

Projected Retained Ability Score (PRAS): A New Methodology for Quantifying Absolute Change in Norm-Based Psychological Test Scores Over Time

Assessment
2021, Vol. 28(2) 367–379
© The Author(s) 2019



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1073191119872250
journals.sagepub.com/home/asm



William G. Kronenberger¹, Magdalena Harrington²,
and Karen S. Yee²

Abstract

A limitation of norm-based ability test scores is that they can only be used to evaluate relative change (compared with change in the norm sample), as opposed to absolute (raw) change in performance from Time 1 to Time 2. To address this limitation, a novel method (Projected Retained Ability Score [PRAS]) was developed to characterize absolute change in norm-based ability test scores. The PRAS method was applied to Differential Ability Scales®–Second Edition (DAS-II) General Conceptual Ability (GCA) scores in three cases of children with the neurodegenerative condition mucopolysaccharidosis type II (MPS II) who were assessed at two visits, 16 to 23 months apart. Although all three cases showed declines in norm-based GCA scores, the PRAS method revealed differences in absolute change in performance. The PRAS method allows for differentiation of slower-than-average improvement or stabilization versus deterioration of cognitive ability when norm-based scores decline from Time 1 to Time 2.

Keywords

cognitive ability, GCA score, Hunter syndrome, mucopolysaccharidosis type II, Projected Retained Ability Score

Ability test scores based on age norms (“norm-based scores”) have well-known statistical and interpretative advantages compared with raw scores (Anastasi, 1988). These advantages include documentation of abnormal performance, comparison of a pattern of ability strengths and weaknesses across subtests, and identification of domains in need of special accommodations or interventions (Anastasi, 1988; Wechsler, 2014). When comparing test results obtained at 2 points in time (from Time 1 to Time 2), norm-based scores also allow for the evaluation of change in performance relative to the typical trajectory of change in the same-aged peer group. However, norm-based scores also have some limitations in the assessment of change in performance from Time 1 to Time 2. Specifically, norm-based scores do not provide information about absolute (raw) change in performance from Time 1 to Time 2 because they are scaled to the performance of a same-aged norm sample.

This limitation of norm-based scores is of particular significance for the clinical and research utility of cognitive ability tests in children who are at risk for decline in cognitive ability with increasing age, such as children with neurodegenerative conditions, severe neurological disorders, and central nervous system insult/injury (e.g., toxic exposure, head injury). In such instances, the question is not only

change relative to the development shown by same-aged peers (represented by norm-based scores) but also change relative to the individual’s earlier level of functioning (absolute change). In this article, we propose a method for using norm-based scores to characterize absolute change from Time 1 to Time 2 and demonstrate the use of this method with case examples of change in Differential Ability Scales®–Second Edition (DAS-II; Elliott, 2007a) General Conceptual Ability (GCA) scores in three patients with the rare X-linked inherited lysosomal disease, Hunter syndrome (mucopolysaccharidosis type II [MPS II]).

Characterization of Change in Cognitive Ability Over Time

Absolute change refers to change over time compared with an unchanging metric or widely accepted benchmark, and is

¹Indiana University School of Medicine, Indianapolis, IN, USA

²Shire, a Takeda company, Lexington, MA, USA

Corresponding Author:

William G. Kronenberger, Department of Psychiatry, Indiana University School of Medicine, 355 West 16th Street, Indianapolis, IN 46202, USA.
Email: wkronenb@iupui.edu

assessed using raw performance on a test (e.g., using raw scores). It can be defined as a change in raw score or absolute performance from an earlier baseline point in time (Time 1) to a later follow-up point in time (Time 2), independent of the typical change seen in the normative population over the same duration (Burgaleta, Johnson, Waber, Colom, & Karama, 2014; Camp et al., 2005; Salthouse, Nesselroade, & Berish, 2006). In contrast, relative change refers to change over time compared with an evolving/dynamic or situation-specific benchmark that also changes during the same time period (most commonly, age). It is assessed using scores that characterize individual performance relative to a same-aged norm group, such as standard scores, *T*-scores, scaled scores, and *z* scores (norm-based scores), and can be defined as a change in performance relative to the average change in performance demonstrated by the normative population over the same period of time (Godfrey & Lee, 2018; Harris, Handleman, Gordon, Kristoff, & Fuentes, 1991; Petersen, 2004).

In early childhood, normal development in cognitive abilities, including attainment of cognitive milestones such as language, memory, and reasoning skills, is rapid and significant. Although the rate of cognitive development slows in adolescence and early adulthood, growth of fluid nonverbal cognitive abilities is well-documented into the early 20s (Luna, Garver, Urban, Lazar, & Sweeney, 2004; Segalowitz & Davies, 2004; Waber et al., 2007), and growth of crystallized verbal abilities persists well into adulthood (Kaufman, 1990). As a result, the typical child shows significant absolute change in performance on ability testing in the form of increases in raw scores with age (Elliott, 2007a; Wechsler, 2014). For example, a child with a raw score of 20 on a measure of global cognitive ability at Time 1 would, on average, be expected to attain a raw score above 20 at Time 2, reflecting normal developmental improvement in performance on ability test items.

In contrast, because age-based norm scores of cognitive ability (such as standard or scaled scores) already take into account developmental growth in the normative population, the typical child shows no relative change in norm-based scores over time. For example, a child who achieves an age-based norm standard score of 100 on a measure of global cognitive ability at Time 1 would, on average, be expected to achieve a standard score of approximately 100 on that measure at a later age (Time 2) if she showed an average amount of improvement in raw score performance over that same time period (mirroring the average raw score improvement of the age-based norm sample). Thus, while children's raw scores on an ability test are expected to increase with age (e.g., children answer more items correctly as they age), the corresponding standard (age-based norm) scores are not expected to increase because age-based norm scores account for normal developmental growth in raw scores.

Based on the differences between absolute change and relative change in cognitive ability during development, five potential patterns of change from baseline (Time 1) to follow-up (Time 2) in cognitive ability are possible: above-average rate of development, average rate of development, below-average rate of development, stabilized development, and deteriorating development (Table 1, Figure 1). Above-average rate of development is characterized by an improvement in absolute (raw) cognitive abilities at a faster rate than observed in normative peers (i.e., both an absolute and relative improvement have occurred). Average rate of development is an observed improvement in absolute (raw) cognitive abilities at the same rate as seen in normative peers (i.e., absolute improvement but relative stability). Below-average rate of development is an improvement in absolute (raw) cognitive abilities at a rate slower than observed in normative peers (i.e., an absolute improvement in cognitive abilities but a relative decline in norm-based scores). Stabilized development is neither an improvement nor a decline in absolute (raw) cognitive abilities compared with baseline (i.e., no absolute change in cognitive abilities despite a relative decline in norm-based scores). Deteriorating development is a decline in absolute (raw) cognitive abilities compared with baseline (i.e., both absolute and relative declines in cognitive abilities have occurred).

Because most composite/global ability test scores (including the DAS-II GCA and intelligence test composite scores such as Wechsler index and Full Scale IQ scores; Wechsler, 2014) are only available as norm-based standard scores and cannot be converted to or represented by raw scores, changes in global ability (e.g., GCA or Full Scale IQ) scores from Time 1 to Time 2 can only be used to evaluate relative change in cognitive abilities. An improvement (above-average rate of development) or no change (average rate of development) in global norm-based ability scores (such as the GCA or Full Scale IQ) from Time 1 to Time 2 indicates that the child has shown positive development in absolute (raw) cognitive skills that exceed or match the growth rate of the age-based norm sample. On the other hand, a decline in global norm-based ability scores from Time 1 to Time 2 indicates that the child's cognitive development is progressing at a slower rate than that of same-aged peers but does not provide information about the type of absolute (raw) change that has occurred in the child's cognitive development. Specifically, a decline in an individual child's norm-based ability scores from Time 1 to Time 2 could potentially reflect (a) an absolute improvement in cognitive abilities at a slower rate than the patient's peer group (below-average rate of development), (b) no change in absolute cognitive abilities (stabilized development), or (c) a decline in absolute cognitive abilities (deteriorating development). The use of norm-based scores alone does not permit differentiation between these three profiles

Table 1. Absolute and Relative Change in Cognitive Ability Test Scores.

Type of change from baseline to follow-up	Absolute change compared with patient's own baseline	Relative change compared with same-age norms	Ability RS change	Ability norm-based SS change	GCA or PRAS GCA
Above-average rate of cognitive development (i.e., development faster than average for patients of the same age)	Absolute positive development	Relative positive development	Follow-up RS > baseline RS	Follow-up SS > baseline SS	Follow-up GCA score > baseline GCA score
Average rate of cognitive development (i.e., development at same rate as average for patients of the same age)		Relative stabilization		Follow-up SS = baseline SS	Follow-up GCA score = baseline GCA score
Below-average rate of cognitive development (i.e., positive development seen relative to baseline but at a slower rate than average for patients of the same age)		Relative decline		Follow-up SS < baseline SS	Follow-up GCA score < baseline GCA score AND follow-up GCA score > PRAS GCA score
Stabilized cognitive development (i.e., no change relative to baseline but slower development than average for patients of the same age)	Absolute stabilization		Follow-up RS = baseline RS		Follow-up GCA score < baseline GCA score AND follow-up GCA score = PRAS GCA score
Deteriorating cognitive development (i.e., a decline relative to baseline)	Absolute decline		Follow-up RS < baseline RS		Follow-up GCA score < baseline GCA score AND follow-up GCA score < PRAS GCA score

Note. GCA = General Conceptual Ability; PRAS = Projected Retained Ability Score; RS = raw score; SS = standard score.

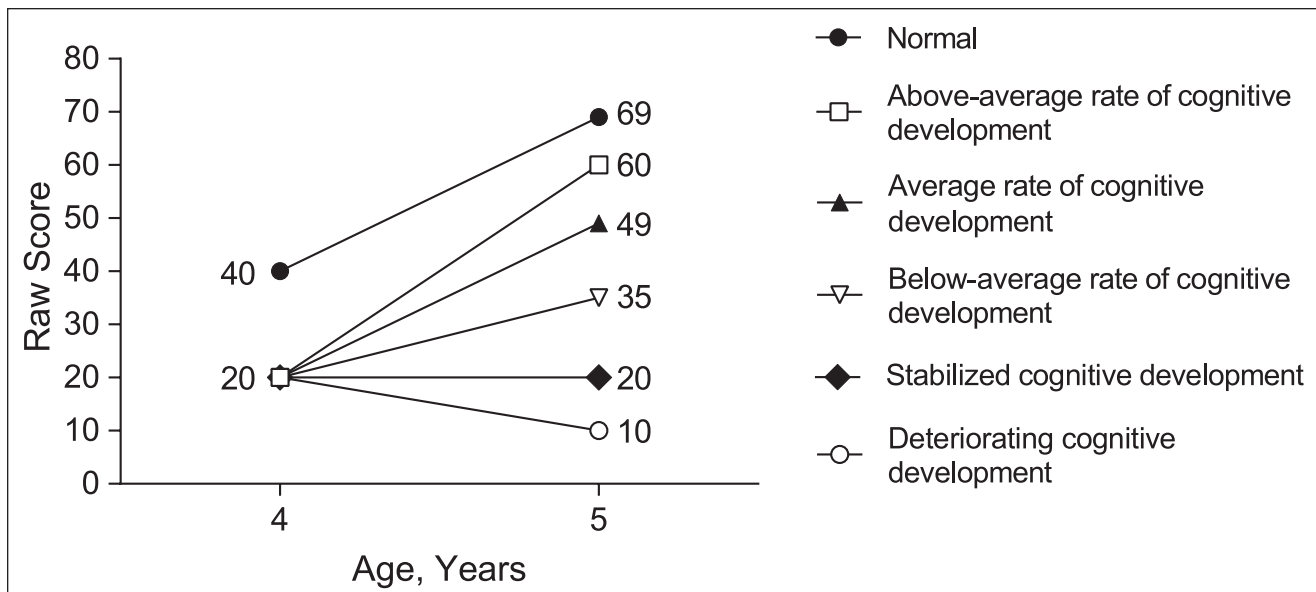


Figure 1. Five scenarios resulting from a raw score change from Time 1 to Time 2, as demonstrated by hypothetical raw scores on a cognitive ability test over a 12-month period from age 4 to 5 years.

Note. "Normal development" reflects development for the median value of the normative sample (e.g., corresponding to a standard score of 100), while "average rate of cognitive development" reflects development at the same rate as the normative sample but at a consistent lag (of 20 points in this example) from Time 1 to Time 2.

of absolute change, any one of which could result in a decline in norm-based ability scores. A decline in norm-based global ability scores from Time 1 to Time 2 is, therefore, ambiguous with regard to absolute change. Consequently, norm-based global ability scores alone are insufficient to fully characterize and understand changes from baseline when a decline in norm-based scores is present.

Assessment of Cognitive Change in Hunter Syndrome (MPS II)

Significant declines in norm-based ability scores are commonly found in children with neurodegenerative conditions such as MPS II (Hunter syndrome). MPS II is characterized by a deficiency of iduronate-2-sulfatase (Scarpa et al., 2011) and affects approximately one in 93,000 to 200,000 live births (Tomatsu et al., 2013). The disease, which varies widely in terms of age at first symptoms and rate of progression, affects multiple organs and systems (Scarpa et al., 2011). Somatic symptoms of MPS II include skeletal deformities, joint stiffness, hepatosplenomegaly, and heart disease (Guffon et al., 2015; Link et al., 2010). Some MPS II patients have initial symptoms that include mild to moderate learning difficulties during late childhood or adolescence without cognitive decline, but approximately two thirds of patients show neurocognitive impairment that appears at 1 year of age and then worsens around 2 to 4 years of age (Burton & Giugliani, 2012; Guffon et al., 2015; Shapiro, Escolar, Delaney, & Mitchell, 2017).

The progressive cognitive decline associated with MPS II can be assessed using neurocognitive ability testing (Janzen, Delaney, & Shapiro, 2017). Several different ability tests, including the DAS-II, Wechsler Intelligence Scale for Children (WISC; Wechsler, 2014), Kaufman Assessment Battery for Children (KABC-II; Kaufman & Kaufman, 2004), Bayley Scales of Infant Development (Bayley III; Bayley, 2006), and Mullen Scales of Early Learning (Mullen, 1995) have been used to assess neurocognitive ability in patients with MPS II (Janzen et al., 2017; Shapiro et al., 2017). Among these tests, the DAS-II has several advantages for assessing children with MPS II, including a broad age range (2.5-17.9 years), a low floor (>4 standard deviations below the normative mean for most composite scores), and flexible use of items specific to different levels of functioning. Several studies have used the DAS-II to assess cognitive ability in samples of children with MPS II, including a natural history study to monitor disease progression (Holt, Poe, & Escolar, 2011) and a study to examine the use of brain magnetic resonance imaging to identify cognitive impairment (Fan, Styner, Muenzer, Poe, & Escolar, 2010). Furthermore, the DAS-II was used to identify eligible patients for clinical trial studies investigating the safety and efficacy of intrathecal administration of idursulfase in pediatric patients with the neuronopathic phenotype of MPS II and early cognitive impairment (Muenzer et al., 2016; Muenzer et al., 2018).

Studies using neurocognitive measures such as the DAS-II, KABC-II, Bayley III, and Mullen scales to investigate change in cognitive functioning in children with neuronopathic MPS II have documented significant declines in

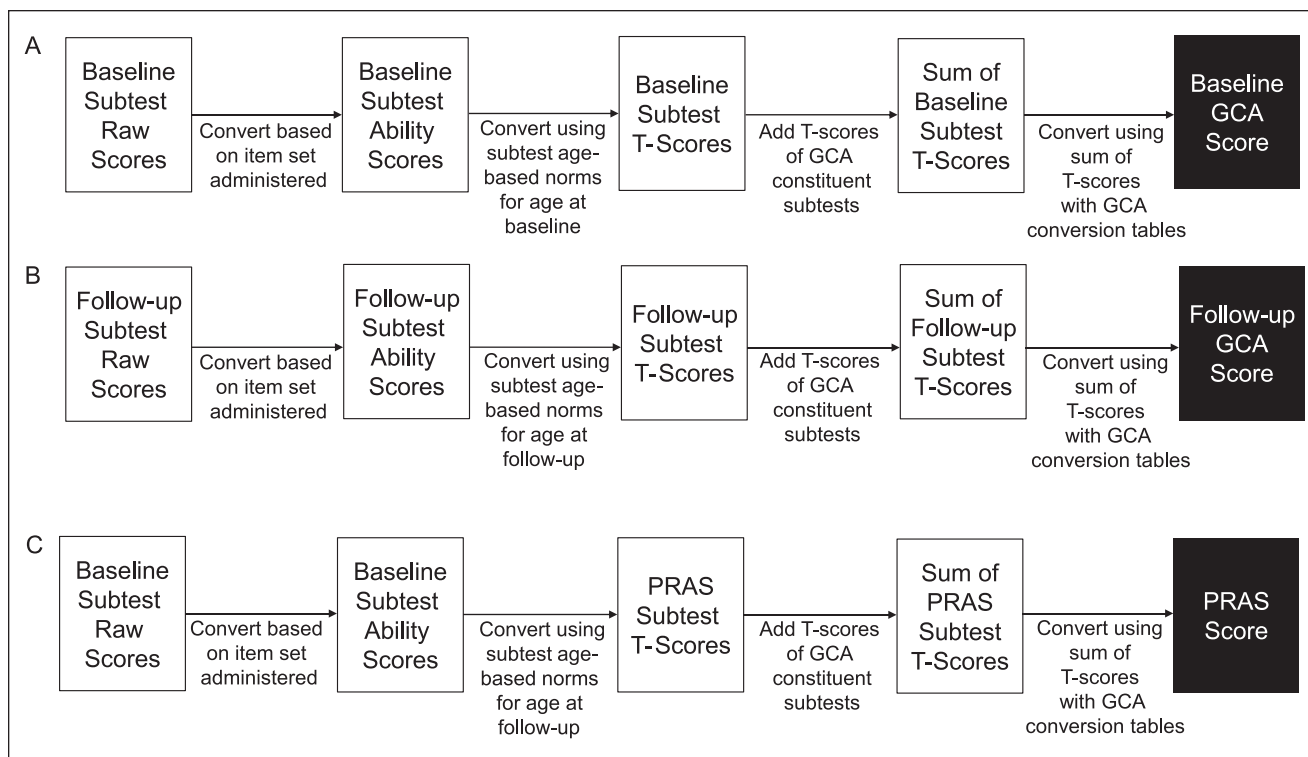


Figure 2. Calculation method for generating the DAS-II GCA scores and PRAS GCA scores from raw scores at baseline (Time 1) and at follow-up (Time 2); (A) baseline GCA; (B) follow-up GCA; and (C) PRAS GCA.

Note. DAS-II = Differential Ability Scales®—Second edition; GCA = General Conceptual Ability; PRAS = Projected Retained Ability Score. The only difference between the calculation of PRAS and GCA scores is that PRAS uses Time 1 (baseline) raw and ability scores but Time 2 (follow-up) age norms to establish the patient's follow-up T-scores and GCA that would be obtained if the patient achieved exactly the same raw scores at follow-up as at baseline. Follow-up score is compared with baseline GCA score to establish relative change from baseline or with PRAS score to establish absolute change from baseline.

norm-based scores of 30 to 50 standard score points at 2 to 4 years of age (Holt et al., 2011; Shapiro et al., 2017). Importantly, this decline in norm-based ability test scores occurred during a period of slowing or plateauing development for children with the neurodegenerative form of MPS II: “. . . the ‘plateauing’ seen in many patients aged between 2 and 4 years actually represents a decline of 50 points in IQ” (Holt et al., 2011, p. e1263). Shapiro et al. (2017) review data showing deficits in norm-based cognitive ability scores in MPS II patients during those ages, with even more significant declines in both raw and norm-based ability scores thereafter. Shapiro et al. (2017) further note that during this period, “standard scores (do not) impart knowledge of whether the child is still developing, is at a developmental standstill, or is losing milestones” (p. 9). Thus, norm-based scores have significant limitations in the assessment of children with MPS II because the decline in norm-based scores during critical early ages is ambiguous with regard to absolute change.

Differentiating between improvement and decline in raw/absolute cognitive abilities is vital in the assessment of children at neurodevelopmental risk, such as those with MPS II. Improvement in raw/absolute cognitive abilities,

even at a slower rate than same-aged peers, suggests that some positive development is continuing to occur, whereas a decline in raw/absolute cognitive abilities raises concerns about neurological and neuropsychological deterioration. Hence, a significant need exists for a method to allow the use of norm-based scores to characterize absolute change.

The Projected Retained Ability Score Method Procedure

In order to address limitations arising from the use of norm-based ability test scores in characterizing change from Time 1 to Time 2, particularly when neurodevelopmental risk is present as in MPS II, we propose a new method, the Projected Retained Ability Score (PRAS). The PRAS is an alternative method for calculating any norm-based score to enable the evaluation of an absolute change from baseline to follow-up (Yee, Kronenberger, & Harrington, 2018), which we apply here to the DAS-II GCA score as a specific example. The PRAS GCA score is derived by applying baseline (Time 1) GCA subtest raw/ability scores to norms for the patient's age at follow-up (Time 2), to produce a

projected score reflecting the follow-up GCA that would be obtained if the patient achieved exactly the same subtest raw/ability scores as at baseline. The PRAS GCA score can then be compared with the actual measured follow-up GCA score to evaluate absolute change from baseline. Comparing the follow-up GCA score with PRAS GCA score tests the hypothesis that no absolute change has occurred between baseline (Time 1) and follow-up (Time 2).

The three main steps of DAS-II GCA and PRAS scoring are depicted in Figure 2. Following the standard DAS-II scoring rules, raw subtest scores are first converted to Rasch-scaled ability scores to account for the difficulty of the subset of items administered to the patient for each DAS-II subtest. Subtest ability scores are then converted to *T*-scores using DAS-II Normative Data Tables (Elliott, 2007b) based on a target age (typically the child's chronological age at the time of testing). Therefore, for baseline or follow-up GCA, the target age for normative comparison is chronological age at baseline or follow-up, applied to the child's scores at baseline or follow-up, respectively. In contrast, for PRAS GCA, the target age for normative comparison is age at follow-up, using ability scores from the child's performance at baseline. Finally, subtest *T*-scores are summed and converted to a GCA score (standard score) using normative data tables (Elliott, 2007b). The PRAS GCA score differs from baseline GCA score only in the age used for normative comparison (baseline age for baseline GCA score; follow-up age for PRAS GCA score). The PRAS GCA score thus reflects the patient's projected GCA score at follow-up if raw/absolute performance was identical to that at baseline. By comparing follow-up GCA, baseline GCA, and PRAS GCA scores, five absolute and relative change profiles can be evaluated (Table 1).

Once baseline GCA, follow-up GCA, and PRAS GCA scores are obtained, they can be compared to evaluate the magnitude of relative and absolute change from Time 1 to Time 2. The difference between baseline and follow-up GCA (both of which are norm-based standard scores corresponding to the child's actual age at testing) represents relative change, whereas the difference between follow-up GCA and PRAS GCA (corresponding to the child's actual follow-up score and the child's follow-up score if no change had taken place since baseline) represents absolute change.

Because some change in scores is likely to occur by chance (since test scores are not perfectly reliable from Time 1 to Time 2), it is important to evaluate the statistical and clinical meaningfulness of difference scores in comparison with an index based on the statistical reliability or clinical significance of score change from Time 1 to Time 2. Extensive literature exists for estimating statistical and clinical meaningfulness based on the reliability of change in test scores from Time 1 to Time 2, using reliable change scores or indices (Duff, 2012; Wise, 2004). Reliable change index scores are more frequently used to assess cognitive change in adults, but the principles of reliable change scores can also be

applied to cognitive testing of children (Brooks, Holdnack, & Iverson, 2017; Busch, Lineweaver, Ferguson, & Haut, 2015). Estimation of clinically meaningful differences in child ability and achievement scores, using indices of reliable change, has been proposed as well (Reynolds, 2003).

The standard error of the difference (S_{Diff} ; Iverson, 2001), an index of the reliability of the difference between two test scores at 2 time points, is one measure that can be used to evaluate statistical reliability of GCA/PRAS score differences, and an equivalent mathematical method has been proposed to identify whether the difference between two cognitive test scores is clinically meaningful (Reynolds, 2003; Voress & Maddox, 2013). S_{Diff} is calculated using the formula $S_{\text{Diff}} = \sqrt{\text{SEM}_1^2 + \text{SEM}_2^2}$, where SEM_1 and SEM_2 are the standard errors of measurement of test scores at Time Point 1 and Time Point 2, respectively. SEM_1 and SEM_2 are calculated using $\text{SEM} = \text{SD} \times \sqrt{1-r}$, where SD is the sample's standard deviation and *r* is the test-retest reliability. Thus, using the published SD and test-retest reliability (*r*) from the DAS-II (15 and 0.9, respectively; Elliott, 2007a) for both time points, SEM_1 and SEM_2 both = 4.74 and $S_{\text{Diff}} = 6.70$. Therefore, an S_{Diff} value of 6.7 may be used to evaluate the statistical and clinical significance of the difference in baseline, follow-up, and PRAS GCA scores from Time 1 to Time 2. Furthermore, because test-retest reliability values are similar between the DAS-II and other major cognitive ability tests such as the Wechsler and Kaufman scales (Kaufman & Kaufman, 2004; Wechsler, 2014), an S_{Diff} of approximately 6.7 will be found for most cognitive ability tests with a normative SD of 15; for other cognitive ability tests, S_{Diff} can be easily calculated using the formulae above.

The difference between PRAS and follow-up GCA scores may be expressed as multiples of S_{Diff} , which can be compared with a table of *z* values to evaluate the statistical magnitude of the difference. Alternatively, comparison of the difference with a cutoff score for a clinically meaningful difference may suggest a functionally relevant change in scores (although this method can be somewhat arbitrary and should be used with appropriate caution). To derive a suggested value for this latter purpose of identifying a clinically meaningful GCA difference, we used the 90% confidence interval (CI), obtained by multiplying the S_{Diff} by 1.64, resulting in ± 11 GCA standard score points. Hence, using the S_{Diff} value for the DAS-II GCA score and 90% CI, a difference of $> \pm 10$ points between GCA and PRAS scores is statistically and clinically meaningful, whereas differences of $\leq \pm 10$ points do not meet that criterion (no different from chance). However, the limitations of this cutoff-based method should be taken into account in its application (e.g., it is somewhat arbitrary to conclude that a score of 10 points is not meaningful, but a score of 11 points is meaningful), and concurrent calculation of GCA/PRAS differences as multiples of S_{Diff} (6.70) is also recommended.

In summary, the objective of this study was to develop a new scoring method, the PRAS, in order to quantify absolute change in norm-based psychological test scores from Time 1 to Time 2, augmenting existing comparisons between norm-based global ability test scores (such as GCA) that are used to quantify relative change in norm-based scores from Time 1 to Time 2. The PRAS method was applied to three cases of a neurodegenerative disease, MPS II, to demonstrate its use and application. Although used here on DAS-II GCA scores for patients with MPS II, this method could be applied to any norm-based score and is particularly useful for ability scores in children when concerns of neurodevelopmental deterioration are present.

Method

Participants and Procedure

Participants for the present study were three preschool-aged (3 years 8 months, 4 years 9 months, and 4 years 11 months at time of first assessment) males diagnosed with MPS II, who were selected from an observational study (NCT01822184) that assessed the neurodevelopmental status of pediatric patients with MPS II. The three participants were selected as case studies specifically to demonstrate different profiles of absolute and relative change using GCA and PRAS scores.

Consistent with the inclusion/exclusion criteria of the broader observational study, the three participants had a formal MPS II diagnosis (demonstrated by iduronate-2-sulfatase enzyme activity deficiency together with a documented mutation in the iduronate-2-sulfatase gene or a normal activity level of one other sulfatase), sufficient auditory capacity with or without the use of hearing aids, no other clinically significant non-MPS II-related central nervous system involvement, and no other medical or psychiatric comorbidities that could interfere with the administration and interpretation of the DAS-II. As standard of care, patients received intravenous idursulfase during the study. Written informed consent and/or assent were obtained from patients or parents/legal guardians before any study-related procedures were conducted. Study materials were approved by independent ethics committees or institutional review boards (UNC-CH Office of Human Research Ethics; NRES Committee North West—Greater Manchester Central; Institutional Review Board Children's Hospital & Research Center Oakland; Ann & Robert H. Lurie Children's Hospital of Chicago Institutional Review Board; Ethics Committee Nino Jesus Children's University Hospital; Institutional Assessment Committee Austral University; National Institute of Pediatrics) before study initiation. The observational study was conducted in accordance with the International Conference on Harmonization Good Clinical Practice guidelines, the principles of the Declaration of Helsinki, and other local ethical and legal requirements.

DAS-II tests were administered at baseline and at 3-month intervals for up to 24 months during the observational study; intervals of 16 to 23 months were used to define baseline and follow-up testing for the purposes of the three case studies in this report. The current analyses are case presentations focusing on DAS-II GCA score change during the 16- to 23-month periods to demonstrate the use and application of PRAS in the context of evaluating change in global cognitive ability when neurodegenerative risk is present.

Measure

Differential Ability Scales®—Second Edition (DAS-II). The DAS-II is a neurocognitive measure of broad and specific cognitive abilities in children and adolescents aged between 2 years 6 months and 17 years 11 months (Elliott, 2007a). It has two overlapping batteries; the Early Years battery for children aged between 2 years 6 months and 6 years 11 months, and the School Age battery for children aged between 7 years 0 months and 17 years 11 months (Elliott, 2007a). Subjects in the current study completed core subtests in the Early Years Upper Level battery of the DAS-II, which assess verbal (Verbal Comprehension and Naming Vocabulary), nonverbal reasoning (Picture Similarities and Matrices), and spatial (Pattern Construction and Copying) skills.

Following standard DAS-II scoring rules, subtest raw scores were converted into ability scores based on a Rasch model, and ability scores were transformed to subtest *T*-scores based on a large, representative norm sample by age. The GCA score is a standard score (normative mean of 100 and *SD* of 15) based on the sum of the *T*-scores of all core subtests; it is the broadest and most reliable index of global cognitive ability obtained from the DAS-II. Consistent with standard DAS-II scoring, GCA scores at baseline and follow-up were obtained by applying the sum of *T*-scores of all core subtests to norm tables for the patient's chronological age at the actual time of testing (baseline or follow-up). PRAS GCA scores were obtained based on the sum of *T*-scores of all core subtests using baseline subtest ability scores applied to norm tables for the patient's follow-up chronological age.

Results

Case 1: Deteriorating Cognitive Development

A male patient with MPS II had a baseline GCA score of 70 at age 3 years 8 months, and a follow-up GCA score of 34 when tested 23 months later, at age 5 years 7 months (Table 2). The follow-up GCA score was 36 points lower than the baseline GCA score, indicating a decline relative to the scores of the patient's same-aged peers. Subtest ability scores from baseline (age 3 years 8 months) were then converted to PRAS *T*-scores using normative data tables for the

Table 2. Case Study I: Example of Deteriorating Cognitive Development in a Patient With MPS II.

	Subtest	Ability score	T-score	Age 5 years 7 months PRAS T-score
Baseline, age 3 years 8 months assessment	Verbal comprehension	87	41	27
	Picture similarities	92	57	42
	Naming vocabulary	82	41	27
	Pattern construction	57	31	10
	Matrices	10	23	11
	Copying	10	23	10
	GCA score		70	46
Follow-up, age 5 years 7 months assessment	Verbal comprehension	43	10	
	Picture similarities	53	24	
	Naming vocabulary	58	16	
	Pattern construction	30	10	
	Matrices	10	11	
	Copying	10	10	
	GCA score		34	
		Score	Type of change	
Results summary	Baseline GCA score	70		
	Follow-up GCA score	34		
	PRAS GCA score	46		
	GCA score change from baseline to follow-up	-36	Relative decline	
	Follow-up GCA score - PRAS GCA score	-12	Absolute decline	
	Conclusion		Deteriorating cognitive development	

Note. GCA = General Conceptual Ability; MPS II = mucopolysaccharidosis type II; PRAS = Projected Retained Ability Score.

patient's age at follow-up (5 years 7 months). For example, the Verbal Comprehension subtest ability score of 87 at baseline (3 years 8 months) yielded a *T*-score of 41 compared with 3-year 8-month norms (Table 2, fourth column), but the same ability score of 87 provided a PRAS *T*-score of only 27 compared with 5-year 7-month norms (Table 2, fifth column) because of the growth in cognitive ability of the norm sample between ages 3 years 8 months and 5 years 7 months. A PRAS GCA score of 46 was obtained from the sum of PRAS *T*-scores, reflecting what GCA would be at follow-up (PRAS GCA) if no improvement in ability score performance had occurred since baseline. The patient's follow-up GCA score (34) was >10 points lower than PRAS GCA (46), indicating that his GCA performance had declined compared with the PRAS estimate of no change from baseline performance. In summary, the patient demonstrated an absolute decline in global cognitive ability compared with his own performance at baseline, as well as a relative decline in global cognitive ability, consistent with neurodegeneration (Table 2).

Case 2: Stabilized Cognitive Development

A second male patient with MPS II had a baseline GCA score of 59 at age 4 years 11 months and a follow-up GCA score of 42 at age 6 years 3 months (Table 3). A PRAS GCA

score of 45 was obtained by applying baseline subtest ability scores obtained at age 4 years 11 months to normative data tables for the patient's age at follow-up (6 years 3 months) to obtain PRAS *T*-scores, which were summed and applied to the GCA conversion table. At follow-up, the patient's GCA score (42) was >10 points lower than the baseline GCA score (59), indicating a decline relative to same-aged peers. On the other hand, the patient's follow-up GCA score (42) was only 3 points lower than the PRAS GCA score (45) and within the ± 10 -point difference interval, indicating that his absolute ability comparing follow-up with baseline performance was comparable (stabilized development). Thus, the patient (Case 2) showed no decline in absolute performance on the cognitive ability test even though his performance had declined relative to the positive developmental trajectory shown by same-aged peers.

Case 3: Below Average Rate of Cognitive Development

A third male patient with MPS II had a baseline GCA score of 113 at age 4 years 9 months (Table 4) and a follow-up GCA score of 102 at age 6 years 8 months. A PRAS GCA score of 83 was obtained by applying baseline subtest ability scores obtained at age 4 years 9 months to normative data tables from the patient's age at follow-up (6 years 8 months)

Table 3. Case Study 2: Stabilized Cognitive Development in a Patient With MPS II.

	Subtest	Ability score	T-score	Age 6 years 3 months PRAS T-score
Baseline, age 4 years 11 months assessment	Verbal comprehension	88	31	24
	Picture similarities	74	37	29
	Naming vocabulary	93	37	26
	Pattern construction ^a	70	16	10
	Matrices	27	32	24
	Copying	43	25	10
	GCA score		59	45
Follow-up, age 6 years 3 months assessment	Verbal comprehension	88	24	
	Picture similarities	57	21	
	Naming vocabulary	75	18	
	Pattern construction	70	10	
	Matrices	35	29	
	Copying	23	10	
	GCA score		42	
		Score	Type of change	
Results summary	Baseline GCA score	59		
	Follow-up GCA score	42		
	PRAS GCA score	45		
	GCA score change from baseline to follow-up	-17	Relative decline	
	Follow-up GCA score - PRAS GCA score	-3	Absolute stabilization	
	Conclusion		Stabilized cognitive development	

Note. GCA = General Conceptual Ability; MPS II = mucopolysaccharidosis type II; PRAS = Projected Retained Ability Score.

^aAlternative scoring for pattern construction was used for this patient (untimed).

for conversion to PRAS *T*-scores, and then applying the summed PRAS *T*-scores to the GCA conversion table. The follow-up GCA score of 102 was >10 points lower than the baseline GCA score of 113, demonstrating a decline compared with same-aged peers. At follow-up, the patient's GCA score (102) was >10 points higher than the PRAS GCA score (83), indicating that although the patient had shown slower development than his peers, he demonstrated an absolute improvement in his own cognitive ability from baseline. This patient, therefore, showed positive (relative to his own baseline), but below average (relative to the developmental trajectory of peers), cognitive development.

Discussion

For patients with early onset of cognitive impairment, such as many of those with MPS II, it is imperative to understand absolute change in cognitive ability from Time 1 to Time 2, in addition to relative change in cognitive ability over that time period. A decline in norm-based cognitive ability scores from Time 1 to Time 2, which occurs for many patients with MPS II, could indicate that the patient's cognitive ability has either improved at a slower rate than his or her peer group

(below-average rate of development), has remained unchanged (stabilized development), or has declined (deteriorating development). These three profiles of absolute change are indistinguishable by norm-based scores alone.

In this article, we proposed a method (PRAS) for using norm-based scores to characterize absolute change from Time 1 to Time 2 and demonstrated the use of this method using case examples of change in DAS-II (Elliott, 2007a) GCA scores in three patients with the rare X-linked inherited lysosomal disease, Hunter syndrome (MPS II). Most cognitive assessment measures, such as the WISC-V, DAS-II, and KABC-II, provide only norm-based composite cognitive ability scores and lack the option of using age equivalents or raw scores to assess absolute change in composite ability scores. Even when raw scores or age-equivalents are available, there is no consistent statistic or other metric to evaluate whether a change in raw score or age-equivalent score is statistically or clinically significant. Therefore, the use of raw or age-equivalent scores is not feasible for the assessment of a statistically and clinically significant change in cognitive abilities.

All three cases demonstrated relative declines in norm-based ability scores, but had different outcomes regarding

Table 4. Case Study 3: Below-Average Rate of Cognitive Development in a Patient With MPS II.

		Ability score	T-score	Age 6 years 8 months PRAS T-score
Baseline, age 4 years 9 months assessment	Verbal comprehension	152	66	53
	Picture similarities	118	60	49
	Naming vocabulary	102	41	29
	Pattern construction	175	57	44
	Matrices	67	57	47
	Copying	94	53	33
	GCA score		113	83
Follow-up, age 6 years 8 months assessment	Verbal comprehension	154	54	
	Picture similarities	110	46	
	Naming vocabulary	132	44	
	Pattern construction	191	51	
	Matrices	70	49	
	Copying	145	62	
	GCA score		102	
		Score	Type of change	
Results summary	Baseline GCA score	113		
	Follow-up GCA score	102		
	PRAS GCA score	83		
	GCA score change from baseline to follow-up	-11	Relative decline	
	Follow-up GCA score - PRAS GCA score	+19	Absolute positive development	
	Conclusion		Below-average rate of cognitive development	

Note. GCA = General Conceptual Ability; MPS II = mucopolysaccharidosis type II; PRAS = Projected Retained Ability Score.

absolute change. The first patient had an absolute decline in cognitive abilities, demonstrating deteriorating cognitive development. In contrast, although the second patient demonstrated a decline in norm-based cognitive test scores relative to his peers, his cognitive abilities relative to his own functioning at baseline remained the same, indicating stabilized cognitive development. The third patient showed an improvement in his absolute level of cognitive ability, but this growth occurred at a slower rate than that of his peers, reflecting below-average rate of cognitive development and a decline in norm-based GCA scores. In the absence of the PRAS method, it would be impossible to differentiate between these three cases with respect to their profiles of absolute change in cognitive ability from Time 1 to Time 2.

In clinical practice, the PRAS method can be used to present findings to parents or caregivers of children with neurodegenerative conditions such as MPS II, those with slower cognitive development than norms, or those with other cognitive or learning disabilities. For patients with below-average rate of cognitive development (e.g., an absolute improvement but a relative decline compared with the typical growth of the norm sample), an assessment using norm-based scores alone (such as the GCA) would suggest deterioration. However, the PRAS method can be used to show an improvement in absolute (raw) cognitive ability, compared with baseline, despite a decline in norm-based scores with respect to the patient's

peers. This notable difference could have an impact on expectations and recommended treatment strategies.

Other clinically relevant applications of the PRAS method include roles in decision making relating to further evaluation and intervention because of cognitive deterioration. Examples of pediatric disorders that present risks for cognitive deterioration and/or stagnation of cognitive growth for which the PRAS method could be applied include cerebral palsy, severe seizure disorders, Down syndrome, fragile X syndrome, and autism-spectrum disorders. Declines in norm-based cognitive ability scores may be seen in children with these disorders, and PRAS allows for these declines to be further identified as below average rate of improvement, no change in absolute performance, or deterioration. Understanding whether a child with a neurological disorder is deteriorating cognitively versus developing positively but at a slower than average rate is critically important for treatment decision making, further evaluation, and educational/developmental interventions and planning.

PRAS may be conceptualized as part of a broader family of scoring methods, collectively referred to as Adjusted Age-Referenced Norm (AARN) scores, that compare raw scores with norms that differ systematically from the patient's chronological age. AARN scores, such as grade-based scores, hearing-age scores, and prematurity corrected-age scores, provide information about a patient's

developmental level relative to an age benchmark that is meaningful but different from chronological age. Grade-based scores, for example, compare a child's performance with norms based on the child's current grade in school, independent of chronological age (Schrank, McGrew, & Mather, 2014). On academic achievement tests, grade-based scores are more appropriate than chronological age norms for understanding a child's academic performance relative to grade-level expectations. Hearing-age scores are used to provide a corrected score on language tests for children who are deaf or hard of hearing and who receive augmentative or corrective interventions such as hearing aids or cochlear implants (Fagan & Pisoni, 2010). To obtain hearing-age scores, the child's score is compared with a norm sample of children whose chronological age matches the child's number of hearing years; hearing-age scores, therefore, provide an index of language development relative to norms for children who were subject to the same amount of hearing exposure. Prematurity corrected-age scores use norms for the age that the child would have been if the child had been born at term; when applied to developmental tests, this procedure adjusts for weeks of lost development resulting from premature birth. PRAS is a new type of AARN score that provides a norm-based score corresponding to no change from an earlier baseline score.

The utility of the PRAS method is dependent on a rapid increase in raw scores in the norm sample with increasing age, such that unchanged raw scores would result in large declines in norm-based scores with increasing age. Conversely, if raw scores do not change in the norm sample with increasing age, PRAS scores offer little advantage over norm-based scores. Thus, the PRAS method depends on a sufficient magnitude of change in norm-based scores to enable score differences to be classified into relative and absolute change. For this reason, the PRAS method is likely to be most useful in younger children and for constructs, such as cognitive ability, that increase consistently and significantly with age. For other situations, such as the assessment of different constructs (e.g., behavior problems) or cognitive decline associated with aging, use of the PRAS method may be limited and requires further investigation. In addition, the presence of a below-average rate of development may not be detectable for older children with shorter durations between assessments because of insufficient change associated with normative development over those time periods.

The duration between assessments may affect confidence about and interpretation of the rate of change in cognitive abilities. For example, a drop of 20 standard score points in a year indicates a faster rate of decline than a drop of 20 standard score points in 5 years. Furthermore, the confidence that change is not due to chance is smaller for longer time intervals because of lower test-retest reliability over longer time intervals. Additionally, although our application of the PRAS in this article applies to only 2 time

points, the use of three or more time points for determining whether developmental change is nonlinear would be beneficial. Future research on absolute versus relative change may consider methods to address nonlinear change over 3 or more time points.

Caution in using the PRAS method is also needed when floor or ceiling effects are present or when the PRAS score is very close to the baseline GCA score, as this may mean that it is impossible to detect significant differences between PRAS, baseline, and follow-up GCA scores. The chosen confidence interval and the specific cutoff or threshold value (e.g., of >10 points used here) should be applied with caution; difference scores between two or more time points are a continuous function, and cutoff scores draw dichotomous conclusions from this more nuanced data. For example, in some cases, it may be advantageous or more appropriate to reflect differences between two scores (e.g., PRAS and follow-up GCA) as a multiple of SEM or S_{Diff} values, which could then be tested for statistical abnormality relative to chance. However, the use of cutoff scores can be advantageous for situations in which a dichotomous (or other ordinal) classification is required, such as when estimating response in clinical trials or categorizing patients into different severity ranges.

Additional research involving larger sample sizes, different types of tests, different clinical samples, and diverse populations will be helpful to understand the validity and range of applications of PRAS. For example, in the case study examples used here to illustrate the application of the PRAS to cognitive assessment of patients with MPS II, all three patients were male because MPS II occurs much more frequently in males. Further studies are needed to test the application of PRAS to other disease areas, other norm-based tests, and female as well as male pediatric patients.

Despite some limitations and need for further research, the PRAS method enables a better understanding of a cognitive decline in children at neurodevelopmental risk, such as children with neurodegenerative disorders such as MPS II. Application of the PRAS method to the DAS-II GCA score has been described in this article; however, the PRAS method could be applied to any norm-based score at the subtest or composite level.

Conclusions

PRAS is a novel method to more specifically characterize absolute change in norm-based cognitive ability scores, which is especially important for patients with neurodegenerative conditions such as MPS II. Although illustrated here with data from patients with MPS II, the method may be extended to other clinical applications and other norm-based tests used for the assessment of outcomes from Time 1 to Time 2. In cases where cognitive ability test scores (e.g., DAS-II GCA scores) demonstrate a relative decline,

application of the PRAS method enables the type of absolute change in cognitive development to be identified more specifically. This is important, as slow positive development (albeit at a below-average rate) and stabilization offer better prognoses than deterioration for patients with neurodegenerative conditions such as MPS II.

Acknowledgments

Sally Hassan, PhD, employee of Excel Medical Affairs, provided writing assistance for this article under the direction of the authors. Writing assistance was funded by Shire (a member of the Takeda group of companies). The interpretation of the data was made by the authors independently. A poster of the PRAS methods was presented at the 14th Annual WORLDSymposium, February 5 to 9, 2018, San Diego, CA.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: William G. Kronenberger has received consulting fees from Shire Human Genetic Therapies, a Takeda company. Magdalena Harrington was an employee of Shire at the time this research was conducted. Karen Yee is an employee of Shire, a Takeda company.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the sponsor, Shire Human Genetic Therapies, a Takeda company. Editorial assistance in formatting, proofreading, copy editing, and fact checking was provided by Excel Medical Affairs and funded by Shire International GmbH, a Takeda company.

References

- Anastasi, A. (1988). *Psychological testing* (6th ed.). New York, NY: Macmillan.
- Bayley, N. (2006). *Bayley scales of infant and toddler development* (3rd ed.). San Antonio, TX: Psychological Corporation.
- Brooks, B. L., Holdnack, J. A., & Iverson, G. L. (2017). Reliable change on memory tests is common in healthy children and adolescents. *Archives of Clinical Neuropsychology, 32*, 1001-1009.
- Burgaleta, M., Johnson, W., Waber, D. P., Colom, R., & Karama, S. (2014). Cognitive ability changes and dynamics of cortical thickness development in healthy children and adolescents. *Neuroimage, 84*, 810-819.
- Burton, B. K., & Giugliani, R. (2012). Diagnosing Hunter syndrome in pediatric practice: Practical considerations and common pitfalls. *European Journal of Pediatrics, 171*, 631-639.
- Busch, R. M., Lineweaver, T. T., Ferguson, L., & Haut, J. S. (2015). Reliable change indices and standardized regression-based change score norms for evaluating neuropsychological change in children with epilepsy. *Epilepsy & Behavior, 47*, 45-54.
- Camp, S. J., Stevenson, V. L., Thompson, A. J., Ingle, G. T., Miller, D. H., Borras, C., . . . Langdon, D. W. (2005). A longitudinal study of cognition in primary progressive multiple sclerosis. *Brain, 128*, 2891-2898.
- Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: Relevant concepts and methods. *Archives Clinical Neuropsychology, 27*, 248-261.
- Elliott, C. D. (2007a). *Differential Ability Scales® (DAS)–Second edition: Introductory and technical handbook*. Bloomington, MN: Pearson.
- Elliott, C. D. (2007b). *Differential Ability Scales® (DAS)–Second edition: Normative data tables manual*. Bloomington, MN: Pearson.
- Fagan, M. K., & Pisoni, D. B. (2010). Hearing experience and receptive vocabulary development in deaf children with cochlear implants. *Journal of Deaf Studies and Deaf Education, 15*, 149-161.
- Fan, Z., Styner, M., Muenzer, J., Poe, M., & Escolar, M. (2010). Correlation of automated volumetric analysis of brain MR imaging with cognitive impairment in a natural history study of mucopolysaccharidosis II. *American Journal of Neuroradiology, 31*, 1319-1323.
- Godfrey, M., & Lee, N. R. (2018). Memory profiles in Down syndrome across development: A review of memory abilities through the lifespan. *Journal of Neurodevelopmental Disorders, 10*, 5. doi:10.1186/s11689-017-9220-y
- Guffon, N., Heron, B., Chabrol, B., Feillet, F., Montauban, V., & Valayannopoulos, V. (2015). Diagnosis, quality of life, and treatment of patients with Hunter syndrome in the French healthcare system: A retrospective observational study. *Orphanet Journal of Rare Diseases, 10*, 43.
- Harris, S. L., Handleman, J. S., Gordon, R., Kristoff, B., & Fuentes, F. (1991). Changes in cognitive and language functioning of preschool children with autism. *Journal of Autism and Developmental Disorder, 21*, 281-290.
- Holt, J. B., Poe, M. D., & Escolar, M. L. (2011). Natural progression of neurological disease in mucopolysaccharidosis type II. *Pediatrics, 127*, e1258-e1265.
- Iverson, G. L. (2001). Interpreting change on the WAIS-III/WMS-III in clinical samples. *Archives of Clinical Neuropsychology, 16*, 183-191.
- Janzen, D., Delaney, K. A., & Shapiro, E. G. (2017). Cognitive and adaptive measurement endpoints for clinical trials in mucopolysaccharidoses types I, II, and III: A review of the literature. *Molecular Genetics and Metabolism, 121*, 57-69.
- Kaufman, A. (1990). *Assessing adolescent and adult intelligence*. Needham, MA: Allyn & Bacon.
- Kaufman, A., & Kaufman, N. (2004). *Kaufman Assessment Battery for children* (2nd ed.). Bloomington, MN: Pearson.
- Link, B., de Camargo Pinto, L. L., Giugliani, R., Wraith, J. E., Guffon, N., Eich, E., & Beck, M. (2010). Orthopedic manifestations in patients with mucopolysaccharidosis type II (Hunter syndrome) enrolled in the Hunter Outcome Survey. *Orthopedic Reviews, 2*, e16.
- Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development, 75*, 1357-1372.

- Muenzer, J., Burton, B. K., Harmatz, P., Gutiérrez-Solana, L. G., Ruiz-García, M., Jones, S. A., . . . Alexanderian, D. (2018). Efficacy and safety of intrathecal idursulfase in pediatric patients with mucopolysaccharidosis type II and early cognitive impairment: Design and methods of a controlled, randomized, phase II/III multicenter study. *Molecular Genetics and Metabolism*, 123(2), S99-S100.
- Muenzer, J., Hendriks, C. J., Fan, Z., Vijayaraghavan, S., Perry, V., Santra, S., . . . Barbier, A. J. (2016). A phase I/II study of intrathecal idursulfase-IT in children with severe mucopolysaccharidosis II. *Genetics in Medicine*, 18, 73-81.
- Mullen, E. (1995). *Mullen scales of early learning*. Minneapolis, MN: American Guidance Service.
- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256, 183-194.
- Reynolds, C. R. (2003). Conceptual and technical problems in learning disability diagnosis. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment in children: Intelligence, aptitude, and achievement* (2nd ed., pp. 475-497). New York, NY: Guilford Press.
- Salthouse, T. A., Nesselroade, J. R., & Berish, D. E. (2006). Short-term variability in cognitive performance and the calibration of longitudinal change. *Journal of Gerontology: Series B Psychological Science and Social Science*, 61, P144-151.
- Scarpa, M., Almásy, Z., Beck, M., Bodamer, O., Bruce, I. A., De Meirleir, L., . . . Wraith, J. E. (2011). Mucopolysaccharidosis type II: European recommendations for the diagnosis and multidisciplinary management of a rare disease. *Orphanet Journal of Rare Diseases*, 6, 72.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson IV tests of cognitive abilities*. Rolling Meadows, IL: Riverside.
- Segalowitz, S. J., & Davies, P. L. (2004). Charting the maturation of the frontal lobe: An electrophysiological strategy. *Brain and Cognition*, 55, 116-133.
- Shapiro, E. G., Escolar, M. L., Delaney, K. A., & Mitchell, J. J. (2017). Assessments of neurocognitive and behavioral function in the mucopolysaccharidoses. *Molecular Genetics and Metabolism*, 122(Suppl.), 8-16.
- Tomatsu, S., Fujii, T., Fukushi, M., Oguma, T., Shimada, T., Maeda, M., . . . Orii, T. (2013). Newborn screening and diagnosis of mucopolysaccharidoses. *Molecular Genetics and Metabolism*, 110, 42-53.
- Voress, J. K., & Maddox, T. (2013). *Developmental assessment of young children examiner's manual* (2nd ed.). Austin, TX: Pro-Ed.
- Waber, D. P., De Moor, C., Forbes, P. W., Almli, C. R., Botteron, K. N., Leonard, G., . . . Rumsey, J. (2007). The NIH MRI study of normal brain development: Performance of a population based sample of healthy children aged 6 to 18 years on a neuropsychological battery. *Journal of the International Neuropsychological Society*, 13, 729-746.
- Wechsler, D. (2014). *Wechsler Intelligence Scale for Children®—Fifth edition (WISC®-V): Technical and interpretive manual*. Bloomington, MN: NCS Pearson.
- Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment*, 82, 50-59.
- Yee, K., Kronenberger, W., & Harrington, M. (2018). Projected Retained Ability Score (PRAS): A new methodology applied to DAS-II GCA scores for the longitudinal assessment of cognitive abilities in pediatric and adolescent patients with Hunter syndrome. *Molecular Genetics and Metabolism*, 123(Suppl.), S151.