

Measuring Within-Child Change in Treatment Studies of Low-Functioning Children

Mark Daniel^{1,2}, Louis-Charles Vannier¹, Stephen M. Maricich³, Elsa Shapiro⁴, Adam Scheller¹

¹ Pearson Clinical Assessments; ² Mark Daniel Services, LLC.; ³ formerly of Allievex Corporation*, currently Calibr; ⁴ Shapiro Neuropsychology Consulting, LLC

CLINICAL NEED

An increasing number of clinical trials involve children who have neuropathic or neurodegenerative disease. Measuring change in developmental outcomes for low-functioning children is methodologically challenging (1), because:

- these children often perform below the level of the norms for their chronological age (CA), and
- the natural progression of the disease may involve flat or declining raw performance (e.g., 2).

Thus, it is often helpful to do out-of-level (OOL) testing where the test matches the child's developmental age (DA) rather than their CA, and to use scores that reflect absolute (raw) performance rather than normative scores.

Also, researchers need the standard error of measurement (SEM) to evaluate change. SEM may be group-based (from reliability and SD in a sample) or conditional (score-specific). The group-based SEM is the average conditional SEM in the sample (3).

WHAT KINDS OF SCORES ARE MOST USEFUL?

There are two broad categories of scores:

1. **Absolute scores** reflect raw performance. Examples: raw score, age equivalent, Growth Scale Value
2. **Normative scores** describe the location of the absolute score in the distribution of absolute scores of age peers. Examples: standard score, T score, percentile

Normative scores are problematic for two reasons:

1. Subjects often cannot obtain a normative score because they perform more than 2 or 3 SDs below the mean for their CA.
2. Change in the normative score does not reflect change in absolute performance, but rather a difference between the child's rate of change and the norm sample's rate of change. This can lead to misinterpretation (4).
3. Absolute scores do not have these shortcomings.

FEATURES OF DIFFERENT ABSOLUTE SCORES

Raw Scores

1. Usable at any CA (therefore support OOL testing)
2. Measure accurately at DAs beyond the age range of the norms, from ~2 SDs below the mean of the youngest age to ~2 SDs above the mean of the oldest age
3. Have group-based SEMs based on the norm sample at the DA

Growth Scale Values (GSV)

1. Have the above-noted features of raw scores
2. Are the ability scores from a Rasch (item response theory) calibration of the test
3. GSVs adjust for the uneven distribution of difficulties of test items to provide an interval scale, where a given size change in ability produces a consistent-size change in GSV at all score levels. This feature is valuable for studies of treatment effects.
4. Have conditional SEMs

Age Equivalents (AE)

1. The age at which the raw score is the median
2. Usable at any CA (therefore support OOL testing)
3. Measure a limited range of DAs (from the youngest to oldest CA in the norm sample)
4. Often take on extreme values near minimum or maximum raw scores
5. No SEMs

Table 1. Features of various score types

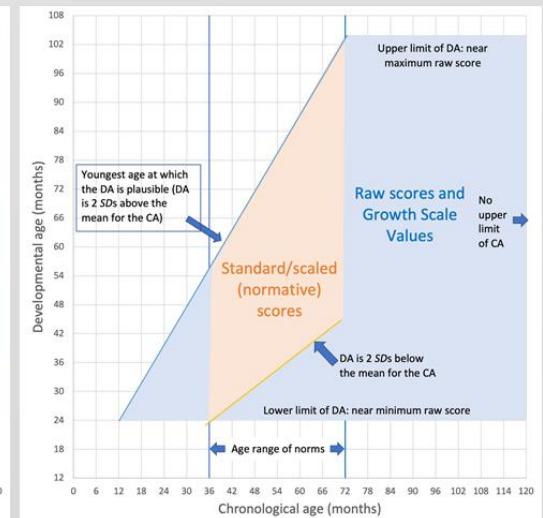
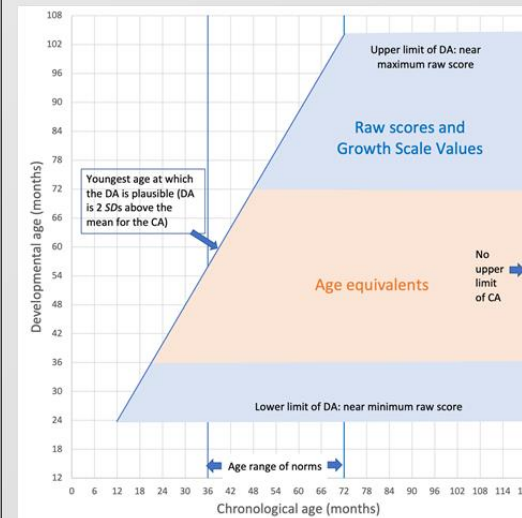
Feature	Absolute score			Normative score
	Raw score	GSV	Age Equiv.	
Reflects raw performance	✓	✓	✓	
Has normative meaning			(✓) ^a	✓
Interval scale		✓		
SEM	Group ^b	Conditional		Group
Measures DAs outside ages of norms	✓	✓		

^a Comparison with CA tells whether raw score is above or below average for CA.

^b May not be reported in test manual.

EFFECTIVE RANGES OF SCORE TYPES

The charts below show how raw scores and GSVs (blue areas) measure a wider range of DAs than age equivalents (left), and wider ranges of CAs and DAs than normative scores (right). Data is for a hypothetical cognitive ability test normed at ages 36 to 72 months, based on norms for Bayley-IV Cognitive, DAS Nonverbal, KABC-II NU Nonverbal Index, and WPPSI-IV Nonverbal Index.



CONCLUSION

In treatment studies of childhood diseases that cause low and flat/declining performance on outcome measures, absolute scores have advantages over normative scores: they support out-of-level testing (i.e., giving developmentally-appropriate tests regardless of CA) and have clearer interpretation. Among absolute scores, raw scores and Growth Scale Values measure a wider range of DAs than do age equivalents. In addition, the GSV scale, which is the ability scale from a Rasch (IRT) test calibration, is an interval scale and provides conditional (score-specific) SEMs.

References

1. Thurm A, Kelleher B, Wheeler A. Outcome measures for core symptoms of intellectual disability: State of the field. *Am J Intellect Dev Disabil.* 2020 Nov;125:418-33.
2. Shapiro EG, Nestrail I, Delaney KA, Rudser K, Kovac V, Nair N, ... Whitley CB. A prospective natural history study of mucopolysaccharidosis type IIIA. *J Pediatr.* 2016 Mar 1;170:278-87.
3. Raju NS, Price LR, Oshima TC, Nering ML. Standardized conditional SEM: A case for conditional reliability. *Appl Psych Meas.* 2007 May;31:169-80.
4. Farmer CA, Kaat AJ, Thurm A, Anselm I, Akshoomoff N, Bennett A, ... Miller JS. Person ability scores as an alternative to norm-referenced scores as outcome measures in studies of neurodevelopmental disorders. *Am J Intellect Dev Disabil.* 2020 Nov;125:475-80.