Assessing language comprehension in motor impaired children needing AAC: validity and reliability of the Norwegian version of the receptive language test C-BiLLT

Sara Ida Fiske , Anne Lise Haddeland , Ingvild Skipar , Jael N. Bootsma , Johanna J. Geytenbeek & Kristine Stadskleiv

Published online: 24 Jul 2020.

Submit your article to this journal

Article views: 64

View related articles

View Crossmark data

RESEARCH ARTICLE

Assessing language comprehension in motor impaired children needing AAC: validity and reliability of the Norwegian version of the receptive language test C-BiLLT

Sara Ida Fiske[a], Anne Lise Haddeland[a], Ingvild Skipar[b], Jael N. Bootsma[c], Johanna J. Geytenbeek[d] and Kristine Stadskleiv[a,e]

[a]Department of Special Needs Education, University of Oslo, Oslo, Norway; [b]Department of Educational Science, University of South-Eastern Norway, Borre, Norway; [c]CanChild Centre for Childhood Disability Research and School of Rehabilitation Science, McMaster University, Hamilton, Canada; [d]Rehabilitation Department, Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands; [e]Department of Clinical Neurosciences for Children, Oslo University Hospital, Oslo, Norway

ABSTRACT

Children with severe motor impairments who need augmentative and alternative communication (AAC) comprise a heterogeneous group with wide variability in cognitive functioning. Assessment of language comprehension will help find the best possible communication solution for each child, but there is a lack of appropriate instruments. This study investigates the reliability and validity of the Norwegian version of the spoken language comprehension test C-BiLLT (computer-based instrument for low motor language testing) – the C-BiLLT-Nor – and whether response modality influences test results. The participants were 238 children with typical development aged 1;2 to 7;10 (years/months) who were assessed with the C-BiLLT-Nor and tests of language comprehension and non-verbal reasoning. There was excellent internal consistency and good test–retest reliability. Tests of language comprehension and non-verbal reasoning correlated significantly with the C-BiLLT-Nor, indicating good construct validity. Factor analysis yielded a two-factor solution, suggesting it as a measure of receptive vocabulary, grammar, and overall language comprehension. No difference in results could be related to response mode, implying that gaze pointing is a viable option for children who cannot point with a finger. The C-BiLLT-Nor, with norms from 1;6–7;6 is a reliable measure of language comprehension.

Children with severe speech and movement impairments comprise a heterogeneous group with wide variability in functioning. Cerebral palsy (CP) is the most common cause of movement impairments in childhood with a prevalence of approximately 2 per 1000 live births (Andersen et al., 2008; Himmelmann, Hagberg, & Uvebrant, 2010). CP is a motor disorder caused by lesions or anomalies in the developing brain. The most recent definition also emphasizes that many children have additional impairments, including communication difficulties that can severely limit their participation in daily life (Rosenbaum et al., 2007). Difficulties with functional communication are frequent in children with more severe motor impairments. Of children classified at Level IV and V on the Gross Motor Function Classification System (GMFCS; Palisano et al., 1997), 85–100% have motor-speech difficulties (Andersen, Mjøen, & Vik, 2010; Himmelmann, Lindh, & Hidecker, 2013; Voorman, Dallmeijer, Van Eck, Schuengel, & Becher, 2009). Children with CP and anarthria, a severe form of motor-speech disorder resulting in the complete absence of speech, are often described as non-verbal. This can be misinterpreted to mean that they do not attain language at all, while in fact, their receptive and expressive language

abilities follow two different development trajectories: (a) expressive language development relating to the degree of motor disability, and (b) receptive language ability to cognitive functioning (Choi, Choi, & Park, 2017; Mei et al., 2016; Vos et al., 2014). Indeed, studies have shown that receptive language abilities in children who do not speak can develop at the same rate as for peers with typical development (Geytenbeek, Heim, Knol, Vermeulen, & Oostrom, 2015; Stadskleiv, Jahnsen, Andersen, & von Tetzchner, 2018). Because there is no one-to-one correspondence between the receptive and expressive language skills of children with severe speech and motor impairments, an assessment of the language comprehension is necessary for planning interventions, including finding the augmentative and alternative communication (AAC) solution that best fits the child's needs. These interventions are fundamental for access to language, literacy, and learning because they enable children to express themselves and provide a means to be communicative and become active participants within their environments (Batorowicz, Stadskleiv, von Tetzchner, & Missiuna, 2016; Murray & Goldbart, 2009).

The assessment of children with severe speech and motor impairment is often considered challenging (Geytenbeek et al., 2010). There are standardized tests available to assess language comprehension in children under 6 years of age (Bishop, 2009; Dunn & Dunn, 1997; Dunn, Dunn, Whetton, & Burley, 1997; Edwards, 1997; Semel, Wiig, & Secord, 2003; Wechsler, 2015) but many of these tests are designed for children with typical development and require either a verbal response, finger-pointing at pictures/objects, or manipulation of small objects (Yin Foo, Guppy, & Johnston, 2013). Such response modes are not feasible for assessing language comprehension in children with severe speech and motor impairments (Geytenbeek et al., 2010; Pirila et al., 2007; Sabbadini, Bonanni, Carlesimo, & Caltagirone, 2001). The lack of standardized tests suited to this population can have serious consequences because interventions will be based on assumptions rather than knowledge of individual capabilities. For example, a register-based study (i.e., a study based on a large representative geographical cohort), reported that only 29% of children with CP had received a standardized assessment of cognitive abilities (Andersen et al., 2008). In particular, those with the most severe speech and motor impairments are the least likely to be assessed (Smits et al., 2011). In many cases, then, interventions are based on clinical evaluations and intuition, something that can lead to an over- or under-estimation of a child's functioning and, potentially, implementation of interventions and AAC systems that are ill-suited to the child's needs and not functional for communication (Andersen et al., 2010).

Despite the challenges, it is possible to assess cognition and language in children with severe speech and motor impairments by, for example, substituting standard response modes with more accessible options. In multiple-choice tests that require finger-pointing as a response mode, alternative ways of pointing can be a reliable adaptation. Using different response modalities has not been shown to influence test results in children with typical development (Kurmanaviciute & Stadskleiv, 2017; Spillane, Ross, & Vasa, 1996); however, the research on alternative response modes and their potential effects on test results is sparse, with small sample sizes, and have included only a few studies where computerized gaze equipment was utilized (Kurmanaviciute & Stadskleiv, 2017).

Research is also scarce when it comes to considering the cognitive and linguistic demands involved in using AAC (Light & McNaughton, 2014). Access to AAC increases the child's communicative possibilities and may further the child's language development (Holyfield, Caron, & Light, 2019; Romski & Sevcik, 2005). In children with typical development, a relationship between neural connectivity and exposure to adult-child conversational experience has been found (Romeo et al., 2018). Given that neuroplasticity decreases with age, it is therefore recommended that interventions be implemented as early as possible (Herskind, Greisen, & Nielsen, 2015). Early interventions might not only promote development and learning, and subsequently, lead to increased activity and participation but also may prevent the development of secondary impairments (Barker et al., 2019; Novak et al., 2017; Romski & Sevcik, 2005).

Test scores are used to draw conclusions about a child's level of language comprehension, and it is essential that these conclusions can be made with confidence. Tests, therefore, need to have acceptable reliability and validity. Reliability pertains to the dependability of test scores and ensuring that the influence of random errors is at an acceptable level (Koo & Li, 2016; Mokkink et al., 2010). This means that a child's result on the test should be stable across test administrators (inter-rater reliability), time (test–retest reliability), and test items (internal consistency) if all test items are designed to measure the same construct. Validity pertains to the degree to which an instrument measures the construct it aims to measure (Mokkink et al., 2010). If the construct is language comprehension it is therefore expected that children's scores on the test at different ages will be in line with what is known about the development of language comprehension (construct validity). There should also be a stronger correlation between test scores on instruments designed to measure the same construct versus those designed to measure different constructs (convergent and discriminant validity). Furthermore, a statistical analysis of the internal structure of the test should be in line with theoretical expectations about the concept (structural validity).

There is a theoretical divide regarding what factors influence the development of language, and this divide also affects the view on the inclusion of receptive vocabulary and receptive grammar in the same test and sum score (Language and Reading Research Consortium, 2015). On the one hand are those who regard the development of grammar and vocabulary as dependent on innate and specialized modules that are independent and governed by different learning mechanisms (Chomsky, 1959; Pinker, 1998). Because children may have selective impairments in vocabulary and grammar, it is recommended that vocabulary and grammar assessments be separated because a sum score would be influenced in an unknown way by the variation in the two phenomena (Slocum-Gori & Zumbo, 2011). On the other hand, there are those who consider vocabulary and grammar closely interrelated and dependent on the same factors and learning mechanisms (Tomasello, 2003), thus making the separation of grammar and vocabulary in assessment less justified. This view is supported by several studies of the factor structure of language tests, which indicate that variations in young children's language ability are best explained with a unidimensional factor solution representing one general language ability (Anthony, Davis, Williams, & Anthony, 2014; Colledge et al., 2002; Hayiou-Thomas et al., 2006; Klem, Gustafsson, & Hagtvet, 2006, 2015). Not until later do grammar and vocabulary become increasingly specialized and independent abilities, reflected in a decrease in the correlation between measures of vocabulary and grammar with increasing age (Justice et al., 2018; Tomblin & Zhang, 2006). This does not necessarily imply that vocabulary and grammar become unrelated factors, but rather that vocabulary is an unconstrained skill where the development is ongoing, compared to the development of grammar, which at some point is completed.

A language comprehension test should cover a wide range of functioning, including for children who are very young or functioning at lower comprehension levels and those who are older or high functioning. In view of the importance of early intervention and follow-up of such interventions, the age-range included should span all of the preschool years (i.e., from comprehension of the first spoken words to comprehension of complex sentence structures), and the test should, therefore, include both assessments of single words (vocabulary) and sentences (grammar). There are language comprehension tests that require only minimal motor skills, such as the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997) and the Test for Reception of Grammar (TROG; Bishop, 2009) but these instruments cover either comprehension of vocabulary or comprehension of grammar – but not both.

Accordingly, there is a need for a measure of receptive language that covers the preschool years, includes the assessment of vocabulary and grammar, requires only pointing as a response mode, and is suitable for children with motor impairments. The Computer-Based instrument for Low motor Language Testing (C-BiLLT; Geytenbeek, Mokkink, Knol, Vermeulen, & Oostrom, 2014) is such a measure. It is a computer-based test with questions (or items) that can be answered (or responded to) using different response modes such as pointing to or touching a screen, eye gaze, eye-gaze computer control, and partner-assisted scanning or scanning with input switches. This test has clear and large photos displayed on a 19-inch (48.26 cm) screen and as such poses lesser demands on the child's ability to point precisely, either with the finger or with eye gaze. Moreover, the large photos make it easier for children with visual impairments to interpret the images. The C-BiLLT has no time limit, which has been shown to give a more correct picture of the comprehension level of children with motor impairments (Sherwell et al., 2014). Visual feedback in the form of a red square around the image is provided for all response modes, making it easy to see which answer option the child has chosen. The test was developed in The Netherlands, and the Dutch version, normed for the age groups $1.5 - 7$ years, has been available since 2014. The pilot version of the Dutch C-BiLLT showed promising validity measures but required small but significant changes to strengthen the reliability and validity of the tool. In the current extended version, more complex items were included. The test continues to show excellent measures of validity and reliability, and the ability to discriminate between children with spastic and dyskinetic cerebral palsy. A sample of children with severe CP and complex communication needs (Geytenbeek, Vermeulen, Becher, & Oostrom, 2015), showed that white matter brain abnormalities were the most important explanatory factors for poor language comprehension (Geytenbeek, Oostrom, et al., 2015) and that sentence comprehension development was delayed rather than deviant in the non-speaking children with severe CP (Geytenbeek, Heim, et al., 2015). The Dutch C-BiLLT has been shown to capture the large individual variety of language comprehension found among children without speech (Geytenbeek, Heim, et al., 2015).

Due to differences in language and culture, the Dutch C-BiLLT was adapted for use in Norway, using guidelines for translating and adapting tests from the International Test Commission (2017). Adaptation of a test involves not only the literal translation of the test items but also a cultural-linguistic adaption process to ensure that difficulty levels are corresponding and test material is culturally appropriate, in addition to investigating the reliability and validity of the adapted version (Hambleton, 2005; Krach, McCreery, & Guerard, 2017).

Accordingly, the primary objective of the current study was to examine the reliability and validity of the C-BiLLT-Nor. The secondary aim was to explore the effect of answering test items using finger-pointing versus gaze-pointing response modes. Specifically, the following were expected: (a) a Cronbach's alpha >0.90, indicating acceptable internal consistency; (b) an intraclass correlation coefficients >0.75, indicating acceptable test-retest and inter-rater reliability; (c) that C-BiLLT-Nor would correlate higher with other measures of language comprehension relative to measures that are dissimilar (i.e., of non-verbal reasoning), thus indicating convergent and discriminant validity respectively; (d) significant differences in raw scores on the C-BiLLT-Nor between children of different ages, indicating construct validity; (e) that the internal structure of the Norwegian version would be the same as in the Dutch investigation of C-BiLLT, resulting in a unidimensional model that represents an overall language comprehension ability, indicating structural validity; and (f) no differences in test results that could be related to the response mode used in answering the test items.

Method

Participants

The participants were children with typical development. For a normed test it is imperative to know the typical functioning in the population in which the instrument is to be applied, and comparing response modalities presupposes that the participants have motor skills to use all modes. The following were the inclusion criteria, which were verified by parent report: (a) aged between 1;0 and 8;0, (b) no history of speech or language delay, (c) no history of auditory or visual problems, (d) no history of developmental delay or neurological or chronic disorders, and (e) Norwegian reported to be the first language.

The goal was to recruit at least eight children per half-year group, totaling 96 participants, because that is required to detect a difference of one standard deviation in a normally distributed sample when power is set at 0.80 and an α-level at 0.05 (Eng, 2003). The participants were recruited from mainstream kindergartens and primary schools in the city of Oslo and the county of Telemark. Care was taken to include participants from both urban and rural areas, both inner city, suburbs and smaller towns, as well as from economically affluent and less affluent areas. A total of 642 invitations were distributed; parents of 256 (39.9%) children – 158 from kindergartens and 98 from schools – gave written consent to participate. In Norway, children typically attend

kindergarten from one year of age and until they start school at 6 years of age. On the consent form, the parents confirmed that their child fulfilled the inclusion criteria.

Six participants were absent on the day of assessment, five consents arrived after data collection was completed, and seven assessments were terminated because the children showed reluctance to being tested, for a final sample of 238 participants: 125 males (53%) and 113 females (47%). The age of the participating children ranged from 14 to 94 months ($M = 4;7$, $SD = 24.4$) Among participants, 40% attended school and 21% came from rural areas.

The parents of the eligible children were given written information about the aim and the procedures of the study and gave written consent to their child's participation. They were informed that consent could be withdrawn at any time without stating any reason. The participating children were given age-appropriate information before assessment. The parents were informed that they would only receive written reports of their child's results if the test results were more than one standard deviation (SD) below the chronological age mean, and invited to provide an address for this purpose. Parents of 11 of the 14 children with scores in the low range had provided an address and consequently received a written report alongside an invitation to consult with the principal investigator (KS) regarding possible referral to the Educational Psychological Services. No parents sought such a consultation. The study was approved by the Norwegian National Data Security agency (#2017/55604; #2019/827474).

Setting

The assessments took place in a quiet room in the child's school or kindergarten. A familiar adult from the kindergarten could accompany the kindergarten participants during the assessment to ensure that they felt secure.

Research design

The internal consistency (research question a) and validity (research questions c, d, and e) of the C-BiLLT-Nor were evaluated using a cross-sectional design. To investigate test-retest and inter-rater-agreement (research question b), and the effect of response modality (research question f), 53 of the participants who were of school-age were retested with C-BiLLT-Nor 2–7 weeks after the first assessment. The reason for retesting only the older children was due to time constraints that resulted from enrolling more participants than planned. To investigate inter-rater-agreement, 28 participants were retested by a different examiner than the one who conducted the first assessment. To examine the effect of response modality, 17 of the 53 retested participants answered with finger-pointing the first time and gaze pointing the second time and 20 participants responded with gaze pointing the first time and finger-pointing the second time.

Materials

The measures to investigate the convergent and discriminant validity of the C-BiLLT-Nor were based on the original Dutch study (Geytenbeek et al., 2014) and their availability in Norwegian. Measures included those used in the original study (Geytenbeek et al., 2014) and those with Norwegian standardization. In Norway, few measures are standardized, thus limiting the available selection for the age groups in question. Comprehension of words and sentences was assessed with the comprehension part of the Reynell Developmental Language Scale (Hagtvet & Lillestølen, 1990), the subtest Word Recognition from the Wechsler Preschool and Primary Scale of Intelligence-IV (WPPSI-IV; Wechsler, 2015) and the Test for Reception of Grammar-2 (TROG-2; Bishop, 2009). Nonverbal reasoning was measured with the Raven's Colored Progressive Matrices (Raven, 2008) and Block Design from WPPSI-IV. The tests were administered as described in the respective administration manuals. Test results are reported as z-scores, where the age average score is zero and the standard deviation is one.

The C-BiLLT is comprised of a pretest with objects and photos, and a computerized part. The maximum score of the pretest is two. The computerized part consists of two segments. Segment 1 has three sections and Segment 2 has nine sections (see Table 1). The 12 sections comprise 86

Table 1. Sections of the Computerized Part of the C-BiLLT, including Description of Content and Examples (Geytenbeek et al., 2014).

Section	No. of items	Item types identified	Item examples
Segment 1			
1	10	Objects	Where is the car?
2	10	Action words	Who sleeps?
3	10	Animal, persons, and objects	Where is the baby?
			Where is the vacuum cleaner?
Segment 2			
4	4	Rarer objects	Where is the umbrella?
5	5	Simple sentences about objects, combined with verbs and prepositions	What can you sleep in?
			The dog lays in the basket
6	5	"Who" questions about persons performing actions	Who shall walk the dog?
7	4	Passive-sentence structure	The boy is pushed by the girl
8	9	More complex sentences	One of the tooth brushes placed in the cup, is long
9	6	Simpler active sentences referring to non-observable situations involving four people (Jacob, Emma, baby, and mother)	Jacob and Emma shall play outside, who stays with mommy?
10	9	Complex sentences with two or more concepts	One small glass of jam stands next to the red glass of jam
11	4	Compound, complex sentences	First, there was an apple on the table, but now there is a banana instead
12	10	Compound, complex sentences	All the spreads are placed on the plates, but it is only on plates with the same color that there is cheese

tasks, of which 34 are questions about objects and actions (vocabulary comprehension). In the remaining 52 tasks, a sentence is read aloud and the participant is asked to identify the photo among four photos that corresponds to that sentence. For the identification of objects and actions, the first 30 tasks involve a choice between two photos and the remaining between four photos. To minimize the impact of guessing, Segment 1 has parallel tasks that should be administered if not all tasks are answered correctly the first time. In the parallel version, the same items are presented, but in a different order and coupled with different photos. For example, in the first test run *doll* is the fourth item with a photo of a doll on the left side of the screen and a TV on the right side; in the parallel version, it is the seventh item, shown together with a photo of a pot on the left side. Participants only receive scores for the items answered correctly in both the original and parallel versions. The maximum score on the C-BiLLT is 88.

The International Test Commission (2017) recommends that adaptation of a test from one language and culture to another be carried out by a team with knowledge of the languages involved, the cultures, the content of the test, and general principles of testing. Per these guidelines, the adaptation process of the C-BiLLT involved both multiple translations from Dutch to Norwegian, carried out independently by professionals with knowledge of both languages and of assessment of language comprehension in children, and a backward translation from Norwegian to Dutch by a certified translator. Furthermore, a consensus discussion to review the translations was held, involving professionals who were native speakers of both Dutch and Norwegian, and also attended by the original developer. There was complete agreement on the pretest items and large agreement between the two initial translations from Dutch to Norwegian, $r(85) = 0.87$, $p < .001$, on the 86 items of the computer task. After the backward translation, 53 items were kept unchanged. For the remaining 33 items, the wording was discussed in the consensus meeting until a complete agreement was reached. The instructions for the tasks underwent the same process. Consideration was taken to ensure similarity to the Dutch version both in the commonality of vocabulary chosen and difficulty of sentence structure.

Several adaptations had to be made in order for the Dutch C-BiLLT to be appropriate for use in Norway. While Dutch and Norwegian are both Germanic languages, there are differences with regard to syntax. For instance, in Norwegian, inflections of nouns are marked by a suffix, not by an article. The Dutch sentence *Waar is de auto?* (Where is the car?) translates as *Hvor er bilen?*, in which the suffix -en indicates definiteness and masculinity of the root of the word "bil" (car). Cultural appropriateness of the stimulus photos was considered, and it was necessary to change two of the original photos. Because children in Norway are not familiar with the Dutch chocolate sprinkles *hagelslag* this image was replaced by a *brunost* (brown cheese). In addition, a Norwegian mailbox replaced the photo of a Dutch mailbox.

Procedures

The participants were randomly assigned to the two different test orders: A (50.8%) and B (49.2%). In both test orders, the C-BiLLT-Nor was administered first to avoid any learning effect or fatigue influencing the test results. Participants assigned to Order A were then administered Reynell, Raven, Word Recognition of the WPPSI-IV, Block Design of the WPPSI-IV, and TROG-2. Participants assigned to Order B were administered the tests in the order Block Design of the WPPSI-IV, TROG-2, Reynell, Word Recognition of the WPPSI-IV, and Raven (see Table 2). In both test orders verbal and visual tasks as well as tasks requiring pointing and manipulation of objects were alternated. This was done to keep the interest of the participants. Independent samples *t*-tests showed no significant difference in the sum of raw scores on the C-BiLLT-Nor or *z*-scores on the other tests attributable to test order ($p > .05$).

Because the C-BiLLT-Nor and the Reynell were the only tests that covered the entire age span, the number of tests administered, and consequently, the duration of the assessment, varied with the child's age. For children between 1;6 and 2;6 the testing lasted approximately 30 min; for children between 2;7 and 4;0 it ranged from 40 min to one hr. For children older than 4;1 the testing lasted up to 2 hr. The oldest children were offered a break halfway through the assessment.

It was decided that Segment 1 of the C-BiLLT-Nor would only be administered to participants aged 1;6 to 4;6 because

Table 2. Sequence of tests administered to the three age groups and test orders (A and B) for the two oldest groups.

	Age groups in years/months				
	1;5–2;5	2;5–4;5	4;5–7;0	2;5–4;5	4;5–7
Sequence		Test order A		Test order B	
1st	C-BiLLT-Nor	C-BiLLT-Nor	C-BiLLT-Nor	C-BiLLT-Nor	C-BiLLT-Nor
2nd	Reynell	Reynell	Reynell	Block design	Block design
3rd			Raven		TROG-2
Short break					
4th		Word recognition	Word recognition	Reynell	Reynell
5th		Block design	Block design	Word recognition	Word recognition
6th			TROG-2		Raven

Note: Raven: Raven's Colored Progressive Matrices; TROG-2: Test for Reception of Grammar (2nd ed.); C-BiLLT-Nor: Computer-Based instrument for Low motor Language Testing, Norwegian version; Reynell: Reynell Developmental Language Scale (Comprehension part); Word Recognition and Block Design are tasks from the Wechsler Preschool and Primary Scale of Intelligence (4th ed.).

the first 30 participants had answered every item of Segment 1 correctly, thus, the duration of the assessment was shortened considerably for the oldest children. Participants who were not administered the pretest and Segment 1 automatically received the maximum score for these parts. To ensure that the performance of the older participants was not overestimated, a reversal rule was established: if children older than 4;6 did not answer the first four vocabulary items of Segment 2 (questions 31 to 34) correctly, Segment 1 would be administered. None of the participants older than 4;6 failed to answer questions 31–34 correctly. The Dutch discontinue rule of eight consecutive errors was not used in this study because it was not known if the order of the item difficulty would be the same in the Norwegian version. Any commenced section was therefore completed for all participants even if the number of consecutive errors exceeded eight. When a participant had more than eight errors in a row and had completed a section, testing with the C-BiLLT-Nor was discontinued.

The participants using eye-gaze did so on computers with eye-gaze technology. Prior to answering using eye gaze, they practiced on tasks included in the C-BiLLT for this purpose (the learning module; see Geytenbeek et al., 2014). In these tasks, an animation (like a ball bouncing) appeared in the part of the screen where the child looked. This was accompanied by an explanation such as *See what happens to the ball when you look at it.*

Data analysis

Statistical analyses were performed using SPSS version 25.0 (IBM Corp., 2017). The significance level was set to 0.05. The distribution of raw scores on the C-BiLLT-Nor in the age groups did not violate assumptions of normality, with Shapiro–Wilk tests $p > .05$, except for the age groups 2;1–2;6 and 4;1–4;6 (see Table 3). Because the data were normally distributed in the 12 other half-year groups, parametric tests were chosen.

Cronbach's alpha was used as a measure of internal consistency. An $\alpha > 0.90$ is regarded as satisfactory for instruments with clinical applications (Bland & Altman, 1997). Because the C-BiLLT-Nor consists of subscales, α was calculated separately for all items together, for the 12 sections of the C-BiLLT-Nor, and separately for test items measuring vocabulary and sentence comprehension. Intraclass correlation coefficients (ICC), using two-way random effects and consistency model, were computed, as these give a measure of both degrees of correlation and agreement between measurements. An ICC between 0.50 and 0.75, 0.75 and 0.90, and above 0.90, respectively, is indicative of good, moderate, and excellent reliability (Koo & Li, 2016).

Pearson's product-moment correlation coefficient was used to investigate convergent and discriminant validity. An exploratory factor analysis was conducted on the raw scores to investigate the structural validity of the C-BiLLT-Nor. As each test item is a dichotomous variable and quite a few test items were answered correctly by every participant, they were merged into larger composite units and the factor analysis was conducted on the 12 sections rather than on the 86 test items. The suitability of the data for factor analysis was investigated with the Kaiser-Meyer-Olkin test and Bartlett's test of sphericity, both of which yielded satisfying results, KMO = 0.93 and $\chi^2(66) = 3677.54$, $p < .001$. The Principle Axis Factoring was used, as it is robust to skew and kurtosis values exceeding ±2. The factors were extracted using the Kaisers criteria of an eigenvalue >1 in combination with a visual inspection of the scree plot. An Oblique rotation (Promax) was used as the method of rotation because this method is open for correlations between factors, which is theoretically expected in a study of mental abilities. Factor loadings >0.50 were considered high (Fabrigar & Wegener, 2012). Factor loadings are considered to give a meaningfull contribution to the overall factor solution only if they exceed .30, therefore factor loadings <.30 are not reported. The difference between a section's loading on several factors was above 0.20.

Differences between age groups on the C-BiLLT-Nor and effect of response mode were investigated using a one-way ANOVA with a Games–Howell post hoc test because equal variances were not assumed. Partial eta squared (η^2) was computed to assess effect size, where values at or below 0.01 are considered small, values between 0.01 and 0.06 are considered medium, and values above 0.14 are deemed to be large (Richardson, 2011).

Results

In all, 26 (10.9%) of the participants were reported to be bilingual, with Norwegian as one of their native languages. They did not differ from the monolingual children with regard to language comprehension, as measured by the Reynell (stanine score M (SD) = 6.1 (1.9) versus 5.6 (2.1), $t(234) = 1.113$, $p = .267$). Table 3 presents the descriptive distribution of raw scores on the C-BiLLT-Nor, while Table 4 presents the z-scores for the other measures of language and nonverbal reasoning used in this study. Mean z-scores varied between 0.2–0.6, with a standard deviation from 0.8 to 1.1.

Table 3. Minimum, maximum, mean (*M*), standard deviation (*SD*), range of scores within one SD from age mean, skewness, and kurtosis of raw scores on the C-BiLLT-Nor.

Age group[a]	n	Min.	Max.	M	SD	95% CI for mean	Skewedness	Kurtosis	p*
1;1–1;6	9	2	22	8.3	6.2	3.5–13.1	1.34	2.24	.163
1;7–2;0	17	3	45	23.8	13.9	16.7–31.0	−0.07	−1.15	.197
2;1–2;6	27	5	48	38.3	9.8	34.5–42.2	−2.61	7.07	.000*
2;7–3;0	27	28	64	47.7	8.1	44.5–50.9	−0.37	0.30	.732
3;1–3;6	12	44	66	54.8	6.0	51.0–58.7	0.17	0.34	.934
3;7–4;0	14	43	70	58.3	6.0	54.8–61.8	−0.83	3.19	.114
4;1–4;6	11	61	71	64.5	4.1	61.7–67.2	0.79	−1.07	.008*
4;7–5;0	6	49	72	63.5	7.8	55.3–71.7	−1.49	3.01	.217
5;1–5;6	17	62	74	69.1	3.6	67.2–70.9	−0.50	−0.75	.305
5;7–6;0	5	68	76	71.4	3.2	67.4–75.4	0.61	−0.68	.794
6;1–6;6	32	68	81	75.1	3.5	73.8–76.3	−0.35	−0.62	.278
6;7–7;0	28	67	82	75.0	3.7	73.6–76.5	−0.36	−0.22	.484
7;1–7;6	28	70	84	77.6	2.9	76.4–78.7	−0.56	0.77	.194
7;7–8;0	5	75	81	78.4	2.4	75.4–81.4	−0.61	−0.95	.787

[a]Years/months.
*Shapiro–Wilk test $p > .05$.

Table 4. Mean (*M*), standard deviation (*SD*), skewness and Kurtosis for *z*-scores on tests of verbal comprehension and non-verbal reasoning.

Test	*n*	*M*	*SD*	Skewness	Kurtosis
Reynell	236	0.50	0.94	−0.32	−0.43
TROG-2	122	0.59	1.07	− 0.61	0.49
Word recognition	176	0.18	0.98	0.25	0.45
Block design	176	0.34	0.84	0.74	1.21
Raven	122	0.25	0.94	0.39	−0.50

Note: Reynell: Comprehension part of the Reynell Developmental Language Scale; TROG-2: Test for Reception of Grammar (2nd ed.); Word Recognition and Block Design: Tasks from the Wechsler Preschool and Primary Scale of Intelligence (4th ed.); Raven: Raven's Colored Progressive Matrices.

Table 5. Pearson correlations (*n*) of raw scores from the tests C-BiLLT-Nor, Reynell, TROG-2, Word Recognition, Block Design, and Raven.

	C-BiLLT-Nor	Reynell	TROG-2	Word recognition	Block design	Raven
C-BiLLT-Nor		0.96** (236)	0.71* (122)	0.84* (176)	0.84* (176)	0.65* (122)
Reynell			0.69* (122)	0.86* (176)	0.79* (176)	0.60* (122)
TROG-2				0.57* (122)	0.53* (122)	0.62* (122)
Word Recognition					0.77* (176)	0.50* (122)
Block Design						0.68* (122)
Raven						

Note. C-BiLLT-Nor: Computer-Based instrument for Low motor Language Testing, Norwegian version; Reynell: Comprehension part of the Reynell Developmental Language Scale; TROG-2: Test for Reception of Grammar (2nd ed.); Word Recognition and Block Design are tasks from the Wechsler Preschool and Primary Scale of Intelligence (4th ed.) Raven: Raven's Colored Progressive Matrices.
*$p < .001$.

Reliability

Cronbach's alpha was calculated for the 86 test items of the C-BiLLT-Nor, $\alpha = 0.98$, for the 12 sections, $\alpha = 0.94$, for the items measuring vocabulary, $\alpha = 0.93$, and for sentence comprehension, $\alpha = 0.95$. The test-retest ICC was 0.88 and the inter-rater ICC was 0.71. There was more variability in test-retest scores among participants younger than 7 years compared to those older than 7 years (*M(SD)* = 2.5 (3.1) versus *M(SD)* = 1.8 (1.1)), but the difference was not significant. There was no difference in the mean increase of scores between first assessment and retest that could be attributed to the order of test administrator, $F(2, 50) = 0.7$, $p = .517$.

Validity

The correlation was highest between the C-BiLLT-Nor score and the Reynell score. The correlation was lower, but still high, between C-BiLLT-Nor and TROG-2, Word Recognition of the WPPSI-IV and Block Design of the WPPSI-IV. The correlation between C-BiLLT-Nor and Raven was moderate. All correlations were significant (see Table 5). The 95% confidence interval of the correlations between C-BiLLT-Nor and the language tests Reynell (0.95–0.97) and Word Recognition of the WPPSI-IV (0.79–0.88) and between C-BiLLT-Nor and the non-verbal test Raven (0.53–0.74) did not overlap. The correlation between C-BiLLT-Nor and Block Design for participants older than 4;1 was moderate, $r(121) = 0.64$, $p < .001$.

Table 6. Differences in mean (*M*) raw scores on the C-BiLLT-Nor between whole-year groups.

			95% confidence interval	
Age group[a]	*M* difference	*p*	Lower limit	Upper limit
1;1–2;0 and 2;1–3;0	24.5	<.001**	15.0	34.0
2;1–3;0 and 3;1–4;0	13.7	<.001**	8.2	19.3
3;1–4;0 and 4;1–5;0	7.4	.003**	1.8	13.0
4;1–5;0 and 5;1–6;0	5.5	.020**	0.6	10.3
5;1–6;0 and 6;1–7;0	5.5	<.001**	2.7	8.2
6;1–7;0 and 7;1–8;0	2.7	.004**	0.6	4.7

[a]Years/months.

The one-way ANOVA on the differences in the mean sum of C-BiLLT-Nor raw scores for different age groups revealed that there were significant differences between different half-year-groups, $F(13, 223) = 153.5$, $p < .001$, partial $\eta^2 = 0.90$, and between whole-year-groups, $F(6, 230) = 254.8$, $p < .001$, partial $\eta^2 = 0.87$ (see Table 6). Figure 1 gives a visual presentation of data from this study of the increase in mean raw score of the C-BiLLT-Nor and Reynell with increasing age, supplemented with data from the Dutch investigation of C-BiLLT for comparison (Geytenbeek et al., 2014).

The exploratory factor analysis resulted in two extracted factors with an eigenvalue larger than Kaiser's criteria, in accordance with a visual inspection of the scree plot. The first factor had a high initial eigenvalue of 8.23 that explained 68.6% of the variance in the data. The second factor had an initial eigenvalue of 1.99 that explained 16.6% of the additional variance in the data. As shown in Table 7, section 7–12 loads on the first factor, sections 1–4 on the second factor, and sections 5 and 6 on both factors. The correlation between factor 1 and factor 2 was moderate, $r = 0.61$.

Response mode

One-way ANOVA revealed no difference in an average increase in raw scores on the C-BiLLT-Nor from the first assessment to retest that could be traced to using different response modes, $F(2, 50) = 0.4$, $p = .694$, partial $\eta^2 = 0.01$ (see Table 8). The ICC between test and retest scores using different response modes was 0.70.

Discussion

The primary aim of the present study was to investigate the reliability and validity of the C-BiLLT-Nor, in order to ensure that content and difficulty level were not altered as a result of the cultural-linguistic adaptation into Norwegian (Hambleton, 2005; Krach et al., 2017). The secondary aim was to explore the effect of answering using two response modes, finger-pointing versus gaze-pointing.

Reliability

The results indicate that the C-BiLLT-Nor has excellent internal consistency, good test-retest reliability, and moderate inter-rater reliability. The reliability of the C-BiLLT-Nor was therefore as expected, with the exception that the ICC for the inter-rater agreement was 0.71, which is acceptable
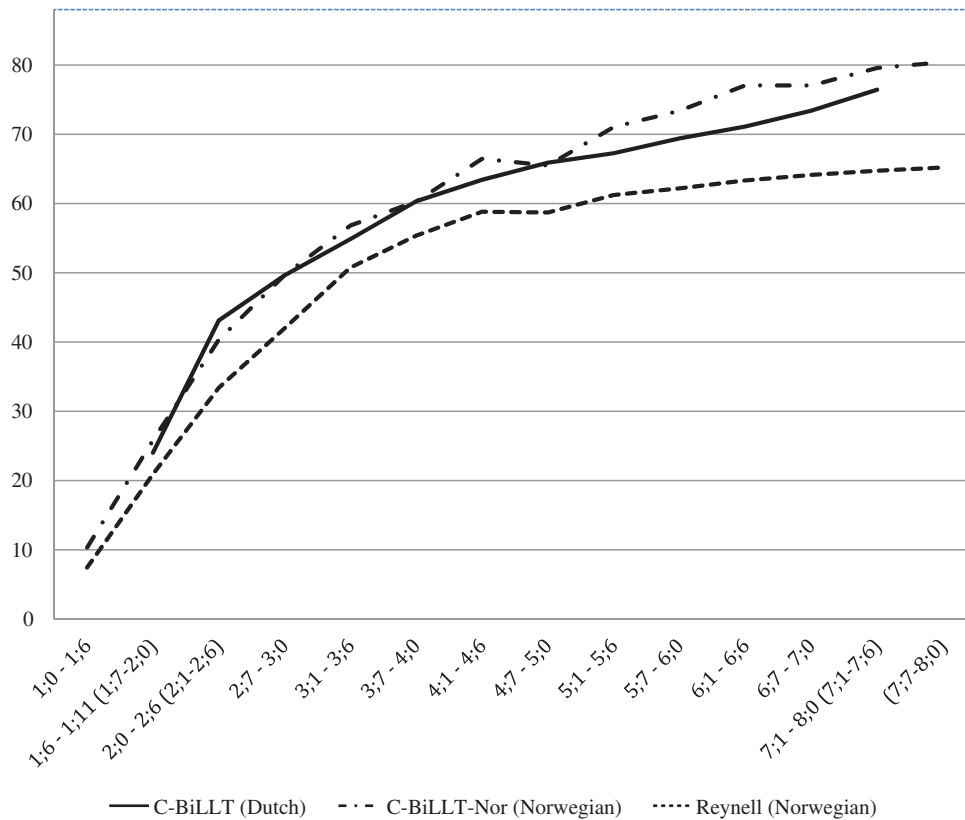
Figure 1. Comparison of raw scores (*y*-axis) in different age groups (*x*-axis) across tests. These include the Dutch version of C-BILLT (maximum score of 88, from Geytenbeek et al., 2014), the Norwegian C-BiLLT-Nor (maximum score of 88) and language comprehension raw scores on the Norwegian version of Reynell Developmental Language Scales (maximum score of 67).

Table 7. Pattern matrix from the exploratory factor analysis. C-BiLLT-Nor sections *n* of test items Factors Communalities (extraction) Factor 1 Factor 2 (*continued*) 1: Nouns 10 0.94 0.78 2: Verbs 10 1.00 0.91 3: Animals, objects and persons 10 1.00 0.90 4: Rarer objects 4 0.90 0.88 5: Simple sentences about objects in combination with verbs and prepositions 5 0.46 0.54 0.81 6: "Who" questions about persons performing activities 5 0.55 0.44 0.79 7: Passive and active sentences with objects and prepositions combined with events 4 0.71 0.69 8: Sentences with more complexity in grammar and semantics 9 0.85 0.90 9: Simple active sentences referring to non-observable situations of four persons 6 0.93 0.81 10: Complex sentences with two or more concepts 9 1.00 0.91 11: Compound, complex sentences 4 0.95 0.80 12: Compound, complex sentences – continuation 10 0.93 0.71 Eigen values 8.23 1.99 % Variance explained 68.6 16.6 *Note*: Factor loadings <0.30 are not reported.

Table 8. Mean (*M*) raw scores on C-BiLLT-Nor for participants using finger pointing and gaze pointing at time of first assessment (test) and second assessment (retest).

Response mode at test and retest	*n*	*M* raw score at test assessment	*M* raw score at retest	Difference in mean raw scores
FP–FP	16	74.8	76.6	1.8
FP–GP	17	77.1	79.4	2.3
GP–FP	20	76.2	78.8	2.6

Note: FP: finger pointing; GP: gaze pointing.

but somewhat lower than the expected >0.75. One reason for this might have been that the 19 participants assessed first by author one and then by author two were somewhat older than the nine assessed in the opposite rater order, (*M* (*SD*) = 79.2 (2.4) versus 75.0 (2.3) months). As there is a tendency toward more variability in scores from test to retest among the younger children, this might have confounded the results.

Validity

The high correlations between the raw scores on the C-BiLLT-Nor and the raw scores on the other tests of language comprehension indicate that the test can be regarded as a measure of language comprehension, and as evidence of convergent validity. These findings are also in line with studies showing that different language abilities, such as the comprehension of vocabulary and grammar, are closely related abilities in children (Colledge et al., 2002; Justice et al., 2018; Tomblin & Zhang, 2006). The C-BiLLT-Nor correlated the highest with the Reynell, which was expected because the C-BiLLT is based on the Reynell in its sequencing of linguistic complexity (Geytenbeek et al., 2014), whereas the TROG-2 correlates only moderately (0.57–0.71) with the other tests of language comprehension. Multidimensionality in children's language ability is an emerging rather than an innate quality of the language system, something that is reflected in a lower correlation between measures of vocabulary and grammar as children grow up (Justice et al., 2018; Tomblin & Zhang, 2006). Because the C-BiLLT-Nor, the Reynell, and the Word Recognition subtest of the WPPSI-IV are normed for younger age groups than the TROG-2, their commonality may reflect more of the unidimensionality of language in the youngest children.

The correlation between the C-BiLLT-Nor and Raven was lower than the correlation between the C-BiLLT-Nor and the tests of language comprehension, supporting the test's discriminant validity. Some correlations between scores on language tests and tests of non-verbal reasoning are to be expected as cognitive abilities are not unrelated modular abilities, and indeed the high correlation between the C-BiLLT-Nor and the Block Design subtest of the WPPSI-IV could be due to cognitive abilities being less differentiated in younger children (Karmiloff-Smith, 2009; Tideman & Gustafsson, 2004). In accordance with this explanation, the correlation between the C-BiLLT-Nor and the Block Design subtest of the WPPSI-IV decreases from high to moderate when children below 4;1 were removed from the analysis.

The results from the analysis of variance support the C-BiLLT-Nor as a test capable of differentiating between the receptive language levels of different age groups; the group means were significantly different between different age groups and the effect size was large. Variation in language abilities of children with typical development is natural and the existence of variability in scores within age groups is therefore expected. A narrow distribution of scores or a mean score of an age group close to a test's minimum or maximum score could be an indication of floor and ceiling effects and that the test is ill-suited to differentiate between linguistically low-functioning and high-functioning children. There was a steady increase, and no overlap, in mean scores over the half-year age groups up to 4-years-old and the whole-year age groups above 4-years-old. In addition, the scores within one standard deviation of the age mean on the C-BiLLT-Nor did not overlap with the age mean of the year groups below or above (results not shown). This supports the use of the test as an instrument to identify children

with a language delay. In addition, no participant obtained the maximum score of 88, supporting the absence of ceiling effects. Thus, the C-BiLLT-Nor may also be used for older groups of children.

In the light of research on the dimensionality of children's language ability, it was unexpected that the factor analysis undertaken to investigate the structural validity of the C-BiLLT-Nor resulted in two extracted factors and not in one. The two factors can be interpreted as measures of receptive vocabulary (Factor 2; Section 1 to 4) and receptive grammar (Factor 1; Section 7 to 12); however, the findings that the factors also correlated significantly and that two sections (Sections 5 and 6) loaded on both factors, indicate the presence of a higher-order factor (Fabrigar & Wegener, 2012). It, therefore, seems that the C-BiLLT-Nor can be construed both as a measure of unitary language abilities (receptive vocabulary and receptive grammar, respectively), as well as a measure of a unidimensional language ability reflected in the presence of a higher-order factor. Interestingly, in the original Dutch study, only one factor was identified. The difference is probably due to the proportionally larger number of older children included in the current study and that in the Dutch study the 12 most complicated items had not yet been added at the time when the factor analysis was performed (Geytenbeek et al., 2014). Children's language ability becomes more differentiated with age, reflected in a lower correlation between measures on grammar and vocabulary (Hayiou-Thomas et al., 2006; Justice et al., 2018).

Response modes

In order to assess cognition in children with severe motor impairments, it is recommended to use instruments that require minimal motor involvement, such as tests with a multiple-choice format where the answer is given just by pointing (Yin Foo et al., 2013); however, some children do not have the necessary fine motor skills to point and need alternative ways of responding, such as gaze pointing. In the current study, there were no differences between the scores on the C-BiLLT-Nor that could be attributed to the different response modes. This indicates that the answering mode (i.e., pointing with an eye gaze or with a finger) does not affect the child's score in any way. This finding has value beyond the administration of the C-BiLLT-Nor. As the C-BiLLT-Nor has a multiple-choice format with a fixed number of answer alternatives, the results of the investigation of response modality can be generalized to other tests of language comprehension and cognition that have a similar design. This non-significant effect of response mode on the test result is in line with previous research (Kurmanaviciute & Stadskleiv, 2017; Spillane et al., 1996). It underlines the importance of offering children with severely restricted motor functioning alternative ways of responding in order to assess their capabilities, as these might not be otherwise easily recognized due to the severity of motor and speech impairments (Stadskleiv et al., 2018).

Clinical implications

Children in need of AAC and with severe motor impairments comprise a heterogeneous group with wide variability in language comprehension and cognitive functioning. They have an obvious need for interventions, most fundamentally to provide a means of expression that matches their level of comprehension. Without a suitable communication aid, there will inevitably be unnecessary restrictions on not only communication but also possibilities of play, academic learning, and participation within the family, with peers, and in the community. There are numerous types of communication aids; finding the one best suited for the particular child requires knowledge of that child's strengths and challenges. Having a reliable method of assessing a child, such as the C-BiLLT-Nor, is an important first step toward achieving this goal.

Another important clinical implication is the finding that response mode did not influence test results. This is clinically relevant, as many children in need of AAC are considered "non-assessable" using standardized tests, purely due to the fact that they cannot move their arms or hands voluntarily. They are often not assessed with these tests, and cognitive functioning and language comprehension is simply assumed (e.g., Andersen et al., 2008). That answering tests with a multiple-choice format, such as the C-BiLLT-Nor, can be done reliably with eye-gaze pointing as well as with finger-pointing, supports the use of different response modalities to answer other questions in other types of tests, such as in the assessment of nonverbal reasoning in children with severe speech and motor impairments.

Compared to other tests of language comprehension, the C-BiLLT/C-BiLLT-Nor has some noteworty benefits. It offers comprehension of both vocabulary and sentences – not just one or the other. It is also designed specifically for children with disabilities by, for example, including photos that reflect their experiences (e.g., children are shown in wheelchairs). Additionally, the test has norms for children as young as 1.5-years-old, thereby making it suitable also for older children and adults with severe intellectual impairments. In addition, it uses transparent photos, which are particularly suited for children with cerebral visual impairment. Furthermore, different access methods are an integral part of the test, thereby avoiding potential copyright issues related to digitalizing other tests for computerized eye-gaze responding (Geytenbeek et al., 2010, 2014).

Limitations and future directions

The study presents a few limitations that can be tackled in future research. One limitation is that response modality and test-retest reliability were examined only in the oldest participants. Therefore, the results may not reflect the sample as a whole, especially infants and toddlers; however, somewhat more variability in performance is accounted for because the standard deviations are larger in the younger compared to the older age groups. A further limitation is the moderate inter-rater agreement, likely caused by a combination of the short time between test and retest and differing age groups assessed by the different raters. The finding that the mean increase in test scores did not differ according to the order of test administrator minimizes this concern.

Although far more participants than originally aimed for (238 versus 96) were recruited, their ages were not evenly distributed. There were three half-year groups (4;7–5;0, 5;7–6;0 7;7–8;0), with less than the desired eight participants; however, for the children between 4- and 6-years-old, all whole-year groups included more than 16 participants. It is therefore recommended that, for children older than 4-years-old, the mean and standard deviations for the whole-year groups be applied.

Although data on socio-economic status and parental education/employment were not collected, it should be noted that, in general, Norway is regarded as a highly homogeneous society. For example, over 92% of all children attend kindergarten (data from Statistics Norway at https://www.ssb.no/barnehager), and all kindergartens follow a similar pedagogy. Similarly, all schools follow the same curriculum and have similar teacher norms (https://www.udir.no/). Therefore, there is not a significant difference in school quality between affluent and less affluent areas.

The C-BiLLT-Nor can be used to differentiate between children functioning below, at, or above what is expected for their age. A larger norm group is desirable before differentiating between more subtle variations in language ability. In the near future, it will also be clinically important to present results of the assessment of Norwegian children with severe motor disabilities, comparing the results they obtain on the C-BiLLT-Nor with other test results and follow the aided development of the group. This can be used to determine whether the test is also a valid measure of language comprehension for children with severe speech and motor impairments and to be able to discuss the test's ecological validity.

Conclusions

This study supports the use of C-BiLLT-Nor as a reliable and valid measure of language comprehension in children. It has norms for children between 1;1 and 7;6. For children between 1;1 and 3;11, half-year norms can be used. For children above 4 years of age, applying full-year norms is recommended.

In order to compute national norms, it is important to know how children in a population are typically performing. It is therefore a strength that the sample is representative of Norwegian children with typical development, as reflected in mean standardized scores on tests of both verbal comprehension and non-verbal reasoning being within one *SD* of the age mean. A strength of this study is that the C-BiLLT-Nor is validated against several standardized tests of language comprehension and nonverbal reasoning which makes it possible to reach a nuanced understanding of what the test measures. The Norwegian version of the test shows high internal consistency and test-retest reliability, moderate inter-rater reliability, and high convergent and discriminant validity. The findings suggest that the standardized administration

procedures of the C-BiLLT-Nor are easy to use, that the scoring is consistent, that the test translated well from Dutch to Norwegian, and that the adaptions that were made to the stimulus material were appropriate.

This study reports on the translation and adaptation of the receptive language test C-BiLLT from Dutch to Norwegian. However, the implications extend to procedures for translating from one language to another, of investigating the appropriateness of the stimulus material, and of investigating the reliability and validity of the test. Moreover, the implications are relevant for professionals who are not only adapting tests to new linguistic and cultural settings but also working with language minorities and children with multilinguistic backgrounds. Furthermore, the finding that response modality did not influence test results has implications for professionals working with children and adults with severe speech and motor impairments who may not be able to use finger-pointing. It shows how tests with a multiple-choice format can be adapted, and how interventions can be based upon assessment results instead of upon assumptions about functioning.

ORCID

Kristine Stadskleiv http://orcid.org/0000-0002-5478-5689

References

Andersen, G., Irgens, L. M., Haagaas, I., Skranes, J. S., Meberg, A. E., & Vik, T. (2008). Cerebral palsy in Norway: Prevalence, subtypes and severity. European Journal of Paediatric Neurology, 12(1), 4–13. doi:10.1016/j.ejpn.2007.05.001

Andersen, G. L., Mjøen, T., & Vik, T. (2010). Prevalence of speech problems and the use of augmentative and alternative communication in children with cerebral palsy: A registry-based study in Norway. Perspectives on Augmentative and Alternative Communication, 19(1), 12–20. doi:10.1044/aac19.1.12

Anthony, J. L., Davis, C., Williams, J. M., & Anthony, T. I. (2014). Preschoolers' oral language abilities: A multilevel examination of dimensionality. Learning and Individual Differences, 35, 56–61. doi:10.1016/j.lindif.2014.07.004

Batorowicz, B., Stadskleiv, K., von Tetzchner, S., & Missiuna, C. (2016). Children who use communication aids instructing peer and adult partners during play-based activity. Augmentative and Alternative Communication, 32(2), 105–119. doi:10.3109/07434618.2016.1160150

Barker, M. R., Romski, M. A., Sevcik, R. A., Adamson, L. B., Smith, A. L., & Bakeman, R. (2019). Intervention focus moderates the association between initial receptive language and language outcomes for toddlers with developmental delay. Augmentative and Alternative Communication, 35, 263–273. doi:10.1080/07434618.2019.1686770

Bishop, D. V. M. (2009). Test for reception of grammar (2nd ed., Norwegian version). London, UK: Pearson.

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. BMJ, 314(7080), 572. doi:10.1136/bmj.314.7080.572

Choi, J. Y., Choi, Y. S., & Park, E. S. (2017). Language development and brain magnetic resonance imaging characteristics in preschool children with cerebral palsy. Journal of Speech Language and Hearing Research, 60(5), 1330–1338. doi:10.1044/2016_JSLHR-L-16-0281

Chomsky, N. (1959). On certain formal properties of grammars. Information and Control, 2(2), 137–167. doi:10.1016/S0019-9958(59)90362-6

Colledge, E., Bishop, D. V. M., Koeppen-Schomerus, G., Price, T. S., Happé, F. G. E., Eley, T. C., … Plomin, R. (2002). The structure of language abilities at 4 years: A twin study. Developmental Psychology, 38(5), 749–757. doi:10.1037//0012-1649.38.5.749

Dunn, L. M., & Dunn, L. M. (1997). PPVT-III: Peabody picture vocabulary test. Bloomington, MN: American Guidance Service.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). The British Picture Vocabulary Scale (2nd ed.). Berkshire, UK: NFER Nelson.

Edwards, S. (1997). Reynell Developmental Language Scales III. Berkshire, UK: NFER Nelson.

Eng, J. (2003). Sample size estimation: How many individuals should be studied? Radiology, 227(2), 309–313. doi:10.1148/radiol.2272012051

Fabrigar, L. R., & Wegener, D. T. (2012). Exploratory factor analysis. New York, US: Oxford University Press.

Geytenbeek, J. J., Harlaar, L., Stam, M., Ket, H., Becher, J. G., Oostrom, K., & Vermeulen, R. J. (2010). Utility of language comprehension tests for unintelligible or non-speaking children with cerebral palsy: A systematic review. Developmental Medicine and Child Neurology, 52, 267–277. doi:10.1111/j.1469-8749.2010.03807.x

Geytenbeek, J. J., Heim, M. J. M., Knol, D. J., Vermeulen, R. J., & Oostrom, K. J. (2015). Spoken language comprehension of phrases, simple and compound-active sentences in non-speaking children with severe cerebral palsy. International Journal of Language & Communication Disorders, 50(4), 499–515. doi:10.1111/1460-6984.12151

Geytenbeek, J. J., Mokkink, L. B., Knol, D. L., Vermeulen, R. J., & Oostrom, K. J. (2014). Reliability and validity of the C-BiLLT: A new instrument to assess comprehension of spoken language in young children with cerebral palsy and complex communication needs. Augmentative and Alternative Communication, 30(3), 252–215. doi:10.3109/07434618.2010.482445

Geytenbeek, J. J., Oostrom, K. J., Harlaar, L., Becher, J. G., Knol, D. L., Barkhof, F., … Vermeulen, R. J. (2015). Language comprehension in nonspeaking children with severe cerebral palsy: Neuroanatomical substrate? European Journal of Paediatric Neurology, 19(5), 510–520. doi:10.1016/j.ejpn.2015.06.001

Geytenbeek, J. J., Vermeulen, R. J., Becher, J. G., & Oostrom, K. J. (2015). Comprehension of spoken language in non-speaking children with severe cerebral palsy: An explorative study on associations with motor type and disabilities. Developmental Medicine & Child Neurology, 57(3), 294–300. doi:10.1111/dmcn.12619

Hagtvet, B., & Lillestølen, R. (1990). Reynells språktest. Oslo, Norway: Gyldendal Akademisk.

Hambleton, R. K., (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In. R. K. Hambleton, P. F. Merenda, & C. D. Spielberger, (Eds.), Adapting educational and psychological tests for cross-cultural assessment (pp. 3–38). Mahwah, New Jersey: Psychology Press.

Hayiou-Thomas, M. E., Kovas, Y., Harlaar, N., Plomin, R., Bishop, D. V. M., & Dale, P. S. (2006). Common aetiology for diverse language skills in 4 1/2-year-old twins. Journal of Child Language, 33(2), 339–368. doi:10.1017/S0305000906007331

Herskind, A., Greisen, G., & Nielsen, J. B. (2015). Early identification and intervention in cerebral palsy. Developmental Medicine and Child Neurology, 57(1), 29–36. doi:10.1111/dmcn.12531

Himmelmann, K., Hagberg, G., & Uvebrant, P. (2010). The changing panorama of cerebral palsy in Sweden. Prevalence and origin in the birth-year period 1999-2002. Acta Paediatrica, 99(9), 1337–1343. doi:10.1111/j.1651-2227.2010.01819.x

Himmelmann, K., Lindh, K., & Hidecker, M. J. (2013). Communication ability in cerebral palsy: A study from the CP register of western Sweden. *European Journal of Paediatric Neurology*, 17(6), 568–574. doi:10.1016/j.ejpn.2013.04.005

Holyfield, C., Caron, J., & Light, J. (2019). Programing AAC just-in-time for beginning communicators: The process. *Augmentative and Alternative Communication*, 35(4), 309–318. doi:10.1080/07434618.2019.1686538

IBM Corp. (2017). IBM SPSS Statistics for Windows (Version 25.0) [Computer software]. Armonk, NY: IBM Corp.

International Test Commission. (2017). ITC guidelines for translating and adapting tests (2nd ed., version 2.4) [Webpage]. Retrieved from www.InTestCom.org

Justice, L., Cain, K., Jiang, H., Logan, J., Jia, R., Jiang, H., … Jia, R. (2018). Modeling the nature of grammar and vocabulary trajectories from prekindergarten to third grade. *Journal of Speech, Language, and Hearing Research*, 61(4), 910–923. doi:10.1044/2018_JSLHR-L-17-0090

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: Rethinking the study of developmental disorders. *Developmental Psychology*, 45(1), 56–63. doi:10.1037/a0014506

Klem, M., Gustafsson, J.-E., & Hagtvet, B. (2015). Dimensionality of language ability in four-year-olds: Construct validation of a language screening tool. *Scandinavian Journal of Educational Research*, 59(2), 195–213. doi:10.1080/00313831.2014.904416

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. doi:10.1016/j.jcm.2016.02.012

Krach, S. K., McCreery, M. P., & Guerard, J. (2017). Cultural-linguistic test adaptations: Guidelines for selection, alteration, use, and review. *School Psychology International*, 38(1), 3–21. doi:10.1177/0143034316684672

Kurmanaviciute, R., & Stadskliev, K. (2017). Assessment of verbal comprehension and cognitive reasoning when standard response mode is challenging: A comparison of different response modes and an exploration of their clinical usefulness. *Cogent Psychology*, 4(1), 17. doi:10.1080/23311908.2016.1275416

Language and Reading Research Consortium. (2015). The dimensionality of language ability in young children. *Child Development*, 86, 1948–1965. doi:10.1111/cdev.12450

Light, J., & McNaughton, D. (2014). From basic to applied research to improve outcomes for individuals who require augmentative and alternative communication: Potential contributions of eye tracking research methods. *Augmentative and Alternative Communication*, 30(2), 99–105. doi:10.3109/07434618.2014.906498

Mei, C., Reilly, S., Reddihough, D., Mensah, F., Pennington, L., & Morgan, A. (2016). Language outcomes of children with cerebral palsy aged 5 years and 6 years: A population-based study. *Developmental Medicine and Child Neurology*, 58(6), 605–611. doi:10.1111/dmcn.12957

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D., … de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737–745. doi:10.1016/j.jclinepi.2010.02.006

Murray, J., & Goldbart, J. (2009). Cognitive and language acquisition in typical and aided language learning: A review of recent evidence from an aided communication perspective. *Child Language Teaching and Therapy*, 25(1), 31–58. doi:10.1177/0265659008098660

Novak, I., Morgan, C., Adde, L., Blackman, J., Boyd, R. N., Brunstrom-Hernandez, J., … Badawi, N. (2017). Early, accurate diagnosis and early intervention in cerebral palsy: Advances in diagnosis and treatment. *JAMA Pediatrics*, 171(9), 897–907. doi:10.1001/jamapediatrics.2017.1689

Palisano, R., Rosenbaum, P., Walter, S., Russell, D., Wood, E., & Galuppi, B. (1997). Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine and Child Neurology*, 39(4), 214–223. doi:10.1111/j.1469-8749.1997.tb07414.x

Pinker, S. (1998). Words and rules. *Lingua*, 106(1–4), 219–242. doi:10.1016/S0024-3841(98)00035-7

Pirila, S., van der Meere, J., Pentikainen, T., Ruusu-Niemi, P., Korpela, R., Kilpinen, J., … Kilpinen, R. (2007). Language and motor speech skills in children with cerebral palsy. *Journal of Communication Disorders*, 40(2), 116–128. doi:10.1016/j.jcomdis.2006.06.002

Raven, J. (2008). *Colored Progressive Matrices and Crichton Vocabulary Scale manual*. London, UK: Pearson.

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6(2), 135–147. doi:10.1016/j.edurev.2010.12.001

Romeo, R. R., Segaran, J., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., … Gabrieli, J. D. E. (2018). Language exposure relates to structural neural connectivity in childhood. *The Journal of Neuroscience*, 38(36), 7870–7877. doi:10.1523/JNEUROSCI.0484-18.2018

Romski, M., & Sevcik, R. A. (2005). Augmentative communication and early intervention: Myths and realities. *Infants & Young Children*, 18(3), 174–185. doi:10.1097/00001163-200507000-00002

Rosenbaum, P., Paneth, N., Leviton, A., Goldstein, M., Bax, M., Damiano, D., … Jacobsson, B. (2007). A report: The definition and classification of cerebral palsy April 2006. *Developmental Medicine and Child Neurology*, 49, 8–14. doi:10.1111/j.1469-8749.2007.tb12610.x

Sabbadini, M., Bonanni, R., Carlesimo, G. A., & Caltagirone, C. (2001). Neuropsychological assessment of patients with severe neuromotor and verbal disabilities. *Journal of Intellectual Disability Research*, 45(2), 169–179. doi:10.1046/j.1365-2788.2001.00301.x

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals* (4th ed.). London, UK: Pearson.

Sherwell, S., Reid, S. M., Reddihough, D. S., Wrennal, J., Ong, B., & Stargatt, R. (2014). Measuring intellectual ability in children with cerebral palsy: Can we do better? *Research in Developmental Disabilities*, 35(10), 2558–2567. doi:10.1016/j.ridd.2014.06.019

Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicators Research*, 102(3), 443–461. doi:10.1007/s11205-010-9682-8

Smits, D. W., Ketelaar, M., Gorter, J. W., Van Schie, P. E., Becher, J. G., Lindeman, E., & Jongmans, M. J. (2011). Development of non-verbal intellectual capacity in school-age children with cerebral palsy. *Journal of Intellectual Disability Research*, 55(6), 550–562. doi:10.1111/j.1365-2788.2011.01409.x

Spillane, M. M., Ross, K. K., & Vasa, S. F. (1996). A comparison of eye-gaze and standard response mode on the PPVT – R. *Psychology in the Schools*, 33(4), 265–271. doi:10.1002/(sici)1520-6807(199610)33:4<265::aid-pits1>3.3.co;2-h

Stadskliev, K., Jahnsen, R., Andersen, G. L., & von Tetzchner, S. (2018). Neuropsychological profiles of children with cerebral palsy. *Developmental Neurorehabilitation*, 21(2), 108–120. doi:10.1080/17518423.2017.1282054

Tideman, E., & Gustafsson, J. E. (2004). Age-related differentiation of cognitive abilities in ages 3-7. *Personality and Individual Differences*, 36(8), 1965–1974. doi:10.1016/j.paid.2003.09.004

Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research*, 49(6), 1193–1208. doi:10.1044/1092-4388

Voorman, J. M., Dallmeijer, A. J., Van Eck, M., Schuengel, C., & Becher, J. G. (2009). Social functioning and communication in children with cerebral palsy: Association with disease and characteristics and personal and environmental factors. *Developmental Medicine & Child Neurology*, 52(5), 441–447. doi:10.1111/j.1469-8749.2009.03399.x

Vos, R. C., Dallmeijer, A., Verhoef, M., Van Schie, P. E., Voorman, J. M., Wiegerink, D. J., … Becher, J. G. (2014). Developmental trajectories of receptive and expressive communication in children and young adults with cerebral palsy. *Developmental Medicine and Child Neurology*, 56(10), 951–959. doi:10.1111/dmcn.12473

Wechsler, D. (2015). *Wechsler Preschool and Primary Scale of Intelligence* (4th ed., Norwegian version). London, UK: Pearson Assessment.

Yin Foo, R., Guppy, M., & Johnston, L. M. (2013). Intelligence assessments for children with cerebral palsy: A systematic review. *Developmental Medicine and Child Neurology*, 55(10), 911–918. doi:10.1111/dmcn.12157