2019-4

Using Measurement Invariance to Study Social Withdrawal in Children With Developmental Language Disorders

Martin Fujiki
Brigham Young University - Provo

Bonnie Brinton
Brigham Young University - Provo

Craig H. Hart
Brigham Young University - Provo, craig_hart@byu.edu

Joseph Olsen
Brigham Young University - Provo

Maille Coombs
Brigham Young University - Provo

LSHSS

Research Article

Using Measurement Invariance to Study Social Withdrawal in Children With Developmental Language Disorders

Martin Fujiki,[a] Bonnie Brinton,[a] Craig H. Hart,[b] Joseph Olsen,[c] and Maille Coombs[a]

Purpose: Teacher ratings were used to compare children with developmental language disorders (DLD) and their typically developing peers on 2 subtypes of social withdrawal (shyness and unsociability). Measurement invariance analysis was utilized to determine if teachers rated the 2 groups using the same underlying construct for each of the rating scale items that have been designed to assess withdrawn behavior.
Method: The Teacher Behavior Rating Scale (TBRS; C. H. Hart & Robinson, 1996) was administered to the teachers of 173 children with DLD and 182 typically developing children (age range: 5;0–12;11 years;months) to compare 2 subtypes of withdrawal, shyness and unsociability. Measurement invariance analysis was used to establish an appropriate basis for comparing the latent group means and other structural parameters, and partial invariance models were used to compare the groups.
Results: For the TBRS, shyness and unsociability were measured by 4 and 5 items, respectively. The measurement invariance analysis indicated that classroom teachers approached 1 item on each of these scales using a different underlying construct when rating the 2 groups of children. Taking this into account, the groups were compared. Teachers rated children with DLD as significantly more withdrawn on both shyness and unsociability in comparison with typically developing children. Age and gender were not significant factors.
Conclusions: When conducting assessments, it should not be assumed that teachers or other raters approach even commonly used rating scale items using the same underlying perception for typically developing children and children with disorders. However, the analysis of invariant items on the TBRS reconfirmed that children with DLD are at risk for social withdrawal as operationalized by assessments of both shyness and unsociability.
Supplemental Material: https://doi.org/10.23641/asha.7789472

Children with developmental language disorders (DLD) often experience a variety of social challenges, including loneliness, isolation, poor peer acceptance, difficulty making friends, and victimization. These difficulties may be evident in the early elementary grades and persist through secondary school (Conti-Ramsden & Botting, 2004, 2008; Durkin & Conti-Ramsden, 2007; Lindsay & Dockrell, 2012; Wadman, Durkin, & Conti-Ramsden, 2008; Yew & O'Kearney, 2013). Many of the poor social outcomes that children with DLD experience involve a lack of social interaction or what is often referred to as *withdrawal*.

Withdrawal is often considered to have negative consequences. It is important to recognize, however, that there are different ways in which one may withdraw from social interactions, and different subtypes of withdrawal can have different social motivations and outcomes (Coplan & Rubin, 2010; Rubin & Coplan, 2004). It is also the case that withdrawal can be difficult to assess because withdrawn behaviors do not lend themselves well to objective testing using traditional standardized tasks. Rating scales are widely used to assess important aspects of social, emotional, cognitive, and language development, including withdrawal. These scales are particularly valuable in that they capitalize on the observations of individuals who know children well and interact with them in a variety of contexts. A basic question in the use of these instruments, however, is whether raters approach items using the same underlying construct for the various population groups that they evaluate. In other words, do raters conceptualize withdrawn behaviors

[a]Department of Communication Disorders, Brigham Young University, Provo, UT
[b]School of Family Life, Brigham Young University, Provo, UT
[c]College of Family, Home, and Social Sciences, Brigham Young University, Provo, UT
Correspondence to Martin Fujiki: martin_fujiki@byu.edu

Editor-in-Chief: Shelley Gray
Editor: Marleen Westerveld
Received May 16, 2018
Revision received September 13, 2018
Accepted October 29, 2018
https://doi.org/10.1044/2018_LSHSS-18-0071

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

in the same way for children with DLD as they do for typically developing children? In practice, this assumption is rarely tested. In the current investigation, we examine teacher ratings of two subtypes of withdrawal (shyness and unsociability) in children with DLD and their typically developing peers to further consider the nature of social withdrawal in children with impaired language. Measurement invariance analysis was used to determine if the items used to measure shyness and unsociability were invariant. Statistical invariance would provide evidence that the raters evaluated children from both groups using the same underlying construct for each item used to measure shyness and unsociability. This validation of item stability would provide a more valid assessment of these behaviors.

Social Withdrawal

For some time, researchers and educators who are interested in optimal child development have been concerned about children who withdraw from social interaction. Historically, however, the study of withdrawal has been hampered by the fact that the term *withdrawal* has been used to refer to a variety of distinct behaviors, including "temperamental and personality traits, motivational and interpersonal processes, and/or observable behaviors" (Coplan & Rubin, 2010, p. 6). Rubin and Asendorpf addressed this concern by beginning with the concept of being alone, or solitude (Rubin, 1982; Rubin & Asendorpf, 1993). Starting from this point, these researchers then considered different ways in which a child might be alone. On one hand, a child could be excluded by his or her peers. On the other, a child might remove himself/herself from social interaction (Rubin & Asendorpf, 1993). Although these behaviors are likely to be intertwined, our current focus is on the latter type, referred to as social withdrawal in this study.

The term *social withdrawal* is an umbrella term that can encompass different ways in which an individual could withdraw from social interaction. Recent investigations have focused on the motivation for different types of social withdrawal and the different outcomes associated with those motivations (Rubin & Coplan, 2004). Two commonly identified subtypes of withdrawal are motivated either by an individual's apprehension of interacting with others or by an individual's preference for solitary activities.

The first subtype of social withdrawal is associated with anxiety or fear of interacting with others that is generally rooted in a temperamental/physiological predisposition for increased fearfulness during social situations (Fox, Henderson, Marshall, Nichols, & Ghera, 2005). This subtype can take several forms. Shyness describes children who would like to interact with others but are fearful of doing so. This fear may stem from temperament or anxiety about being judged by others. Shy children have been characterized as having a high approach–high avoidance conflict when interacting with other people. Other closely related types of withdrawal include "behavioral inhibition," "social reticence," and "anxious solitude" (Coplan & Rubin, 2010).

Although there are theoretical reasons to separate this subtype of social withdrawal into fine-grained categories, reliably identifying the specific behaviors associated with each category may be challenging (Coplan & Rubin, 2010). In addition, these subtypes of behavior are likely to be closely connected. To illustrate, temperamental shyness, when measured separately at home, has been linked to child socioemotional adjustment problems (e.g., peer exclusion) as mediated by observations of anxious solitude at school (Balkaya, Cheah, Yu, Hart, & Sun, 2018). Shy children who approach interactions with anxiety and wariness often display reticence, which is characterized as onlooking, unoccupied behavior (Arbeau, Coplan, & Weeks, 2010; Coplan & Armer, 2007; Nelson, Hart, Wu, Yang, & Roper, 2006). For present purposes, we refer to this general subtype of social withdrawal as *shyness.*

The second subtype of social withdrawal categorizes children who spend time alone because they may enjoy solitude and are not particularly motivated to interact with others. These individuals can be characterized as having a low approach "but not necessarily a high social avoidance motive" (Rubin & Asendorpf, 1993, p. 13). Thus, they do not desire social interaction, but at the same time, they do not actively avoid it either. Although several labels have been used to refer to this subtype of withdrawn behavior (e.g., social disinterest), we use the term *unsociability.* Unsociability is often characterized by solitary-passive withdrawal, which consists of constructive, solitary activities such as quietly playing alone (Rubin, 1982; Rubin, Coplan, Fox, & Calkins, 1995).[1]

There is considerable evidence that shyness is linked to a variety of negative outcomes. These include problems with attachment (Bohlin, Hagekull, & Andersson, 2005), perspective taking (Banergee & Henderson, 2001), peer acceptance and emotion regulation (C. H. Hart et al., 2000; Rubin et al., 1995), and anxiety disorders in later childhood and adolescence (e.g., White et al., 2017). It is not surprising then that children who display high levels of shyness have difficulty interacting with their peers and that their withdrawn behavior is associated with poor performance in school (Arbeau et al., 2010; Rudasill & Rimm-Kaufman, 2009), victimization (Gazelle & Ladd, 2002), anxiety, depression, and poor self-esteem (Coplan et al., 2013; Luster, 2015; Nelson, Rubin, & Fox, 2005).

Unsociability or social disinterest has not been studied to the same extent as shyness, and the available data are more equivocal (e.g., Nelson, Hart, Yang, Wu, & Jin, 2012; Nelson et al., 2005). There is a body of evidence indicating that unsociability is not as problematic as shyness (Coplan et al., 2013; Coplan & Weeks, 2010; Nelson, 2013; Ojanen, Nostrand, Bowker, & Markovic, 2017). For example, Coplan and Weeks (2010) examined unsociability and shyness in first and second graders. Based on ratings from teachers and mothers and child self-ratings,

[1]See Coplan and Ooi (2014) for a discussion of solitary-passive withdrawal questioning whether it is indicative of unsociability.

it was concluded that these two subtypes of social withdrawal remained distinct in middle childhood. Whereas shyness was associated with a variety of social and emotional difficulties, children who were rated as displaying unsociability did not generally differ on socioemotional parameters from their typically developing peers. A closer examination, however, indicated that unsociability may not be entirely without risk. Coplan and Weeks also reported that boys who were rated as high in unsociability were more likely to have peer difficulties (see also Nelson et al., 2005).

Although parents and teachers may not view unsociability as problematic, it has been argued that this subtype of withdrawal merges with shyness as children mature, resulting in negative social consequences in later development (Asendorpf, 1993). Supporting this notion, researchers have reported that unsociability is linked to adverse outcomes in preschoolers (Nelson et al., 2009; Spinrad et al., 2004), elementary school children (Nelson et al., 2005), and adolescents (Wang, Rubin, Laursen, Booth-LaForce, & Rose-Krasnor, 2013).

The relationship between shyness and unsociability may be complex. For example, Henderson, Marshall, Fox, and Rubin (2004) examined physiological, behavioral, and parent report data from groups of 4-year-old children classified as displaying either reticence (indicative of shyness), solitary-passive withdrawal (used as an indication of unsociability), or high social abilities. The groups with reticent and solitary-passive withdrawal engaged in solitary behavior, but only the group exhibiting reticent behaviors performed in a way suggesting that their withdrawn behavior was linked to a "fearful temperament" and difficulty in regulating emotion. Because of the high level of behavioral variability observed in the children with solitary-passive withdrawal, this group was further divided based on resting low heart period. During the evaluation period, the children in the group with solitary-passive withdrawal who also had a high resting heart period behaved more like the children in the group with reticence. The researchers suggested that these children were using solitary play behaviors as a means of coping with their social anxiety. Thus, it may be the case that not all children who are characterized as displaying unsociability have the same motivations for their withdrawn behavior. For some children, solitary behavior may be motivated by social anxiety and may merge with reticence.

Social Withdrawal and Language Disorder

Given the role of language in social behavior, it is not surprising that children and adolescents with DLD demonstrate high levels of withdrawal (Conti-Ramsden & Botting, 2004; Coster, Goorhuis-Brouwer, Nakken, & Spelberg, 1999; Fujiki, Brinton, Morgan, & Hart, 1999; Fujiki, Spackman, Brinton, & Hall, 2004; K. I. Hart, Fujiki, Brinton, & Hart, 2004; Redmond & Rice, 1998; Wadman et al., 2008; Wadman, Durkin, & Conti-Ramdsen, 2011). Often, this withdrawal takes the form of shyness, manifested behaviorally as reticence. For example, teachers consistently rate children with DLD as more reticent than their typically developing peers (Fujiki et al., 1999; K. I. Hart et al., 2004). These teacher ratings have also been supported by direct observations of children with DLD on the playground (Fujiki, Brinton, Isaacson, & Summers, 2001).

Findings regarding unsociability and children with DLD are more ambiguous. Fujiki et al. (1999) found that the same elementary school teachers who reported high levels of reticent behavior in children with DLD did not rate these children as demonstrating higher levels of solitary-passive withdrawal than their typically developing peers. Wadman et al. (2008) found similar results with adolescents. The participants with DLD (ages 16–17 years) produced higher self-ratings of shyness than their typically developing peers. Unsociability ratings, however, did not separate the two groups. Thus, the adolescents with DLD spent more time alone than their peers, but their withdrawn behavior appeared to be motivated by social anxiety rather than lack of desire to interact with others.

K. I. Hart et al. (2004) examined social withdrawal in elementary school–age children between the ages of 6–9 and 10–13 years. Using the same scale as Fujiki et al. (1999), these researchers found that teachers rated children with DLD as exhibiting higher levels of both reticence and solitary-passive withdrawal than their typically developing peers. K. I. Hart et al. also reported that confirmatory factor analysis produced a moderately high latent construct correlation between reticence and solitary-passive withdrawal (.52). This finding provided support for the notion that reticence and solitary-passive withdrawal may merge into a common behavior as children age.

Results from previous research suggest that children with DLD consistently demonstrate high levels of shy behavior. Unsociability has also been observed, but sometimes at no greater levels than that displayed by typically developing children. This would suggest that children with DLD withdraw from social interactions because of social anxiety rather than because of a preference for solitude. However, interpreting the motivations for this withdrawal is complicated by the language limitations these children experience. Rather than being inherently fearful, children with DLD may be motivated to withdraw from language intensive social interactions because they recognize their own linguistic limitations (Redmond & Rice, 1998). However, given the complexities observed in the withdrawn behavior of typically developing children (e.g., Henderson et al., 2004), other interpretations are possible. Although limited language ability would be expected to contribute to social difficulties, language alone does not fully explain withdrawal in the affected children. For example, K. I. Hart et al. (2004) found that severity of language disorder was not related to levels of reticence in children with DLD.

This backdrop of motivational factors for social withdrawal sets the context for the primary goal of this study, which was to extend previous work involving small samples

by using advanced statistical modeling to compare teacher ratings of a large number of children with DLD and their typically developing peers. It was anticipated that this analysis would provide insight into how teachers might construe individual and grouped indicators of social withdrawal that have been observed in children with DLD relative to their typically developing peers. Such findings can inform future studies that examine specific motivations for shyness and unsociable behavioral indicators that may be construed similarly and differently by teachers for these two groups of children.

Measurement Invariance Analysis

Measurement invariance analysis for (ordinal) categorical items (Bowen & Masa, 2015; Pendergast, von der Embse, Kilgus, & Eklund, 2017) was used to establish an appropriate basis for comparing latent means and other structural parameters in ways that can help meet the primary aim of this study.[2] Recently, this statistical tool has been applied to a range of research questions important to educators, speech-language pathologists, and others interested in children from special populations. These applications include considering whether a test will produce stable results across different samples of children (Greenwood, Buzhardt, Walker, McCune, & Howard, 2013), whether individuals with aphasia and their family members and friends are answering items on a questionnaire using the same conceptualization of the items (Doyle et al., 2013), and whether children with and without disorders would complete assessment measures in a similar manner (Kapantzoglou, Thompson, Gray, & Restrepo, 2016).

In the current study, measurement invariance analysis was used to address the question of whether teacher ratings of children with DLD and their typically developing peers were invariant across groups. The current study employed the Teacher Behavior Rating Scale (TBRS; C. H. Hart & Robinson, 1996) to examine shyness and unsociability. A basic question was whether teachers applied the rating scale items in the same way when rating typically developing children and children with DLD. For example, when presented with an item such as "reserved around other children," did raters conceptualize this item in the same way for both groups of children? To examine this issue, measurement invariance analysis was employed to determine if teacher ratings were invariant across the two groups being studied. In addition to group, gender and age were also considered to determine if these variables were related to item functioning.

It was suspected that some items on the TBRS might be non-invariant, reflecting the way that teachers view children who do not communicate well. At the same time, it was expected that the majority of the TBRS items would remain invariant across the groups, supporting partial measurement invariance of teacher ratings of reticence and solitary-passive behavior for children with DLD and their typically developing peers. It was also hypothesized that latent mean comparisons with partially invariant models would confirm previous findings that children with DLD are often at risk for withdrawn behaviors. In accordance with the objectives outlined above, the specific aims of the study were as follows:

1. to evaluate which items on the TBRS used to measure shyness and unsociability were and were not invariant across children with DLD and their typically developing peers, and
2. to use appropriate (perhaps partial) measurement invariance analysis in considering whether children with DLD and typically developing children differ in teacher ratings of shy and unsociable behavior utilizing a much larger comparative sample (355 total children) than has been employed in past research. This larger sample allowed for invariance testing that was not possible in prior studies.

Method

The data used in this study were drawn from samples originally collected in association with projects by Brinton, Fujiki, Hurst, Jones, and Spackman (2015), Brinton, Spackman, Fujiki, and Ricks (2007), Fujiki et al. (1999, 2001, 2004), and Fujiki, Brinton, and Clarke (2002).[3] Although all of these studies involved the TBRS, only Fujiki et al. (1999) and Fujiki et al. (2004) previously examined teacher ratings of social withdrawal. Measurement invariance analysis was not used to analyze the data in either of these studies. Additional participants were also included from a current on-going project. Children were sampled from elementary schools in northern and central Utah. All data were gathered and analyzed using methods approved by a university institutional review board.

Children With DLD

Participants included 98 boys and 75 girls with DLD, for a total sample of 173. The participants ranged in age from 5;0 to 12;11 (years;months) and were categorized by age into older and younger groups. The younger group ranged in age from 5;0 to 9;11 for both boys (M = 7;10, SD = 1;3) and girls (M = 7;9, SD = 1;4). The older group ranged in age from 10;0 to 12;11 for both boys (M = 10;11, SD = 0;7) and girls (M = 10;10, SD = 0;8). All participants were monolingual English speakers.

[2]See https:/www.frontiersin.org/research-topic/1695/measurement-invariance for additional discussion of this topic.

[3]Number of participants from each study: Brinton et al. (2015), n = 44; Brinton et al. (2007), n = 38; Fujiki et al. (1999), n = 50; Fujiki et al. (2004), n = 86; Fujiki et al. (2002), n = 77; Fujiki et al. (2001), n = 16. Other participants, n = 44. The number of participants varied from the number of reported participants in some of the studies because, in these cases, only a subset of the data were available for inclusion.

The participants were initially referred by school speech-language pathologists. All of these children had a formal diagnosis of DLD and were receiving intervention services at the time they were evaluated. All participants were required to have scored more than 1 *SD* below the mean on a standardized language measure. In addition, school district personnel ruled out intellectual disability and psychiatric or emotional conditions, including autism spectrum disorder. All participants had a score from a standardized test indicating nonverbal intelligence within the typical range (see Supplemental Material S1 for specific IQ and language measures). All of the children also passed a hearing screening administered by school personnel.

Typically Developing Children

As the participants in the group with DLD were selected, typically developing children were also identified to provide a control group. The resulting sample consisted of 105 boys and 77 girls, for a total sample of 182 children. The participants were also categorized by age into older and younger groups, with the younger group ranging in age from 5;0 to 9;11 for both boys (M = 7;11, SD = 1;1) and girls (M = 7;10, SD = 1;5). The older group ranged in age from 10;0 to 12;11 for both boys (M = 10;10, SD = 0;8) and girls (M = 10;9, SD = 0;7). All children were monolingual English speakers.

Children were selected on the basis of (a) enrollment in the same classroom as the participant with DLD, (b) same gender as the participant with DLD, (c) within 7 months in age of the participant with DLD, (d) expected school placement with no current or history of enrollment in special services, and (e) teacher judgment of typical development and academic achievement. On a few occasions, the teachers did not complete the TBRS for both the children in the classroom. In other cases, permission was only received to administer the TBRS to one of the two children in a classroom. Thus, the total number of children with DLD was slightly different from the number of typically developing children. Additional descriptive information on both groups of participants is provided in Supplemental Material S1.

Instrumentation

Behavioral Measure

The TBRS (C. H. Hart & Robinson, 1996) is a questionnaire that consists of subscales to measure various types of internalizing and externalizing social behaviors (e.g., impulsivity, aggression, withdrawal, and sociability). The subscales measuring two subtypes of withdrawn behavior, reticence (indicative of shyness) and solitary-passive withdrawal (indicative of unsociability), were used. Teachers were asked to rate each child while "thinking about the child's present behavior relative to others in this age group that you know or have known" (C. H. Hart & Robinson, 1996, p. 1). In most cases, a teacher rated both a child with typical language and a child with DLD in his or her classroom. The reticence subscale was composed of behaviors indicating that the child would like to interact but was afraid to do so. It also contained items indicating that the child was reserved or unengaged around others. The four items on the Reticent subscale of the TBRS were as follows:

- R1: Stares at other children without interacting
- R2: Reserved around other children
- R3: Unoccupied even when there is plenty to do
- R4: Fearful when approaching other children

The second subscale of the TBRS used to measure unsociability focused on solitary-passive withdrawal. This subscale included items describing a child engaged in constructive activities in solitude. It also included behaviors indicating that the child did not seek out interactions with others but preferred playing alone. The five specific items on the Solitary-Passive Withdrawal subscale of the TBRS were as follows:

- SP1: Reads books alone away from peers
- SP2: Does constructive activities alone (e.g., blocks, legos, puzzles)
- SP3: Builds things by self rather than with other children
- SP4: Likes to play alone
- SP5: Plays with toys by self rather than with other children

The psychometric properties of the TBRS for elementary school–age children have been evaluated in depth and are described in detail in Fujiki et al. (1999) and K. I. Hart et al. (2004). To briefly summarize, teachers completed TBRS questionnaires for 382 school-age children ranging in age from 6;4 to 12;6 (M = 8;10, SD = 1;6). After dropping several withdrawal items with relatively little variance, substantial cross-loadings (> .40), or low item total correlations for factors derived in preliminary analyses, a final principal components analysis produced three reliable factors for withdrawal with eigenvalues greater than 1, accounting for 55% of the item variance. The items on two of these factors were selected for use in the current study.[4]

Two versions of the TBRS were used in this study: a full-length version (160 items) used to assess 273 participants and a shortened version (79 items) used to assess 82 participants. The items of interest were randomly dispersed throughout both questionnaires in order to minimize rater bias. In addition, classroom teachers were not informed as to the specific purposes of the research before completing the TBRS.

[4]The third factor, solitary-active withdrawal, involved solitary pretend play and thus was considered distinct from the two forms of social withdrawal studied, in which children removed themselves from the interaction.

Procedure

Administration of TBRS

The TBRS was distributed to all of the participating teachers with the following instructions.

> Think about (child's name)'s present behavior relative to other children you know or have known of the same age. Decide how often the child does the things described. If you are not sure about a particular item, use your best judgment based on your knowledge of the child's personality.

Teachers rated all items using a 3-point scale (0 = *child never displays this behavior,* 1 = *child sometimes displays this behavior,* 2 = *child very often displays this behavior*). All ratings were taken at least 2 months into the school year to give the teachers a chance to get to know the children they were rating.

Analysis

Multiple-Group Confirmatory Factor Analysis

Measurement invariance comparing typically developing children and children with DLD was initially assessed using multiple-group confirmatory factor analysis (MGCFA) for items with three ordered categories. The default delta parameterization in Mplus Version 8 with weighted least squares (WLSMV) estimation was used. In this analysis, children with DLD were designated as the *focal group* and typically developing children were designated as the *reference group*. An overall test of measurement invariance was obtained by comparing the *configural* invariance model estimating group-specific factor loadings and item thresholds, with the *scalar* invariance model equating the factor loadings and the thresholds across the groups. The scale and location of the latent variables for the scalar invariance model were set by fixing the latent variances to 1 and the latent means to 0 in the reference group. The latent variances and means for the *configural* invariance model were set to 1 and 0, respectively, in both groups. Estimating a categorical item factor analysis with WLSMV for polytomous items is equivalent to estimating a probit or normal ogive graded response model in item response theory. In item response theory, measurement non-invariance is also known as differential item functioning (DIF). Non-invariant factor loadings are seen as evidence of what is often called *non-uniform DIF*, and non-invariant thresholds are seen as evidence of the so-called *uniform DIF*.

Modification Indices

With evidence of general measurement non-invariance or DIF, it can be useful and informative to identify which of the items are responsible for the lack of invariance. In order to identify the source of the non-invariance, the modification indices (MIs) for the factor loadings and thresholds of the scalar invariance model were examined. With two groups, an item's factor loading MI is the estimated chi-square model fit improvement (with 1 *df*) that would be expected if group-specific factor loadings were estimated rather than equating the item's loading across the groups. MIs are also provided for group differences in the (second) threshold of each item. When testing measurement invariance with ordinal items having more than two categories using WLSMV in Mplus, the first threshold is equated in order to properly identify the model.

Model Comparisons and Partial Invariance

MIs for the scalar model provide approximations of parameter-level DIF one parameter (loading or threshold) at a time. We also assessed possible item-level DIF by jointly testing the equality of the factor loading and threshold of each item using model comparison chi-square difference tests with 2 *df* obtained with Mplus' DIFFTEST procedure. This provides a general assessment of DIF incorporating both uniform and non-uniform DIF. Because of the possibility of elevated Type I error for such multiple tests, we applied a Bonferroni adjustment by dividing a standard p value of .05 by the number of tests (nine), giving an adjusted critical p value of .0056. When complete invariance is not attained, group-specific measurement parameters may be estimated for some of the items, although these parameters are equated across groups for the rest of the items. If such a partially invariant model fits the data adequately, it can provide a sufficient basis for valid latent mean comparisons.

Latent Mean Differences and Impact

Although measurement invariance testing is generally focused on the factor loadings and thresholds, we also have a strong substantive interest in the relative means and variances of the reticence and solitary-passive latent variables. Because the model was identified and scaled by fixing the reference group means and variances to 0 and 1, respectively, the estimated means and variances of the focal group can be used to evaluate group differences in these parameters using the partially invariant model. Latent mean differences are sometimes referred to as the "impact" of the group membership variable on the true underlying latent trait.

Multiple-Indicator Multiple-Cause Models

MGCFA testing of measurement invariance can readily test for both uniform (threshold) and non-uniform (loading) invariance based on group comparisons of these parameters. When a potential source of measurement non-invariance is a continuous variable or when there are multiple potential sources of measurement invariance, multiple-group invariance assessment can become challenging and multiple-indicator multiple-cause (MIMIC) models are sometimes preferred instead. To examine measurement invariance by age with MGCFA, it would be necessary to categorize the continuous variable age into two or more groups. Group size shrinks as the number of such groups increases, and MGCFA estimation can become unstable and difficult to interpret. Treating age as a continuous quantitative variable rather than a nominal categorical variable can

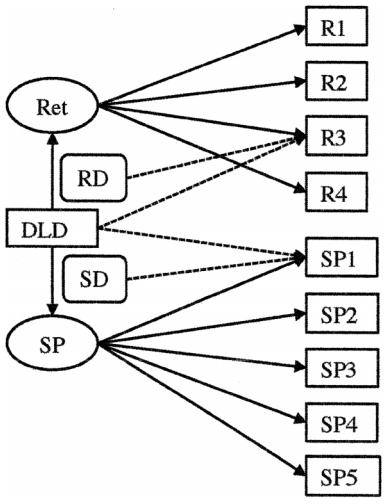
provide a more parsimonious model, simplifying detection of DIF.

In this study, age and gender could be considered as potential additional sources of DIF, along with group membership (typically developing vs. children with DLD). MIMIC models are most commonly and readily used to analyze uniform DIF. To assess uniform DIF due to group membership, for example, we can regress the withdrawn behavior latent variables on a binary group membership variable (DLD) with typically developing children coded 0 and children with DLD coded 1. Whereas the latent variables in MGCFA are often identified and scaled by setting their variances to 1, the referent indicator method where one of the loadings for each factor is fixed to 1 is usually employed to identify and scale the latent variables in MIMIC models.

Uniform and Non-Uniform DIF via MIMIC Models

An illustrative MIMIC model for the current study is presented in Figure 1. Uniform DIF is represented by the regression of the selected indicator on the binary group membership variable, reflecting intercept differences between the focal and reference groups. Non-uniform DIF is represented by the regression of the selected indicator on the associated interaction variable (RD & SD), reflecting differences in the factor loadings between the focal and reference groups.

Figure 1. Multiple-indicator multiple-cause model for subtypes of withdrawn behavior (Ret = Reticence; SP = Solitary-Passive Withdrawal). The usual latent variable path diagramming conventions are used, showing latent variables as circles or ovals, observed variables as squares or rectangles, and factor loadings and regression coefficients represented with directed straight arrows. A nonstandard element in the diagram is the use of rectangles with rounded corners to represent interactions between the binary group membership variable and the two latent variables, with the model parameters for differential item functioning testing represented with dotted arrows. Here we have used RD to indicate the interaction between reticence and the developmental language disorders (DLD) dummy variable and SD to indicate the interaction between solitary-passive and the DLD dummy variable.



We used WLSMV estimation in the MGCFA approach for both uniform and non-uniform DIF. Because MIMIC models for jointly estimating uniform and non-uniform DIF require the use of latent interactions (Woods, 2009), which are not available with WLSMV in Mplus, MIMIC analyses with maximum likelihood estimation were used instead. We note, however, that WLSMV has the advantage of directly providing a variety of standard structural equation modeling fit indices (CFI, RMSEA, etc.) that are not available with latent interaction modeling in Mplus, and WLSMV estimation is generally more computationally efficient for large models.

Results

Descriptive Data

Descriptive data regarding rate of response to the three potential answers for each item in the Reticence and Solitary-Passive Withdrawal subscales are presented in Table 1. Because the items on the TBRS have only three response options, the descriptive information is presented in terms of the response proportions rather than item means and standard deviations. The following analyses were performed to address the research questions.

Measurement Invariance via MGCFA

The multigroup configural model provided a reasonable fit to the data, $\chi^2(56) = 142.479$, $p < .0001$, CFI = .963, TLI = .949, RMSEA = .099, WRMR = 1.423, and fit significantly better than the corresponding scalar invariance model, $\chi^2_{diff}(14) = 49.065$, $p = .0001$, indicating measurement non-invariance for one or more of the items. The loadings and thresholds for the third reticence item (R3, unoccupied even when there is plenty to do) and the first solitary-passive item (SP1, reads books alone away from peers) had the largest MIs (see Table 2) and were likely candidates as sources of the non-invariance. With the Bonferroni-adjusted critical p value, only items R3 and SP1 displayed strong evidence of differential functioning. A partially invariant model allowing group-specific loadings and thresholds for only items R3 and SP1 fit the data as well as the original configural model, $\chi^2_{diff}(10) = 15.168$, $p = .1261$, indicating that these two items were the main source of measurement non-invariance for these data.

Partial Invariance via MGCFA

The factor loadings and thresholds of the items for the two groups from the partially invariant model are given in Table 3. In this case, seven of the nine items function similarly for typically developing children and children with DLD. However, items R3 and SP1 exhibited quite discrepant item parameters across the groups. For item R3, the focal group factor loading and (second) threshold parameters are much larger than the corresponding reference group parameters. Based on the factor loading or discrimination parameter, item R3 is more strongly related to the latent trait in the focal group than in the

Table 1. Proportions of children with developmental language disorders (DLD) and typically developing children rated in each category on reticence and solitary-passive items.

Subscale	Children with DLD			Typically developing children		
	Never	Sometimes	Very often	Never	Sometimes	Very often
Reticence						
R1: Stares at other children without interacting	.593	.262	.145	.906	.088	.006
R2: Reserved around other children	.382	.405	.214	.659	.297	.044
R3: Unoccupied even when there is plenty to do	.407	.343	.250	.823	.166	.011
R4: Fearful when approaching other children	.581	.355	.064	.862	.133	.006
Solitary-Passive Withdrawal						
SP1: Reads books alone away from peers	.573	.351	.076	.506	.428	.067
SP2: Does constructive activities alone (e.g., blocks, legos, puzzles)	.453	.476	.071	.572	.350	.078
SP3: Builds things by self rather than with other children	.491	.404	.105	.722	.261	.017
SP4: Likes to play alone	.453	.465	.082	.756	.233	.011
SP5: Plays with toys by self rather than with other children	.535	.365	.100	.855	.134	.011

reference group and also shows a larger threshold. However, the standard errors for both of these parameters are also very large. Such large parameter differences and standard errors are evidence of potential item–model misfit and instability in parameter estimation. The focal group SP1 factor loadings and thresholds are much smaller than the corresponding reference group parameters. SP1 "Reads books alone away from peers" does not appear to measure the solitary-passive latent variable well in the focal group. Rather than discarding the item, we have chosen to interpret the possible meaning of the non-invariance and retain the item under conditions of partial invariance (Hu & Cheung, 2008). This requires estimating separate (loading and threshold) measurement parameters for the item when comparing the solitary-passive latent means of the children with DLD and typically developing children.

Latent Mean Differences via MGCFA

The estimated latent reticence mean (with its standard error) in the focal group was 1.060 (0.165), $Z = 6.426$, $p < .0001$, indicating that children with DLD were judged to be substantially more reticent than their typically developing peers. Because the latent variables were standardized to the mean and variance of the reference group, the focal group latent mean value of 1.060 is analogous to Glass's delta (Δ), indicating a difference of a little more than a standard deviation on the latent variable. Similarly, the estimated focal group solitary-passive latent mean (with standard error) was 0.798 (0.136), $Z = 5.782$, $p < .0001$, indicating that children with DLD were about eight tenths of a standard deviation higher on solitary-passive behavior than typically developing children. The estimated focal group variances were 1.259 and 1.101 for the reticence and solitary-passive latent variables, respectively, although neither of these differed significantly from the expected value of 1.0. The estimated correlations between the two latent variables were also very similar for the two groups (.640 in the reference group and .686 in the focal group). Children with DLD exhibited substantially more reticence and solitary-passive behavior than their typically developing peers. However, the variances of these two constructs and the correlation between them were very similar across the groups.

Table 2. Item factor loading and threshold modification indices, and item differential functioning chi-square tests ($df = 2$) for the multiple-group scalar invariance model.

Subscale	Modification indices		Differential Functioning	
	Loadings	Thresholds	χ^2	*p*
Reticence				
R1	0.715	0.821	0.857	.6513
R2	0.527	1.287	1.225	.5420
R3	16.532	15.233	***11.754***	***.0028***
R4	3.113	2.140	3.283	.1936
Solitary-Passive Withdrawal				
SP1	9.034	21.896	***17.771***	***.0001***
SP2	5.720	3.418	5.262	.0720
SP3	0.114	0.011	0.267	.8751
SP4	7.069	7.396	9.789	.0075
SP5	5.233	6.764	8.451	.0146

Note. Items with statistically significant (Bonferroni adjusted critical $p = .0056$) chi-square tests are in bold italics.

Table 3. Factor loadings and thresholds with standard errors for reticence and solitary-passive behavior among typically developing children and children with developmental language disorders (DLD) from the partially invariant model.

	Factor loadings				Threshold 1		Threshold 2			
	Typical		DLD		Typical & DLD		Typical		DLD	
Subscale	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*	Est	*SE*
Reticence										
R1	0.919	0.058	0.919	0.058	1.312	0.118	2.481	0.297	2.481	0.297
R2	0.715	0.060	0.715	0.060	0.413	0.093	1.673	0.149	1.673	0.149
R3	***0.586***	***0.090***	***1.662***	***0.787***	0.928	0.109	***2.289***	***0.267***	***4.150***	***2.166***
R4	0.809	0.055	0.809	0.055	1.098	0.109	2.671	0.339	2.671	0.339
Solitary-Passive Withdrawal										
SP1	***0.599***	***0.076***	***0.012***	***0.083***	0.014	0.093	***1.501***	***0.144***	***0.041***	***0.274***
SP2	0.574	0.056	0.574	0.056	0.284	0.076	1.573	0.126	1.573	0.126
SP3	0.925	0.034	0.925	0.034	0.648	0.094	2.112	0.221	2.112	0.221
SP4	0.822	0.045	0.822	0.045	0.610	0.090	2.073	0.201	2.073	0.201
SP5	0.976	0.035	0.976	0.035	0.962	0.106	2.238	0.224	2.238	0.224

Note. The first thresholds of each item are equated across groups in order to identify the model. Group varying parameters are in bold italics.

Summary—MGCFA

In a well-fitting, partially invariant model, the latent variable group mean differences are based primarily on the contributions of the invariant items. Three of the four reticence items and four of the five solitary-passive behavior items were found to be invariant, therefore providing an appropriate basis for latent mean comparisons. Such comparisons correct for measurement error and are generally more accurate and less biased than comparisons based on unit-weighted composite scores. Although all of the items appeared to function adequately for typically developing children, one reticence item and one solitary-passive behavior item did not function consistently for children with DLD.

Measurement Invariance via MIMIC Models

As with MGCFA, it is possible to examine overall measurement non-invariance by comparing the traditional MIMIC model with no DIF parameters to a MIMIC model with uniform and non-uniform DIF parameters for all but the designated referent indicators, $\chi^2_{diff}(14) = 52.700$, $p < .0001$, using a likelihood ratio model comparison chi-square difference test. In the latter model, DIF parameters are not estimated for the referent indicators in order to identify the model. The final partial invariance model with DIF effects for R3 and SP1 fit considerably better than the scalar invariance model, $\chi^2_{diff}(4) = 33.484$, $p < .0001$, and nearly as well as the configural model, $\chi^2_{diff}(10) = 19.216$, $p = .0376$, without invariance constraints. These overall invariance tests with ML MIMIC models largely confirm the WLSMV MGCFA results reported earlier. DIF due to group membership appears generally localized to items R3 and SP1, again supporting a partial invariance model with group-specific measurement parameters for these two items.

Partial Invariance and Group Membership Impact via MIMIC

Rather than testing only the individual MIMIC model parameters, we examine DIF for an item more generally by jointly testing both uniform and non-uniform DIF using Wald chi-square tests with 2 *df* for each item (see Table 4). As with the corresponding MGCFA tests in Table 2, strong DLD DIF effects were found for items R3 and SP1. In addition to these DLD DIF effects, DLD had a strong impact on both reticence ($Z = 4.919$, $p < .001$) and solitary-passive behavior ($Z = 4.271$, $p < .001$) for the parital invariance model. As with the MGCFA models, items R3 and SP1 exhibit substantial DIF, and there remain strong effects of group membership on reticence and solitary-passive behavior under conditions of partial measurement invariance.

Invariance and Impact by Age and Gender via MIMIC

In this study, children's age ranged from just over 5 years to just over 12 years, with a mean and median of about 9 years of age. We centered age at 9 years to facilitate interpretation of the results. DIF by demographic variables such as age and gender is often examined to test for potential item bias by these factors. None of the reticence or solitary-passive items exhibited age or gender DIF when applying the Bonferroni-adjusted p value of .0056 (see Table 4). In addition, age had no impact on reticence ($Z = 1.331$, $p = .183$) or solitary-passive behavior ($Z = -1.677$, $p = .094$). Gender was also unrelated to reticence ($Z = -.658$, $p = .510$) and had only a small effect on solitary-passive behavior ($Z = 2.031$, $p = .042$).

Summary for MGCFA and MIMIC

There was little evidence of gender or age DIF for the items, but both the MGCFA and MIMIC approaches confirmed the presence of group membership (typically

Table 4. Statistical tests for item factor loading and threshold differences, and Item differential functioning Wald chi-square tests (*df* = 2) for the multiple-indicator multiple-cause invariance model.

Subscale	DLD		Age		Gender	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
Reticence						
R1	0.450	.7985	7.225	.0270	1.512	.4695
R2	1.696	.4283	1.503	.4717	1.491	.4745
R3	***12.457***	***.0020***	2.428	.2970	3.923	.1406
R4	2.759	.2517	.513	.7738	3.154	.2066
Solitary-Passive Withdrawal						
SP1	18.614	***< .0001***	2.912	.2332	2.292	.3179
SP2	6.493	.0389	2.451	.2936	0.577	.7494
SP3	1.452	.4838	0.561	.7554	3.920	.1409
SP4	3.108	.2114	0.636	.7276	2.018	.3646
SP5	8.113	.0173	0.042	.9792	0.738	.6914

Note. Items with statistically significant (Bonferroni adjusted critical *p* = .0056) chi-square tests are in bold italics. DLD = developmental language disorders.

developing children vs. children with DLD) DIF for items R3 and SP1. As expected, we replicated the group membership (typically developing children vs. children with DLD) measurement invariance and DIF results with two different model specification and estimation approaches (MGCFA and MIMIC).

All of the items appeared to function appropriately regardless of age or gender, but two of the items functioned differently for children with DLD than for typically developing children. There was little evidence of age or gender impact on reticent or solitary-passive behaviors, but both the MGCFA and MIMIC approaches found substantial impact of group membership on latent reticence and solitary-passive behavior. In this study, older children did not differ from younger children and girls did not generally differ from boys in levels of reticent and solitary-passive behavior, but children with DLD were dramatically different from typically developing children in these withdrawn behaviors.

Discussion

The purpose of this study was to examine two subtypes of social withdrawal in a large sample of children with DLD. As in most previous research, withdrawal was assessed using a rating scale format completed by observers familiar with the children. In this case, the observers were classroom teachers. To enhance the validity of teacher observations, measurement invariance analysis was used to determine if teachers were evaluating the withdrawn behaviors of both groups of children in a similar manner. Group differences were examined after establishing the invariance of items on the measurement scale.

Measurement Invariance Analysis

Examination of the measurement invariance analysis revealed that one item on each subscale was non-invariant. For the reticence subscale, the non-invariant item was R3, "unoccupied even when there is plenty to do." Teachers saw this as more strongly indicative of reticence among children with DLD than among typically developing children. Although the teachers reported that children with DLD demonstrated more of this behavior than did their typically developing peers, those teachers evidently did not think about or conceptualize the behavior in the same way for the two groups. In other words, "unoccupied even when there is plenty to do" meant something different for children with DLD than it did for typically developing children. For example, it may be the case that teachers perceived a typically developing child who was not occupied as wasting time. That same teacher might have perceived a child with DLD as overwhelmed or unable to participate in activity. Future studies can ascertain whether this is the case.

It is interesting to note that the invariant items R1 (stares at other children without interacting), R2 (reserved around other children), and R4 (fearful when approaching other children) focused specifically on being anxious or fearful when interacting with other children. In contrast, the non-invariant item R3, "unoccupied when there is plenty to do," did not refer directly to the child's interaction with other children. We might speculate that for the children with DLD, teachers viewed unoccupied behavior as reflecting something other than shyness. Perhaps teachers considered unoccupied behavior in these children as reflecting poor emotion regulation rather than social anxiety or fearfulness. Although emotion regulation is often associated with being able to calm or control emotions, regulation can also include the need to elevate emotions. For example, a child may need to "gear up"—or intensify—emotions to perform laborious tasks, such as doing household chores or completing difficult homework. Children who cannot gear up their emotions to perform such tasks may appear to be off-task and unoccupied. It is notable that links have been demonstrated between shyness and emotion regulation

(C. H. Hart et al., 2000; Rubin et al., 1995) and that children with DLD have poorer emotion regulation skills than typically developing children (Fujiki et al., 2002). Teachers may have conceptualized item R3, "unoccupied even when there is plenty to do," as indicative of poor emotion regulation in children with DLD.

The one non-invariant solitary-passive withdrawal item was SP1, "reads books alone away from peers." Teachers also viewed this behavior differently for children with DLD than they did for typically developing children. For typically developing children, this item appeared to reflect the same socially passive latent construct as the other four subscale items, but it did not similarly reflect underlying social passivity for children with DLD. We could speculate that teachers viewed the item in light of the literacy skills that children with DLD demonstrated in the classroom. It has long been recognized that children with DLD lag well behind their typically developing peers in literacy learning (Catts, Fey, Tomblin, & Zhang, 2002). Although not formally assessed, it is likely that most of the children with DLD in the current sample had difficulties with reading, as it is unlikely that they would be referred for language intervention if their academic performance were unremarkable.

The teachers may have had different expectations for the children with DLD than they did for their typically developing peers. Perhaps, the teachers perceived that the children with DLD read alone because they were slower at the task and were completing unfinished work, whereas typically developing children read alone for pleasure or information. Whatever the explanation, teachers viewed "reading alone" as a different behavior for each group. This finding can inform future studies that examine the underlying basis for this difference in teacher perceptions of nonsocial behavior between these two groups. As measures are developed and applied in various research settings, we agree with Greiff and Scherer (2018) that evidence against invariance can be a theoretically informative and substantively valuable finding.

The non-invariance of these two items is consequential for the current study and may also have important implications for the use of rating scales with children with DLD in general. The TBRS items used to measure shyness (as indicated by reticent behavior) and unsociability (as indicated by solitary-passive withdrawal) are widely accepted as being central to the two subtypes of social withdrawal. Therefore, it might not be anticipated that raters would utilize specific items differently for the two groups under study. However, differences were revealed through the measurement invariance analysis. Generalizing these findings to other rating scales, measurement invariance is a potential concern when these measures are used to evaluate children with DLD or other disabilities. Rating scales are commonly used to assess a range of social and emotional behaviors (Merrell, 2003). However, it is rarely the case that individual items on rating scales are examined to determine if observers (teacher, caretakers, etc.) view the targeted behaviors consistently across different populations. If items on rating scales are non-invariant, established norms or expectations based largely on the performance of typically developing children may not be completely appropriate for children with disabilities (e.g., DLD, autism spectrum disorders, intellectual disability). This concern is an important consideration for clinical practice and should be the topic of future study.

Group Differences on the Subtypes of Social Withdrawal

Even after considering the question of measurement invariance, there was sufficient item invariance on both subscales to make valid group comparisons. The largest group differences were found on the Reticence subscale, but the two groups of children also differed significantly on the Solitary-Passive Withdrawal subscale. Using a more stringent statistical modeling analysis with a larger sample, the current study provides added support for previous research documenting differences in social withdrawal between children with DLD (e.g., Fujiki et al., 1999) and their typically developing peers.

These findings may also offer additional insight into the motivation underlying the reported withdrawal in children with DLD that can be explored in future research. In the current study, the teacher reports presented a profile of children with DLD who did not easily participate in the classroom, stared at other children without joining them, and were reserved and fearful of interaction. In addition, they often played or worked alone in less structured activities. The reported reticent or shy behaviors suggest that the children with DLD did not work and play alone simply because they enjoyed solitude or preferred to do things by themselves. It is far more likely that these children were often reduced to doing things alone because they were either too shy to enter into interaction with their peers or they were unable to gear themselves up to engage in activity within the classroom. The combination of shy and unsociable behaviors suggests that, for these school-age children with DLD, solitary-passive withdrawal was merging with reticence, reminiscent of the previously discussed work of Henderson et al. (2004).

Regardless of the motivating factors that underpinned the shyness and unsociability observed in the children with DLD, the current study emphasizes the vulnerability of many of these children. Children with DLD particularly need academic support and experience in social interaction with their peers. The shyness and unsociability reported could limit their full inclusion in school activities and their access to important learning contexts. The social withdrawal demonstrated by children with DLD may sabotage their social well-being and academic progress in the classroom. Particular interventions designed to bolster positive peer interactions and prosocial behaviors warrant further attention with regard to their effectiveness to head off the isolation and rejection that children with DLD often experience (Adams, Lockton, Gaile, Earl, & Freed, 2012; Fujiki & Brinton, 2017).

Limitations

In this work, solitary-passive withdrawal was used as a measure of unsociability. Although this is a generally accepted practice (Rubin, 1982; Rubin et al., 1995), some researchers have questioned whether solitary-passive behavior is a sufficient measure of unsociability (Coplan & Armer, 2007). It is recognized that it may be the case that a child who spends considerable time alone may still be able to interact appropriately when called upon to do so. As discussed above, however, the fact that the children with DLD were also rated high in shy behaviors would suggest that this was not common for the children studied. This issue will warrant closer attention in future studies.

It is not clear how some participant characteristics influenced teacher perceptions of social withdrawal. As noted in Supplemental Material S1, data on race were not reported for all participants. Although the children all attended schools in similar suburban districts, specific data on socioeconomic status were incomplete. Further study examining a wider population could clarify this issue.

As an additional consideration, the current study relied on teacher reports of shy and unsociable behaviors rather than on actual observation of behaviors. Thus, the results are subject to the limitations of rating scale measures beyond consideration of measurement invariance. Factors such as teacher characteristics, biases, dispositions, and experience may have influenced ratings (cf. Nelson, Hart, & Evans, 2008). In addition, further research is warranted to determine how teachers' perceptions of withdrawal both reflect and affect children's actual behavior. Future study could probe the underlying motivations for the withdrawn behavior and the interplay between type and severity of language deficits and withdrawal.

Conclusions

Rating scales provide an efficient means to evaluate behaviors such as social withdrawal in children. These scales can take advantage of both the expertise and observations of teachers who spend a great deal of time with children and understand the social context in which those children interact. The results of this study, however, underscore the importance of determining if raters are viewing different populations of children through the same lens. There may be two ways to address this issue. First, rating scales should be designed to assure that individual items are invariant across typically developing and clinical populations. Second, following the completion of a rating scale, it may be helpful for practitioners to follow up with teachers to ask for reflections and behavior examples relating to rating scale items. This may provide insight into individual cases.

The results of this study underscore the fact that many children with DLD demonstrate social withdrawal at school. It does not appear to be the case that these children prefer to be alone or choose to be alone. Rather, they demonstrate shyness associated with fear and anxiety in interacting with their peers. It is critical to recognize this socioemotional component associated with DLD. Most important, it will be important to learn how best to design curricula and implement interventions to support children with DLD to become full participants within their academic and social worlds.

Acknowledgments

This work was funded, in part, by the David O. McKay School of Education and a David O. McKay Fellowship from Brigham Young University awarded to Bonnie Brinton and Martin Fujiki. We are grateful to Miranda Miller for her assistance in data organization.

References

Adams, C., Lockton, E., Gaile, J., Earl, G., & Freed, J. (2012). Implementation of a manualized communication intervention for school-aged children with pragmatic and social communication needs in a randomized controlled trial: The social communication intervention project. *International Journal of Language & Communication Disorders, 47,* 245–256. https://doi.org/10.1111/j.1460-6984.2012.00147.x

Arbeau, K., Coplan, R. J., & Weeks, M. (2010). Shyness, teacher-child relationships, and socioemotional adjustment in Grade 1. *International Journal of Behavioral Development, 34,* 259–269. https://doi.org/10.1177/0165025409350959

Asendorpf, J. B. (1993). Abnormal shyness in children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 34,* 1069–1081. https://doi.org/10.1111/j.1469-7610.1993.tb01774.x

Balkaya, M., Cheah, C. S. L., Yu, J., Hart, C. H., & Sun, S. (2018). Maternal encouragement of modest behavior, temperamental shyness, and anxious withdrawal linkages to Chinese American children's social adjustment: A moderated mediation analysis. *Social Development, 27,* 876–890. https://doi.org/10.1111/sode.12295

Banergee, R., & Henderson, L. (2001). Social–cognitive factors in childhood social anxiety: A preliminary investigation. *Social Development, 10,* 558–572. https://doi.org/10.1111/ 1467-9507.00180

Bohlin, G., Hagekull, B., & Andersson, K. (2005). Behavioral inhibition as a precursor of peer social competence in early school age: The interplay with attachment and nonparental care. *Merrill-Palmer Quarterly, 51,* 1–19. https://doi.org/10.1353/mpq.2005.0001

Bowen, N. K., & Masa, R. D. (2015). Conducting measurement invariance tests with ordinal data: A guide for social work researchers. *Journal of the Society for Social Work and Research, 6,* 229–249. https://doi.org/10.1086/681607

Brinton, B., Fujiki, M., Hurst, N. Q., Jones, E., & Spackman, M. P. (2015). The ability of children with language impairment to dissemble emotion in hypothetical scenarios and natural situations. *Language, Speech, and Hearing Services in Schools, 46,* 325–336. https://doi.org/10.1044/2015_LSHSS-14-0096

Brinton, B., Spackman, M. P., Fujiki, M., & Ricks, J. (2007). What should Chris say? The ability of children with specific language impairment to recognize the need to dissemble emotions in social situations. *Journal of Speech, Language, and Hearing Research, 50,* 798–811. https://doi.org/10.1044/1092-4388(2007/055)

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45,* 1142–1157. https://doi.org/10.1044/1092-4388(2002/093)

Conti-Ramsden, G., & Botting, N. (2004). Social difficulties and victimization in children with SDLD at 11 years of age. *Journal of Speech, Language, and Hearing Research, 47,* 145–161. https://doi.org/10.1044/1092-4388(2004/013)

Conti-Ramsden, G., & Botting, N. (2008). Emotional health in adolescents with and without a history of specific language impairment (SLI). *Journal of Child Psychology and Psychiatry and Allied Disciplines, 49,* 516–525. https://doi.org/10.1111/j.1469-7610.2007.01858.x

Coplan, R. J., & Armer, M. (2007). A "multitude" of solitude: A closer look at social withdrawal and nonsocial play in early childhood. *Child Development Perspectives, 1*(1), 26–32. https://doi.org/10.1111/j.1750-8606.2007.00006.x

Coplan, R. J., & Ooi, L. L. (2014). The causes and consequences of playing alone in childhood. In R. Coplan & J. C. Bowker (Eds.), *The handbook of solitude: Psychological perspectives on social isolation, social withdrawal, and the experience of being alone* (pp. 111–128). New York, NY: Wiley-Blackwell.

Coplan, R. J., Rose-Krasnor, L., Weeks, M., Kingsbury, A., Kingsbury, M., & Bullock, A. (2013). Alone is a crowd: Social motivations, social withdrawal, and socioemotional functioning in later childhood. *Developmental Psychology, 49,* 861–875. https://doi.org/10.1037/a0028861

Coplan, R. J., & Rubin, K. H. (2010). Social withdrawal and shyness in childhood: History, theories, definitions, and assessments. In *The development of shyness and social withdrawal* (pp. 3–20). New York, NY: Guilford.

Coplan, R. J., & Weeks, M. (2010). Unsociability in middle childhood: Conceptualization, assessment, and associations with socioemotional functioning. *Merrill-Palmer Quarterly, 56,* 105–130.

Coster, F. W., Goorhuis-Brouwer, S. M., Nakken, H., & Spelberg, H. C. L. (1999). Specific language impairments and behavioural problems. *Folia Phoniatrica et Logopaedica, 51,* 99–107. https://doi.org/10.1159/000021484

Doyle, P. J., Hula, W. D., Austermann Hula, S. N., Stone, C. A., Wambaugh, J. L., Ross, K. B., & Schumacher, J. G. (2013). Self- and surrogate-reported communication functioning in aphasia. *Quality of Life Research, 22,* 957–967. https://doi.org/10.1007/s11136-012-0224-5

Durkin, K., & Conti-Ramsden, G. (2007). Language, social behavior, and the quality of friendships in adolescents with and without a history of specific language impairment. *Child Development, 78,* 1441–1457. https://doi.org/10.1111/j.1467-8624.2007.01076.x

Fox, N. A., Henderson, H. A., Marshall, P. J., Nichols, K. E., & Ghera, M. M. (2005). Behavioral inhibition: Linking biology and behavior within a developmental framework. *Annual Review of Psychology, 56,* 235–262. https://doi.org/10.1146/annurev.psych.55.090902.141532

Fujiki, M., & Brinton, B. (2017). Social communication intervention for children with language impairment. In R. J. McCauley, M. E. Fey, & R. B. Gillam (Eds.), *Treatment of language disorders* (2nd ed., pp. 421–449). Baltimore, MD: Brookes.

Fujiki, M., Brinton, B., & Clarke, D. (2002). Emotion regulation in children with specific language impairment. *Language, Speech, and Hearing Services in Schools, 33,* 102–111. https://doi.org/10.1044/0161-1461(2002/008)

Fujiki, M., Brinton, B., Isaacson, T., & Summers, C. (2001). Social behaviors of children with language impairment on the playground: A pilot study. *Language, Speech, and Hearing Services in Schools, 32,* 101–113. http://doi.org/10.1044/0161-1461(2001/008)

Fujiki, M., Brinton, B., Morgan, M., & Hart, C. H. (1999). Withdrawn and sociable behavior of children with language impairment. *Language, Speech, and Hearing Services in Schools, 30,* 183–195. https://doi.org/10.1044/0161-1461.3002.183

Fujiki, M., Spackman, M. P., Brinton, B., & Hall, A. (2004). The relationship of language and emotion regulation skills to reticence in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 47,* 637–646. https://doi.org/10.1044/1092-4388(2004/049)

Gazelle, H., & Ladd, G. (2002). Interventions for children victimized by peers. In P. A. Schewe (Ed.), *Preventing violence in relationships: Interventions across the life span* (pp. 55–78). Washington, DC: American Psychological Association.

Greenwood, C. R., Buzhardt, J., Walker, D., McCune, L., & Howard, W. (2013). Advancing the construct validity of the early communication indicator (ECI) for infants and toddlers: Equivalence of growth trajectories across two early head start samples. *Early Childhood Research Quarterly, 28,* 743–758. https://doi.org/10.1016/j.ecresq.2013.07.002

Greiff, S., & Scherer, R. (2018). Still comparing apples with oranges? Some thoughts on the principles and practices of measurement invariance testing. *European Journal of Psychological Assessment, 34,* 141–144. https://doi.org/10.1027/1015-5759/a000487

Hart, C. H., & Robinson, C. C. (1996). *Teacher Behavior Rating Scale*. Provo, UT: Brigham Young University.

Hart, C. H., Yang, C., Nelson, L. J., Robinson, C. C., Olsen, J. A., Nelson, D. A., . . . Wu, P. (2000). Peer acceptance in early childhood and subtypes of socially withdrawn behavior in China, Russia, and the United States. *International Journal of Behavioral Development, 24,* 73–81. http://doi.org/10.1080/016502500383494

Hart, K. I., Fujiki, M., Brinton, B., & Hart, C. H. (2004). The relationship between social behavior and severity of language impairment. *Journal of Speech, Language, and Hearing Research, 47,* 647–662. https://doi.org/10.1044/1092-4388(2004/050)

Henderson, H. A., Marshall, P. J., Fox, N. A., & Rubin, K. H. (2004). Psychophysiological and behavioral evidence for varying forms and functions of nonsocial behavior in preschoolers. *Child Development, 75,* 251–263. https://doi.org/10.1111/j.1467-8624.2004.00667.x

Hu, C., & Cheung, G. W. (2008). Eight general questions on measurement equivalence / invariance that bother me most. *Academy of Management Annual Meeting Proceedings, 2008*(1), 1–6.

Kapantzoglou, M., Thompson, M. S., Gray, S., & Restrepo, M. A. (2016). Assessing measurement invariance for Spanish sentence repetition and morphology elicitation tasks. *Journal of Speech, Language, and Hearing Research, 60,* 1–13. https://doi.org/10.1044/2015_JSLHR-L-14-0319

Lindsay, G., & Dockrell, J. E. (2012). Longitudinal patterns of behavioral, emotional, and social difficulties and self-concepts in adolescents with a history of specific language impairment. *Language, Speech, and Hearing Services in Schools, 43,* 445–460. https://doi.org/10.1044/0161-1461(2012/11-0069)

Luster, S. S. (2015). *Social withdrawal and internalizing problems in emerging adulthood: Does parenting matter*? (PhD dissertation), Brigham Young University, Provo, UT.

Merrell, K. W. (2003). *Behavioral, social, and emotional assessment of children and adolescents* (2nd ed.). Mahwah, NJ: Erlbaum.

Nelson, L. J. (2013). Going it alone: Comparing subtypes of withdrawal on indices of adjustment and maladjustment in emerging adulthood. *Social Development, 22,* 522–538. https://doi.org/10.1111/j.1467-9507.2012.00671.x

Nelson, L. J., Hart, C. H., & Evans, C. A. (2008). Solitary-functional play and solitary-pretend play: Another look at the construct of solitary-active behavior using playground observations. *Social Development, 17,* 812–831. https://doi.org/10.1111/j.1467-9507.2008.00470.x

Nelson, L. J., Hart, C. H., Evans, C. A., Coplan, R. J., Olsen Roper, S., & Robinson, C. C. (2009). Behavioral and relational correlates of low self-perceived competence in young children. *Early Childhood Research Quarterly, 24,* 350–361. https://doi.org/10.1016/j.ecresq.2009.06.002

Nelson, L. J., Hart, C. H., Wu, B., Yang, C., & Roper, S. O. (2006). Relations between Chinese mothers' parenting practices and social withdrawal in early childhood. *International Journal of Behavioral Development, 30,* 261–271. https://doi.org/10.1177/0165024060667 61

Nelson, L. J., Hart, C. H., Yang, C., Wu, P., & Jin, S. (2012). An examination of the behavioral correlates of subtypes of nonsocial play among Chinese preschoolers. *Merrill-Palmer Quarterly, 58,* 77–109. https://doi.org.erl.lib.byu.edu/10.1353/mpq.2012.0006

Nelson, L. J., Rubin, K. H., & Fox, N. (2005). Social withdrawal, observed peer acceptance, and the development of self-perceptions in children ages 4–7. *Early Childhood Research Quarterly, 20,* 185–200. https://doi.org/10.1016/j.ecresq.2005.04.007

Ojanen, T., Nostrand, D. F.-V., Bowker, J. C., & Markovic, A. (2017). Examining the distinctiveness and the socio-emotional correlates of anxious-withdrawal and unsociability during early adolescence in Finland. *Journal of Early Adolescence, 37,* 433–446. https://doi.org/10.1177/0272431615609156

Pendergast, L. L., von der Embse, N., Kilgus, S. P., & Eklund, K. R. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology, 60,* 65–82. https://doi.org/10.1016/j.jsp.2016.11.002

Redmond, S. M., & Rice, M. L. (1998). The socioemotional behaviors of children with SLI: Social adaptation or social deviance. *Journal of Speech, Language, and Hearing Research, 41,* 688–700. https://doi.org/10.1044/jslhr.4103.688

Rubin, K. H. (1982). Nonsocial play in preschoolers: Necessary evil. *Child Development, 53,* 651–657. https://doi.org/10.2307/1129376

Rubin, K. H., & Asendorpf, J. B. (1993). Social withdrawal, inhibition, and shyness in childhood: Conceptual and definitional issues. In K. H. Rubin & J. B. Asendorpf (Eds.), *Social withdrawal, inhibition, and shyness in childhood* (pp. 3–28). Hillsdale, NJ: Erlbaum.

Rubin, K. H., & Coplan, R. J. (2004). Paying attention to and not neglecting social withdrawal and social isolation. *Merrill-Palmer Quarterly, 50,* 506–534. https://doi.org/10.1353/mpq.2004.0036

Rubin, K. H., Coplan, R. J., Fox, N., & Calkins, S. D. (1995). Emotionality, emotion regulation, and preschoolers' social adaptation. *Development and Psychopathology, 7,* 49–62. https://doi.org/10.1017/S0954579400006337

Rudasill, K. M., & Rimm-Kaufman, S. E. (2009). Teacher–child relationship quality: The role of child temperament and teacher–child interactions. *Early Childhood Research Quarterly, 24,* 107–120. https://doi.org/10.1016/j.ecresq.2008.12.003

Spinrad, T. L., Eisenberg, N., Harris, E., Hanish, L., Fabes, R. A., Kupanoff, K., . . . Holmes, J. (2004). The relation of children's everyday nonsocial peer play behavior to their emotionality, regulation, and social functioning. *Developmental Psychology, 40,* 67–80. https://doi.org/10.1037/0012-1649.40.1.67

Wadman, R., Durkin, K., & Conti-Ramsden, G. (2008). Self-esteem, shyness, and sociability in adolescents with specific language impairment. *Journal of Speech, Language, and Hearing Research, 51,* 938–952. https://doi.org/10.1044/1092-4388(2008/069)

Wadman, R., Durkin, K., & Conti-Ramsden, G. (2011). Close relationships in adolescents with and without a history of specific language impairment. *Language, Speech, and Hearing Services in Schools, 42,* 41–51. https://doi.org/10.1044/0161-1461(2010/10-0003)

Wang, J. M., Rubin, K. H., Laursen, B., Booth-LaForce, C., & Rose-Krasnor, L. (2013). Preference-for-solitude and adjustment difficulties in early and late adolescence. *Journal of Clinical Child & Adolescent Psychology, 42,* 834–842. https://doi.org/10.1080/15374416.2013.794700

White, L. K., Degnan, K. A., Henderson, H. A., Perez-Edgar, K., Walker, O. L., Schechner, T., . . . Fox, N. A. (2017). Developmental relations between behavioral inhibition, anxiety, and attention biases to threat and positive information. *Child Development, 88,* 141–155. https://doi.org/10.1111/cdev.12696

Woods, C. M. (2009). Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis. *Multivariate Behavioral Research, 44,* 1–27. https://doi.org/10.1080/00273170802620121

Yew, S. G., & O'Kearney, R. (2013). Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: Meta-analyses of controlled prospective studies. *The Journal of Child Psychology and Psychiatry, 54,* 516–524. https://doi.org/10.1111/jcpp.12009