Published in final edited form as:

Child Neuropsychol. 2016 ; 22(5): 556–569. doi:10.1080/09297049.2015.1020775.

Concurrent Validity of the Differential Ability Scales, Second Edition with the Mullen Scales of Early Learning in Young Children with and without Neurodevelopmental Disorders

Cristan Farmer, Christine Golden, and Audrey Thurm

Pediatrics and Developmental Neuroscience Branch, National Institute of Health, Bethesda, Maryland, USA

Abstract

Estimates of intelligence in young children with neurodevelopmental disorders are critical for making diagnoses, in characterizing symptoms of disorders, and in predicting future outcomes. The limitations of standardized testing for children with developmental delay or cognitive impairment are well-known: tests do not exist that provide developmentally appropriate material along with norms that extend to the lower reaches of ability. Two commonly used and interchanged instruments are the *Mullen Scales of Early Learning* (MSEL), a test of developmental level, and the *Differential Ability Scales, 2nd Edition* (DAS-II), a more traditional cognitive test. We evaluated the correspondence of contemporaneous MSEL and the DAS-II scores in a mixed sample of children aged 2 to 10 years with autism spectrum disorder (ASD), non-ASD developmental delays, and typically developing children across the full spectrum of cognitive ability. Consistent with published data on the original DAS and the MSEL, scores on the DAS-II and MSEL were highly correlated. However, curve estimation revealed large mean differences that varied as a function of the child's cognitive ability level. We conclude that interchanging MSEL and DAS-II scores without regard to the discrepancy in scores may produce misleading results in both cross-sectional and longitudinal studies of children with and without ASD, thus, this practice should be implemented with caution.

Keywords

cognitive testing; IQ; developmental quotient; DAS-II; Mullen Scales of Early Learning; autism

Introduction

A major challenge in the assessment of intellectual ability is test selection. This challenge is present for all children, but is especially pronounced in children with autism spectrum disorders (ASD), who vary widely in verbal and cognitive abilities (Magiati & Howlin, 2001). While deficits in cognitive abilities are common, the exact proportion of individuals with ASD and low IQ and/or intellectual disability (ID) is unknown and difficult to estimate.

In fact, a recent review of 14 studies found that estimates ranged from 34% to 84% (Dykens & Lense, 2012). Although previous versions of the *Diagnostic and Statistical Manual* (e.g., DSM-IV-TR; American Psychiatric Association, 2000) indicated that cognitive impairment could complicate the diagnosis of ASD, the DSM-5 states explicitly that the disturbances in social communication and repetitive behavior must not be better explained by ID or global developmental delay (American Psychiatric Association, 2013, p. 51). In practice, this means that the behaviors found deviant on assessment must be abnormal for peers at the child's general developmental level, *not* for chronological-age peers. Thus, accurate assessment of intellectual ability is a necessary component in interpreting results from autism diagnostic instruments in a comprehensive ASD differential diagnosis (Risi et al., 2006).

The selected test must be appropriate for capturing valid scores for all children with potential cognitive deficits, both for one-time evaluations and over time in longitudinal assessments. Often, tests used in typically developing children, such as the *Wechsler Preschool and Primary Scale of Intelligence* (WPPSI; Wechsler, 2012) or the *Stanford-Binet Intelligence Scale* (Roid, 2003) require significant receptive and expressive verbal abilities, even in the so-called non-verbal subtests. For example, directions for WPPSI Block Design, part of the nonverbal scale, are administered verbally. Further, these tests are frequently invalid for the assessment of children with very low IQ, as the development and standardization samples rarely include substantial representation of this segment of the population. For this reason, standard scores of less than 40 are rarely available, and the precision of these tests for individuals with low IQ is limited. Tests standardized for younger children may include developmentally appropriate content for older children with low IQ, but may exclude the assessment of higher-level cognitive processes. These tests are often not useful for older children with average IQ, limiting their utility in a longitudinal or cross-sectional study. In the ASD literature especially, researchers often use multiple tests both within and across time points to ensure that at least some measure of cognitive ability is obtained for each individual, regardless of verbal ability and cognitive level (e.g., Frazier, Georgiades, Bishop, & Hardan, 2014; Lord & Schopler, 1989; Sigman & McGovern, 2005).

The use of multiple tests to assess a single construct requires confidence in the convergent and concurrent validity of the measures. An abundance of data on the convergent and concurrent validity of popular IQ tests is available for the typically developing population, but these data are scarce for children with disabilities. Some studies report differences in the concurrent validity of tests for specific diagnostic groups. For example, one research group found differences of about 30 percentile points between the *Wechsler* scales and the *Raven's Coloured Progressive Matrices* in children and adults with autism but not in the typically developing controls (Dawson, Soulières, Gernsbacher, & Mottron, 2007). Bolte and colleagues (2009) were able to only partially replicate this finding, demonstrating a nine percentile point advantage of the *Raven's* in individuals with low IQ only. A small study comparing the *Wechsler Intelligence Scale for Children, Fourth Edition* to the *Differential Ability Scales, Second Edition* (DAS-II; Elliott, 2007) showed a mean advantage of the DAS-II of only five points on the standard scale (Kuriakose, 2014). Another recent study found a statistically significant but small difference (3.25 points) between the *Stanford-Binet Fifth Edition* and the *Wechsler Intelligence Scale for Children, Fourth Edition*, as well as

significant differences in verbal-nonverbal IQ discrepancies between the tests (Baum, Shear, Howe & Bishop, 2014).

Along with the DAS-II, the *Mullen Scales of Early Learning* (MSEL, Mullen, 1995) has become increasingly popular in studies of ASD (Akshoomoff, 2006), and the two have been used interchangeably (e.g., Lord et al., 2012). However, very few data exist on the concurrent validity of these instruments. In one study that used both measures concurrently in a combined ASD and DD sample, moderate-to-high correlations were found between the various versions of the MSEL (i.e., 1989, 1995) and the original version of the DAS (Bishop et al., 2011). While the correlations between the DAS and DAS-II are reported to be fairly high (Elliott, 2007), there are notable changes between the two editions of this measure, such as the addition of Matrices and removal of Early Number Concepts from the nonverbal composites. Thus, more data are required to support the continued use of the MSEL and DAS-II as exchangeable tests, given the use of their concurrent administration in both large scale and longitudinal studies of ASD (Anderson, Liang, & Lord, 2014; Lord, et al. 2012).

Often, children are outside the standardized age range of the test that is most developmentally appropriate. The ratio IQ (RIQ) may be used in this situation, or in any other situation in which a child is unable to achieve a standard score. The RIQ relies on the original conceptualization of IQ as the ratio of mental age to chronological age [e.g., a 6-year-old child with a mental age of 36 months has an RIQ of $(36 \div 72) * 100 = 50$]. Many studies use MSEL RIQs regardless of whether the child is able to obtain a standard score, because the MSEL does not produce nonverbal and verbal composite scores. These are often calculated as RIQs using the subtests presumed to best measure verbal (Expressive and Receptive) and nonverbal (Fine Motor and Visual Reception) abilities. Historical data indicate that the standard deviation of the distribution of RIQ does not remain constant with age, meaning that the same RIQ has a different meaning at different ages (Sattler, 2008). Still, armed with few other options, researchers, clinicians, and educators in the developmental disabilities field continue to use the RIQ. Thus, more data are needed to lend empirical support to this practice.

In the current study, we address these paucities in the literature: we evaluate the relationship between the MSEL and the DAS-II, and explore the impact of the use of RIQ. We replicate and extend the concurrent validity analyses between the MSEL and DAS done by Bishop and colleagues (2011), using the newer DAS-II and a mixed sample representative of all levels of ability and disability. We hypothesized that the strong correlations between the MSEL RIQ and the DAS IQ would be confirmed; however, based on clinical experience, we hypothesized a mean difference in scores, favoring the DAS-II. To evaluate these hypotheses, we predicted DAS-II IQ and RIQ scores from MSEL IQ and RIQ scores.

Method

Participants

The sample included children aged 2–10 years who either screened for or participated in a longitudinal study of autism at the Pediatrics and Developmental Neuroscience Branch of the National Institute of Mental Health at the National Institutes of Health, Bethesda, MD

(ClinicalTrials.gov ID: NCT00271622). This study was approved by the Institutional Review Board of the Neuroscience Institutes of the National Institutes of Health, and informed consent (and assent from the child when applicable) was obtained for all participants. The study comprised a combined sample of children classified into three study groups: ASD, non-ASD developmental delay (DD), and typical development (i.e., without DD; TD). Children with ASD and DD were recruited based on referral for ASD or delay; typically developing children were recruited based on a lack of developmental concerns.

Inclusion criteria for all groups were that English was the primary language in the home and that the child was independently mobile (required by portions of the assessment battery). An additional inclusion criterion for the ASD group was a diagnosis of DSM-IV-TR autistic disorder or pervasive developmental disorder-not otherwise specified (PDD-NOS). These diagnoses were established by best-estimate clinical judgment of doctoral-level clinicians, using DSM-IV-TR criteria and cutoff scores for autism and PDD-NOS (Risi et al., 2006) on the Autism Diagnostic Interview-Revised (Rutter et al., 2003) and the Autism Diagnostic Observation Schedule (Lord, Rutter, DiLavore, & Risi, 1999). Members of the DD group received the same ASD diagnostic battery, but ASD was ruled out. Inclusion criteria for the DD group included either (a) cognitive scores (domains or portions of domains) more than 1.5 standard deviations below the mean on the MSEL or (b) ADI-R or ADOS scores above the ASD cutoff (indicating social communication impairment). All cognitive and diagnostic evaluations for children in the ASD and DD group were conducted by experienced doctoral-level clinicians, with some cognitive testing in the typical group done by trained research assistants with on-site doctoral-level supervision. All clinicians met standard requirements for research reliability on the relevant instruments.

The diagnostic evaluations performed for the ASD, DD, and TD groups included a hierarchy of cognitive tests. This hierarchy started with the MSEL and progressed through the Early Years and School Age forms of the DAS-II, depending on the child's age and demonstrated language ability. During the course of the larger longitudinal study, an attempt was made to administer the DAS-II to any child that was close to the maximum age range for the MSEL, so as to have data on both measures on at least one time point. Thus, a subgroup of participants in the study had contemporaneous (within 1 month; 93% within 3 days) administration of the MSEL and DAS-II. This was the inclusion criterion for the current analyses. In about 75% of cases, the MSEL was administered prior to the DAS-II, though this was not dictated by the testing protocol.

Measures

Mullen Scales of Early Learning (MSEL; Mullen, 1995)—The MSEL is a standardized developmental test for children birth to 5 years, 8 months. The MSEL was standardized on a sample of 1,849 children, which excluded children with known physical or mental disabilities. The test manual reports moderate correlations with scores from the *Bayley Scales of Infant Development* (Bayley, 1969) and the *Preschool Language Assessment* (PLA; Zimmerman, Steiner, Evatt, & Pond, 1979).

The MSEL yields an overall standard score (the Early Learning Composite, ELC) and does not have empirically derived verbal and nonverbal domain scores. T-scores (range 20–80)

are available for the subdomains: Gross Motor (through 33 months), Visual Reception, Fine Motor, Receptive Language and Expressive Language (the later four are combined for the ELC). In the current study, the Fine Motor and Visual Reception domains were used to calculate a nonverbal composite ratio score (NV-RIQ) and the Receptive Language and Expressive Language subscales were used to calculate a verbal composite ratio score (V-RIQ) using the average age equivalents (reported in 1-month increments). MSEL IQs were also estimated, by combining T-scores for the same areas and converting to standard scores.

Differential Ability Scales, Second Edition (DAS-II; Elliot, 2007)—The DAS-II is an individually administered measure of cognitive functioning for preschool and school-age children, ages 2 years, 6 months to 17 years, 11 months. The DAS-II manual (Elliot, 2007) cites generally high correlations with the Wechsler scales, but moderate correlations with the *Bayley-III*, which produced higher scores than the DAS-II in a sample of 54 typically developing preschoolers. The DAS-II was also reported to be well-correlated with the original DAS (Verbal Composite, $r=.60$; Special Nonverbal Composite, $r=.85$) (Elliot, 2007). The DAS-II produces a General Conceptual Ability score corresponding to full scale IQ, as well as standard scores on Verbal (VIQ) and Nonverbal (NVIQ) domains. On the DAS-II, the nonverbal domain for the Upper ages (3 years, 5 months to 6 years, 11 months, with standardization up to 8 years, 11 months) consists of either the Special Nonverbal Composite (SNC; Picture Similarities, Pattern Construction, Matrices and Copying) or the Nonverbal Reasoning Cluster (Pattern Construction and Matrices). The Nonverbal cluster for the Lower ages (2 years, 6 months to 3 years, 5 months) consists of Picture Similarities and Pattern Construction. Nonverbal Reasoning was used only when the SNC was not available. The Verbal cluster is comprised of the Verbal Comprehension and Naming Vocabulary subtests, regardless of age or ability level. The alternative stop points suggested on the DAS-II were not used. Standard scores (IQs) on the DAS-II range from 30–170. In addition to NVIQ and VIQ, RIQs were calculated for the DAS-II, using the age equivalents that are generally reported in 3 month increments. When age equivalent scores were either at the floor or the ceiling, 3 months were subtracted or added, respectively, from the nearest age equivalent.

Data Analysis

SAS/STAT and SAS/GRAPH Version 9.3 software (SAS Institute, 2012) was used to estimate the best-fitting line to fit DAS-II IQ or RIQ versus MSEL RIQ, separately for verbal and nonverbal scores. Polynomial terms were added to the model where indicated by non-randomly distributed residuals, and these terms were retained if they were statistically significant ($p < .01$). In this sample, the participants with higher IQ were younger than those with lower IQ, generating a significant but irrelevant correlation between IQ and age. Thus, age (in months) was entered as a covariate in all analyses. Centering at some value is often indicated when introducing polynomial terms (to reduce collinearity) or when zero is not a meaningful value for the independent variable (Kraemer & Blasey, 2004). Age was centered at the grand mean. All analyses were run with and without mean-centered DAS/MSEL scores and no meaningful effects on the parameters were observed. Further, we could find no satisfactory value at which to center the DAS/MSEL scores in order to make the results

more interpretable than when they were centered naturally at zero. Therefore, we report results and show figures using uncentered DAS/MSEL scores.

Results

A total of 118 participants had contemporaneous administration of the MSEL and DAS-II: $n = 47$ children with ASD ($n = 44$ autism, $n = 3$ PDD-NOS), $n = 28$ children with DD, and $n = 43$ TD (see Table 1). Participants in the DD group had (a) overall cognitive scores at least 1.5 standard deviations below the mean and below-threshold scores on the ADOS and ADI-R ($n=16$), (b) overall cognitive scores at least 1.5 standard deviations below the mean and either the ADOS or ADI-R was above threshold, but clinicians were confident that the child did not have ASD ($n=8$), or (c) a specific delay in either social communication (score above the threshold on ADOS or ADI-R but no ASD; $n=1$) or scores at least 1.5 standard deviations below the mean on nonverbal, expressive, or receptive domains ($n=3$).

The focus of these analyses was not to demonstrate differences between diagnoses, so we explored the advisability of combining the study groups. In order to determine whether the combination of groups was warranted, we examined by group the mean differences between the MSEL and DAS-II using RIQ and standard scores (Table 2), and found only one out of six comparisons contained a difference between diagnostic groups (DAS NVIQ – MSEL NVIQ).

Pearson correlations between DAS-II IQ and MSEL RIQ were significant ($p < .001$) and strong ($r_{\text{nonverbal}}=.90$; $r_{\text{verbal}}=.91$), though the mean difference in scores, favoring the DAS-II, was about 10 points all comparisons (Table 2). While there were no DAS-II NVIQ scores that were 15 or more points lower than MSEL NV-RIQ scores, 31% ($n=37$) of the full sample had DAS-II NVIQ scores at least 15 points higher than the MSEL NV-RIQ. In the full sample, two (2%) DAS-II VIQ scores were at least 15 points lower than MSEL V-RIQ, while 41% ($n=49$) of DAS VIQ scores were at least 15 points higher than MSEL V-RIQ scores.

Residuals from the linear regression of DAS-II VIQ on MSEL V-RIQ were fan-shaped and indicated that a quadratic term better explained the variation. The best-fit quadratic equation accounted for 87% of the variance [see solution inset in Figure 1, all terms $p < .001$; model $F(3,112)=243.32$, $MSE = 104.94$, $p < .001$]. However, Figure 1 illustrates a mean difference in scores between the DAS-II and the MSEL that varied quadratically by score. Table 3 illustrates the mean predicted DAS-II score, controlling for age, for various MSEL scores. For nonverbal scores, the residuals from the linear regression were again non-randomly distributed and indicated that a quadratic term was necessary. The best-fit quadratic equation for DAS-II NVIQ on MSEL NV-RIQ explained 83% of the variance (see equation inset in Figure 2, $p_{\text{AGE}} = .04$, all other terms $p < .005$; model $F(3,111)=184.30$, $MSE = 106.87$, $p < .001$). As shown in Figure 2, the mean difference between the DAS-II NVIQ and MSEL NV-RIQ varied quadratically by score (predicted values, controlling for age, are found in Table 3).

We explored the possibility that the mean difference in MSEL and DAS-II scores was observed because the MSEL was calculated as a RIQ and the DAS-II was a standard score. We repeated the analyses using DAS-II RIQ instead of DAS-II IQ (Figure 3). Within the DAS-II, the IQ and RIQ were highly correlated ($r_V = .97, p < .001$; $r_{NV} = .98, p < .001$). The linear equation was best-fit to the nonverbal data, explaining 83% of the variance in MSEL NV-RIQ predicting DAS-II NV-RIQ [Figure 3, left panel; model $F(2,115)=275.99, MSE = 127.11, p < .001$; age *n.s.*, all other terms $p < .001$]. Although the slope was nearly 1.0, the intercept was 12.70, indicating a large mean difference between DAS-II NV-RIQ and MSEL NV-RIQ scores (see Table 3 for predicted values controlling for age). For Verbal scores, the linear equation explained 84% of the variance and the residuals did not suggest the need for polynomial terms [Figure 3, right panel; model $F(2,115)=322.77, MSE = 205.05, p < .001$; age $p = .02$, all other terms $p < .001$]. Mean differences were diminished for individuals with lower scores (controlling for age, an MSEL V-RIQ score of 40 corresponded to a DAS-II V-RIQ score of 39), but were maintained for higher scores (Table 3).

It was also possible to evaluate the effect of RIQ versus IQ by using the estimates of MSEL IQ scores for the $n=79$ participants who received them. These participants were aged less than 69 months and had standard scores above the floor: AUT, $n=17$; DD, $n=21$; TYP, $n=41$. Thus, we repeated the analyses using DAS-II IQ and MSEL IQ (Figure 4). MSEL T-scores were converted to standard score for comparison on the standard scale. Within the MSEL, the IQ and RIQ were highly correlated ($r_V = .97, p < .001$; $r_{NV} = .98, p < .001$). The quadratic equation was best-fit to the nonverbal data, explaining 69% of the variance in MSEL NVIQ predicting DAS-II NVIQ (Figure 4, left panel; age *n.s.*, other terms $p < .001$; model $F(3,74) = 53.81, MSE = 103.65, p < .001$). For Verbal scores, the quadratic equation explained 71% of the variance [Figure 4, right panel; all terms $p < .001$; model $F(3,72) = 57.95, MSE = 101.28, p < .001$]. Predicted DAS-II scores, controlling for age, are shown in Table 3.

Discussion

In this study, we evaluated the concurrent validity of the *Mullen Scales of Early Learning* (MSEL) and the *Differential Ability Scales, 2nd edition* (DAS-II), due to the frequency with which these tests are employed, and often combined or compared, in studies of children with ASD and other developmental disorders. The strong correlation between the instruments suggests that they measure similar constructs, and echoes the results using the previous version of the DAS (Bishop et al., 2011). The correlations in this study are of larger magnitude than those reported in the DAS-II manual between the DAS-II and the Bayley-III, which is comparable in content to the MSEL (Elliott, 2007). The DAS-II differs from the original DAS primarily in the composition of the nonverbal composite, but this did not appear to affect the correspondence to the MSEL based on simple correlation.

Based on clinical experience, we hypothesized that the DAS-II would yield higher scores than the MSEL. This hypothesis was supported, suggesting that although the instruments appear to measure the same construct, they do this on a different scale. Notably, this effect was observed across diagnostic groups and ability levels.

No two tests are expected to yield exactly the same scores; derivation IQ scores are estimates that fall somewhere within the range of the standard error of measurement, on either side of the point estimate (Sattler, 2008). Controlling for age, the mean difference between tests did vary by ability level, such that mean differences of approximately 10 to 20 points were observed for children with MSEL scores greater than 70. Mean differences in favor of the DAS-II were still present but smaller in magnitude for individuals with scores less than 70. These differences were observed regardless of whether IQ or RIQ was used for DAS-II or MSEL. It is worth noting that for higher-scoring children who are older, a ceiling effect on the MSEL does exist (Mullen, 1995). Essentially, after the age of 49 months, it becomes impossible to obtain a T-score of 80 on most of the domains. This test limitation could only have contributed to test differences in a small segment of the sample.

These results were different but compatible with those reported by Bishop et al. (2011). In that study of children with and without neurodevelopmental disorders, mean *absolute* differences of 10.20 (nonverbal) and 7.89 (verbal) were observed between the DAS and MSEL. However, the mean *raw* differences (the metric used in this study) would have been much smaller, given that the participants were relatively evenly distributed between DAS>MSEL and MSEL>DAS (with the exception of those with very low or very high scores, where the DAS was generally higher). In the current study, the vast majority of participants with ASD or DD (about 90%), and the majority of the typical development group (about 80%), received a higher score on the DAS-II than on the MSEL. The combined ASD and DD sample in Bishop et al. (2011) was younger and had higher mean scores than the ASD and DD participants in the current sample. These sample characteristics may have contributed to differences between the studies. Further, Bishop et al. (2011) used multiple versions of the MSEL and the original version of the DAS; the DAS-II includes revisions to the nonverbal scale and updated norms.

The consistency with which the DAS-II scores were higher than the MSEL scores is an important finding. As described in a meta-analysis by Floyd, Clark, and Shadish (2008), most of the variation between tests in typical children is not attributable to systematic differences between testing batteries. In other words, Floyd and colleagues did not find evidence in the norms available for several widely used standardized tests that one testing battery was likely to systematically produce higher scores than another. In this study, however, there was evidence for a systematic difference between the DAS-II and MSEL.

The influence of the psychometric properties of ratio scores (RIQ) is hard to quantify but important to consider, given the frequency with which this non-standardized score is used for individuals with ASD and/or ID. The standard deviation of the RIQ is known to be larger than that of standard scores; in one longitudinal study between the ages of 6 years and 18 years, standard deviations of the Stanford-Binet RIQ varied from a low of 15.1 to a high of 23.6 (Bayley, 1949). This evidence was used to call for a transition to deviation IQ scores in the general population. Therefore, the magnitude of difference between the MSEL RIQ and the DAS-II IQ and RIQ may constitute normal variation. This hypothesis was not supported by our analyses of the subgroup of children with standard scores on the MSEL; large differences in favor of the DAS-II IQ were observed even when MSEL IQ was used instead of the RIQ. Still, the known age-related distributional aberrations of the RIQ are important

to consider, especially when analyzing alongside deviation IQ scores. Although a great deal of data on RIQ was produced during the early 20th century, modern psychometric data are required if researchers continue to use the ratio score.

Alternatives to the RIQ are available. For example, Fragile X researchers have accounted for floor effects in some common intelligence tests by developing alternative deviation scoring systems (Hessl et al., 2009; Sansone et al., 2014), though this approach has not yet been applied to pre-school testing. In the context of research studies, a predicted score might be used (Whitaker & Gordon, 2012). For example, standard scores may be regressed on age equivalents for the children able to achieve a standard score. The resulting equation may be used to produce predicted values on the standard scale for children unable to achieve a norm-based standard score, based on their age equivalent.

Because age was correlated with IQ and RIQ scores, age was entered as a covariate in each of the models. The most likely explanation for why age was correlated with IQ and RIQ was that children with TD (and higher cognitive ability) were younger than the remainder of the sample. Although the goal was simply to control for any incidental effects of age, this information is important for those researchers and clinicians working with older children with and without disability. Future research should elucidate the relationship between age and the correspondence between MSEL and DAS-II scores.

Unfortunately, a limitation of this study is that diagnostic group in this study is conflated with both age and ability level, so we were unable to determine the unique contributions of age and diagnosis. Researchers have long noted the possibility that deficits in social communication associated with ASD may obfuscate the true ability of a child as measured by traditional IQ tests (Akshoomoff, 2006; Rapin, 2003). Generally, the results of this study do not support the notion that the relationship between the MSEL and the DAS-II is a function of diagnosis (ASD versus DD versus TYP). Large mean differences were observed across the entire sample, regardless of diagnosis. Still, most of the best-fit lines were curvilinear, indicating that the mean difference was smaller in the lower and higher reaches of ability, controlling for age. Future studies with neurodevelopmental disorder and typical groups that overlap in both age and ability level will further elucidate this issue.

Generalization of these results is limited by the composition of the sample. By necessity, this sample only included children to whom both the MSEL and DAS-II were administered successfully within a very short time period. Thus, we excluded children for whom behavioral problems (including autism symptom severity) or cognitive limitations contraindicated administration of both tests. It is worth noting that the ASD group had a high proportion of low IQ (regardless of which test score was used) and may therefore not be representative of the broad ASD population. Likewise, many members of the typically developing group achieved very high scores. Finally, the order of the tests was not balanced; in most cases, the MSEL was administered first. Although this may have had an effect on the outcome, one would expect fatigue to cause lower scores on the second test, contrary to what we observed.

These results have considerable implications for clinical practice, research design, and data analysis. Clinically, while test selection is always an important decision when assessing young children, choosing between the DAS-II or MSEL may have a greater impact than a decision between two other tests. This choice is especially common when working with children who have developmental delay. The successful administration of the MSEL may be more likely when testing children with ASD or other developmental disorders, but successful administration of the DAS-II is likely to produce higher scores. Clinicians or educators may initially assess a child with the MSEL, owing to age or ability, and then switch to the DAS-II at later assessments. The results of this study, while requiring replication, suggest that this may result in the false impression that a child has improved considerably over time. In a research setting where the two tests are interchanged, if the test type is not balanced between groups, group differences in IQ could be misattributed to an external variable or treatment. The degree of difference seems to be associated with ability, such that children on either extreme of ability are less affected. This fact, combined with the significant inter-individual variability, makes a simple correction for the general tendency of higher DAS-II scores ill-advised.

Acknowledgments

This work was supported by the Intramural Program of the National Institute of Mental Health of the National Institutes of Health, NCT00298246, 06-M-0102. The views expressed in this paper do not necessarily represent the views of the NIMH, NIH, HHS, or the United States Government. The authors extend their gratitude to the children and their families who volunteered their time and efforts during the research.

References

- Akshoomoff N. Use of the Mullen Scales of Early Learning for the assessment of young children with Autism Spectrum Disorders. *Child Neuropsychology*. 2006; 12(4–5):269–277. [PubMed: 16911972]
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders, DSM-IV-TR. Washington, DC: American Psychiatric Publishing; 2000.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5. Arlington, VA: American Psychiatric Publishing; 2013.
- Anderson DK, Liang JW, Lord C. Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. 2014; 55(5):485–494.
- Baum KT, Shear PK, Howe SR, Bishop SL. A comparison of WISC-IV and SB-5 intelligence scores in adolescents with autism spectrum disorder. *Autism*. 2014;1–10.
- Bayley N. Consistency and variability in the growth of intelligence from birth to 18 years. *Journal of Genetic Psychology*. 1949; 75(2):165–196. [PubMed: 15403674]
- Bayley, N. Manual for the Bayley Scales of Infant Development. New York: Psychological Corporation; 1969.
- Bishop SL, Guthrie W, Coffing M, Lord C. Convergent validity of the Mullen Scales of Early Learning and the differential ability scales in children with autism spectrum disorders. *American Association on Intellectual and Developmental Disabilities*. 2011; 116(5):331–343. DOI: 10.1352/1944-7558-116.5.331
- Bölte S, Dziobek I, Poustka F. Brief Report: The Level and Nature of Autistic Intelligence Revisited. *Journal of autism and developmental disorders*. 2009; 39(4):678–682. [PubMed: 19052857]
- Dawson M, Soulières I, Gernsbacher M, Mottron L. The Level and Nature of Autistic Intelligence. *Psychological Science*. 2007; 18(8):657–662. [PubMed: 17680932]

- Dykens, EM., Lense, M. Intellectual disabilities and autism spectrum disorder: A cautionary note. In: Amaral, D., Geschwind, D., Dawson, G., editors. *Autism Spectrum Disorders*. New York: Oxford; 2012. p. 263-284.
- Elliott, CD. *Manual for the Differential Ability Scales, Second Edition*. San Antonio, TX: Harcourt Assessment; 2007.
- Floyd RG, Clark M, Shadish WR. The exchangeability of IQs: Implications for professional psychology. *Professional psychology: research and practice*. 2008; 39(4):414.
- Frazier TW, Georgiades S, Bishop SL, Hardan AY. Behavioral and Cognitive Characteristics of Females and Males With Autism in the Simons Simplex Collection. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2014; 53(3):329–340. e323. [PubMed: 24565360]
- Hessl D, Nguyen D, Green C, Chavez A, Tassone F, Hagerman R, ... Hall S. A solution to limitations of cognitive testing in children with intellectual disabilities: the case of fragile X syndrome. *Journal of Neurodevelopmental Disorders*. 2009; 1(1):33–45. [PubMed: 19865612]
- Kraemer HC, Blasey CM. Centring in regression analyses: A strategy to prevent errors in statistical inference. *International Journal of Methods in Psychiatric Research*. 2004; 13(3):141–151. [PubMed: 15297898]
- Kuriakose S. Concurrent Validity of the WISC-IV and DAS-II in Children With Autism Spectrum Disorder. *Journal of Psychoeducational Assessment*. 2014; 32(4):283–294.
- Lord C, Petkova E, Hus V, Gan W, Lu F, Martin DM, ... Risi S. A multisite study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry*. 2012; 69(3): 306–313. [PubMed: 22065253]
- Lord, C., Rutter, M., DiLavore, PC., Risi, S. *Autism Diagnostic Observation Schedule (ADOS)*. Los Angeles, California: Western Psychological Services; 1999.
- Lord C, Schopler E. The role of age at assessment, developmental level, and test in the stability of intelligence scores in young autistic children. *Journal of Autism and Developmental Disorders*. 1989; 19(4):483–499. [PubMed: 2606880]
- Magiati I, Howlin P. Monitoring the progress of preschool children with autism enrolled in early intervention programmes: problems in cognitive assessment. *Autism*. 2001; 5(4):399–406. [PubMed: 11777256]
- Mullen, EM. *Mullen scales of early learning*. Circle Pines, MN: American Guidance Service; 1995.
- Rapin I. Value and limitations of preschool cognitive tests, with an emphasis on longitudinal study of children on the autistic spectrum. *Brain and Development*. 2003; 25(8):546–548. [PubMed: 14580667]
- Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, ... Pickles A. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2006; 45(9):1094–1103. [PubMed: 16926617]
- Roid, GH. *Stanford-Binet Intelligence Scales*. 5. Itasca, IL: Riverside Publishing; 2003.
- Rutter, M., LeCouteur, A., Lord, C. *Autism Diagnostic Interview-Revised (ADI-R)*. Los Angeles, CA: Western Psychological Services; 2003.
- Sansone S, Schneider A, Bickel E, Berry-Kravis E, Prescott C, Hessl D. Improving IQ measurement in intellectual disabilities using true deviation from population norms. *Journal of Neurodevelopmental Disorders*. 2014; 6(16)
- SAS Institute I. *SAS (Version 9.3)*. Cary, NC: SAS Institute, Inc; 2012.
- Sattler, JM. *Assessment of children: Cognitive foundations*. 5. La Mesa, CA: Jerome M. Sattler, Publisher, Inc; 2008.
- Sigman M, McGovern C. Improvement in cognitive and language skills from preschool to adolescence in autism. *Journal of Autism and Developmental Disorders*. 2005; 35(1):15–23. [PubMed: 15796118]
- Wechsler, D. *Wechsler Preschool and Primary Scale of Intelligence*. 4. San Antonio, TX: Psychological Corporation; 2012.
- Whitaker S, Gordon S. Floor effects on the WISC-IV. *International Journal of Developmental Disabilities*. 2012; 58(2):111–119.

Zimmerman, IL., Steiner, VG., Evatt, RL., Pond, RE. Preschool Language Assessment. Columbus, Ohio: Charles E. Merrill; 1979.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Regression lines reflect age (in months; centered at grand mean of 63.67) entered as a covariate. Dotted line represents a 1:1 relationship between MSEL and DAS-II.

Figure 2.

Regression lines reflect age (in months; centered at grand mean of 63.67) entered as a covariate. Dotted line represents a 1:1 relationship between MSEL and DAS-II.

Figure 3.

Regression line reflects age (in months; centered at grand mean of 63.67) entered as a covariate for Verbal; age not significant in Nonverbal equation. Equation for Nonverbal (left panel): $Y = 15.30 + 0.98X$; $R^2 = .83$. Equation for Verbal (right panel): $Y = -14.53 + 0.31AGE + 1.33X$; $R^2 = .85$. Dotted line represents perfect 1:1 relationship between DAS-II and MSEL.

Figure 4.

Nonverbal, N=78 (AUT=16, DD=21, TYP=41); Verbal, N=76 (AUT=14, DD=21, TYP=41).

Regression line reflect age (in months; centered at grand mean of 55.32) entered as a covariate for Verbal; age not significant in Nonverbal equation. Equation for Nonverbal (left panel): $Y = -19.51 + 2.04X - 0.0071X^2$; $R^2=.69$. Equation for Verbal (right panel): $Y = -26.58 + 0.59AGE + 2.20X - 0.0077X^2$; $R^2=.71$. Dotted line represents perfect 1:1 relationship between DAS-II and MSEL.

Table 1

Participant characteristics

	TOTAL	ASD	DD	TYP
N	118	47	28	43
In MSEL Age Range (<69 mo.)	84 (71%)	21 (45%)	22 (77%)	41 (95%)
Age, months [/]	63.67 ± 15.13	74.20 ± 14.77	61.74 ± 12.44	53.42 ± 8.04
Male	85 (72%)	39 (83%)	19 (68%)	27 (63%)
DAS-II Scores				
NVIQ	92.47 ± 24.95	78.93 ± 21.95 _a	77.64 ± 17.1 _a	115.95 ± 10.23 _b
NVDQ	94.65 ± 26.92	78.77 ± 22.37 _a	80.95 ± 16.72 _a	120.94 ± 13.25 _b
VIQ	87.74 ± 27.71	67.20 ± 23.65 _a	81.11 ± 18.31 _b	113.56 ± 11.77 _c
VDQ	86.69 ± 36.51	59.89 ± 25.31 _a	74.49 ± 22.44 _b	123.94 ± 19.07 _c
MSEL Scores				
NVDQ	81.14 ± 25.04	64.40 ± 20.8 _a	70.99 ± 13.63 _a	106.05 ± 11.65 _b
VDQ	76.17 ± 27.89	54.76 ± 20.0 _a	66.74 ± 15.1 _b	105.72 ± 10.7 _c
Race				
White	93 (79%)	37 (79%)	19 (68%)	37 (86%)
Black/African-American	10 (9%)	5 (11%)	4 (14%)	1 (2%)
Asian-American	6 (5%)	3 (6%)	2 (7%)	1 (2%)
Other/multiple	9 (8%)	2 (4%)	3 (11%)	4 (9%)
ADOS Calibrated Severity Score	--	6.98 ± 1.89	1.75 ± 0.89	n/a
Social Affect CSS	--	6.62 ± 2.08	1.68 ± 0.77	n/a
Restricted/Repetitive Behavior CSS	--	7.79 ± 1.9	4.96 ± 2.27	n/a

[/] Groups differ significantly on age (all pairwise $p < .01$).

^{a,b,c} Different subscripts denote means that differ significantly ($p < .01$) from one another between diagnostic groups. Lack of subscripts indicates that the omnibus test was not statistically significant (i.e., no group differences).

Note: All participants were in the DAS-II age range. ASD participants were diagnosed with either DSM-IV-TR autistic disorder ($n=44$, 94%) or pervasive developmental disorder-not otherwise specified ($n=3$, 6%). The omnibus test of differences between groups was not statistically significant for sex or race (White versus non-White). Sample size for DAS-II in the ASD group differs: $n=44$ DAS-II NVIQ and $n=45$ DAS-II VIQ.

Table 2

Mean differences between scores on the MSEL and DAS-II

	TOTAL	ASD	DD	TYP
DAS NVIQ - MSEL NV-RIQ	10.34 ± 10.93	13.11 ± 10.30	6.65 ± 9.95	9.90 ± 11.60
DAS NVDO - MSEL NV-RIQ	13.51 ± 11.19	14.36 ± 10.74	9.96 ± 9.93	14.89 ± 12.18
DAS VIQ - MSEL V-RIQ	10.69 ± 11.56	11.13 ± 10.23	14.37 ± 9.99	7.84 ± 13.21
DAS VDO - MSEL V-RIQ	10.52 ± 15.56	5.12 ± 11.53 _a	7.75 ± 13.80 _a	18.22 ± 17.55 _b
DAS NVIQ - MSEL NVIQ [/]	12.06 ± 12.24	17.19 ± 10.07	12.43 ± 10.04	9.88 ± 13.56
DAS VIQ - MSEL VIQ [/]	11.20 ± 12.21	11.36 ± 9.04	16.62 ± 9.87	8.37 ± 13.44

[/] Sample size differs where MSEL standard scores were not available. Verbal: ASD, n = 14; DD, n=21; TYP, n=41. Nonverbal: ASD, n=16; DD, n=21; TYP, n=41.

^{a,b,c} Different subscripts denote means that differ significantly ($p < .01$) from one another between diagnostic groups. Lack of subscripts indicates that the omnibus test was not statistically significant (i.e., no group differences).

Note: IQ = Intelligence Quotient (SS); RIQ = Ratio IQ; NV = Nonverbal; V = Verbal.

Table 3

Predicted values of DAS-II scores given MSEL standard scores, controlling for age

	MSEL = 40	MSEL = 70	MSEL = 100
1. DAS-II NVIQ : MSEL NV-RIQ	45	83	112
2. DAS-II VIQ : MSEL V-RIQ	46	86	113
3. DAS-II NV-RIQ : MSEL NV-RIQ	54	84	114
4. DAS-II V-RIQ : MSEL V-RIQ	39	79	118
5. DAS-II NVIQ : MSEL NVIQ	--	87	112
6. DAS-II VIQ : MSEL VIQ	--	85	111

Note: Using the equations generated, DAS-II scores are calculated while age is held constant at the mean value of 63.67 months (comparisons 1–4) and 55.32 months (comparisons 5 and 6).