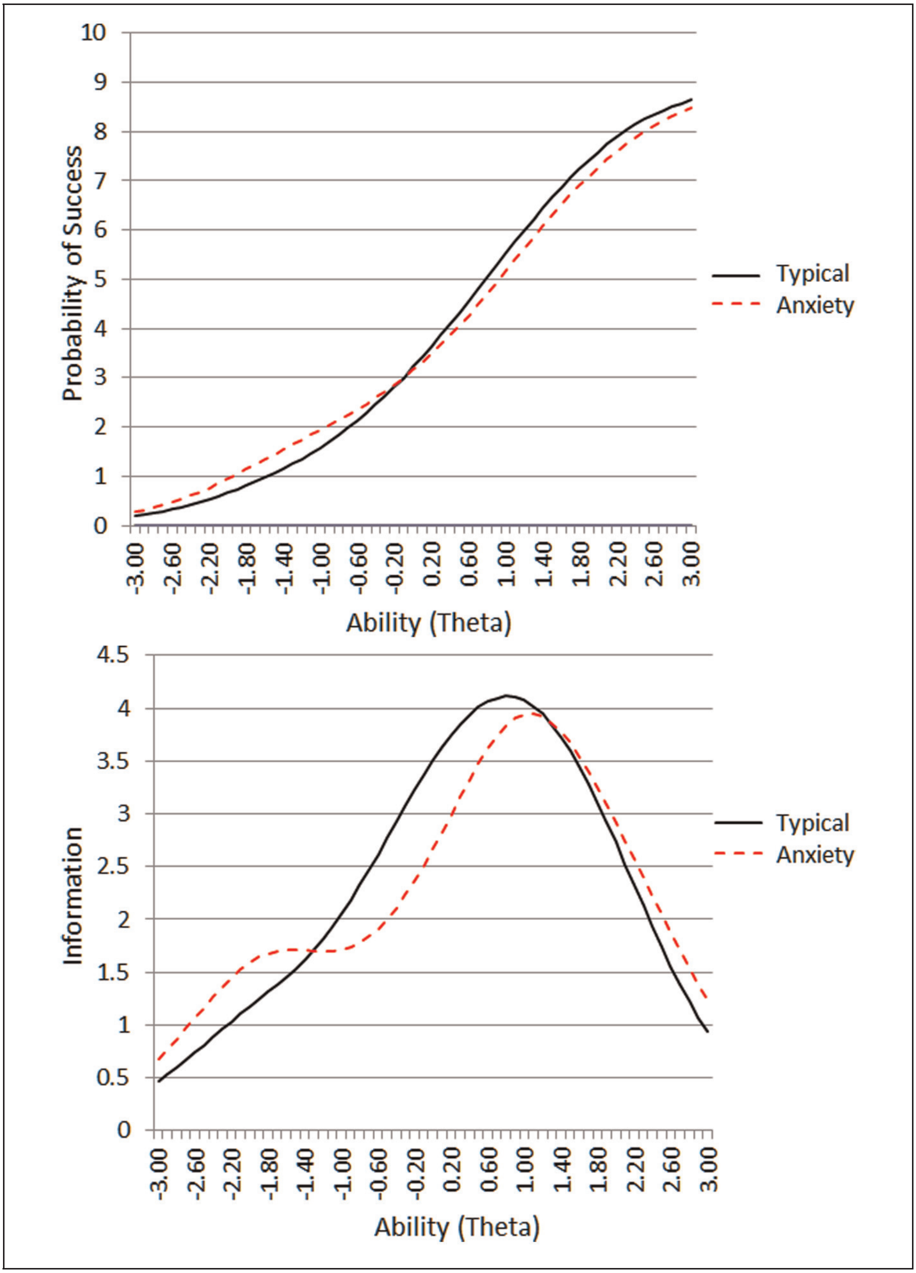

Figure 3. Test characteristic curves in the presence or absence of anxiety (upper panel) and test information functions (lower panel) in Study 3.

restrictive placements (e.g., special schools or special classes) have oftentimes been associated with low expectations, ridicule, humiliation, discrimination, and the development of psychopathological tendencies (Adelman et al., 1989; Sideridis, 2003, 2009). The present findings suggested that negative emotional tendencies (in the form of anxiety) or positive attitudes (in the form of motivation) are associated with saliently different estimates of ability for individuals with learning disabilities. Specifically, higher probabilities of correct responding were observed in the presence of a positive attribute (such as motivation), whereas the opposite effect was observed in the presence of a negative attribute (such as anxiety).

The present findings have significant implications for the psychological assessment literature. Research on motivation suggests that the motivational state that possesses when entering a testing situation determines achievement levels since the former *regulates* behavior in the presence of boredom, disappointment, feelings of shame, and so forth. Several motivational theories have described motivational patterns that have been associated with adaptive or maladaptive self-regulation and achievement (e.g., achievement goal theory, Dweck & Leggett, 1988). Specifically for learning disabilities, the role of motivation has been substantial (Adelman et al., 1989; Sideridis, 2009). Nevertheless, its role has rather been underestimated in this population. In Study 2 of the present article, the actual estimates of difficulty of the scale (at the item level or as a whole) were significantly different in the presence of adaptive motivation (hardiness) or maladaptive (ought avoidance). When accounting for hardiness the levels of ability required by the LD student group were significantly lower compared to not accounting for the levels on that trait. The opposite was true for maladaptive motivation in that high levels of ought avoidance were associated with required levels of ability for a given item/scale much higher compared to ignoring its presence (i.e., compared with the baseline Rasch model). Obviously, the effects of motivation in Study 2 were substantial and suggest that its effects should not be ignored, particularly since learning disabilities are defined in the absence of motivation.

The above findings with regard to students with LD took on an interesting twist when comparing estimates between typical and LD student groups. As shown in Figure 3, the actual Rasch estimates were similar between the two groups with the typical student group requiring lower levels of ability by almost half a logit (.44 logits) in order to perform the correct Stroop matching. However, the levels of difficulty of the scale are reversed in favor of the LD student group when accounting for their levels of hardiness. Thus, in the presence of high levels of adaptive motivation, the scale appears to be significantly easier for the LD student group, compared to their typical peers (by 1.69 logits), which is considered a large effect (Steinberg and Thissen, 2006; Wang & Chen, 2004). This finding agrees with early research work on learning disabilities in that when students with LD were given ample time they achieved at levels that were equal to those of typical students (Runyan, 1991; Alster, 1997), with few exceptions (e.g., Cohen, Gregg, & Deng, 2005). Similarly, when accounting for accommodations for disabilities, results on cognitive testing for students with and without disabilities were comparable, compared to not accounting for specific accommodations (Zurcher & Bryant, 2001). This finding suggests that by applying the methodology of HGLM and PHGLM the resultant student abilities may fall short from meeting the threshold of a learning disability by being comparable to typical students. This finding may greatly change identification rates, and consequently, students placed and served in special education.

The present studies have implications about the validity of the measures. Apparently, a scale is not valid if its item or scale properties change over time (it is not reliable) with the actual trait being unchanged. That certainly was the case with the present studies' findings. These present findings suggest that the effects of positive or negative self-attributes (such as adaptive or maladaptive motivation and anxiety) exert *systematic effects* on the propensities of a scale. Systematic because negative attributes tended to require higher levels of ability from the part of the participants to achieve a certain level of success (at the item or scale level) and the opposite was true for positive characteristics. These findings certainly require some thought on how to account for effects previously ignored. Schools and clinicians that are involved in the assessment of students with learning disabilities may allocate resources to providing the means to reduce anxiety and provide environments free of extraneous sources of stress. Programs that may enhance motivation in the form of self-determination may be particularly useful. A last resort will involve the adjustment of student abilities using the present models as a means of controlling for those extraneous sources of influence. The present findings reflect a first attempt to weigh assessments of achievement and ability for the presence of motivation and anxiety. The results suggest that the effects of these characteristics are substantial and the methodology of HGLM for both dichotomous and a polytomous measure provides a quantification of that effect and eventually a control for those extraneous influences. More work is needed, however, to test the methodology for other sources of error such as unreliability of measurement, regression from the mean, or other sources of systematic error and/or simple random error.

Conclusions and Recommendations

Methodologically speaking, one needs to establish that all aspects of the environment are appropriate for the administration of achievement tests. When either the measurement conditions are altered during testing or the personal characteristics of the participants are not accounted for, then valid administration is at stake. Statistically speaking, several researchers have suggested models to account for the effects of motivation and other sources of variance during testing. Within that line of literature, motivation is considered a source of systematic measurement error, that is, a nuisance in testing. Corrections for its presence could potentially involve Ferrando's (2011) pseudo-congeneric model in which individual differences due to motivation could be accounted for. Magis, Beland, and Raiche (2011) suggested a Bayesian approach that could potentially account for motivation by adjusting test-length. Other models that have corrected for various sources of systematic error have been described by Belov (2011), Van der Linden (2011), and St-Onge, Valois, Abdous, and Germain (2011) and could potentially aid our assessment technology.

The present studies provided overwhelming evidence on the fact that positive and negative motivational processes and anxiety exert significant effects over actual achievement levels of students with LD. When controlling for positive motivation the actual difficulty levels of a scale became significantly lessened compared to ignoring its presence. This finding was reversed for the presence of maladaptive motivational processes and anxiety. As the NJCLD has recently suggested, these effects should, somehow, be factored in our assessment procedures. Ignoring them will only lead to invalid assessments and the related consequences of stress, anxiety, and anger in the families of the affected individuals.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. The differences between the two groups at the item level (item DIF) involved application of the following HGLM:

$$
\begin{aligned}
\eta_{ij} = \gamma_{00} + \gamma_{01} * \mathrm{OAV}_j \\
&+ \gamma_{11} * \mathrm{TYPICAL}_j * \mathrm{I3}_{ij} + \gamma_{11} * \mathrm{LD}_j * \mathrm{I3}_{ij} \\
&+ \gamma_{21} * \mathrm{TYPICAL}_j * \mathrm{I5}_{ij} + \gamma_{21} * \mathrm{LD}_j * \mathrm{I5}_{ij} \\
&+ \gamma_{31} * \mathrm{TYPICAL}_j * \mathrm{I6}_{ij} + \gamma_{31} * \mathrm{LD}_j * \mathrm{I6}_{ij} \\
&+ \gamma_{41} * \mathrm{TYPICAL}_j * \mathrm{I18}_{ij} + \gamma_{41} * \mathrm{LD}_j * \mathrm{I18}_{ij} \\
&+ \gamma_{51} * \mathrm{TYPICAL}_j * \mathrm{I20}_{ij} + \gamma_{51} * \mathrm{LD}_j * \mathrm{I20}_{ij} \\
&+ \gamma_{61} * \mathrm{TYPICAL}_j * \mathrm{I25}_{ij} + \gamma_{61} * \mathrm{LD}_j * \mathrm{I25}_{ij} \\
&+ \gamma_{71} * \mathrm{TYPICAL}_j * \mathrm{I27}_{ij} + \gamma_{71} * \mathrm{LD}_j * \mathrm{I27}_{ij} \\
&+ \gamma_{81} * \mathrm{TYPICAL}_j * \mathrm{I29}_{ij} + \gamma_{81} * \mathrm{LD}_j * \mathrm{I29}_{ij} \\
&+ u_{0j}
\end{aligned}
$$

The comparisons of slopes between groups for each Stroop item involved the use of a chi-square test which runs with 1 *df*. When comparing coefficients in the presence of ought avoidance, significant differences were observed for four items using Steinberg and Thissen's (2006) suggestions of large effect sizes in IRT models (i.e., larger than 0.50 logits). The respective findings in the presence of hardiness reflected DIF in eight Stroop items.

References

Adams, R. J., Wilson, M. R., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, *21*, 1-23.

Adelman, H., Lauber, B., Nelson, P., & Smith, D. (1989). Toward a procedure for minimizing and detecting false positive diagnoses of learning disability. *Journal of Learning Disabilities*, *22*, 234-244.

Alster, E. (1997). The effects of extended time on algebra test scores for college students with and without learning disabilities. *Journal of Learning Disabilities*, *30*, 222-227.

Bailey, N. M., & Andrews, T. M. (2003). Diagnostic Criteria for Psychiatric Disorders for Use with Adults with Learning Disabilities/Mental Retardation (DC–LD) and the diagnosis of anxiety disorders. *Journal of Intellectual Disability Research*, *47*(Suppl. 1), 50-61.

Belov, D. (2011). Detection of answer copying based on the structure of a high stakes test. *Applied Psychological Measurement*, *35*, 495-517.

Beretvas, N., Meyers, J., & Rodriguez, R. (2005). The cross-classified multilevel measurement model: An explanation and demonstration. *Journal of Applied Measurement*, *6*, 322-341.

Beretvas, S. N., & Williams, N. J. (2004). The use of hierarchical generalized linear model for item dimensionality assessment. *Journal of Educational Measurement*, *41*, 379-395.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Carver, C. S., Lawrence, J. W., & Scheier, M. F. (1999). Self-discrepancies and affect: Incorporating the role of feared selves. *Personality and Social Psychology Bulletin*, *25*, 783-792.

Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. Cambridge, England: Cambridge University Press.

Cheong, Y. F. (2006). Analysis of school context effects on differential item functioning using hierarchical generalized linear models. *International Journal of Testing*, *6*, 57-79.

Chorpita, B. F., Moffitt, C. E., & Gray, J. (2005). Psychometric properties of the revised child anxiety and depression scale in a clinical sample. *Behaviour Research and Therapy*, *43*, 309-322.

Cohen, A. S., Gregg, N., & Deng, M. (2005). The role of extended time and item content on a high-stakes mathematics test. *Learning Disabilities Research and Practice*, *20*, 225-233.

Cole, D. A., Hoffman, K., Tram, J. M., & Maxwell, S. E. (2000). Structural differences in parent and child reports of children's symptoms of depression and anxiety. *Psychological Assessment*, *12*, 174-185.

Cooper, S. A. (1997). Psychiatry of elderly compared to younger adults with intellectual disability. *Journal of Applied Research in Intellectual Disability*, *10*, 303-311.

Cooray, S. E., & Bakala, A. (2005). Anxiety disorders in people with learning disabilities. *Advances in Psychiatric Treatment*, *11*, 355-361.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*, 256-273.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*, 336-353.

Feldman, B. J., Masyn, K. E., & Conger, R. D. (2009). New approaches to studying problem behaviors: A comparison of methods for modeling longitudinal, categorical adolescent drinking data. *Developmental Psychology*, *45*, 652-676.

Ferrando, P. (2011). A linear variable-θ model for measuring individual differences in response precision. *Applied Psychological Measurement*, *35*, 200-216.

Fletcher, J. M., Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia*, *54*, 304-331.

Fletcher, J. M., Morris, R. D., & Lyon, G. R. (2003). Classification and definition of learning disabilities: An integrative perspective. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 30-56). New York, NY: Guilford Press.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41*, 93-99.

Galway, T., & Metsala, J. (2011). Social cognition and its relation to psychosocial adjustment and in children with nonverbal learning disabilities. *Journal of Learning Disabilities*, *44*, 33-49.

Genov, A., Shay, I., & Boone, R. T. (2002). *Genov modified Stroop task (GMST)* [Computer software and manual]. Retrieved from http://facpub.stjohns.edu/~booner/GMSTsite/index.htm

Greulich, L., Al Otaiba, S., Schatschneider, C., Wanzek, J., Ortiz, M., & Wagner, R. K. (2014). Understanding inadequate response to first-grade multi-tier intervention: Nomothetic and ideographic perspectives. *Learning Disability Quarterly*, *37*, 204-217.

Guadagnoli, E., & Velicer, W. F. (1991). A comparison of pattern matching indices. *Multivariate Behavioral Research*, *26*, 323-343.

Hebbler, S. W. (1989). A BASIC program for computing the coefficient of concordance, r_c. *Educational and Psychological Measurement*, *49*, 615-618.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, *52*, 1280-1300.

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Kamata, A. (1998, April). *One-parameter hierarchical generalized linear logistic model: An application of HGLM to IRT*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Kamata, A. (1999). *Some generalizations of the Rasch model: An application of the hierarchical generalized linear model* (Unpublished doctoral dissertation). Michigan State University, East Lansing, MI.

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, *38*, 79-93.

Kamata, A. (2002, April). *Procedure to perform item response analysis by hierarchical generalized linear model*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Kobasa, S. C. (1979). Stressful life events, personality, and health: An inquiry into hardiness. *Journal of Personality and Social Psychology*, *37*, 1-11.

Leplow, B., Murphy, R., & Nutzinger, D. (2002). Specificity of conditional associative-learning deficits in obsessive-compulsive disorder (OCD) and non-OCD anxiety disorders. *Journal of Clinical and Experimental Neuropsychology*, *24*, 792-805.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84-99.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163-203.

MacMillan, D. L., & Siperstein, G. N. (2002). Learning disabilities as operationally defined by schools. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 287-333). Mahwah, NJ: Lawrence Erlbaum.
Maddi, S. R., Harvey, R. H., Khoshaba, D. M., Lu, J. L., Persico, M., & Brow, M. (2006). The personality construct of hardiness, III: Relationships with repression, innovativeness, authoritarianism, and performance. *Journal of Personality*, *74*, 575-598.
Maddi, S. R., & Hightower, M. (1999). Hardiness and optimism as expressed in coping patterns. *Consulting Psychology Journal*, *51*, 95-105.
Maddi, S. R., & Khoshaba, D. M. (1994). Hardiness and mental health. *Journal of Personality Assessment*, *63*, 265-274.
Maddi, S. R., & Khoshaba, D. M. (2001). *Personal Views Survey III–R: Test development and internet instruction manual*. Irvine, CA: Hardiness Institute.
Magis, D., Beland, S., & Raiche, G. (2011). A test length correction to the estimation of extreme proficiency levels. *Applied Psychological Measurement*, *35*, 91-109.
Masi, G., Favilla, L., & Mucci, M. (2000). Generalised anxiety disorder in adolescents and young adults with mild mental retardation. *Psychiatry*, *63*, 54-64.
Mellard, D., Deshler, D., & Barth, A. (2004). LD identification: It's not simply a matter of building a better mousetrap. *Learning Disability Quarterly*, *27*, 229-242.
Muckle, T. J., & Karabatsos, G. (2009). Hierarchical generalized linear models for the analysis of judge ratings. *Journal of Educational Measurement*, *46*, 198-219.
Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, *14*, 59-71.
National Joint Committee on Learning Disabilities. (2010, June). *Comprehensive assessment and evaluation of students with learning disabilities*. Retrieved from http://www.nasponline.org/advocacy/NJCLD-Comprehensive_Assessment_6-3-10.pdf
Nelson, J. M., & Harwood, H. (2011). Learning disabilities and anxiety: A meta-analysis. *Journal of Learning Disabilities*, *44*, 3-17.
Padeliadu, S., & Antoniou, F. (2008). *Development of the Test of Reading Ability (TERA)* (in Greek). Patras, Greece: Press.
Padeliadu, S., & Sideridis, G. D. (2008). *Learning Disabilities Screening for Teachers*. EPEAEK II Action 1.1.3a, Ministry of Education.
Pastor, D. (2003). The use of multilevel item response theory modeling in applied research: An illustration. *Applied Measurement in Education*, *16*, 223-243.
Pastor, D., & Beretvas, N. (2006). Longitudinal Rasch modeling in the context of psychotherapy outcomes assessment. *Applied Psychological Measurement*, *30*, 100-120.
Protopapas, A., & Skaloumbakas, C. (2008). *Software for screening learning skills and difficulties (LAMDA)*. EPEAEK II Action 1.1.3a, Ministry of Education.
Raghavan, R. (1997). Anxiety disorders in people with learning disabilities: a review of the literature. *Journal of Learning Disabilities for Nursing, Health and Social Care*, *2*, 3-9.
Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago, IL: University of Chicago Press.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage.
Raudenbush, S. W., Johnson, C., & Sampson, R. J. (2003). A multivariate, multilevel Rasch model with application to self–reported criminal behavior. *Sociological Methodology*, *33*, 169-211.

Raykov, T., & Marcoulides, G. (2008). *An introduction to applied multivariate analysis*. New York, NY: Routledge.

Reynolds, C. R., & Richmond, B. O. (1978). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology*, *6*, 271-280.

Roberts, K., & Herrington, R. (2005). Demonstration of software programs for estimating multilevel measurement model parameters. *Journal of Applied Measurement*, *6*, 255-272.

Runyan, M. K. (1991). The effects of extra time on reading comprehensions scores for university students with and without learning disabilities. *Journal of Learning Disabilities*, *24*, 104-108.

Ryan, R., & Deci, E. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54-67.

Sideridis, G. D. (2003). On the origins of helpless behavior in students with learning disabilities: Avoidance motivation?*International Journal of Educational Research*, *39*, 497-517.

Sideridis, G. D. (2005). Goal orientation, academic achievement, and depression: Evidence in favour of a revised goal theory framework. *Journal of Educational Psychology*, *97*, 366-375.

Sideridis, G. D. (2008). Feeling obliged to ''do well'' or ''not to fail''? The distinction between approach and avoidance dimensions of oughts. *Learning and Individual Differences*, *18*, 176-186.

Sideridis, G. D. (2009). Motivation and learning disabilities: Past, present and future. In A. Wigfield & K. Wentzel (Eds.), *Handbook on motivation* (pp. 605-626). Hillsdale, NJ: Lawrence Erlbaum.

St-Onge, C., Valois, P., Abdous, B., & Germain, S. (2011). Accuracy of person-fit statistics. A Monte Carlo study of the influence of aberrance rates. *Applied Psychological Measurement*, *35*, 419-432.

Stark, K. D., & Laurent, J. (2001). Joint factor analysis of the children's depression inventory and the revised children's manifest anxiety scale. *Journal of Clinical Child Psychology*, *30*, 552-567.

Steinberg, L., & Thissen, D. (2006). Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychological Methods*, *11*, 402-415.

Swerdzewski, P., Harmes, C., & Finney, S. (2011). Two approaches for identifying low-motivated students in a low-stakes assessment context. *Applied Measurement in Education*, *24*, 162-188.

Van der Linden, W. (2011). Setting time limits on tests. *Applied Psychological Measurement*, *35*, 183-199.

Wang, W., & Chen, H. (2004). The standardized mean difference within the framework of item response theory. *Educational and Psychological Measurement*, *64*, 201-223.

Wiener, J. (2004). Do peer relationships foster behavioral adjustment in children with learning disabilities?*Learning Disability Quarterly*, *27*, 21-30.

Williams, N. J., & Beretvas, S. N. (2006). DIF identification using HGLM for polytomous items. *Applied Psychological Measurement*, *30*, 22-42.

Wolfe, E., & Smith, E. (2007). Instrument development tools and activities for measure validation using Rasch models. In R. Smith & E. Smith (Eds.), *Rasch measurement: Advanced and specialized applications* (pp. 243-290). Maple Grove, MN: JAM Press.

Woodcock, S., & Vialle, W. (2011). Are we exacerbating students' learning disabilities? An investigation of preservice teacher's attributions' of the educational outcomes of students with learning disabilities. *Annals of Dyslexia*, *61*, 223-241.
Zhao, J., Zhang, B., & Yu, G. (2008). Effects of concealable stigma for learning disabilities. *Social Behavior and Personality*, *36*, 1179-1188.
Zurcher, R., & Bryant, D. P. (2001). The validity and comparability of entrance examination scores after accommodations are made for students with LD. *Journal of Learning Disabilities*, *34*, 462-471.