Id: 1899349

Participatory Educational Research (PER)
Vol.13(5), pp. 226-244, September 2026
Available online at http://www.perjournal.com
ISSN: 2148-6123
http://dx.doi.org/10.17275/per.26.73.13.5

Development and Initial Validation of an Expressive Vocabulary Test for Movement Concepts and Skills in Adults with Down syndrome

Georgios Sivvas*
Department of Physical Education & Sport Sciences, Democritus University of Thrace, Komotini, Greece
ORCID: 0009-0005-8678-4065

Kyriaki Emmanouilidou
Department of Physical Education & Sport Sciences, Democritus University of Thrace, Komotini, Greece
ORCID: 0000-0003-3446-1799

Athina Pitsi
Department of Physical Education & Sport Sciences, Democritus University of Thrace, Komotini, Greece
ORCID: 0000-0002-6693-7678

Grigorios Masadis
Department of Physical Education & Sport Sciences, Democritus University of Thrace, Komotini, Greece
ORCID: 0000-0002-2964-7241

Nerantzoula Koufou
Department of Physical Education & Sport Sciences, Democritus University of Thrace, Komotini, Greece
ORCID: 0000-0001-7790-5981

Konstantinos Filippou
Sauder's Robert H. Lee Graduate School. Master of Management. University of British Columbia. Vancouver, Canada
ORCID: 0009-0002-6325-8781

Panagiotis Varsamis
Department of Educational & Social Policy, University of Macedonia, Greece, Greece
ORCID: 0000-0001-9657-2684

* Correspondency: ffilippo@phyed.duth.gr

Vasiliki Derri
Department of Physical Education & Sport Sciences, Democritus University of Thrace, Komotini, Greece
ORCID: 0000-0001-9487-9586

Filippos Filippou
Department of Physical Education & Sport Sciences, Democritus University of Thrace, Komotini, Greece
ORCID: 0000-0003-3116-7930

Article history

Received:
27.02.2026

Received in revised form:
02.05.2026

Accepted:
17.06.2026

Key words:
expressive vocabulary; movement concepts; movement skills; item analysis; content validity

The present study aimed to develop and psychometrically evaluate the Expressive Vocabulary Test - Movement Concepts Skills - Down syndrome (EVT-MCS-DS) and to examine possible differences in expressive vocabulary performance according to demographic characteristics, including sex, living arrangement, and age group. The sample comprised 48 men and 18 women with Down syndrome (M age = 26.56 ± 2.80 years).The instrument was developed through a multi-phase procedure including the selection of 50 movement concepts and skills, their transformation into pictorial stimuli by an expert committee, item analysis (difficulty, discrimination, and point-biserial correlation indices), content validity evaluation, and reliability assessment (Cronbach's α, Kuder–Richardson 20, split-half reliability, and test–retest stability). Item analysis resulted in the exclusion of 18 items due to unacceptable psychometric indices, yielding a final 32-item test. Internal consistency was high (Cronbach's α = .952, KR-20 = .80, Spearman–Brown = .94), while test–retest reliability demonstrated excellent stability (ICC Single Measures = .967; ICC Average Measures = .983). Content validity was also supported (I-CVI = .80–1.00; S-CVI/Ave = .92). The EVT–MCS-DS demonstrates satisfactory initial evidence of validity and reliability and may serve as a useful tool for assessing expressive naming of movement concepts and motor skills in adults with Down syndrome.

Introduction

The development and preservation of expressive vocabulary is a key indicator of language ability and cognitive function in people with Down syndrome (DS). The term "expressive vocabulary" refers to the set of words that an individual can actively use to communicate with their environment. Cognitive function, which includes a wide range of cognitive skills, plays a key role in effective communication (Abbeduto, Warren, & Conners, 2007; Filipe, Cruz, Veloso, & Frota, 2022; Onnivello et al., 2022). In individuals with DS, language development often follows a different developmental trajectory compared to the typical population, which necessitates the systematic assessment of their language skills (Cristea et al., 2025).

Language ability directly affects functional communication, autonomy and quality of life of individuals with DS. Particularly during adulthood, a gradual decline in expressive vocabulary and overall language performance has been reported, which may negatively affect their participation in daily activities and social inclusion (Deckers, Van Zaalen, Van Balkom, & Verhoeven, 2017; Martin, Klusek, Estigarribia, & Roberts, 2009). Therefore, continuous monitoring of the development of expressive vocabulary is considered essential in order to identify any language difficulties in time and to design appropriate educational and intervention programmes aimed at maintaining and improving language skills.

The demands of daily life during adulthood differ substantially from those encountered during childhood and adolescence. Consequently, assessing the expressive vocabulary of adults with DS is both necessary and important. For these adults to achieve independence, navigate their environment safely, participate in vocational and community settings, or maintain a social life, they must be able to understand and use relevant functional concepts. Competence in comprehending and verbally expressing spatial and movement-related concepts, such as "left", "right", "back", "forward", "slow", and "fast", is essential for following instructions and participating safely in everyday activities (Carr, 2008; Deckers et al., 2017).

In this context, several standardized vocabulary and language assessment tools have been developed and widely used, including the Expressive Vocabulary Test (EVT-3), the Peabody Picture Vocabulary Test (PPVT-5), the Test of Auditory Comprehension of Language (TACL-4) and the Clinical Evaluation of Language Fundamentals (CELF). Although these instruments are reliable for the general population, their application to adults with DS and mild intellectual disability has important limitations. According to the literature, most have not been standardized for clinical populations, resulting in norms that do not adequately reflect the cognitive and language profiles of individuals with DS (Aksen-Cengizhan, Ahmet Dogan, & Isık, 2021; Loveall, Channell, Mattie, & Barkhimer, 2022; Thurman et al., 2021).

Most existing language assessment tools are designed to evaluate the expressive vocabulary of children and adolescents. As a result, their content does not correspond to the cognitive, social, or functional levels of adults. Many of these instruments include vocabulary and pictorial stimuli developed primarily for children, which may not adequately reflect the functional communication needs and everyday experiences of adults with DS. Consequently, these tools may not always accurately reflect the language abilities and functional communication needs of this population (Finestack, Sterling, & Abbeduto, 2013; Loveall et al., 2022).

Furthermore, existing instruments often do not take into account key characteristics of individuals with DS, such as limited auditory memory, difficulties in auditory information processing, and specificities in language development. Thus, the low performance of individuals with DS on language tests is likely not a reflection of their actual language abilities, but rather the inability of these tools to detect small but meaningful variations in the low performance range (Loveall et al., 2022; Thurman et al., 2021). This highlights the need to develop specialized assessment tools tailored to the characteristics of adults with DS population.

A further important gap in the literature concerns the absence of tools that assess expressive vocabulary in relation to motor concepts and skills. Motor concepts and skills play a crucial role in the daily functioning of people with DS, as they are related to safe movement, spatial orientation, and understanding instructions related to space and time (Bull, 2020; Fidler, 2005; Palisano et al., 2001). Their mastery facilitates the participation of people with DS in physical activities, such as dance and Physical Education classes, while contributing to improved autonomy and social inclusion (te Kaat, Volman, Jongmans, & Lauteslager, 2016).

Understanding motor and spatial concepts is important for adults with DS not only to participate in physical activity programs but, more importantly, to meet the demands of daily life and ensure personal safety. Concepts such as "in–out," "front–back," "stop–go," "slow–fast," and "near–far" are frequently used in everyday life and are essential for understanding instructions related to movement, work environments, and public spaces (te Kaat et al., 2016).

Therefore, assessing the level of verbal expression of these concepts, along with other similar concepts, may provide useful information about the level of functional communication and autonomy of adults with DS.

Previous research indicates that participation in physical and movement activities is strongly associated with the cognitive and language development of people with DS. Kinesthetic and visuomotor experiences appear to facilitate comprehension and use of new words, enhancing both expressive and receptive vocabulary. Receptive vocabulary refers to the words an individual understands but does not necessarily use in speech (Bedford, Pickles, & Lord, 2016; Custode & Tamis-LeMonda, 2020; Hwang & Lee, 2022). Furthermore, motor skills have been identified as important predictors of language and communication performance, highlighting the close relationship between motor and language development (Karimi & Nelson, 2023; Will & Roberts, 2021; Yamauchi, Aoki, Koike, Hanzawa, & Hashimoto, 2019).

Iverson (2010) and Oudgenoeg-Paz, Volman, and Leseman (2012) argued that developments in developmental and educational psychology in recent years have highlighted the interdependence of motor development and language development. Similarly, the theory of embodied cognition proposes that an individual's concepts and linguistic expression depend to a significant extent on their physical interaction with their social environment (Barsalou, 2008; Wilson, 2002). A similar view is expressed by Thelen and Smith (1994), who argue that an individual's ability to perceive directions and spatiotemporal relationships, as well as to orient themselves in space, is inextricably linked to the full range of their bodily and motor experiences.

Taken together, the above findings highlight the need for the development of a specialized and reliable instrument capable of assessing the expressive vocabulary of adults diagnosed with DS and mild intellectual disability. Such an instrument should also take into account their distinctive learning and cognitive characteristics and focus on vocabulary related to motor concepts and everyday life skills. Such an instrument can support both the early identification of language difficulties and the evaluation of the effectiveness of intervention programs in physical activity (Karakuş & Akbaş, 2026).

Such a specialized assessment tool, supported by evidence of validity and reliability, may be useful for special education teachers, speech therapists, and Physical Education instructors. It would enable professionals to identify specific difficulties not only in the understanding but also in the verbal expression of the aforementioned concepts. At the same time, it may support the design of individualized intervention programs and the evaluation of their effectiveness in improving the expressive vocabulary of adults with DS (Florian & Black-Hawkins, 2011).

Therefore, the present study aimed to develop and psychometrically evaluate the EVT-MCS-DS and to examine possible differences in expressive vocabulary performance according to demographic characteristics, including sex, living arrangement, and age group.

Methodology

Initial Test Development

The development of the initial item pool followed international guidelines. These guidelines recommend that the selection of initial items be based on theoretical and analytical criteria (Boateng, Neilands, Frongillo, Melgar-Quiñonez, & Young, 2018; Korkmaz & Bai, 2019; Worthington & Whittaker, 2006).

To develop the initial EVT-MCS-DS and ensure content validity, the Greek National Curriculum for students with mild and moderate intellectual disabilities, including the 18 thematic axes of instruction (Polychronopoulou, et al. & Greek Ministry of Education & Religious Affairs, 2004), was used as a reference. Additionally, guidelines from relevant literature on how people with DS learn new motor skills were considered. Visual stimuli, such as pictures, contribute to information recall even when verbal difficulties are present. A similar view is expressed by Baadte and Meinhardt-Injac (2019), who argue that the presentation of pictures as stimuli is more effective for recall and naming than words alone. Based on these considerations, fifty concepts and skills related to the daily life of people with DS were initially selected and then illustrated as 50 pictorial stimuli by the researchers.

Development of the Initial List of Concepts, Skills, and Images

In the next stage, sixty Physical Education teachers were asked to rate which of the fifty concepts and skills they used in their teaching and considered relevant to the daily life of these individuals. The teachers were randomly selected from a pool of 200 PE teachers, all with at least ten years of professional experience, working in special schools in Greece. Of the sixty teachers selected, fifty (83.33%) agreed to participate.

The fifty teachers received an online form containing the fifty words and corresponding images. They were asked to propose additional words related to motor concepts and skills used in their teaching and identify images that best represented them. Based on the teachers' responses, a total of 200 images were collected, approximately four images per word.

Expert Panel Review of Images

The set of two hundred images was evaluated by a panel of ten experts in special Physical Education, each with at least fifteen years of teaching experience and holding a master's or doctoral degree. The panel was tasked with selecting the image that most accurately depicted each of the fifty concepts and/or skills. Panel members could also recommend alternative images.

The committee reached consensus on forty-two images and suggested alternatives for eight movement concepts or skills (slow, fast, on the spot, zig-zag, turn, in, out, leading). In addition, three new pictures were proposed by two members for three of the movement concepts (on the spot, zig-zag and leading). This process resulted in a list comprising forty-two unanimously selected images and twenty-six candidates for selection for the remaining eight. From these, the committee selected eight, so that the final list consisted of fifty images.

The aim of this process was to optimize the materials and ensure that the selected images accurately represented the relevant concepts and skills. Representative examples of the visual stimuli used in the EVT-MCS-DS are presented in Figure 1. The images consisted of simplified illustrations with minimal background detail in order to facilitate recognition and

naming by adults with DS.

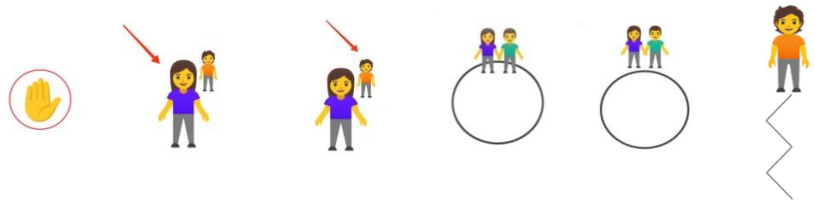

Figure 1. Representative visual stimuli used in the EVT-MCS-DS. Examples include body-part, spatial, directional, and movement-related concepts assessed using simplified visual illustrations for adults with DS.

Final version of the test

Inclusion criteria were defined to ensure that participants met the requirements regarding cognitive abilities and functional capacity for participation in the evaluation process. Participants had to meet all five of the following criteria: (a) Type T21 (Typical Trisomy 21), (b) diagnosed with mild intellectual disability, (c) no visual or auditory impairments, (d) ability to communicate effectively, and (e) aged 20–35 years.

The study included 66 participants (48 males, 18 females) diagnosed with DS, aged 20–32 years ($M = 26.56 \pm 2.80$). Of these, 42 lived in residential facilities and attended classes based on the Ministry of Education curriculum (Polychronopoulou, et al. & Greek Ministry of Education & Religious Affairs, 2004), while 24 lived with their families and attended a local special education school.

Ethics approval and data collection procedure

The study was approved by the Ethics Committee of the Democritus University of Thrace (Decision 1st/26–09-2024). Permission was also granted by the managers of the facilities and by the parents/guardians of participants living with their families. Written informed consent was obtained from parents or legal guardians, and verbal assent was obtained from all participants prior to data collection. Parents and guardians received written assurance of full anonymity and data protection. The researchers also committed to using the data exclusively for scientific purposes.

Data collection took place at the participants' schools or residential facilities in the presence of their PE teachers. Evaluation was conducted individually. Each PE teacher presented the pictures, and participants named the depicted concept or skill. Participants were given up to two attempts to correctly name each concept or skill. Each correct response received one point, regardless of whether it was provided on the first or second attempt; incorrect responses received zero points. Thus, the highest score a participant could obtain was 50 points and the lowest was 0 points.

The assessment procedure was conducted by the participants' own Physical Education teachers. This approach was selected with the dual aim of minimizing participants' anxiety and maximizing their performance. This decision was based on findings from previous research indicating that the presence of familiar assessors significantly reduces anxiety during the assessment process and improves the performance of individuals with intellectual disabilities (Fuchs & Fuchs, 1986; Szarko et al., 2013). Furthermore, to minimize assessor bias, all PE teachers were provided with a standardized assessment protocol, which was used

to evaluate all participants.

Item analysis of the test

Initially, an item analysis was conducted based on Classical Test Theory (Crocker & Algina, 1986; Lord & Novick, 1968). The analysis included the calculation of the difficulty (p), discrimination (D), and point-biserial correlation (pbsc) indices.

According to Popham (2017), when designing a criterion-referenced test aimed at assessing mastery in specific cognitive domains, such as the EVT–MCS–DS, emphasis should be placed on individual item performance rather than developing a model for a latent construct. Furthermore, sample size dictated the appropriateness of choosing Classical Test Theory for the initial stage of the instrument's validation, rather than Item Response Theory, which potentially provides greater measurement precision (Bond & Fox, 2015).

The difficulty index (p = R/N) was calculated using the total number of correct responses (R) and the total number of participants (N). In line with established psychometric guidelines (Ebel & Frisbie, 1991), and supported by recent applications in educational measurement (Kim et al., 2024; Konakci, 2024; Rezigalla et al., 2024), values between .30 and .80 were considered acceptable, following established test construction guidelines (Haladyna & Rodriguez, 2013).

The number of correct responses given by participants in the high-performing group (Ru), the number of correct responses given by those in the low-performing group (Rl), and the total number of participants in the high-performing group (Nu) were used to calculate the discrimination index (D) {(Ru - Rl)/Nu}. Values between .00 and .19 were considered poor and unacceptable, values between .20 and .29 were considered merely acceptable, values between .30 and .39 were considered good, and values > .40 were considered excellent (Kim et al., 2024; Konakci, 2024; Rezigalla et al., 2024).

The Point-Biserial Correlation (pbsc) index reflects the relationship between each item and the total test score. Values of pbsc were interpreted in accordance with classical item analysis recommendations (Allen & Yen, 2002; Bonnet, 2020).

In addition, possible differences in expressive vocabulary performance according to sex, age group, and living arrangement were examined using non-parametric Mann–Whitney U tests because the data deviated from normality and the groups were unequally distributed.

Content validity

A further assessment of content validity was performed after item analysis. An expert panel comprising five Physical Education teachers holding doctoral degrees and having more than 15 years of professional experience evaluated the 32 retained items on a four-point Likert scale (1 = not at all relevant, 4 = highly relevant).

The Item-Content Validity Index (I-CVI) and Scale-Content Validity Index/ Average (S-CVI/Ave) were calculated. Following established recommendations (Alexandre & Coluci, 2022; Luo et al., 2022), values ≥ .80 for I-CVI and ≥ .90 for S-CVI/Ave were considered acceptable.

Reliability Assessment

Reliability was assessed using Cronbach's α, Kuder-Richardson 20 (KR-20), and split-half reliability with Spearman-Brown correction. Values greater than .80 were considered indicative of good internal consistency, in line with recommendations for scale development (DeVellis, 2016; DeVellis & Thorpe, 2022; George & Mallery, 2019; Tavakol & Dennick, 2011a; Tavakol & Dennick, 2011b).

Test–Retest Reliability

Test-retest reliability was evaluated using the intra-class correlation coefficient (ICC). Twenty-seven participants were randomly selected to complete the test again 15 days after the initial assessment (Johanson & Brooks, 2010; Streiner & Norman, 2008). The reassessment followed the same procedure as the initial assessment. Participants were allowed two attempts per item, and the same scoring procedure was applied as in the initial assessment.

Results

Item Analysis

The original EVT–MCS-DS consisted of fifty images representing movement concepts and skills. To test item difficulty and discriminative ability, the relevant indices were considered. Thus, for each image, the difficulty index (p), discrimination index (D), and corrected item–total correlation (CITC) were calculated.

Based on these indices, eighteen images were removed. Thirteen items were removed because they were too easy (boy, girl, head, lift, catch, let go, big, small, jump, walk, run, sideways, and step on), whereas five items were removed because they were too difficult (partners, balance, leading, hop, and gallop).

The remaining 32 images demonstrated a difficulty index of 0.3 to 0.6, a discrimination index of 0.3 to 1.00, and a corrected item–total correlation (CITC) of 0.19 to 0.88 (Table 1).

Table 1. Item analysis and internal consistency for each image of the EVT-MCS-DS

Item	p	D	CITC
Hands	0.4	1.0	.846
Shoulder	0.3	0.7	.703
Palm	0.4	0.4	.188
Fingers	0.4	0.8	.652
Chest	0.3	0.7	.624
Knee	0.3	0.7	.523
Elbow	0.4	0.8	.673
Waist	0.3	0.8	.771
Stop	0.3	0.8	.615
Start	0.3	0.8	.623
Curved	0.3	0.8	.762
Straight	0.3	0.8	.762
Opposite to	0.3	0.8	.679
Far	0.3	0.8	.646
Near	0.3	0.8	.646
Inside	0.3	0.8	.768
Outside	0.3	0.8	.768
Right	0.3	0.7	.628
Left	0.3	0.7	.628

In front	0.3	0.4	.353
Behind	0.3	0.3	.353
Turning	0.6	0.7	.490
On the spot	0.6	0.3	.275
Zig-zag	0.3	0.6	.453
High	0.3	0.7	.632
Up	0.3	1.0	.882
Down	0.3	1.0	.882
Slow	0.3	0.3	.390
Fast	0.3	0.3	.390
Stepping	0.6	1.0	.554
Next to	0.3	0.7	.546
Low	0.3	0.7	.632

p: Difficulty Index; D: Discrimination Index; CITC: Corrected Item-Total Correlation

Following item analysis, the EVT–MCS–DS consists of 32 images representing skills (4) and movement concepts related to directions (6), pathways (3), extensions (2), level (2), time (2), social interactions (5), and body awareness (8).

Reliability and Validity of the Final Test

To evaluate the reliability of the final version of the test, the following analyses were conducted: (a) Internal consistency of the test: The internal consistency was assessed using Cronbach's α, which was found to be .952, indicating high consistency among the thirty-two items. Because the item responses were binary (correct/incorrect), the KR-20 index was also calculated and found to be .80, indicating good internal consistency. Split-half reliability was also high, as demonstrated by the Spearman-Brown split-half index of .94, indicating high consistency between the two halves of the test. (b) Evidence of content validity: As shown in Table 2, the individual values of the I-CVI ranged from .80 to 1.00, and the overall S-CVI/Ave was .92, providing strong evidence of content validity. (c) Stability: the test was performed by the test-retest method taking into account the Intraclass Correlation Coefficient (ICC) using the two-way mixed effects model with absolute agreement. The results show an excellent stability between the two measurements. The ICC (Single Measures) was found .967 with 95% Confidence Interval [.895, .987], $F(26, 26) = 79.940$ and $p < .001$ while the ICC (Average Measures) was found .983 with 95% Confidence Interval [.945, .993], $F(26, 26) = 79.940$ and $p < .001$

Table 2. Results of content review by the expert panel

Item ID	Expert1	Expert2	Expert3	Expert4	Expert5	I-CVI
Hands	3	4	3	2	3	.80
Shoulder	4	4	4	4	4	1
Palm	4	4	4	4	4	1
Fingers	4	4	4	4	4	1
Chest	2	4	4	4	3	.80
Knee	3	4	4	4	2	.80
Elbow	4	3	4	4	2	.80
Waist	4	2	4	3	3	.80
Stop	4	4	3	3	4	1
Start	3	4	4	3	2	.80
Curved	4	4	4	4	4	1
Straight	4	4	4	4	4	1
Opposite to	4	4	4	4	4	1
Far	4	4	4	4	4	1
Near	3	4	2	4	3	.80
Inside	3	4	4	2	3	.80

Outside	4	4	4	4	4	1
Right	4	4	4	4	4	1
Left	4	3	4	4	2	.80
In front	4	4	4	4	4	1
Behind	3	4	4	2	4	.80
Turning	4	4	4	4	4	1
On the spot	4	4	4	4	4	1
Zig-zag	4	4	4	4	4	1
High	4	4	4	4	4	1
Up	4	4	4	4	4	1
Down	4	4	4	2	3	.80
Slow	4	4	2	4	3	.80
Fast	4	4	4	4	4	1
Stepping	4	4	4	4	4	1
Next to	4	4	4	4	4	1
Low	3	4	4	2	3	.80

I-CVI: Item-Content Validity Index

Sex Differences in Expressive Vocabulary Performance

To investigate sex differences in expressive vocabulary, the non-parametric Mann–Whitney U test was applied due to the non-normal distribution of the data and the unequal distribution of the sample between the two groups. Men demonstrated slightly higher performance (M = 10.58, SD = 9.68, Mdn = 7.00) compared to women (M = 8.56, SD = 8.68, Mdn = 7.50); however, the difference was not statistically significant, U = 388.50, z = −0.63, p = .529, r = .08.

Differences by Living Environment

To investigate whether living environment was associated with differences in expressive vocabulary, a non-parametric Mann–Whitney U test was conducted due to the non-normal distribution of the data and the simultaneous presence of many zero values (living in institutions: W = .807, p< .001; participants living with their families: W = .873, p = .006). Participants living with their families demonstrated higher EVT-MCS-DS scores (Mdn = 13.5, IQR = 20) compared to those residing in residential institutions (Mdn = 5, IQR = 13). The difference between the two groups was statistically significant, U = 331.5, z = −2.31, p = .021, with a small-to-moderate effect size (r = .28). As shown in Figure 2, participants living with their families demonstrated higher performance and correctly named more items than those living in residential institutions.

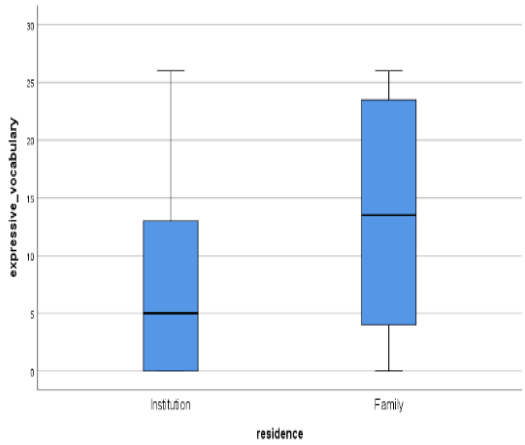

Figure 2. Box plot of EVT-MCS-DS scores by living environment

Differences by Age Group

To investigate whether age was associated with differences in expressive vocabulary, a non-parametric Mann–Whitney *U* test was conducted due to the non-normal distribution of the data and the simultaneous presence of many zero values. As shown in Table 3, participants aged 20–25 years correctly named more items (*Mdn* = 23, *IQR* = 11) compared with those aged 26–32 years (*Mdn* = 4, *IQR* = 7). The Mann–Whitney *U* test showed that the difference between the two age groups was statistically significant, $U = 109$, $z = -5.37$, $p < .001$, with a large effect size ($r = .66$).

Table 3. Descriptive statistics and results of the Mann–Whitney U test for expressive vocabulary by age group

Age group	*n*	*Mdn*	*IQR*	*Mean Rank*	*U*	*z*	*p*
20-25	25	23	11	49.64	109	-5.37	< .001
26-32	41	4	7	23.66			

Mdn = median, *IQR* = interquartile range

Figure 3 illustrates the distribution of EVT-MCS-DS scores across the two age groups, showing higher scores among participants aged 20–25 years.

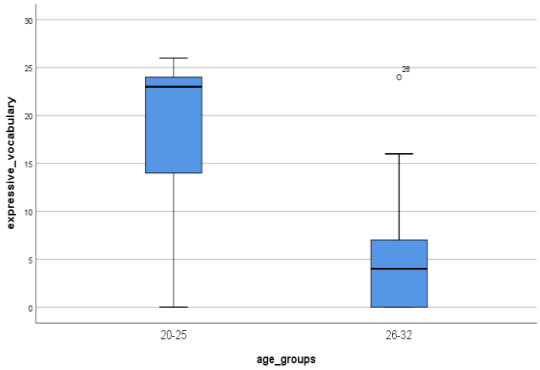

Figure 3. Box plot of EVT-MCS-DS scores by age group.

Discussion

The present study aimed to develop and psychometrically evaluate the EVT-MCS-DS and to examine possible differences in expressive vocabulary performance according to demographic characteristics, including sex, living arrangement, and age group. A sequential and rigorously structured procedure was implemented to develop and validate the EVT–MCS–DS. Following international guidelines for the development of measurement instruments (Boateng et al., 2018; Menteş Tanaydın & Akman, 2026), the initial items were evaluated by expert panels, followed by item analysis based on difficulty, discrimination, and point-biserial correlation indices. The final stage of the validation process involved content validity assessment.

The EVT–MCS–DS was designed as a criterion-referenced cognitive test that enables teachers to assess examinees' mastery of specific motor concepts and skills. In this type of assessment, each item —whether visual or verbal— corresponds to a distinct cognitive domain. This means that each test item represents a discrete area of knowledge rather than a latent psychological construct. According to the American Educational Research Association (AERA, 2014) and DeVellis and Thorpe (2022), the existence of latent variables is a prerequisite for conducting factor analysis. Consequently, factor analysis was not considered the most appropriate approach for evaluating the validity of criterion-referenced cognitive tests. The development of the EVT–MCS–DS was informed by the perspective of Black and Wiliam (2009), who argue that assessment is a process that supports learning rather than merely a means of measuring performance. A similar perspective was presented by Florian and Black-Hawkins (2011), who argued that assessment, within the framework of inclusive education, should provide the information necessary to support the design of individualized intervention programs and facilitate participation.

The development of the EVT–MCS–DS is consistent with the views of Pruden, Levine, and Huttenlocher (2011) and Oudgenoeg–Paz et al. (2012), who suggested that knowledge of spatial and motor concepts is important for participation in educational programs because it facilitates the understanding of instructions. Furthermore, learning and using these concepts is closely associated with cognitive skills and the learning process. Considering that individuals with DS often experience motor difficulties and have a limited expressive vocabulary, improving functional communication, in combination with the acquisition of movement concepts, may contribute to social participation (Fidler, 2005; Martin et al., 2009). Thus, concepts such as "right", "left", "front", "back", "slow", "fast", and "next to", which are included in the EVT–MCS–DS, represent functional concepts closely related to autonomy and safety in individuals with DS, as they are important for understanding instructions required for participation in physical activities and activities of daily living.

Following the guidelines for criterion-referenced test development (DeVellis & Thorpe, 2022; Kline, 2015), this study focused primarily on indicators of difficulty, discriminative power, content validity, and reliability. The systematic development process enabled progressive refinement of the test by excluding items that did not meet acceptable indicator values and ensuring the alignment of the retained items with the conceptual framework of expressive vocabulary related to movement concepts and skills in adults with DS. Although the item "Palm" showed a relatively low Corrected Item–Total Correlation value (.188), it was retained in the final version of the EVT–MCS–DS based on its content relevance and functional importance. Body-awareness concepts represent important components of motor instruction, everyday communication, and participation in physical activities among adults with DS. In criterion-referenced assessments, item retention is guided not only by statistical indicators but

also by theoretical relevance and instructional value (DeVellis & Thorpe, 2022; Popham, 2017). The combined methodological approach —incorporating expert judgement and empirical data— enhanced the validity and reliability of the final test and is in line with international guidelines for the development of psychometric tests in the field of special needs Physical Education (Kline, 2015).

The findings indicate that the final version of the test demonstrates strong content validity, high clarity, and satisfactory reliability indicators. The high level of agreement among experts in evaluating the images suggests that the concepts related to motor activities and the body were accurately represented. Furthermore, the observed high inter-expert agreement (I-CVI and S-CVI) confirms that the selected verbal items adequately covered the basic spatial movement concepts and skills. The findings support the potential applicability of the test in research contexts and in the evaluation of Physical Education intervention programs. The content validity findings further suggest that the majority of images were considered appropriate and clear, and that participants recognized them relatively easily.

The internal consistency indices indicate strong coherence among the items assessing expressive vocabulary related to movement concepts and skills. Considering that the EVT–MCS–DS is a criterion-referenced cognitive test rather than a tool designed to measure latent variables, its internal consistency reflects content homogeneity rather than the existence of a one-dimensional latent structure. This finding is relevant because expressive language performance in individuals with DS is often heterogeneous. The high value of Cronbach's α was expected, given that the EVT–MCS–DS is a cognitive assessment tool with clearly defined and homogeneous content (DeVellis & Thorpe, 2022; Tavakol & Dennick, 2011b). The observed difference between Cronbach's α and KR-20 values is likely due to the relatively small number of participants and the distribution of item difficulty, two factors that may affect reliability indices in dichotomous data.

These findings suggest that the coherence of the test content reflects thematic consistency rather than redundancy among individual items (Streiner, 2003). Although Cronbach's α was high, the retained items represented distinct but conceptually related skills and movement concepts, supporting content homogeneity rather than item redundancy. The findings further indicate that movement concepts and skills represent a distinct category that can be reliably measured (Koo & Li, 2016).

The test–retest results indicated strong temporal stability across repeated measurements. The observed temporal stability supports the use of the EVT–MCS–DS in longitudinal intervention programmes, where accurate and consistent measurements are required to evaluate changes in participants' performance. The ICC values for Single Measures indicate that the observed variability is due to real differences between participants rather than measurement error. The reliability of the estimate is also confirmed by the F and p values. Another element reflecting the consistency of the test in repeated assessments is the ICC for Average Measures, which demonstrates consistency when the average of multiple measurements is used. Finally, the agreement-based ICC indicates that the test not only consistently captures the relative ranking of participants but also accurately reflects their absolute values.

No significant differences according to sex were observed in expressive vocabulary scores. This finding suggests that, within the present sample, expressive vocabulary related to skills and movement concepts did not differ according to sex among individuals with DS.

Furthermore, significant differences were found with regard to living environment, as individuals living with their families demonstrated higher performance in expressive vocabulary compared to those living in institutions. This finding may be interpreted as suggesting that family environments could provide more opportunities for daily verbal interaction, personalized communication, and participation in community activities.

The findings of this study reveal significant age-related differences in participants' expressive vocabulary. Specifically, younger participants (aged 20–25) achieved significantly higher performance in expressive vocabulary compared to older participants (aged 26–32). This finding is consistent with the results of other studies (Carr, 2008; Deckers et al., 2017), which suggest that language performance, verbal memory, and certain cognitive functions in individuals with DS may show a gradual decline with increasing age. These differences may be associated with age-related changes in cognitive processes, including working memory and information processing speed. This finding supports the need for timely and continuous educational and language intervention programs aimed at maintaining functional communication and expressive vocabulary in adults with Down syndrome.

The sample size should be interpreted in light of the recruitment challenges associated with this clinical population. Recruitment difficulties were related to the challenges of obtaining consent from caregivers and responsible authorities while strictly adhering to the predefined participation criteria. Furthermore, according to the recommendations in the literature (AERA, 2014), in studies targeting specialized or clinical populations, such as the present study, sample homogeneity, in addition to sample size, is also considered important.

Limitations

This study has several limitations:

(1) The relatively small sample size may limit the generalizability of the findings.
(2) The study was conducted within a specific cultural and educational context, which may limit the applicability of the findings to other populations.
(3) Although indicators proposed by the scientific community were taken into account to investigate reliability and content validity, further testing using larger and more heterogeneous samples may contribute to strengthening the psychometric characteristics of the test.
(4) Participants' verbal performance may have been influenced by individual differences, including verbal fluency, attention, and previous experience with the specific concepts.
(5) The relationship between participants' general cognitive level or mental age and their performance on the EVT–MCS–DS was not examined. Consequently, further research is needed to investigate the extent to which expressive vocabulary performance on the test is associated with broader cognitive functioning, particularly among lower-performing individuals.
(6) The EVT–MCS–DS was not compared with existing standardized language assessment instruments. Therefore, further research is needed to examine its relationship with other established linguistic tools and provide additional evidence of convergent validity

Conclusions

The EVT–MCS–DS appears to be a practical instrument that demonstrates satisfactory evidence of reliability and content validity for the assessment of expressive vocabulary in adults with Down syndrome. The test may be useful in supporting the design and evaluation of educational programmes targeting both vocabulary and motor skills and movement concepts.

In addition, due to its criterion-referenced nature, the EVT–MCS–DS may provide teachers and therapists with individualized information regarding the specific skills and movement concepts that adults with DS have mastered and those requiring further support, thereby facilitating targeted educational planning and intervention.

Future Research Directions

Future studies should include larger and more heterogeneous samples in order to further investigate the psychometric properties and cross-validation of the instrument in different educational and cultural contexts. Additional psychometric investigations using contemporary measurement approaches, such as Rasch analysis, may provide additional evidence regarding the measurement properties of the test. Furthermore, future longitudinal studies using the EVT–MCS–DS may help monitor age-related changes in expressive vocabulary and functional communication skills in adults with DS.

Declarations

Acknowledgments: *N/A*
Funding: *This research received no external funding.*
Ethics Statements: *The study was approved by the Ethics Committee of the D.U.Th. (Decision 1st/26-09-2024).*
Conflict of Interest: *The authors declare no conflict of interest.*
Informed Consent*: Permission was also granted by the managers of the facilities and by the parents/guardians of participants living with their families. Parents and guardians received written assurance of full anonymity and data protection.*
Data availability: *The data that support the findings of this study are available upon request from the authors.*

References

Abbeduto, L., Warren, S.F., & Conners, F.A. (2007). Language development in Down syndrome: From the prelinguistic period to the acquisition of literacy. *Mental Retardation and Developmental Disabilities Research Reviews, 13*(3), 247–261. https://doi.org/10.1002/mrdd.20158

Aksen-Cengizhan, P., Ahmet Dogan, A., & Isık, O. (2021). Avoiding isolation: Physical activity for persons with disabilities and Covid-19. *Baltic Journal of Health and Physical Activity, 13*(Suppl. 1), 65–73. https://doi.org/10.29359/BJHPA.2021.Suppl.1.08

Allen, M.J., & Yen, W.M. (2002). *Introduction to measurement theory*. Waveland Press.

Alexandre, N.M., & Coluci, M.Z. (2011). Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas [Content validity in the development and adaptation processes of measurement instruments]. *Ciência & Saúde Coletiva, 16*(7), 3061–3068.

https://doi.org/10.1590/s1413-81232011000800006
American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
Baadte, C., & Meinhardt-Injac, B. (2019). The picture superiority effect in associative memory: A developmental study. *British Journal of Developmental Psychology, 37*(3), 382–395. https://doi.org/10.1111/bjdp.12280
Barsalou L.W. (2008). Grounded cognition. *Annual review of psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639
Bedford, R., Pickles, A., & Lord, C. (2016). Early gross motor skills predict the subsequent development of language in children with autism spectrum disorder. *Autism Research, 9*(9), 993–1001. https://doi.org/10.1002/aur.1587
Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5
Boateng, G.O., Neilands, T.B., Frongillo, E.A., Melgar-Quiñonez, H.R., & Young, S.L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149
Bond, T.G., & Fox, C.M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge. https://doi.org/10.4324/9781315814698
Bonett, D.G. (2020). Point-biserial correlation: Interval estimation, hypothesis testing, meta-analysis, and sample size determination. *British Journal of Mathematical and Statistical Psychology, 73*(Suppl. 1), 113–144. https://doi.org/10.1111/bmsp.12189
Bull, M.J. (2020). Down syndrome. *New England Journal of Medicine, 382*(24), 2344–2352. https://doi.org/10.1056/NEJMra1706537
Carr, J. (2008). The Everyday Life of Adults with Down Syndrome. *Journal of Applied Research in Intellectual Disabilities*, *21*, 389-397. https://doi.org/10.1111/j.1468-3148.2007.00418.x
Cristea, D.I., Pop, A.C., Tatar, C.F., Buhas, R., Olszewski-Strzyzowski, D.J., & Stef, M. (2025). Physical activity, cultural tourism and social inclusion for persons with Down syndrome. *Baltic Journal of Health and Physical Activity, 17*(3), Article 1. https://doi.org/10.29359/BJHPA.17.3.01
Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.
Custode, S.A., & Tamis-LeMonda, C. (2020). Cracking the code: Social and contextual cues to language input in the home environment. *Infancy, 25*(6), 809–826. https://doi.org/10.1111/infa.12361
Deckers, S.R.J.M., Van Zaalen, Y., Van Balkom, H., & Verhoeven, L. (2019). Predictors of receptive and expressive vocabulary development in children with Down syndrome. *International Journal of Speech-Language Pathology, 21*(1), 10–22. https://doi.org/10.1080/17549507.2017.1363290
DeVellis, R.F., & Thorpe, C.T. (2022). *Scale development: Theory and applications* (5th ed.). Sage.
Ebel, R.L., & Frisbie, D.A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.
Fidler, D.J. (2005). The emerging Down syndrome behavioral phenotype in early childhood: *Implications for practice. Infants & Young Children*, *18*(2), 86–103. https://doi.org/10.1097/00001163-200504000-00003
Filipe, M.G., Cruz, S., Veloso, A.S., & Frota, S. (2022). Early predictors of language

outcomes in Down syndrome: A mini-review. *Frontiers in Psychology, 13*, 934490. https://doi.org/10.3389/fpsyg.2022.934490

Finestack, L. H., Sterling, A. M., & Abbeduto, L. (2013). Discriminating Down Syndrome and Fragile X syndrome based on language ability. *Journal of child language*, *40*(1), 244–265. https://doi.org/10.1017/S0305000912000207

Florian, L., & Black-Hawkins, K. (2011). Exploring inclusive pedagogy. *British Educational Research Journal*, *37*(5), 813–828. https://doi.org/10.1080/01411926.2010.501096

Fuchs, D., & Fuchs, L.S. (1986). Test procedure bias: A meta-analysis of examiner familiarity effects. *Review of Educational Research, 56*(2), 243–262. https://doi.org/10.3102/003465430560022

George, D., & Mallery, P. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge. https://doi.org/10.4324/9780429056765

Polychronopoulou, St., Alevriadou, A., Anastasiou, D., Antonopoulou, K., Asteri, Th., Belitsou, N., Papadopoulos, A., Pappa, I., Stamboltzi, A., Stavrousi, P. (2004). *Comprehensive curricula of compulsory education for students with moderate and mild mental retardation*. Greek Ministry of Education & Religious Affairs.

Haladyna, T.M., & Rodriguez, M.C. (2013). *Developing and validating test items* (3rd ed.). Routledge.

Hwang, B.H., & Lee, D. (2024). Association between motor and language skills development in children with autism spectrum disorder: A scoping review. *International Journal of Disability, Development and Education, 71*(2), 135–149. https://doi.org/10.1080/1034912X.2022.2092081

Johanson, G.A., & Brooks, G.P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement, 70*(3), 394–400. https://doi.org/10.1177/0013164409355692

Iverson J.M. (2010). Developing language in a developing body: the relationship between motor development and language development. *Journal of child language*, *37*(2), 229–261. https://doi.org/10.1017/S0305000909990432

Karakuş, S., & Akbaş, Y. (2026). Water literacy of secondary school students: A scale development study. *Participatory Educational Research, 13*(1), 256–275. https://doi.org/10.17275/per.26.14.13.1

Karimi, A., & Nelson, E.L. (2023). Motor-language links in children with Down syndrome: A scoping review to revisit the literature with a developmental cascades lens. *Frontiers in Psychology, 14*, 1275325. https://doi.org/10.3389/fpsyg.2023.1275325

Kim, Y.H., Kim, B.H., Kim, J., Jung, B., & Bae, S. (2023). Item difficulty index, discrimination index, and reliability of the 26 health professions licensing examinations in 2022, Korea: A psychometric study. *Journal of Educational Evaluation for Health Professions, 20*, 31. https://doi.org/10.3352/jeehp.2023.20.31

Kline, P.A. (2015). *A handbook of test construction: Introduction to psychometric design*. Routledge.

Koo, T.K., & Li, M.Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Konakcı, S. (2024). Item analysis in multiple choice questions: A study on question difficulty and authors' evaluation. *Journal of Basic and Clinical Health Sciences, 8*(2), 490–497. https://doi.org/10.30621/jbachs.1427464

Korkmaz, Ö., & Bai, X. (2019). Adapting computational thinking scale (CTS) for Chinese high school students and their thinking scale skills level. *Participatory Educational Research, 6(*1), 10–26. https://doi.org/10.17275/per.19.2.6.1

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Addison-

Wesley.
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*, 929433. https://doi.org/10.3389/fpsyg.2022.929433
Luo, L., Song, N., Huang, J., Zou, X., Yuan, J., Li, C., Yang, J., Zhou, L., Zhang, L., Luo, S., & Gao, X. (2022). Validity evaluation of the college student physical literacy questionnaire. *Frontiers in Public Health, 10*, 856659. https://doi.org/10.3389/fpubh.2022.856659
Martin, G.E., Klusek, J., Estigarribia, B., & Roberts, J.E. (2009). Language characteristics of individuals with Down syndrome. *Topics in Language Disorders, 29*(2), 112–132. https://doi.org/10.1097/tld.0b013e3181a71fe1
Menteş Tanaydın, H.S., & Akman, B. (2026). Computational thinking in early childhood: Psychometric properties of the BD-HAM test with Rasch model. *Participatory Educational Research, 13*(1), 276–296. https://doi.org/10.17275/per.26.15.13.1
Onnivello, S., Pulina, F., Locatelli, C., Marcolin, C., Ramacieri, G., Antonaros, F., Vione, B., Caracausi, M., & Lanfranchi, S. (2022). Cognitive profiles in children and adolescents with Down syndrome. *Scientific Reports, 12*(1), 1936. https://doi.org/10.1038/s41598-022-05825-4
Oudgenoeg-Paz, O., Volman, M.C., & Leseman, P.P. (2012). Attainment of sitting and walking predicts development of productive vocabulary between ages 16 and 28 months. *Infant behavior & development*, *35*(4), 733–736. https://doi.org/10.1016/j.infbeh.2012.07.010
Palisano, R.J., Walter, S.D., Russell, D.J., Rosenbaum, P.L., Gémus, M., Galuppi, B.E., & Cunningham, L. (2001). Gross motor function of children with Down syndrome: Creation of motor growth curves. *Archives of Physical Medicine and Rehabilitation, 82*(4), 494–500. https://doi.org/10.1053/apmr.2001.21956
Popham, W.J. (2017). *Classroom assessment: What teachers need to know* (8thed.). Pearson.
Pruden, M.S., Levine, C.S., & Huttenlocher, J. (2011). Children's spatial thinking: does talk about the spatial world matter? *Developmental Science*, *14*(6) 1417-1430. https://doi.org/10.1111/j.1467-7687.2011.01088.x
Rezigalla, A.A., Eleragi, A.M.E.S.A., Elhussein, A.B., Alfaifi, J., AlGhamdi, M.A., Al Ameer, A.Y., Yahia, A.I.O., Mohammed, O.A., & Adam, M.I.E. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education, 24*(1), 445. https://doi.org/10.1186/s12909-024-05433-y
Streiner, D.L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18
Streiner, D.L., & Norman, G.R. (2008). *Health measurement scales: A practical guide to their development and use* (4thed.). Oxford University Press.
Szarko, J., Brown, A., & Watkins, M. (2013). Examiner familiarity effects for children with autism spectrum disorders. *Journal of Applied School Psychology, 29*(1), 37–51. https://doi.org/10.1080/15377903.2013.751475
Tavakol, M., & Dennick, R. (2011a). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd
Tavakol, M., & Dennick, R. (2011b). Post-examination analysis of objective tests. *Medical Teacher, 33*(6), 447–458. https://doi.org/10.3109/0142159X.2011.564682
te Kaat, D., Volman, M., Jongmans, M., & Lauteslager, P.E.M. (2016). Expressive

vocabulary development in children with Down syndrome: A longitudinal study. *Journal of Policy and Practice in Intellectual Disabilities, 14*(4), 311–318. https://doi.org/10.1111/jppi.12212

Thelen, E., & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press.

Thurman, A.J., Edgin, J.O., Sherman, S.L., Sterling, A., McDuffie, A., Berry- Kravis, E., Hamilton, D., & Abbeduto, L. (2021). Spoken language outcome measures for treatment studies in Down syndrome. *Journal of Neurodevelopmental Disorders, 13*(1), 13. https://doi.org/10.1186/s11689-021-09361-6

Will, E.A., & Roberts, J.E. (2021). Motor influences on communication: Comparisons between Down syndrome and Fragile X syndrome. *American Journal on Intellectual and Developmental Disabilities, 126*(6), 460–476. https://doi.org/10.1352/1944-7558-126.6.460

Worthington, R.L., & Whittaker, T.A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806–838. https://doi.org/10.1177/0011000006288127

Yamauchi, Y., Aoki, S., Koike, J., Hanzawa, N., & Hashimoto, K. (2019). Motor and cognitive development of children with Down syndrome. *Brain & Development, 41*(4), 320–326. https://doi.org/10.1016/j.braindev.2018.11.008