**Title**: Exploring the Administration and Scoring Practices of Standardized Language Assessments for Youth with Down Syndrome

**Authors**: Abby Lutjeharms1, Susan J. Loveall1, & Frances A. Conners2

**Introduction:** Standardized, norm-referenced language assessments are used by researchers and practitioners for a variety of purposes, including measuring achievement, identifying areas of strength and difficulty within language, and tracking growth over time (Loveall et al., 2022). To interpret standardized scores from these measures, they must be administered and scored in a specific, standardized way, as outlined in the published manual. This includes rules about start points, establishing basals and ceilings, and how to calculate scores. Although many norm-referenced assessments indicate that start points can be adjusted for examinees with intellectual disabilities such as Down syndrome (DS), a population that often has difficulties with language, there is limited guidance on how to make these adaptations. These assessments also often have floor effects and lack variability in scores when used with individuals with DS (Esbensen et al., 2017). Understanding the patterns of performances by individuals with DS on standardized language measures, including how examiner administration and scoring adaptations impact scores, is an important first step to providing more precise guidelines for using norm-referenced assessments with this population. Thus, our aims were to (1) document the percentage of participants with DS who meet basal criteria at different start points of three language assessments, (2) describe scores on those assessments when using standard start points, and (3) compare if and how scores change when scoring procedures are altered.

**Method**: As part of a larger study, 45 participants with DS ranging from 9-18 years (M=14.3 years; SD=2.5; 25 females) completed a series of standardized, norm-referenced assessments. Language comprehension was assessed via the Listening Comprehension subtest of the OWLS-ll, receptive vocabulary via the PPVT-5, and receptive syntax and grammar via the Elaborated Phrases and Sentences (EPS) and Grammatical Morphemes (GM) subtests of the TACL-4. For Aim 1, the percentage of participants who met basal at their chronological age start point, one start point below, two start points below, etc. was documented until accounting for all participants. For Aim 2, each assessment was evaluated for raw and standard scores when using standard start points, including means, SDs, ranges, and number of participants at floor (i.e., receiving the lowest possible standard score). For Aim 3, a paired-samples *t*-test was used to compare if participants’ scores changed on the OWLS-ll based on whether the raw score was calculated using the highest basal and lowest ceiling (as is common for many norm-referenced assessments) versus the lowest basal and highest ceiling (as specified in the OWLS-ll manual).

**Results**: For Aim 1, on the OWLS-ll, 4% of participants would have met basal had they started at their chronological age; 20% never met basal (i.e., had to return to item 1), and 76% had to move to 1-7 start points below their chronological age to meet basal. On the PPVT-5, 60% of participants would have met basal at or above their chronological age, and 40% had to move 1-4 start points below chronological age to meet basal. Because the TACL-4 is only normed up to age 12;11, the highest start point was considered “at chronological age” for participants at or over this age. For EPS, 14% of participants met basal at or above their chronological age; 9% never met basal, and 77% had to move to 1-2 start points below their chronological age to meet basal. For GM, 20% of participants would have met basal had they started at their chronological age; 5% never met basal, and 75% had to move to 1-2 start points below their chronological age to meet basal.

For Aim 2, raw score descriptives were: (1) OWLS-ll M=45.0, SD=17.5, range=19-84; (2) PPVT-5 M=113.7, SD=27.7, range=57-168; (3) TACL-4 EPS M=18.9, SD=10.4, range=0-51; and (4) TACL-4 GM M=29.2, SD=13.3, range=1-54. Standard score descriptives were: (1) OWLS-ll M=44.4, SD=10.4, range=40-89; (2) PPVT-5 M=53.0, SD=11.5, range=40-82; but could not be calculated for the TACL-4 given its more limited standardized age range. Standard score floor effects were observed on both the OWLS-2 (76% of the sample) and the PPVT-5 (31%).

For Aim 3, on the OWLS-II there was a statistically significant difference in the mean raw scores when calculated using the lowest basal and highest ceiling (M=45.0, SD=17.5) versus highest basal and lowest ceiling (M=38.0, SD=21.7), *t*(44)=5.76, *p*<.001, Cohen’s *d* = .84.

**Discussion:** These results offer important information that researchers and practitioners can use when administering norm-referenced language assessments to participants with DS. Adjusting start points allows practitioners to tailor assessments to the specific needs and abilities of each client. Lowering the start point may not be needed as often for the PPVT-5 or TACL-4 as it is for the OWLS-ll. The results also indicate that standard scores should be used with caution, given the substantial floor effects and limited variability when used with individuals with DS. Future research should develop and test additional measurements of language with this population.

**Key References**

Esbensen, A. J., Hooper, S. R., Fidler, D., Hartley, S. L., Edgin, J., d'Ardhuy, X. L., Capone, G., Conners, F. A., Mervis, C. B., Abbeduto, L., Rafii, M. S., Krinsky-McHale, S. J., Urv, T., & Outcome Measures Working Group (2017). Outcome measures for clinical trials in Down syndrome. *American Journal on Intellectual and Developmental Disabilities*, *122*(3), 247–281.

Loveall, S. J., Channell, M. M., Mattie, L. J., & Barkhimer, A. E. (2022). Inclusion of individuals with neurodevelopmental disorders in norm-referenced language assessments. *Frontiers in Psychology*, *13*.

1University of Nebraska-Lincoln

2 University of Alabama