



# NIH Public Access

## Author Manuscript

*Res Dev Disabil.* Author manuscript; available in PMC 2015 February 01.

Published in final edited form as:

*Res Dev Disabil.* 2014 February ; 35(2): 429–438. doi:10.1016/j.ridd.2013.11.016.

## Matching Variables for Research Involving Youth with Down Syndrome: Leiter-R versus PPVT-4

**B. Allyson Phillips, M.A.<sup>a</sup>, Susan J. Loveall, Ph.D.<sup>a,b</sup>, Marie Moore Channel, Ph.D.<sup>a,c</sup>, and Frances A. Conners, Ph.D.<sup>a</sup>**

<sup>a</sup>Department of Psychology, University of Alabama, Box 870348, Tuscaloosa, AL 35487-0348

<sup>b</sup>Life Span Institute, University of Kansas, 1000 Sunnyside Ave., Lawrence, KS 66045

<sup>c</sup>MIND Institute, University of California, Davis, 2825 50<sup>th</sup> St., Sacramento, CA 95817

### Abstract

Much of what is known about the cognitive profile of Down syndrome (DS) is based on using either receptive vocabulary (e.g., PPVT-4) or nonverbal ability (e.g., Leiter-R) as a baseline to represent cognitive developmental level. In the present study, we examined the relation between these two measures in youth with DS, with non-DS intellectual disability (ID) and with typical development (TD). We also examined the degree to which these two measures produce similar results when used as a group matching variable. In a cross-sectional developmental trajectory analysis, we found that the relation between PPVT-4 and Leiter-R was largely similar across groups. However, when contrasting PPVT-4 and Leiter-R as alternate matching variables, the pattern of results was not always the same. When matched on Leiter-R or PPVT-4, the group with DS performed below that of the groups with ID and TD on receptive grammar and below the group with TD on category learning. When matched on the PPVT-4, the group with ID performed below that of the group with TD on receptive grammar and category learning, but these differences between the groups with ID and TD were not found when matched on the Leiter-R. The results of the study suggest that the PPVT-4 and Leiter-R are interchangeable at least for some outcome measures for comparing youth with DS and TD, but they may produce different results when comparing youth with ID and TD.

### Keywords

Down syndrome; nonverbal ability; receptive vocabulary; matching variables

Much of what is known about the cognitive profile of individuals with Down syndrome (DS) is based on using either receptive vocabulary or nonverbal ability as a baseline to represent cognitive developmental level. These are thought to be appropriate choices because both are relatively unaffected by the expressive and grammatical language impairments characteristic of this population. However, little work has directly compared differences in using these two types of measures as matching variables, and most researchers do not provide a rationale for selecting one type of measure over the other. The aim of the current study was to examine the relation between a receptive vocabulary measure, the Peabody Picture Vocabulary Test-4<sup>th</sup> edition (PPVT-4; Dunn & Dunn, 2007), and a

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

nonverbal ability measure, the Leiter International Performance Test-Revised (Leiter-R; Roid & Miller, 1997), in youth with DS compared to youth with non-DS intellectual disability (ID) and to children with typical development (TD). Further, we examined the degree to which these two measures produce similar results when used as a group matching variable.

## 1. 1 Methodological Issues in Matching

When conducting behavioral research in the field of ID, major methodological challenges exist. With an inability to randomly assign participants to either the target group or the comparison group, quasi-experimental designs are required, which limit causal inferences. Additionally, small sample sizes due to low prevalence rates restrict methodological designs and statistical procedures. Consequently, one of the most common methods utilized in ID research is the group-match design, but this design brings its own host of concerns (Beeghly, 2006; Burack, 2004; Eigsti, de Marchena, Schuh, & Kelley, 2011; Kover & Atwood, 2013). In a group-match design, participants from two or more pre-existing groups (e.g., Down syndrome and typically developing) are matched on one or more variables such as IQ, chronological age (CA), or mental age (MA); once matched, participants are compared on the dependent variable(s). The matching variable is selected based on its supposed relation to the dependent variable(s), and when matched, groups are considered equivalent on the matching variable. Therefore, when differences between groups are found, one can conclude whether or not that particular construct is a strength or weakness for the group with ID relative to the matching variable, which is often interpreted as a proxy for general developmental level (e.g., whether or not visuospatial ability is a strength or weakness for individuals with DS relative to their MA). Beyond the scope of the current paper but still an issue of major concern is how to determine group equivalence on the matching variable (e.g., using  $p$  values versus effect sizes; see Kover & Atwood, 2013).

The other primary issue with the group-match design, and the focus of the current discussion, is how to appropriately match participants (e.g., Burack, Iarocci, Bowler, & Mottron, 2002; Mervis & Klein-Tasman, 2004; Mervis & Robinson, 1999; Silverman, 2007; Strauss, 2001). The broadest decision about matching participants is whether to select a comparison group based on CA or MA (for review, see Burack et al., 2002; Chapman & Hesketh, 2000). When matching a group with ID to a TD group on MA, researchers can eliminate the expected delays in development due to the group with ID's lower cognitive functioning. By setting groups equivalent on a general level of cognitive functioning, researchers can determine relative strengths and weaknesses after accounting for the known general delay. Matching on MA typically results in group comparisons with significantly different CA, which means different biological maturation and life experiences that can influence task performance. However, matching on developmental level is usually preferred to matching on CA for detecting relative strengths and weaknesses and providing information about cognitive behavioral profiles.

Another key consideration for matching participants is the developmental profile of the target population (for review, see Burack et al., 2002; Chapman & Hesketh, 2000). When selecting a variable on which to match participants, researchers must consider the developmental strengths and weaknesses of the target population. If not accounted for, the matching variable may underestimate or overestimate the cognitive ability of the target group. For example, verbal tests may underestimate the cognitive abilities of individuals with autism while nonverbal ability measures may overestimate their cognitive abilities (Shah & Frith, 1993). Matching on different variables has the potential to influence the results of a study, as shown by Ozonoff, Pennington, and Rogers (1990) when they examined emotion perception in individuals with autism compared to TD individuals. They

found that emotion perception was delayed in individuals with autism when they were matched on nonverbal MA but not when they were matched on mean length of utterance (a verbal measure). Additionally, if the matching variable is not related to the variable of interest, the results may be affected (Burack et al., 2002). For example, it makes more sense to use a nonverbal ability measure rather than a verbal ability measure to match groups when examining visuospatial skills. While this will diminish the likelihood of finding significant differences between groups, researchers can be more confident that their results truly demonstrate the relative strength or weakness of the target skill. Therefore, researchers are encouraged to consider both the participant characteristics and the research question when selecting a matching variable.

One suggested method for handling discrepancies associated with the selection of different matching variables is to include more than one comparison group and match on several measures of cognitive development (e.g., one comparison group matched on verbal ability and one comparison group matched on nonverbal ability; Hobson, 1991). This allows researchers to better determine strengths and weaknesses within specific domains of cognitive abilities and ideally provide an enhanced understanding of the target group's level of functioning; however, such a design places an increased burden on the researcher in terms of recruitment and testing. The current study utilized such a technique by including separate analyses where participants were either matched on nonverbal ability or receptive vocabulary to determine how results may be affected based on these different matching variables.

## 1.2 Cognitive-linguistic Profile of Down Syndrome

These issues with matching are particularly relevant in DS research, where one must take into consideration the unique profile of individuals with DS when selecting a matching variable. DS is the most common genetic disorder that results in ID. It is caused by an extra copy of chromosome 21 (i.e., Trisomy 21) and affects approximately one in 691 live births (Parker et al., 2010). Individuals with DS experience impairments in cognitive, emotional, and physical development including a moderate to severe intellectual delay with an average IQ range of 30 to 70. They also have a distinct cognitive-linguistic profile. Based on verbal and nonverbal MA comparisons, speech, language, and verbal short-term memory are all areas of clear impairment in DS (for reviews, see Abbeduto, Warren, & Connors, 2007; Baddeley & Jarrold, 2007; Chapman & Hesketh, 2000; Kent & Vorperian, 2013; Næss, Lyster, Hulme, & Melby-Lervåg, 2011). While verbal abilities are a clear weakness for individuals with DS, visuospatial processing is not quite as impaired (Jarrold & Baddeley, 1997; Jarrold, Baddeley, & Hewes, 1999; Silverstein, Legutki, Friedman, & Takayama, 1982). For example, on short-term memory tasks, individuals with DS perform better when the task involves visual or spatial materials (e.g., pictures, block locations) than when the task involves verbal materials (e.g., letters, digits).

Such clear distinctions between verbal and visuospatial abilities cause one to question whether using a receptive vocabulary measure as a matching variable is appropriate. While individuals with DS perform relatively better on receptive vocabulary measures than other language measures such as expressive vocabulary, grammar, and verbal short-term memory measures (Naess, et al., 2011), there is still concern that using a receptive vocabulary measure like the PPVT-4 to match individuals with DS to a comparison group may, in fact, underestimate the cognitive abilities of individuals with DS. This could lead to overestimation of other abilities if receptive vocabulary is used as the matching variable. Researchers such as Chapman and Hesketh (2000) strongly discourage the use of any verbal test to match participants with DS. In contrast, using nonverbal ability measures (e.g., Leiter-R) as matching variables may lead to underestimation of other abilities when

assessing the cognitive functioning of individuals with DS. However, because the Leiter-R in particular requires no verbal responses from the participants and all instructions are given to the participant nonverbally, this test may provide a fairly accurate measure of cognitive ability in individuals with language impairments, such as those with DS.

### 1.3 Nonverbal Ability versus Receptive Vocabulary

Several studies have examined receptive vocabulary skills in individuals with DS compared to nonverbal MA-matched TD individuals. The results of these studies can inform the issue of whether receptive vocabulary and nonverbal ability are different enough in DS to produce different results when used as matching variables. In a meta-analysis, Naess and colleagues (2011) analyzed the results of ten such studies, and after the removal of one outlier, they found that the aggregate effect size did not differ significantly from zero. In other words, they found that across studies groups with DS showed similar performance on receptive vocabulary to nonverbal MA-matched TD groups. This suggests that perhaps receptive vocabulary and nonverbal ability are not very discrepant in DS.

However, Naess et al.'s (2011) findings are not conclusive. Two limitations relevant to the interpretation of findings exist. First, there was significant heterogeneity in the set of effect sizes, which was not explained. Second, two of the studies included in the meta-analysis used the Leiter (Leiter, 1969) or Leiter-R to match on nonverbal MA and the PPVT (or British Picture Vocabulary Scale, BPVS, the UK standardized version of the PPVT) as the measure of receptive vocabulary (Hick, Botting, & Conti-Ramsden, 2005; Roberts et al., 2007). Both studies were longitudinal and matched groups on Leiter at Time 1. Hick and colleagues found no significant difference in BPVS-2 (Dunn, Dunn, & Whetton, & Burley, 1997) scores between the group with DS (mean CA = 9.75) and the TD group matched at Time 1. However, Roberts and colleagues found that the group with DS (mean CA = 8.33) scored lower on PPVT-3 (Dunn, Dunn, & Dunn, 1997) than the TD group. Notably, the sample in Roberts et al. (2007) consisted of only boys, and this could have influenced this particular comparison.

Two other studies also inform this issue (Carr, 1995; Glenn & Cunningham, 2005). Each compared receptive vocabulary and nonverbal ability scores within a sample of participants with DS. Carr (1995) used the British Picture Vocabulary Scales (BPVS; Dunn, Dunn, Whetton, & Pintilie, 1982), and the non-revised version of the Leiter (Leiter, 1980). In a within-groups analysis, she found that individuals with DS (all age 21 years) had a lower mean MA on the BPVS (4.50 years) than on the Leiter (5.39 years). These results could be demonstrating that receptive vocabulary measures are, in fact, underestimating the cognitive abilities of individuals with DS. Glenn and Cunningham (2005) attempted to replicate this study using revised versions of the measures, the BPVS-II (Dunn, Dunn, Whetton, & Burley, 1997) and the Leiter-R. In contrast with Carr (1995), they found that, in their group with DS (mean CA = 19.83) the verbal MA of the BPVS-II (6.50 years) was significantly higher than the nonverbal MA of the Leiter-R brief IQ (5.20 years). The authors concluded that the higher verbal MA is possibly an overestimation of the general cognitive ability and that this overestimation could be due to environmental factors related to the individuals with DS being adults and having increased opportunities to learn vocabulary. However, while the MA scores were not equivalent for the two measures, the correlation between the verbal MA and nonverbal MA was significant ( $r = 0.61, p < .001$ ), and therefore, Glenn and Cunningham (2005) concluded that both measures were appropriate measures for matching on general cognitive ability. However, they recommended that the overall research question in the study be taken into consideration when selecting a matching variable. For example, if the study involves verbal communication, Glenn and Cunningham recommend the use of a receptive vocabulary measure such as the BPVS-II or PPVT-4, but if the study does not

involve verbal communication, they would recommend a nonverbal ability measure such as the Leiter-R.

In the present study, we extended the work of Carr (1995) and Glenn and Cunningham (2005) in two ways. First, we further examined the relation between receptive vocabulary and nonverbal ability. Whereas Glenn and Cunningham (2005) reported a strong correlation between these two abilities in a single group with DS, we examined the nature of this relation in a group of youth with DS compared to a group of youth with mixed-etiiology ID and a group of TD children in the same range of ability. Specifically, we used cross-sectional developmental trajectory analysis (Thomas et al., 2009) to do so. If the slopes of the trajectories are similar across the three groups, this would support Glenn and Cunningham's (2005) contention that either receptive vocabulary or nonverbal ability can be used equally well as group matching variables.

The second way we extended the work of Carr (1995) and Glenn and Cunningham (2005) was in directly assessing the outcomes of group matching on receptive vocabulary versus nonverbal ability. Although Carr (1995) and Glenn and Cunningham (2005) discussed the issue of whether study results would depend on which of these matching variables were used, they did not actually use these variables to match participants. Thus, in the present study, we compared study outcomes under two matching scenarios – matching on PPVT-4 and matching on Leiter-R across groups with DS, ID, and TD. Further, we used both a verbal outcome measure (i.e., Test for Reception of Grammar, 2<sup>nd</sup> Version) and a nonverbal outcome measure (i.e., the Modified Card Sort Task) in this analysis. This allowed us to directly test Glenn and Cunningham's (2005) assertion that the research topic (verbal versus nonverbal) should be considered when selecting a matching variable.

## 2 Methods

### 2.1 Participants

Fifty-four youth with DS, 29 youth with ID, and 35 TD children were included in the current study. All participants were in a similar range of nonverbal ability as determined by the Leiter-R growth score value. The current study was a part of a larger study examining the cognitive predictors of language impairment in DS in which participants were recruited through local schools, agencies, and research participant registries in Alabama, Wisconsin, and California. All participants had to be native English speakers, use speech as their primary means of communication, be without a prior diagnosis of autism, and have full use of their hands to be eligible for the larger study. Additionally, for the current study participants had to be able to complete the nonverbal ability, the receptive vocabulary, the receptive grammar, and the rule-based category learning measures that were administered. They also had to score between 4 and 9 years in nonverbal MA on the Leiter-R to be included in the current analyses. See Table 1 for participant characteristics.

**2.1.1 Participants with DS**—In addition to the general eligibility criteria, participants with DS had to be between 10 and 21 years and pass an autism screener (i.e., Social Communication Questionnaire) or autism evaluation (i.e., Autism Diagnostic Observation Schedule) if the participant screened high on the autism screener. Of 54 participants with DS in the larger study who met the general eligibility criteria, all were included in the final data analysis (24 males; 41 Caucasian, 2 African American, 7 White Hispanic, 2 More Than One Race, 2 Other Race).

**2.1.2 Participants with ID**—In addition to the general eligibility criteria, participants in the ID group had to be between 10 and 21 years old, have a school classification or clinical diagnosis of intellectual disability and score below 75 on the Leiter-R Brief IQ test

administered in the study. Also, they had to be without a diagnosis of autism and pass an autism screener (i.e., Social Communication Questionnaire) or autism evaluation (i.e., Autism Diagnostic Observation Schedule) if the participant screened high on the autism screener. Of 30 participants with ID in the larger study who met the general eligibility criteria for the present study, 29 were included in the final data analysis (14 males; 25 Caucasian, 3 African American, 1 White Hispanic). The one participant was excluded because of a nonverbal IQ score above 75.

**2.1.3 TD participants**—In addition to the general eligibility criteria, to be included in the TD group, participants had to be at least 4 years old, ineligible for special services in school (including those for learning disability, speech and language services, and giftedness), and not have a parent-reported diagnosis of attention deficit hyperactivity disorder or autism spectrum disorder. Of 43 participants with TD in the larger study who met the general eligibility criteria for the present study, 35 were included in the final data analysis (21 males; 19 Caucasian, 9 African American, 1 Black Hispanic, 4 White Hispanic, 2 More Than One Race). Eight participants were excluded because they were older participants (CA range = 14.17 – 18.42) with very low MA scores (MA range = 8.63 – 9.75).

## 2.2 Measures

### 2.2.1 Primary measures

**Nonverbal ability (30 minutes):** The Leiter International Performance Test- Revised brief form (Leiter-R; Roid & Miller, 1997) was used to measure nonverbal ability. The Leiter-R is a published standardized norm-referenced test for individuals aged 2 years 0 months through 20 years 11 months. Administration of the Leiter-R is completely nonverbal. No verbal response is required from the participant, and instructions are communicated via examiner pantomime. The four subtests that make up the Brief IQ battery—Figure Ground, Form Completion, Sequential Order, and Repeated Patterns—were administered. Together, these subtests measure visual spatial and inductive reasoning skills that are typically classified as fluid intelligence. TD participants started each subtest at the appropriate starting point based on CA, but participants with DS or ID were started at the beginning of each subtest regardless of CA. The MA (or age equivalence) scores were used to determine eligibility for the present analyses and to match participants across groups, and the growth score values (GSV) were used in the developmental trajectory analysis. A GSV is the conversion of a raw score in which scale corrections are made for variability in item difficulty. It is reported that the Leiter-R brief form correlates .85 with both the full version of the Leiter-R and the WISC-III IQ test, and reported Cronbach's alpha for the present subtests ranges from .75 to .88 (Roid & Miller, 1997).

**Receptive vocabulary (15 minutes):** The Peabody Picture Vocabulary Test – 4<sup>th</sup> ed. (PPVT-4; Dunn & Dunn, 2007) was used to measure receptive vocabulary. The PPVT-4 is a standardized norm-referenced test that is appropriate for functioning levels equivalent to as young as 2.5 years old and older. The PPVT-4 requires participants to point to the picture that corresponds with a spoken word. The test covers 20 content categories and includes nouns, verbs, and adjectives. TD participants began the task at the appropriate starting point based on CA, but participants with DS or ID began the task at the test block corresponding to half their CA or at the test block corresponding to a CA of 8 years, whichever resulted in a lower start point. Reported split-half reliability ranges from .94 to .95. We used age equivalence scores to match participants across groups and GSVs in the developmental trajectory analysis. While GSVs are used for both the Leiter-R and the PPVT-4, the GSVs for each measure are not on the same scale and are thus not comparable.

## 2.2.2 Outcome measures

**Receptive grammar (15 minutes):** The Test for Reception of Grammar, 2<sup>nd</sup> Version (TROG-2; Bishop, 2003) was used to measure receptive grammar. The TROG-2 is a standardized norm-referenced test for ages 4 – 86 years. The participant listens to a sentence and points to the picture that best fits the sentence. The TROG-2 presents 20 different grammatical contrasts in a series of trial blocks. Reported internal consistency reliability is .88. We used raw scores in the current analyses.

**Category learning (10 minutes):** The Modified Card Sort Task (MCST; Nelson, 1976) was used to measure rule-based category learning. It is an adaptation of the Wisconsin Card Sorting Test (Berg, 1948; see also Heaton, Chelune, Talley, Kay, & Curtis, 1993), introduced by Nelson (1976) and developed by Cianchetti, Corona, Foscolano, Contu, and Siannio-Fancello (2007). The MCST uses cards containing different shapes (triangles, stars, crosses, or circles), different colors (red, green, yellow or blue), and different numbers of shapes (one, two, three or four) on each card. Each card has from one to four shapes on it, all of which are the same color and the same shape (i.e. three red triangles or four yellow crosses). After showing the participant the different colors, shapes, and number of shapes on the cards, the examiner lays out four stimulus cards on the table. The following four stimulus cards are used - (1) one red triangle, (2) two green stars, (3) three yellow crosses, and (4) four blue circles. These stimulus cards are used for the participant to “match” additional cards according to one of three rules: color, shape or number. An additional forty-eight cards with various combinations of the colors, shapes, and numbers are used for matching. After laying out the stimulus cards, the experimenter hands the participant one of the forty-eight and asks the participant to match the card to one of the four stimulus cards in order to figure out the examiner’s ‘rule’ (matching according to color, shape or number). The examiner provides feedback after each trial, telling the participant if the match was correct or incorrect. The participant completes a ‘category’ when he/she correctly matches six cards in a row. The cards are then picked up and the examiner asks the participant to find a new rule. The game is played six times with each of the three categories (color, shape, and number) being a matching rule two times. Each category (shape, color, number) is used once as the rule in games 1-3, and again in games 4-6 in the same order. The task ends when the participant has run out of cards to match and/or completes all six games, whichever is first. A learning efficiency score (Cianchetti et al., 2007) is used as the main measure from this task. It is calculated by taking the total number of categories completed (i.e., 1-6) multiplied by six plus the number of un-played cards (if any). In an analysis based on the same larger study, the MCST had a Spearman-Brown split-half reliability of .96 (Phillips, Conners, Merrill, & Klinger, in press).

## 2.3 Procedures

In the larger study, participants were tested individually during two to four testing sessions (the number of sessions varied depending on the individual characteristics of the participant). They completed a battery of implicit learning, explicit learning, phonological memory, and language tests. The Leiter-R, TROG-2, and MCST were always administered during the first session, and the PPVT-4 was administered during the second half of tasks ordered randomly among the remaining tasks.

# 3 Results

## 3.1 Analytic Approach

To determine the relationship between the Leiter-R and PPVT-4, we examined performance on the PPVT-4 relative to nonverbal ability by performing a developmental trajectory analysis as described by Thomas et al. (2009). This method adapts the Analysis of

Covariance (ANCOVA) function with SPSS's General Linear Model. When *traditionally* using ANCOVA, one must satisfy the assumption that the covariate has the same relation to the dependent variable in each group. This assumption is met by finding that there is not a significant group  $\times$  covariate interaction. For the present trajectory analysis, we specifically tested whether nonverbal ability has a *different* relation to receptive vocabulary across groups. The ANCOVA, therefore, tested group differences in slope, which was indicated by the group  $\times$  Leiter-R interaction.

To determine the degree to which the Leiter-R and PPVT-4 produce similar results when used as matching variables, we matched subgroups of participants by mean score on each variable ( $p > .5$ ) and compared group performance on a verbal measure (i.e., TROG-2) and a nonverbal measure (i.e., MCST). In order to derive equivalent groups from the larger sample, the highest and lowest age equivalence scores on the matching variable (either the Leiter-R or PPVT-4) were removed, and a one-way ANOVA was run on the age equivalence scores across the three groups (DS, ID, and TD). Extreme scores continued to be removed until group equivalence was determined by  $p > .5$  in the ANOVA. When matching on the Leiter-R, 39 participants with DS, 25 participants with ID, and 29 TD participants were included in the analyses,  $F(2, 90) = 0.41$ ,  $p = .664$ . When matching on the PPVT-4, 40 participants with DS, 19 participants with ID, and 33 TD participants were included in the analyses,  $F(2, 89) = 0.56$ ,  $p = .572$ . See Table 2 for participant characteristics for both the Leiter-R and PPVT-4 matches.

### 3.2 Preliminary Analyses

Means and standard deviations for all variables are listed in Tables 1 (full sample) and 2 (matched samples). Two participants with DS had a CA that is outside the upper CA range of the Leiter-R (CA = 21.91 and 21.08). All analyses were run with and without these two participants, and the pattern of results was the same. It was decided to retain these two participants in the final sample because (1) all of the analyses were based on raw scores, not standard scores, so CA was not a factor, (2) though these two participants fell outside the CA range of the Leiter-R, they still fell within the cognitive ability range, and (3) retaining these two participants maximized statistical power. There were no serious violations of normality for any variable. To prepare for the cross-sectional developmental trajectory analysis, scatterplots for receptive vocabulary over nonverbal ability as well as nonverbal ability over receptive vocabulary were created within each group to visually examine linearity. No clear nonlinear trends were found. Cook's D values were also calculated separately within each group for the PPVT-4 GSV and for the Leiter-R GSV to determine whether any points were exerting undue influence on the analysis. A value greater than 1.00 would indicate undue influence, but no such values were found. Finally, the Leiter-R GSV variable and PPVT-4 GSV variable were transformed so that the lowest score in the analysis was set equal to zero. This allowed for greater interpretability of the intercepts of the trajectory analyses.

### 3.3 Main Analyses

**3.3.1 Developmental trajectory analysis**—The cross-sectional developmental trajectory for the PPVT-4 over Leiter-R performance is shown in Figure 1. There was no overall effect of Group,  $F(2, 112) = 1.31$ ,  $p = .274$ ,  $\eta_p^2 = .023$ . Thus, at the lowest Leiter-R score in the data set (equivalent to about a MA of 4 years), the regression lines for each group were not significantly different. Additionally, there was no significant Group  $\times$  Leiter-R interaction,  $F(2, 112) = 1.31$ ,  $p = .273$ ,  $\eta_p^2 = .023$ . Therefore, growth in PPVT-4 as a function of Leiter-R was similar across the three groups.

**3.3.2 Group match on Leiter-R**—When matched on the Leiter-R, separate one-way analyses of variance (ANOVA) comparing the three groups were conducted to compare group performance on the TROG-2 and MCST. The first ANOVA indicated that groups were significantly different for the TROG-2,  $F(2, 90) = 23.88, p < .001, \eta_p^2 = .347$ . Post hoc analyses using Tukey post hoc criterion for significance indicated that TROG-2 scores were significantly lower for participants with DS than for participants with ID ( $p < .001$ ) and for participants with TD ( $p < .001$ ). Scores for participants with ID did not differ significantly from those with TD ( $p = .266$ ). The second ANOVA indicated that group scores were marginally significantly different for the MCST,  $F(2, 90) = 3.07, p = .051, \eta_p^2 = .064$ . Post hoc analyses using Tukey post hoc criterion for significance indicated that MCST scores were significantly lower for participants with DS than for participants with TD ( $p = .045$ ). Scores for participants with ID did not differ significantly from those with DS or those with TD ( $p$ 's = .329 and .546, respectively). These results have been reported previously (Phillips et al., in press).

**3.3.3 Group match on PPVT-4**—When matched on the PPVT-4, two one-way ANOVAs were conducted to compare performance of the three groups on the TROG-2 and MCST. The first ANOVA indicated that groups performed significantly different on the TROG-2,  $F(2, 89) = 29.23, p < .001, \eta_p^2 = .396$ . Post hoc analyses using Tukey post hoc criterion for significance indicated that TROG-2 scores were significantly lower for participants with DS than for participants with ID ( $p = .030$ ) and for participants with TD ( $p < .001$ ). Additionally, TROG-2 scores were significantly lower for participants with ID than for participants with TD ( $p = .001$ ). The second ANOVA indicated that scores were significantly different for the MCST,  $F(2, 89) = 7.92, p = .001, \eta_p^2 = .151$ . Post hoc analyses using Tukey post hoc criterion for significance indicated that MCST scores were significantly lower for participants with DS than for participants with TD ( $p = .002$ ). Additionally, MCST scores were significantly lower for participants with ID than for participants with TD ( $p = .004$ ). Scores for participants with DS did not differ significantly from those with ID ( $p = .896$ ).

## 4 Discussion

Previous research has indicated that, when conducting research on groups with ID in comparison to TD groups, the variable selected on which to match groups can have an impact on the study's findings (e.g., Ozonoff et al., 1990). However, this may depend on several factors including the cognitive-linguistic profile of the target population. The study of DS presents a particular challenge for group comparisons, in that many verbal abilities are very poor compared with nonverbal abilities. In the present study we examined two commonly used matching variables in studies of this population – receptive vocabulary as measured by the PPVT-4 and nonverbal ability as measured by the Leiter-R. The first aim of the study was to examine the nature of the relation between these two matching variables in a group of youth with DS compared to groups of youth with ID and TD children, using a cross-sectional developmental trajectory analysis. The second aim was to determine whether these two matching measures produce different group comparison outcomes, and whether this depends on if the outcome variable is verbal or nonverbal.

The cross-sectional developmental trajectory analysis indicated that the relation between receptive vocabulary (PPVT-4) and nonverbal ability (Leiter-R) was largely similar across groups. Slopes were not different across groups, indicating that increases in receptive vocabulary as a function of nonverbal ability were similar across groups. This result extends Glenn and Cunningham's (2005) finding that receptive vocabulary and nonverbal ability correlated strongly in their group with DS. It shows that the nature of the relation between

these two variables is not drastically different for youth with DS than it is for those with ID or those with TD.

In addition to slopes, intercepts were also similar across groups. In other words, at the lowest level of nonverbal ability measured in the present study, the groups were similar in receptive vocabulary. This is somewhat at odds with Carr's (1995) finding that participants with DS had lower MAs on receptive vocabulary than on nonverbal ability, as well as with Glenn and Cunningham's (2005) findings in the opposite direction. Findings from these two studies might lead us to expect participants with DS to score differently on the PPVT-4 than participants with TD at any given point on the Leiter-R (in one direction or the other depending on the study). However, in the present study there was no difference. Possibly, sample differences across the studies affected the comparisons. Also, in the present study, we used more psychometrically sound GSV scores and a contrasting control group, rather than MA scores compared within group. Regardless, our cross-sectional developmental trajectory analysis suggests that the nature of the relation between receptive vocabulary and nonverbal ability is more similar than different across groups with DS, TD, and ID who are performing in the same range of nonverbal ability.

When contrasting receptive vocabulary and nonverbal ability as alternate matching variables, however, the patterns of results were not always the same. When the groups were matched on nonverbal ability (Leiter-R), the group with DS performed below the groups with ID and TD on receptive grammar and below the group with TD on category learning only. However, the group with ID performed similarly to the group with DS on category learning and similarly to the TD group on both measures. When the groups were matched on receptive vocabulary (PPVT-4), the group with DS again performed below that of the groups with ID and TD on receptive grammar. However, the group with ID also performed significantly below the group with TD on receptive grammar. Also, when measuring category learning, the groups with DS and ID *both* performed below the group with TD, and similarly to each other. Thus, for the group with DS, the results were the same regardless of matching variable. However, for the group with ID, results depended on the matching variable. The results were the same for the comparison between the groups with DS and ID with the group with ID performing better than the group with DS on receptive grammar and the group with ID performing similarly to the group with DS on category learning. However, the results differed based on the matching variable for the comparison between the groups with ID and TD. When matched on the Leiter-R, the group with ID did not perform significantly different than the group with TD on either outcome measure, but when matched on the PPVT-4, the group with ID performed significantly below that of the group with TD on both outcome measures.

Glenn and Cunningham (2005) suggested that though either Leiter or PPVT can be used as a viable matching variable, the two matching measures could yield different outcomes for participants with DS depending on whether performance on a verbal or a nonverbal task was being assessed. They therefore recommended that the Leiter be used for matching when examining a nonverbal ability and the PPVT be used for matching when examining a verbal ability. In the present study, however, regardless of the matching measure, the group with DS performed below both comparison groups on receptive grammar and below the TD group only on category learning. Based on these results, and contrary to the recommendation of Glenn and Cunningham (2005), the use of the PPVT-4 and the Leiter-R seem roughly interchangeable as matching variables for comparing performance of youth with DS and TD. However, further refinement of this conclusion is needed, as the present study examined only one verbal and one nonverbal outcome measure.

On the other hand, for the group with ID, the matching variable did impact the outcome, though this did not depend on whether the performance being compared was verbal or nonverbal. When matched on Leiter-R the group with ID did not differ significantly from the TD group on either receptive vocabulary or category learning, yet when matched on PPVT-4, they performed more poorly than the TD group on both measures. Examination of the group means for the Leiter-R and PPVT-4 under the two matching scenarios provides some insight into why there were different outcomes for the group with ID only. When the groups were matched on Leiter-R, the PPVT-4 mean for the group with ID was higher than for the other groups. When the groups were matched on PPVT-4, the Leiter-R means were similar. This suggests that the group with ID was relatively strong in receptive vocabulary. Consequently, when matched on their relatively strong ability (PPVT-4), they showed poor performance on both receptive grammar and category learning. Whether this pattern is externally valid for groups with mixed-etiiology ID or is an idiosyncrasy of our sample or matching constraints will have to be borne out in future research. Nevertheless, our results suggest that the outcome of an MA-match comparison that includes a group with mixed etiology ID can be different for receptive vocabulary versus nonverbal ability matching measures. Researchers should be aware that when including a group with ID in their study, the matching variable could affect the outcome of group comparisons. Further, they should not necessarily assume that the difference relates to whether the outcome variable is verbal or nonverbal.

A number of limitations in the present study warrant mention. First, as already noted, the present study only included one verbal and one nonverbal outcome variable, while both verbal and nonverbal skills are multi-faceted. Second, though the sample size of the group with DS was relatively large, the sample sizes for the groups with ID and TD were smaller. Setting up the two matching scenarios required reducing the sample sizes further. A future study that uses more than one verbal and nonverbal outcome measure and is not as restricted in sample size would be helpful to substantiate the present results. A third limitation of the present study is that our samples with DS and ID ranged widely not only in the variables being studied, but also in age. Participants in these groups were pre-adolescent through young adults. On one hand, this is similar to many studies in the DS literature. On the other hand, it would be valuable to examine the impact of choice of matching measure in narrower age bands, without the additional variability present in a sample ranging widely in age. Further, it is unclear based on the current study if the results would generalize to younger preschool or school aged children with DS and ID or to older adults with DS and ID. Future research is needed to understand the generalizability of the results and caution should be taken when seeking methodological advice for younger or older age groups.

Many questions remain about the impact of matching variable on study outcome, particularly for studying populations such as DS, in which there are large disparities in cognitive and linguistic abilities. We hope that the present study will renew attention to this issue in the research community. The results of the study suggest that the PPVT-4 and Leiter-R are interchangeable at least for some outcome measures when comparing groups with DS and TD. However, they may produce different results when comparing groups with ID and TD. We await further such studies that add to and refine these conclusions.

## References

- Abbeduto L, Warren SF, Conners FA. Language development in Down syndrome: From the prelinguistic period to the acquisition of literacy. *Mental Retardation and Developmental Disabilities Research Reviews*. 2007; 13:247–261. [PubMed: 17910087]
- Baddeley AD, Jarrold C. Working memory and Down syndrome. *Journal of Intellectual Disability Research*. 2007; 51:925–931. [PubMed: 17990999]

- Beeghly M. Translational research on early language development: Current challenges and future directions. *Development and Psychopathology*. 2006; 18:737–757. [PubMed: 17152398]
- Berg EA. A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*. 1948; 39:15–22. [PubMed: 18889466]
- Bishop, DVM. The Test for Reception of Grammar, Version 2 (TROG-2). London: Psychological Corporation; 2003.
- Burack J. Editorial preface. *Journal of Autism and Developmental Disorders*. 2004; 34:3–5.
- Burack J, Iarocci G, Bowler D, Mottron L. Benefits and pitfalls in the merging of disciplines: The example of developmental psychopathology and the study of persons with autism. *Development and Psychopathology*. 2002; 14:225–237. [PubMed: 12030689]
- Carr, J. Down's syndrome: Children growing up. Cambridge: Cambridge University Press; 1995.
- Chapman RS, Hesketh LJ. Behavioral phenotype of individuals with Down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews*. 2000; 6:84–95. [PubMed: 10899801]
- Cianchetti C, Corona S, Foscoliano M, Contu D, Siannio-Fancello G. Modified Wisconsin Card Sorting Test (MCST, MWCST): Normative data in children 4–13 years old, according to classical and new types of scoring. *The Clinical Neuropsychologist*. 2007; 21:456–478. [PubMed: 17455031]
- Dunn, L.; Dunn, L. Peabody Picture Vocabulary Test- Revised. Circle Pines, MN: American Guidance Service; 1981.
- Dunn, L.; Dunn, L. Peabody Picture Vocabulary Test- Fourth Edition. Bloomington, MN: Pearson Assessments; 2007.
- Dunn, L.; Dunn, L.; Dunn, D. Peabody Picture Vocabulary Test- Third Edition. Circle Pines, MN: American Guidance Service; 1997.
- Dunn, L.; Dunn, L.; Whetton, C.; Burley, J. British Picture Vocabulary Scale II. Windsor: NFER-Nelson; 1997.
- Dunn, L.; Dunn, L.; Whetton, C.; Pintilie, D. British picture Vocabulary Scale. Windsor: NFER-Nelson; 1982.
- Eigsti IM, de Marchena AB, Schuh JM, Kelley E. Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders*. 2011; 5:681–691.
- Glenn S, Cunningham C. Performance of young people with Down syndrome on the Leiter-R and British Picture Vocabulary Scales. *Journal of Intellectual Disability Research*. 2005; 49:239–244. [PubMed: 15816810]
- Heaton, RK.; Chelune, GJ.; Talley, JL.; Kay, GG.; Curtis, G. Wisconsin Card Sorting Test Manual: Revised and Expanded. Odessa, FL: Psychological Assessment Resources; 1993.
- Hick RF, Botting N, Conti-Ramsden G. Short-term memory and vocabulary development in children with Down syndrome and children with specific language impairment. *Developmental Medicine and Child Neurology*. 2005; 47:532–538. [PubMed: 16108453]
- Hobson P. Methodologic issues for experiments on autistic individuals' perception and understanding of emotion. *Journal of Child Psychology and Psychiatry*. 1991; 32:1135–1158. [PubMed: 1787140]
- Jarrold C, Baddeley A. Short-term memory for verbal and visuospatial information in Down's syndrome. *Cognitive Neuropsychiatry*. 1997; 2:101–122.
- Jarrold C, Baddeley A, Hewes A. Genetically dissociated components of working memory: Evidence from Down's and Williams syndrome. *Neuropsychologica*. 1999; 37:637–651.
- Kent RD, Vorperian HK. Speech impairment in Down syndrome: A review. *Journal of Speech, Language, and Hearing Research*. 2013; 56:178–210.
- Kover ST, Atwood AK. Establishing equivalence: Methodological progress in group-matching design and analysis. *American Journal on Intellectual and Developmental Disabilities*. 2013; 118:3–15. [PubMed: 23301899]
- Leiter, RG. The Leiter International Performance Scale. Chicago: Stoelting; 1969.
- Leiter, RG. Leiter International Performance Scale. Chicago: Stoelting; 1980.

- Mervis CB, Klein-Tasman BP. Methodological issues in group-matching designs:  $\alpha$  levels for control variable comparisons and measurement characteristics of control and target variables. *Journal of Autism and Developmental Disorders*. 2004; 34:7–17. [PubMed: 15098952]
- Mervis CB, Robinson BF. Methodological issues in cross-syndrome comparisons: Matching procedures, sensitivity (Se) and specificity (Sp). *Monographs of the Society for Research in Child Development*. 1999; 64:115–130. [PubMed: 10412223]
- Naess KB, Lyster SH, Hulme C, Melby-Lervag M. Language and verbal short-term memory skills in children with Down syndrome: A meta-analytic review. *Research in Developmental Disabilities*. 2011; 32:2225–2234. [PubMed: 21628091]
- Nelson HE. A modified card sorting test sensitive to frontal lobe defects. *Cortex*. 1976; 12:313–324. [PubMed: 1009768]
- Ozonoff S, Pennington BF, Rogers SJ. Are there emotion perception deficits in young autistic children? *Journal of Child Psychology and Psychiatry*. 1990; 31:343–361. [PubMed: 2318918]
- Parker SE, Mai CT, Canfield MA, Rickard R, Wang Y, Meyer RE, et al. Updated national birth prevalence estimates for selected birth defects in the United States, 2004–2006. *Birth Defects Research (Part A): Clinical and Molecular Teratology*. 2010; 88:1008–1016.
- Phillips BA, Conners F, Merrill E, Klinger M. Rule-based category learning in Down syndrome. *American Journal on Intellectual and Developmental Disabilities*. in press.
- Roberts J, Price J, Barnes E, Nelson L, Burchinal M, Hennon EA, et al. Receptive vocabulary, expressive vocabulary, and speech production of boys with fragile X syndrome in comparison to boys with Down syndrome. *American Journal on Mental Retardation*. 2007; 112:177–193. [PubMed: 17542655]
- Roid, G.; Miller, L. *Leiter International Performance Scale- Revised*. Wood Dale, IL: Stoelting; 1997.
- Shah A, Frith U. Why do autistic individuals show superior performance on the block design task? *Journal of Child Psychology and Psychiatry*. 1993; 34:1351–1364. [PubMed: 8294523]
- Silverman W. Down syndrome: Cognitive phenotype. *Mental Retardation and Developmental Disabilities Research Reviews*. 2007; 13:228–236. [PubMed: 17910084]
- Silverstein AB, Legutki G, Friedman SL, Takayama DL. Performance of Down syndrome individuals on the Stanford-Binet Intelligence Scale. *American Journal of Mental Deficiency*. 1982; 86:548–551. [PubMed: 6214950]
- Strauss ME. Demonstrating specific cognitive deficits: A psychometric perspective. *Journal of Abnormal Psychology*. 2001; 110:6–14. [PubMed: 11261401]
- Thomas MS, Annaz D, Ansari D, Scerif G, Jarrold C, Karmiloff-Smith A. Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language, and Hearing Research*. 2009; 52:336–358.

## Highlights

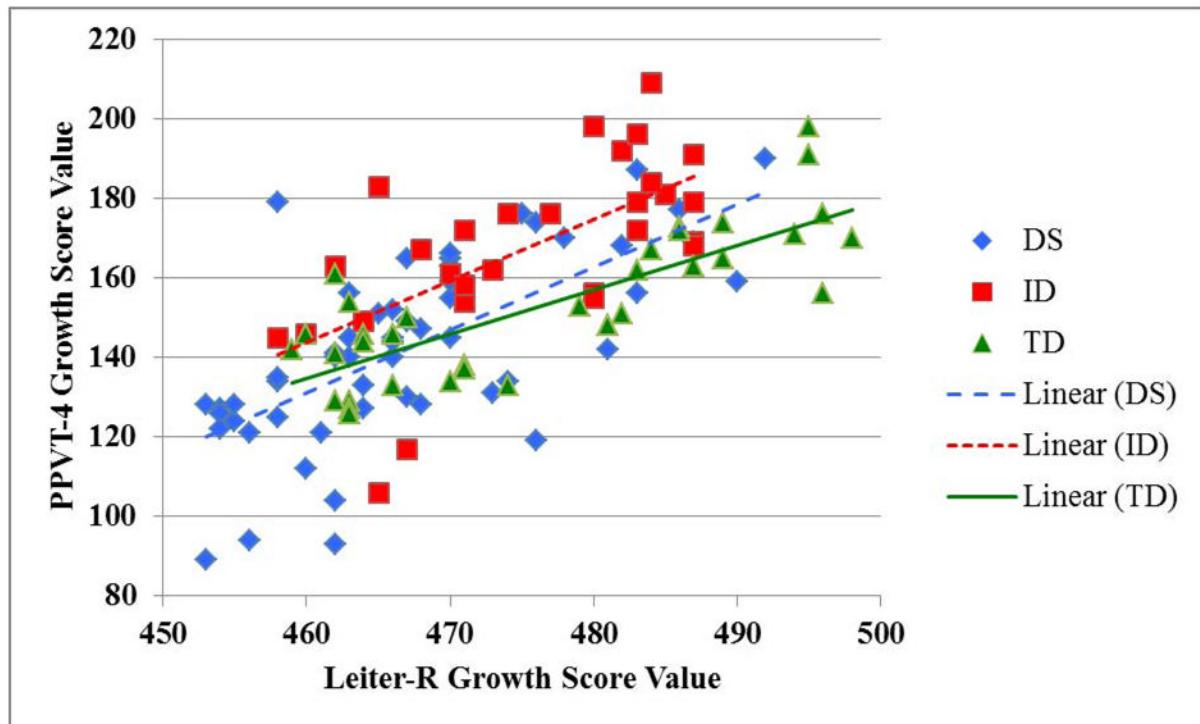
We compared the Leiter-R and PPVT-4 as matching variables for Down syndrome research.

We found that the relation between the two was largely similar across groups.

When contrasting the two as alternate matching variables, the results were not the same.

The two are interchangeable for comparing groups with DS and TD.

The two may produce different results when comparing groups with ID and TD.



**Figure 1.**

Growth rate of PPVT-4 GSV over Leiter-R GSV.

Note. GSVs on the PPVT-4 and GSVs on the Leiter-R are not on the same scale; therefore, the individual GSVs for each measure are not directly comparable.

**Table 1**  
**Participant Characteristics for Entire Sample**

	Mean	SD	Range
<b>Down Syndrome (N = 54)</b>			
<b>Chronological Age (years)</b>	15.14	3.18	10.25 – 21.92
<b>Leiter-R (nonverbal IQ)*</b>	44.90	8.40	36 – 71
<b>Leiter-R (GSV)</b>	467.22	10.04	453 – 492
<b>PPVT-4 (GSV)</b>	142.59	23.35	89 – 190
<b>Intellectual Disability (N = 29)</b>			
<b>Chronological Age (years)</b>	15.89	2.59	10.25 – 20.67
<b>Leiter-R (nonverbal IQ)</b>	52.54	10.84	36 – 77
<b>Leiter-R (GSV)</b>	475.45	9.16	458 – 487
<b>PPVT-4 (GSV)</b>	167.72	22.41	106 – 209
<b>Typically Developing (N = 35)</b>			
<b>Chronological Age (years)</b>	6.99	2.56	4.08 – 13.50
<b>Leiter-R (nonverbal IQ)</b>	101.31	16.17	76 – 135
<b>Leiter-R (GSV)</b>	476.31	12.90	459 – 498
<b>PPVT-4 (GSV)</b>	152.97	18.12	126 – 198

Note. GSV = Growth Score Value.

\* Note. Mean and SD for Leiter-R (nonverbal IQ) for group with DS is based on a *n* of 52. The two participants who fell outside of the CA range of the Leiter-R were excluded from these analyses because the IQ score is calculated from age-based norms.

**Table 2**  
**Descriptive Statistics for Leiter-R Matches and PPVT-4 Matches**

Leiter-R Matches: Means (Standard Deviations)						
Group	CA	Leiter AE	Leiter GSV	PPVT GSV	TROG	MCST
<b>DS</b>	15.62	5.76	471.03	148.85	32.23	14.97
<b>N = 39</b>	(3.06)	(1.00)	(8.80)	(21.41)	(15.58)	(10.06)
<b>ID</b>	15.82	6.00	473.60	166.28	51.28	19.36
<b>N = 25</b>	(2.60)	(0.94)	(8.49)	(23.57)	(16.93)	(14.82)
<b>TD</b>	6.45	5.89	472.31	148.00	57.21	21.97
<b>N = 29</b>	(2.25)	(1.17)	(10.25)	(14.42)	(14.48)	(10.90)
PPVT-4 Matches: Means (Standard Deviations)						
Group	CA	PPVT AE	PPVT GSV	Leiter GSV	TROG	MCST
<b>DS</b>	15.39	7.23	151.85	469.67	32.78	14.90
<b>N = 40</b>	(3.11)	(1.88)	(17.80)	(9.86)	(15.02)	(9.94)
<b>ID</b>	15.19	7.65	156.42	472.00	43.37	13.53
<b>N = 19</b>	(2.76)	(1.72)	(18.49)	(8.60)	(14.12)	(11.12)
<b>TD</b>	6.83	7.11	151.06	475.21	59.15	23.73
<b>N = 33</b>	(2.45)	(1.75)	(16.48)	(12.45)	(14.58)	(11.56)

Note. GSV = Growth Score Value. AE = Age Equivalent.