Article publié dans American Journal of Speech-Language Pathology et qui peut être retracé à l'adresse suivante : https://doi.org/10.1044/2023_AJSLP-23-00101

Adaptation and Reliability of the Cinderella Story Retell Task in Canadian French Persons Without Brain Injury

Amélie Brisebois[1, 2], Simona Maria Brambati[2,3,4], Claudie Jutras[2], Elizabeth Rochon[5, 6, 7, 8], Carol Leonard[5,7,9], Anna Zumbansen[9,10], Carole Anglade[1], Karine Marcotte[1, 2]

[1] École d'orthophonie et d'audiologie, Faculté de médecine, Université de Montréal, Montréal, Québec, Canada.

[2] Centre de recherche du Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Île-de-Montréal, Montréal, Québec, Canada.

[3] Département de psychologie, Faculté des arts et des sciences, Université de Montréal, Montréal, Québec, Canada

[4] Centre de recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, Québec, Canada.

[5] Department of Speech-Language Pathology, Temerty Faculty of Medicine, University of Toronto, Toronto, Canada

[6] KITE Research Institute, Toronto Rehabilitation Institute, Toronto, Canada

[7] Heart and Stroke Foundation, Canadian Partnership for Stroke Recovery, Ontario, Canada

[8] Rehabilitation Sciences Institute, University of Toronto

[9] School of Rehabilitation Sciences, University of Ottawa

[10] Music and Health Research Institute, University of Ottawa

***Correspondence:**
Karine Marcotte, Ph.D.
Address: Centre de recherche du Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Île-de-Montréal, 5400 Gouin Ouest, Montréal, Québec, Canada, H4J 1C5
Phone number: 514-338-2222 extension 7710
Fax number: 514-340-2115
E-mail: karine.marcotte@umontreal.ca

Article publié dans American Journal of Speech-Language Pathology et qui peut être retracé à l'adresse suivante : https://doi.org/10.1044/2023_AJSLP-23-00101

Abstract

Purpose: Main Concept Analysis (MCA) is a well-documented method of discourse analysis in adults with and without brain injury. This study aims to develop a main concepts (MCs) checklist culturally and linguistically adapted for Canadian French speakers and examine its reliability. We also documented microstructural properties and provide a normative reference in persons not brain injured (PNBI).

Method: Discourse samples from 43 PNBI were collected. All participants completed the Cinderella story retelling task twice. Manual transcription was performed for all samples. The 34 MCs for the Cinderella story retelling task were adapted into Canadian French and used to score all transcripts. In addition, microstructural variables were extracted using CLAN. Intraclass correlation coefficients were computed to assess inter-rater reliability for MC codes and microstructural variables. Test-retest reliability was assessed using intraclass correlations, Spearman rho correlations, and Wilcoxon rank test. Bland Altman plots were used to examine the agreement of the discourse measures between the two sessions.

Results: The MC checklist for the Cinderella story retell task adapted for Canadian French speakers is provided. Good to excellent inter-rater reliability was obtained for most MC codes; however, reliability ranged from poor to excellent for the incorrect and incomplete code. Microstructural variables demonstrated excellent inter-rater reliability. Test-retest reliability ranged from poor to excellent for all variables, with the majority falling between moderate to excellent. Bland Altman plots illustrated the limits of agreement between test and retest.

Conclusion: The present study provides the MC checklist for clinicians and researchers working with Canadian French speakers when assessing discourse with the Cinderella retell task. It also addresses the gap in available psychometric data regarding test-retest reliability in PNBI.

Introduction

The study of discourse, which is language beyond a simple clause (Armstrong, 2000), has become an increasingly important area of interest in acquired neurogenic disorders. Discourse abilities are fundamental for the realization of a large range of everyday needs and social participation (Dipper & Pritchard, 2017), which support their increasing importance both from a clinical and a research point of view. According to Frederiksen's model of discourse (Frederiksen & Stemmer, 1993), discourse production is divided into three distinct stages: 1) conceptual preparation (i.e., idea generation and macrostructural processes), 2) linguistic formulation (i.e., microstructural processes which relate to sentence processing) and 3) articulation and monitoring of the verbal message. Most studies conducting discourse analysis have focused on conceptual preparation, which represents the macrostructural level of discourse, or on linguistic formulation, also known as the microstructural level, and less on the articulation of the verbal level. Macrostructural measures refer to discourse-level processing features such as informativeness, coherence, and cohesion; while microstructural measures refer to within-sentence features and depict discourse's lexical and grammatical components.

Main Concept Analysis

Considered a hybrid macro- and micro-structural approach, the Main Concept Analysis (MCA) focuses on the proposition level of knowledge expression (Richardson & Dalton, 2016a). A main concept (MC) is an utterance that contains a subject, one main verb (and its subordinate clauses), and an optional object (Nicholas & Brookshire, 1993b). Each MC consists of semantic elements considered to be essential to the story gist, and the accuracy and completeness achieved in formulating these elements by the speakers are coded using a multilevel coding system (Dalton & Richardson, 2015; Kong, 2009, 2011; Nicholas & Brookshire, 1993b, 1995; Richardson & Dalton, 2016a). The MCA is useful to assess discourse in constrained discourse tasks either in clinical or research setting. MCA documents the ability to convey conceptual information at the macrostructural level of discourse processing as well as the accuracy of the words and sentences used to express these conceptual elements, which is at the microstructural level. Closed sets of MCs have been developed for specific discourse tasks and cultures. Namely, they were developed for: the Cookie Theft picture description task (BDAE; Goodglass et al., 2000) for English (Nicholas & Brookshire, 1993b) and Japanese speakers (Yazu et al., 2022); Cinderella story retell task, the Broken Window picture sequence narrative and the Peanut Butter and Jelly sandwich procedural tasks (Richardson & Dalton, 2016a) as well as the picture description scene of the cat in the tree and sequence-pictures description of the refused umbrella (Richardson & Dalton, 2020) for English speakers; four sets of sequential stimuli for Cantonese speakers (Kong, 2009), and adapted to Taiwanese Mandarin speakers (Kong & Yeh, 2015), American English speakers (Kong et al., 2016), Japonese speakers (Yazu et al., 2022), Spanish speakers (Kong, 2021), Dutch speakers (Criel et al., 2021); a set of four discourse tasks, two picture scenes and two picture sequences, developed by Nicholas & Brookshire (Nicholas & Brookshire, 1993b) in young English-Spanish bilinguals (Rivera et al., 2018); the Cat in the tree (Nicholas & Brookshire, 1993b) for English speakers (Hameister & Nickels, 2018; Richardson & Dalton, 2020).

MC coding used for Cinderella story retell task of the present study appears in Table 1. The first aspect scored is a concept's presence or absence (AB). If present, the concept receives one of the four following codes: accurate and complete (AC); accurate but incomplete (AI); inaccurate but complete (IC); and inaccurate and incomplete (II). AC, AI, IC, and II codes allow the examiner to analyze the quality of the information and provide more details on the overall informativeness. Detailed scoring guidelines for the Cinderella story retell task appear on the AphasiaBank website (*AphasiaBank*, 2022; MacWhinney et al., 2011; Richardson & Dalton, 2016b).

[Table 1 should be inserted here]

MCA has been largely used to assess the discourse of adults with neurogenic language disorders (e.g., Adams, 2021; Dalton et al., 2020; Fromm et al., 2017; Kong et al., 2016; Kong & Yeh, 2015; Nicholas & Brookshire, 1995). Namely, persons with aphasia have demonstrated less accurate and complete MCs than PNBI without significant differences in the overall production of MCs (e.g., Kong et al., 2016; Kong & Yeh, 2015; Nicholas & Brookshire, 1995). Compelling results have also been obtained in persons with neurocognitive disorders and primary progressive aphasia. For instance, in a sample of individuals with fluent and non-fluent aphasia, with Alzheimer's disease, and PNBI, a lower degree of presence, completeness, accuracy, and efficiency of producing MCs was identified in all clinical groups compared to the PNBI (Kong et al., 2016). Similarly, 17 persons with primary progressive aphasia demonstrated less accurate and complete MCs than PNBI (Dalton et al., 2020). MCA demonstrated high diagnostic sensitivity in 27 persons with subclinical aphasia (Fromm et al., 2017). Less accurate and complete MCs and more absent codes were also observed in 60 persons with latent aphasia compared to persons with anomic aphasia and PNBI (Adams, 2021). Aging effects have also been observed in a large sample of 92 PNBI; speakers less than 59 years of age produced more accurate and complete MCs in the Cinderella retell task than speakers over 60 years of age (Richardson & Dalton, 2016b).

Reliability of Main Concept Analysis

MCA is easy and relatively rapid to score, which supports its clinical feasibility. Microstructural analyses rely on long transcriptions which explains largely why discourse analysis is less used in clinical settings (Bryant et al., 2017). Conversely, MC scoring is based on a finite set of themes, which makes it quicker to analyze, and thus reconciles quantifiable measures with clinical practical requirements. It is also strongly recommended due to its psychometric strengths, including good inter- and intra-rater reliability (Boyle, 2014; Dalton & Richardson, 2015; Kong, 2011; Kong et al., 2016; Nicholas & Brookshire, 1995; Richardson & Dalton, 2016a) and test-retest reliability (Kong, 2011; Kong et al., 2016; Nicholas & Brookshire, 1995). More precisely, studies have reported above 80% point-to-point intra-rater reliability in transcripts of PNBI and participants with aphasia (Nicholas & Brookshire, 1995) and above 90% in PNBI (Richardson & Dalton, 2016a). Good point-to-point inter-rater reliability (i.e., > 80%) was also demonstrated in PNBI and participants with aphasia (Boyle, 2014; Nicholas & Brookshire, 1993b, 1995; Richardson & Dalton, 2016a). Additionally, MCA demonstrated good test-retest reliability for close (i.e., < 3 weeks between sessions; Boyle, 2014; Kong, 2009; Nicholas & Brookshire, 1993b, 1995) and distant (i.e., 12-16 months between sessions; Kong, 2011) assessment for some MC codes. AC and AB codes reached sufficient test-retest reliability for use in research (>.70 recommended for research studies; Fitzpatrick et al., 1998) (Boyle, 2014) and in clinical decision-making (>.90; Kong, 2011). In contrast, poor reliability was obtained in statements including one or more pieces of inaccurate information (IN code in Nicholas and Brookshire's (1995) scoring system) possibly because of the limited number of IN statements for this category (Boyle, 2014). However, the test-retest reliability of MC codes was mainly adequate when tested by combining multiple tasks into one sample (Boyle, 2014, 2015; Brookshire & Nicholas, 1994a, 1994b). Similarly, test-retest reliability of microstructural variables has been mainly assessed using combination of various discourse tasks (e.g, Boyle, 2014; Brookshire & Nicholas, 1994a). However, it has been recently reported for both the five separate monologic tasks and for the combination of the five tasks (Stark, Alexander, et al., 2022) in persons with aphasia. Test-retest reliability was lower for PNBI (Stark, Alexander, et al., 2022), which supports the need to determine psychometric properties of MCA for specific populations.

Cinderella Story retell task

The retell task of the Cinderella story is a semi-spontaneous discourse elicitation method that has been primarily studied in English speakers with and without brain injury (e.g., Fergadiotis & Wright, 2011; Fromm et al., 2017; Greenslade et al., 2020; Richardson & Dalton, 2016a; Stark, 2019). The procedure (see AphasiaBank website; Richardson & Dalton, 2016b) requires the participant to generate a story after looking at a wordless book of the Cinderella tale. Compared to single pictorial stimuli, sequential pictures elicited more relational ideas in PNBI (Capilouto et al., 2005) and more story grammar episodes in individuals with and without closed head injury (Coelho, 2002). Also, the Cinderella task has elicited unique microstructural features compared to expositional and procedural discourse tasks in a large aphasia group (*n* = 90) and a matched PNBI group (Stark, 2019). For instance, contrary to expositional and procedural tasks, the Cinderella task elicited the densest but the least lexically diverse speech in participants with aphasia and matched PNBI, and the most tokens in PNBI (Stark, 2019). These results highlight the importance of investigating the different types of discourse separately as they are mediated by different variables, such as long-term memory and executive functions in the case of Cinderella, and some tasks might be more sensitive than others on different language aspects (Stark, Alexander, et al., 2022). Similarly, in a group of 27 PNBI, lexical diversity was significantly larger in the Cinderella retell compared to results obtained with single and sequential picture descriptions (Fergadiotis & Wright, 2011). In addition, using the Cinderella retell task, a group of 27 post-stroke participants who were not aphasic according to the Western Aphasia Battery-Revised (Kertesz, 2006) performed significantly different than 92 participants with anomic aphasia and 177 PNBI on several measures, including the number of words per minute, the Moving-Average Type Token ratio (MATTR; a measure of lexical diversity) and the MCs (Fromm et al., 2017). In sum, discourse performance in the Cinderella story retell task has been documented at the micro- and macro-structural levels of discourse processing in adults with and without brain damage, including people with subclinical language difficulties in English.

A recent international survey of current practices in discourse assessment identified a lack of linguistic and culturally specific discourse assessment methods (Stark et al., 2021). Indeed, the scarcity of discourse protocols and normative data, including psychometric properties, was identified as a barrier to discourse assessment in non-dominant languages. Although using other tasks, the Main Concept Analysis (Kong, 2009) has been adapted, along with its respective stimuli, from Cantonese to Taiwanese Mandarin speakers (Kong & Yeh, 2015), American English speakers (Kong et al., 2016), Japanese speakers (Yazu et al., 2022), Spanish speakers (Kong, 2021) and Dutch speakers (Criel et al., 2021). However, no such MC list exist in Canadian French. The Cinderella story is well known in the Canadian French culture; thus MCA of the Cinderella story retell task is well suited for cultural and linguistic adaptation. There is also a growing need to document the psychometric properties of discourse measures, which are often influenced by the nature of discourse tasks (e.g., Capilouto et al., 2005; Stark, 2019; Stark, Alexander, et al., 2022). Additionally, knowledge about typical variability in performance in both micro- and macro-structural measures allows clinicians to differentiate 'normal' fluctuations between two assessments from variations attributed to significant language changes (Boyle, 2014). Hence, the main aim of this study is to adapt the MCA for the

Cinderella story retell task for PNBI speakers of Canadian French and to examine its reliability. We also extend our work with the secondary aims of reporting microstructural measures and providing Canadian French norms for these measures in PNBI. Similar to previous studies, good inter-rater reliability is expected, but lower test-retest reliability is probable in PNBI (Stark, Alexander, et al., 2022). We believe making this information available will improve future studies using MCA with Canadian French speakers and will also contribute to the advances in culturally adapted psychometrically sound discourse analysis methods for both research and clinical settings.

Methods

This project is part of a larger study approved by the ethics review board of the Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Ile-de Montréal (CIUSSS-NÎM; #2020-1900) which sought to investigate longitudinal discourse changes following a stroke and to include PNBI. Written informed consent was obtained from all participants. We report best practice guidelines for spoken discourse research in aphasia (Stark, Bryant, et al., 2022; see Supplementary Table 1). Currently, our ethics committee does not grant permission to share individual raw data (i.e., videos and language sample transcriptions).

Participants

Initial recruitment was performed between May and August 2020 in the Montreal (Quebec) area. Forty-three participants were included: 28 females, 15 males; age (M = 64.2, SD = 6.5); education (M = 16.4, SD = 2.7). All participants performed an online assessment twice (days between sessions: M = 241.8, SD = 56.6). The inclusion criteria for this study were: 1) to be at least 50 years of age; 2) have Canadian (Quebec) French as their primary language. The exclusion criteria for this study were: 1) presenting a severe mental illness; 2) presenting an acquired or developmental language impairment; 3) suffering from a neurological impairment; 4) having suffered from a traumatic brain injury; 5) self-reporting uncorrected visual or auditory deficits. Cognitive screening using the videoconference version Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) with instructions for remote administration (https://www.mocatest.org/remote-moca-testing/) was completed (M = 27.7, SD = 1.6). All participants scored within normal range on the videoconference version of the MoCA according to French-Quebec normative data of videoconference administration adjusted to age and education (Gagnon et al., 2022). Participant characteristics appear in Table 2.

[Table 2 should be inserted here]

Data collection

The procedures for virtual assessment are reported in a previously published article by our team (see Marcotte et al., 2022). The story retell task of Cinderella was administered following the AphasiaBank protocol (https://aphasia.talkbank.org/protocol/english/materials-aphasia/instructions.pdf). Participants were shown wordless images of the Cinderella book on their computer screen and asked to remember the story as they went on. The research assistant oversaw sharing and advancing the pictures, which were each presented for 10 seconds. Pictures were presented a second time if participants wanted to revisit previously shown pictures. Then, images were withdrawn from the screen, and participants were asked to retell the story. The instruction was: 'Racontez-moi l'histoire de Cendrillon du mieux que vous pouvez. Vous pouvez utiliser tous les détails que vous connaissiez déjà de l’histoire ainsi que les images que vous venez de regarder.' [Tell me the Cinderella story as well as you can. You can use any details you know about the story as well as the pictures you just looked at]. In cases where participants produced less than three utterances or remained silent for more than 10 seconds, the examiner prompted them: ‘Que s’est-il ensuite passé?’ [What happened next?] or ‘Allez-y.’ [Go on.) Participants' productions were recorded via the Zoom platform (www.zoom.us).

Transcription

Video/audios of each discourse sample were imported and transcribed in ELAN (Sloetjes & Wittenburg, 2008) using CHAT conventions. Complete orthographic transcriptions were conducted, and the transcription was verbatim. The CHAT manual (MacWhinney, 2000) was used for utterance segmentation, transcription and scoring, with additional guidance for French speakers (Colin & Le Meur, 2016). Transcriptions were performed by an experienced speech-language pathologist (A.B.) and an undergraduate student in psychology (C.J.). The same transcriber transcribed both test and retest samples from the same participant for consistency. Transcribers were blind to patient identity.

Microstructural variables extraction

Once the transcription was completed, the morphological and grammatical information coding was conducted using the CLAN program called *mor*, which tags morphemes and words under each utterance in the transcripts. Subsequently, all microstructural variables were extracted for each sample using the program EVAL of CLAN (MacWhinney, 2000). Specific CLAN commands for each variable are provided in Table 1 of Supplementary Material 2.

Main Concept list adaptation in Canadian French

MCA of the Cinderella story retell task was developed originally for American English speakers (Richardson & Dalton, 2016b), and cultural adaptation requires that the target population shares a similar cultural background with the initial sample. Cultural and linguistic sound adaptations usually involve modifications, i.e., developing an entirely new task (Kong, 2009) or refining the scoring protocol (Criel et al., 2021; Yazu et al., 2022). Considering that Canadian French speakers share a similar cultural background with American English speakers regarding Cinderella, an adaptation was made by refining the scoring protocol. Thus, the MC checklist was translated and adapted from Richardson and Dalton's (2016) original list. First, we used the online free version of DeepL Translator (*DeepL Traduction – DeepL Translate*, 2022) to translate the first draft of the 34 MCs in French. Second, a research assistant (C.J.), who was a native Canadian French speaker with advanced knowledge of written English, reviewed the first draft to ensure that each element was as semantically similar as possible to the original version as possible. Third, final adjustments were made via discussion between the research assistant (C.J.), the first author (A.B.) and the principal investigator (K.M.). The final reconciled translation of Main Concept list is reported in the Results section.

MC Scoring

MC scoring was performed with the training materials and scoring guidelines (Richardson & Dalton, 2016a) provided on the AphasiaBank website (https://aphasia.talkbank.org/discourse/MainConcepts/) including video training sessions. The transcripts were used to score MCs manually using a Microsoft Excel spreadsheet (the template is available in Supplementary Material 3). All transcripts were reviewed to identify potentially relevant MCs which were not present in the original list. None were identified.

Dependent variables

Main Concepts

We used Richardson and Dalton's MC scoring system (2016) as depicted in Table 1. The variables are MC Composite, AC, AI, IC, II, and AB.

Microstructural variables

The initial selection of microstructural variables was based on Stark (2019) and recent literature reviews on neurocognitive disorders (Filiou et al., 2020; Slegers et al., 2018). These variables are described in Table 3 and include the mean length of utterance (MLU), duration of samples, the propositional density (Fromm et al., 2016), the number of words per minute, the number of verbs per utterance, the open-closed class ratio, the noun-to-verb ratio, the number of tokens, the percentage of Correct Information Units (CIUs; Nicholas & Brookshire, 1993a), and the percentage of CIUs and Moving Average Token-Type Ratio (MATTR; Covington, 2007).

[Table 3 should be inserted here]

Data analysis

Analysis of MC frequency

Previous test adaptation in Canadian French has demonstrated cultural differences in performance on specific task items (e.g., Callahan et al., 2010). Hence, the frequency of each MC was computed at test and retest. As recommended by Richardson and Dalton (2016), only the MCs which were produced by a minimum of 33% of the sample were kept in the final adaptation of the MC checklist.

Inter-rater reliability

To determine inter-rater reliability in transcription, 19 transcripts per rater (representing 22% of the transcripts each) were selected for each of the two raters randomly. In other words, C.J. transcribed samples that were initially transcribed by A.B. and vice-versa. Two-way mixed intraclass correlation coefficients (ICCs) with absolute agreement were calculated on the tokens, the total number of utterances, and the percentage of CIUs. The total number of Tokens represents the accuracy of the transcription. The number of utterances is critical in CHAT format since it relies uniquely on the transcriber's competence to distinguish utterance boundaries. Reliability on this measure suggests consistency in utterance segmentation throughout the samples.

To determine test and retest consistency between the two raters (A.B. and C.J.) who scored the MCs, samples from 10 participants were randomly selected. ICCs with complete agreement were calculated for all Main Concept codes: MC Composite, AC, AI, IC, II, and AB.

Statistical analysis of test and retest reliability

Data distribution was analyzed using Kolmogorov-Smirnov test for all dependant variables, for each session. Consistent with Stark et al. (2022), more than 70% of the data were not normally distributed; as a result, non-parametric tests were used throughout. Although correlation is one of the most common statistical methods used to investigate test-retest reliability, the sole use of correlations in studies dealing with replicate data is insufficient as it does not test agreement (Bland & Altman, 1986). Test-retest reliability refers to the capacity of a test or measure to replicate the same ordering between participants when tested twice (Kottner et al., 2011), whereas agreement refers to the capacity to provide the same result twice (Berchtold, 2016). Following the guidelines of Koo and Li (2016) to select the appropriate ICC, reliability between test and retest sessions was evaluated using two-way mixed ICC with absolute agreement.

Agreement was tested using Wilcoxon signed rank test to evaluate if there was a statistically significant difference between test and retest. We also measured the strength of association using Spearman's rho to assess the similarity between test and retest. The significance level was set at $p < .05$. Regarding agreement, Bland-Altman plots were produced to allow visual inspection of the data by examining the limits of agreement between testing points (Altman & Bland, 1983). Bland-Altman plots are scatterplots with the Y axis representing the difference between results at test and retest and the X axis representing the mean test and retest results. The scatterplot also illustrates the limits of agreement with horizontal dashed lines at ±1.96 standard deviations of the mean of differences. A good agreement between test and retest is obtained if 95% of the data falls between these limits (Bland & Altman, 1999). These plots were created for the variables that obtained the best test-retest ICC.

As MCA could be useful for detecting subclinical language or cognitive deficits, we also provided minimal detectable change (MDC) for each dependent variable. MDC at a 90% confidence interval (CI) (MDC90) was computed to assess the approximate change needed to be associated with clinical change, given the variance from the test-retest result (Donoghue et al., 2009). MDC90 includes the standard error of measurement (SEM), computed with the following formula: $SEM = SD\sqrt{1-r}$, where SD is the standard deviation for the obtained score distribution and r is the correlation coefficient (i.e., ICC). The formula to calculate MDC90 is $MDC90 = SEM*1.65*\sqrt{2}$.

Analysis software

All statistical analyses were performed using SPSS® v26.0. Bland-Altman plots were computed using RStudio 2022.07.2.

Results

Development of the adapted MC list

The frequency of each MC was computed at test and retest and appear in Table 4. MC #9, #11, and #12 did not reach the 33% frequency threshold suggested by Richardson and Dalton (2016) and, therefore, were not included in the statistical analyses (see the Excel sheet ‘Modèle à copier’ in Supplementary Material 3 for the checklist adapted in Canadian French). The final adapted list of MCs with the detailed scoring guide appears in Supplemental Material 4.

[Table 4 should be inserted here]

Inter-rater reliability

Koo and Li (2016) interpretation guidelines were used for all ICCs (inter-rater and test-retest reliability): below .50 = poor; between .50 and .75 = moderate; between .75 and .90 = good; and above .90 = excellent.

Transcription reliability on the first assessment was excellent for the total number of utterances ($ICC_{[2,1]}$ = 0.901, 95% CI [0.732, 0.963]) and tokens ($ICC_{[2,1]}$ = 0.997, 95% CI [0.991, 0.999]), and %CIU ($ICC_{[2,1]}$ = 0.985, 95% CI [0.861, 0.994]). MC Composite scoring reliability was excellent at both test ($ICC_{[2,1]}$ = 0.941, 95% CI [0.783, 0.985]) and retest ($ICC_{[2,1]}$ = 0.965, 95% CI [0.866, 0.991]). Excellent inter-rater reliability was also found for AC at both test ($ICC_{[2,1]}$ = 0.932, 95% CI [0.753, 0.983]) and retest ($ICC_{[2,1]}$ = .976, 95% CI [.906-.994]). IC scoring reliability was excellent at both test ($ICC_{[2,1]}$ =0 .951, 95% CI [0.815, 0.987]) and retest ($ICC_{[2,1]}$ = 0.915, 95% CI [0.696, 0.978]). AB scoring reliability was excellent at both test ($ICC_{[2,1]}$ =0 .952, 95% CI [0.821, 0.988]) and retest ($ICC_{[2,1]}$ = 0.950, 95% CI [0.813, 0.987]). Good inter-rater reliability was found at test ($ICC_{[2,1]}$ = 0.800, 95% CI [0.382, 0.983]), whereas it was excellent at retest ($ICC_{[2,1]}$ = 0.914, 95% CI [0.694, 0.978]). Inter-rater reliability of II was on average poor, with the confidence interval, at test ($ICC_{[2,1]}$ =0.533, 95% CI [-0.101, 0.859]), but excellent at retest ($ICC_{[2,1]}$ = 0.950, 95% CI [0.813-.987]). Table 2 of Supplementary Material 2 provides ICC inter-rater reliability results for MC Composite, AC, AI, IC, II, and AB codes.

Test-retest reliability

Considering the extensiveness of the results, a summary is presented in Table 5. No systematic differences were obtained for all of the MC codes and microstructural variables, except for the coding of II, which showed a significant test-retest difference (p=.007). The strengths of the relationship between sessions ranged from weak to strong. The MC codes of AC, IC, and AB as well as MC Composite obtained moderate associations between test and retest, demonstrating the highest strength of relationship. Microstructural variables demonstrated associations ranging from very weak to strong relationships between test and retest. Duration, tokens, number of words per minute, density, noun/verb ratio and CIU per minute demonstrated strong associations, and the number of verbs per utterance demonstrated moderate association between sessions.

[Table 5 should be inserted here]

A summary of test-retest results, ICCs, Spearman rho correlations, and absolute value differences is reported in Table 6. The Minimal Detectable Change at 90% CI (MDC90) is also presented in Table 6. The MC codes AC and AB as well as MC Composite obtained good inter-rater reliability. The inter-rater reliability for IC ranged from moderate to good. For the inter-rater reliability of the coding of AI and II poor ICC was obtained. As for the microstructural variables, the percentage of CIUs obtained an excellent ICC with CI ranging between moderate and excellent. In addition, the measures of duration, number of tokens, and number of words per minute all obtained good ICC with 95% CI ranging from moderate to good.

[Table 6 should be inserted here]

Bland-Altman plots were created for the MC variables and the microstructural variable that obtained the best test-retest ICCs. Figure 1 illustrates the limits of agreement for the variables MC Composite, AC, and AB, whereas Figure 2 represents the limits of agreement for %CIU. Mean differences of agreement were close to zero for both AC and AB, respectively at 0.95 and 0.81. MC Composite presented a mean of differences of 2.21 between test and retest. MC Composite and AC demonstrated good agreement according to the standards of Bland and Altman (1999), with 95% of data (i.e., 41 out of 43) within ±1.96 standard deviations of the mean of differences. AB obtained 90% of the values (i.e., 40 out of 43) within limits of agreement of ±1.96 standard deviations. The mean difference of agreement between test and retest was also close to zero for %CIU, more precisely at -0.22. The %CIU also obtained good agreement according to the standards espoused by Bland and Altman (1999), with 95% of the data (i.e., 41 out of 43) within ±1.96 standard deviations of the mean of differences.

[Figures 1 and 2 should be inserted here]

Discussion

This study aimed to document the test-retest reliability of Main Concept Analysis (MCA) for the Cinderella story retell task by Canadian French speakers. To begin, a cultural and linguistic adaptation of the main concept checklist of Richardson and Dalton (2016) was constructed to reflect speakers of Canadian French. We also reported microstructural measures and provided a normative reference for PNBI. Similar to the adaptation of the Pyramid and Palm Trees Test for

Canadian French speakers (Callahan et al., 2010), our adaptation of the MC list (Richardson & Dalton, 2016a) for the Cinderella story retell task led to the removal of 3 infrequent items is now freely available (see Supplementary Material 3 and 4). Inter-rater reliability results ranged from good to excellent for MC Composite, AC, AI, IC and AB and poor for II. Analyses of systematic differences, evaluation of the strength of the relationship, and ICCs confirmed test-retest reliability for MC variables MC Composite, AC and AB and microstructural variables of duration, number of tokens, number of words per minute, and the percentage of CIU. Conversely, MC codes of AI, IC and II as well as the microstructural variables of MLU, verbs per utterance, propositional density, noun/verb ratio, open/closed class ratio, CIU/minute and MATTR demonstrated poor to moderate test-retest reliability. MDC90 is reported for all variables, thus providing guidelines that are culturally and linguistically adapted to Canadian French speakers for the Cinderella story retell task. As a result, this discourse assessment has the potential to detect preclinical language and/or cognitive deficits.

MCA test-retest reliability

Concerning test-retest reliability, all MC codes except II demonstrated no significant differences between test and retest, thus supporting our hypothesis of the stability of the coding. The relationship between test and retest sessions for the MC codes AC, IC and AB as well as MC Composite was moderate. For the AI and II codes a weak test-retest association was found. These results are, in fact, in line with previous reports of incorrect MC codes being less reliable than others in persons with aphasia (Boyle, 2014; Kong, 2011). Our study demonstrated a lower strength of relationship between test-retest sessions than studies conducted with participants with aphasia (Boyle, 2014; Kong, 2011) This is also consistent with expectations of higher performance variability in PNBI (Stark, Alexander, et al., 2022).

Prior research indicates that the codes of AC, AI, and AB are reliable for research in discourse processing (Boyle, 2014; Fitzpatrick et al., 1998). The codes of AC, AB, and MC Composite also obtained sufficient stability over time for clinical decisions concerning persons with aphasia (Kong, 2011). In our study with PNBI, MC Composite as well as the MC codes AC and AB were sufficiently stable, thus supporting the use of these codes to conduct group research studies (i.e., ICC>.70) in Canadian French speakers. This confirms and extends previous findings with samples of English speakers. However, the incorrect and incomplete code (II) evidenced quite different psychometric properties than the other MC codes with poor to excellent inter-rater reliability, significant systematic difference, and a weak relationship between test and retest. Similar to the present results, Boyle (2014) found that the inaccurate information code (IN) in persons with aphasia obtained poor test-retest correlations across three sessions. Boyle (2014) suggested that the restricted range of IN responses may have influenced such low correlations; this is also a plausible explanation concerning our dataset. However, it is important to note that, contrary to this finding, Nicholas and Brookshire (1995) found high test-retest correlations for the coding of IN responses in three discourse tasks across three sessions in persons with aphasia. The inclusion of more than one task in the calculation of stability may have affected the results (Boyle, 2014). Overall, in the present study, the MC variables of MC Composite, AC, and AB demonstrated the best psychometric properties with no systematic difference between test and retest, moderate associations between sessions, good ICC quality, and more than 90% of data within limits of agreement of ±1.96 standard deviations. These results suggest that MC Composite, AC, and AB are the most reliable codes to assess discourse production in Canadian French speaking PNBI.

Microstructural test-retest reliability

Regarding the microstructural variables assessed, those of duration, number of tokens, and number of words per minute obtained good test-retest reliability, and the percentage of CIU obtained excellent test-retest reliability. Notably, the duration, number of tokens, and number of words per minute reached the criteria for inclusion in research studies (>.70 ICC) and the percentage of CIU attained the criterion for clinical use (>.90 ICC). Interesting, however, conflicting evidence has been found for some microlinguistic variables. For instance, lexical diversity demonstrated moderate to good test-retest reliability in persons with aphasia (Boyle, 2014), and moderate test-retest reliability in PNBI (Stark, Alexander, et al., 2022); however, in the present study, the test-retest reliability of this measure was poor. The nature of the metrics themselves may help to explain these conflicting results. Specifically, we chose the MATTR variable to assess lexical diversity because its calculation considers the variation in the length of samples; whereas the Type-Token ratio used by Stark, Alexander et al. (2022) and VocD used by Boyle (2014) do not consider this potential confound. Another point to consider is that the test-retest reliability of microstructural variables has mainly been reported for the combination of discourse tasks, as per clinical guidelines for people with aphasia (Boyle, 2015). This practice is based on the view that the assessment of multiple discourse tasks is necessary to provide a comprehensive picture of an individual's discourse abilities. With respect to word retrieval measures, including MC codes, a combination of discourse tasks has been reported to improve the test-retest reliability of measures in persons with aphasia (Boyle, 2014). This method also increases the sample size, with a minimum of 300 to 400 words recommended to improve test-retest reliability (Brookshire & Nicholas, 1994a). While the present study included only one task, we collected mean samples of 758 words at test and 738 words at retest, which is well above the recommended minimum length of samples to investigate test-retest reliability.

Assessing discourse performance on multiple tasks was beyond the scope of the current study. Nonetheless, the sample sizes and discourse task combination are considerations to keep in mind when assessing test-retest reliability of discourse measures and in future investigations.

Minimal Detectable Change

Our report of expected variability and Minimal Detectable Change (MDC90) allows future studies, including subclinical or clinical population comparisons, to provide reference data for speakers of Canadian French. In literature reviews of discourse measures in people with neurocognitive diseases, microstructural variables were identified to be different in people with mild cognitive impairment compared to PNBI in picture description tasks (Filiou et al., 2020; Slegers et al., 2018). Indeed, the number of words per minute, the mean length of utterances, the propositional density, the lexical informativeness, and the lexical diversity were variables that differentiated people with mild cognitive impairment or mild Alzheimer's disease from controls. To our knowledge, MCA has not been studied in people with subjective cognitive impairment, which is the subjective presence of cognitive decline without evidence of objective cognitive impairment (Jessen et al., 2020). Subtle cognitive decline is usually not detected by standard cognitive testing, and its identification requires highly sensitive measures with robust psychometrical features (Jessen et al., 2014). We would expect MCA to be able to detect early signs of cognitive decline because it demonstrated good diagnostic sensitivity with latent (Adams, 2021) and subclinical aphasia (Fromm et al., 2017), and also healthy aging individuals. (Richardson & Dalton, 2016a).

Clinical implications

The present psychometric data in Canadian French will allow future studies to test the potential use of MCA in identifying subtle language changes and subjective cognitive decline. As mentioned previously, MCA demonstrated good diagnostic sensitivity (Adams, 2021; Fromm et al., 2016; Richardson & Dalton, 2016a), which suggests that it could be a sensitive measure, with robust psychometrical features, to detect subtle cognitive decline in older adults. The cultural and linguistic adaptation of any test or list is critical to avoid any potential bias when analyzing the results. Accordingly, three MCs were removed from the original list because they were used infrequently in our group of people speaking Canadian French. Another important reason for the adaptation of the MC list for the Cinderella retell task was that, as a measure, it is relatively easy and quick to implement in language assessments, including both PNBI and people with aphasia. Microstructural analyses typically rely on long transcriptions which are less used in clinical settings (Bryant et al., 2017). Similar to our TU list (Brisebois et al., 2020) developed for the picnic scene of the Western Aphasia Battery – Revised (Kertesz, 2006), the MC scoring of the Cinderella retell task is based on a finite set of themes that are more easily quantified and thus more suitable for clinical settings. In addition to providing reference data regarding the longitudinal changes in discourse of PNBI for MCA, our study also enriches the data available on the microstructure of discourse for the Cinderella story retell task (Stark, Alexander, et al., 2022; Stark & Fukuyama, 2021).

Study limitations

This study is not without some limitations. First, concerning inter-rater analyses of transcriptions, we conducted the analysis at only one time point. We agree that, like others, samples from both test and retest could have been included in the analysis (Stark &Alexander, 2022). Nonetheless, interrater reliability was calculated in 22% of the total samples, which is consistent with previous studies (e.g., Kong, 2011; Stark, Alexander, et al., 2022). Second, the sample size is relatively small. However, this sample size is comparable to other similar studies (e.g., Richardson & Dalton, 2016b) considering the population of reference. Third, the sample may not be representative of the older population since we only included speakers from 55 to 79 years of age. Fourth, in contrast to previous studies (e.g., Stark, Alexander, et al., 2022), we chose a longer period between testing sessions ranging from 162 to 373 days that may better reflect changes associated with typical aging (Mueller et al., 2018). Our sample's age range does fall well within the age range whereby the first signs of some degenerative disorders appear, such as primary progressive aphasia (Mouton et al., 2022) and subjective cognitive impairment (Jessen et al., 2014). We did not administer a second cognitive screening because the MoCA was conducted at the follow-up session. Finally, no vision nor auditory screenings were conducted to ensure all participants had intact and sufficient vision and hearing abilities.

Conclusion

To conclude, the assessment of discourse abilities is considered an essential part of a comprehensive language and communication evaluation for people with acquired language difficulties (Bryant et al., 2017). Studying language abilities beyond the level of the utterance may be particularly useful in identifying performance differences in people with more covert language impairments (Kong, 2011). The current study focused on the development of a linguistically and culturally adapted, psychometrically sound discourse measure – that of the Cinderella story retell task -- for speakers of Canadian French. The scarcity of discourse protocols and normative data in Canadian French, a non-dominant language in North America, is a barrier to discourse assessment both for research and clinical purposes, as reported for other non-dominant languages (Stark et al., 2021). The Cinderella story is well-known to speakers of

Canadian French (as it is to Canadian speakers of English). Thus, the cultural adaptation of the MC list of the Cinderella story retell task (Richardson & Dalton, 2016a) was well suited for the present cultural and linguistic adaptation. Detailed information on MCA is available on the AphasiaBank website (https://aphasia.talkbank.org/discourse/MainConcepts/); however no such data yet exists for Canadian French. The overall results provide insight into typical performance and variation, which is crucial to differentiate language changes due to pathology (Boyle, 2014).

Acknowledgments

We are very grateful to all the participants who contributed to this study. Also, we wish to thank Marianne Désilets-Barnabé and Noémie Desjardins for their help in data collection. This project was funded by the Heart and Stroke Foundation (grant-in-aid numbers G-16-00014039 and G-19-0026212) to K.M. and S.M.B. S.M.B. holds a Career Award from the "Fonds de Recherche du Québec – Santé." And A.B. holds a scholarship from the "Fonds de Recherche du Québec – Santé."

Data availability statement

The raw data presented in this article are not readily available because of the sensitivity of the video materials. The datasets analysed during the current study are available from the corresponding author upon reasonable request.

References

Adams, J. B. (2021). *Main Concept, Sequencing, and Story Grammar (MSSG) Analyses of the Cinderella Story in Latent Aphasia.* [University of New Mexico - Main Campus]. https://digitalrepository.unm.edu/shs_etds/30

Altman, D. G., & Bland, J. M. (1983). Measurement in Medicine: The Analysis of Method Comparison Studies. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *32*(3), 307–317. https://doi.org/10.2307/2987937

AphasiaBank. (2022). https://aphasia.talkbank.org/

Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, *14*(9), 875–892. https://doi.org/10.1080/02687030050127685

Berchtold, A. (2016). Test–retest: Agreement or reliability? *Methodological Innovations*, *9*. https://doi.org/10.1177/2059799116672875

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet (London, England)*, *1*(8476), 307–310.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*(2), 135–160. https://doi.org/10.1177/096228029900800204open_in_new

Boyle, M. (2014). Test-retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research*, *57*(3), 966–978. https://doi.org/10.1044/2014_JSLHR-L-13-0171

Boyle, M. (2015). Stability of Word-Retrieval Errors With the AphasiaBank Stimuli. *American Journal of Speech-Language Pathology*, *24*(4), S953–S960. https://doi.org/10.1044/2015_AJSLP-14-0152

Brisebois, A., Brambati, S. M., Désilets-Barnabé, M., Boucher, J., García, A. O., Rochon, E., Leonard, C., Desautels, A., & Marcotte, K. (2020). The importance of thematic informativeness in narrative discourse recovery in acute post-stroke aphasia. *Aphasiology*, *34*(4). https://doi.org/10.1080/02687038.2019.1705661

Brookshire, R. H., & Nicholas, L. E. (1994a). Speech sample size and test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech and Hearing Research*, *37*(2), 399–407.

Brookshire, R. H., & Nicholas, L. E. (1994b). *Test-Retest Stability of Measures of Connected Speech in Aphasia* [Clinical Aphasiology Paper]. Clinical Aphasiology; Pro-Ed. http://aphasiology.pitt.edu/163/

Bryant, L., Spencer, E., & Ferguson, A. (2017). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology*, *31*(10), 1105–1126. https://doi.org/10.1080/02687038.2016.1239013

Callahan, B. L., Macoir, J., Hudon, C., Bier, N., Chouinard, N., Cossette-Harvey, M., Daigle, N., Fradette, C., Gagnon, L., & Potvin, O. (2010). Normative data for the pyramids and palm trees test in the quebec-french population. *Archives of Clinical Neuropsychology*, *25*(3), 212–217. https://doi.org/10.1093/arclin/acq013

Capilouto, G., Wright, H. H., & Wagovich, S. A. (2005). CIU and main event analyses of the structured discourse of older and younger adults. *Journal of Communication Disorders*, *38*(6), 431–444. https://doi.org/10.1016/j.jcomdis.2005.03.005

Coelho, C. A. (2002). Story narratives of adults with closed head injury and non-brain-injured adults: Influence of socioeconomic status, elicitation task, and executive functioning. *Journal of Speech, Language, and Hearing Research: JSLHR*, *45*(6), 1232–1248. https://doi.org/10.1044/1092-4388(2002/099)

Criel, Y., Deleu, M., De Groote, E., Bockstael, A., Kong, A. P.-H., & De Letter, M. (2021). The Dutch Main Concept Analysis: Translation and Establishment of Normative Data. *American Journal of Speech-Language Pathology*, *30*, 1750–1766. https://doi.org/10.1044/2021_AJSLP-20-00285

Dalton, S. G. H., Hubbard, H. I., & Richardson, J. D. (2020). Moving Toward Non-transcription based Discourse Analysis in Stable and Progressive Aphasia. *Seminars in Speech and Language*, *41*(1), 32–44. https://doi.org/10.1055/s-0039-3400990

Dalton, S. G. H., & Richardson, J. D. (2015). Core-Lexicon and Main-Concept Production During Picture-Sequence Description in Adults Without Brain Damage and Adults With Aphasia. *American Journal of Speech-Language Pathology*, *24*(4), S923-938. https://doi.org/10.1044/2015_AJSLP-14-0161

DeepL Traduction – DeepL Translate: Le meilleur traducteur au monde. (2022). https://www.DeepL.com/translator

Dipper, L. T., & Pritchard, M. (2017). Discourse: Assessment and Therapy. In F. D. M. Fernandes (Ed.), *Advances in Speech-language Pathology*. InTech. https://doi.org/10.5772/intechopen.69894

Donoghue, D., Physiotherapy Research and Older People (PROP) group, & Stokes, E. K. (2009). How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. *Journal of Rehabilitation Medicine*, *41*(5), 343–346. https://doi.org/10.2340/16501977-0337

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, *25*(11), 1414–1430. https://doi.org/10.1080/02687038.2011.603898

Filiou, R.-P., Bier, N., Slegers, A., Houzé, B., Belchior, P., & Brambati, S. M. (2020). Connected speech assessment in the early detection of Alzheimer's disease and mild cognitive impairment: A scoping review. *Aphasiology*, *34*(6), 723–755. https://doi.org/10.1080/02687038.2019.1608502

Fitzpatrick, R., Davey, C., Buxton, M. J., & Jones, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials: A review. *Health Technology Assessment*, *2*(14). https://doi.org/10.3310/hta2140

Frederiksen, C. H., & Stemmer, B. (1993). Conceptual processing of discourse by a right hemisphere brain-damaged patient. In H. Brownell & Y. Joanette (Ed.), *Narrative discourse in neurologically impaired and normal aging adults* (pp. 239–278). Singular Publishing Group.

Fromm, D., Forbes, M., Holland, A., Dalton, S. G. H., Richardson, J., & MacWhinney, B. (2017). Discourse Characteristics in Aphasia Beyond the Western Aphasia Battery Cutoff. *American Journal of Speech-Language Pathology*, *26*(3), 762–768. https://doi.org/10.1044/2016_AJSLP-16-0071

Fromm, D., Greenhouse, J., Hou, K., Russell, G. A., Cai, X., Forbes, M., Holland, A., & MacWhinney, B. (2016). Automated Proposition Density Analysis for Discourse in Aphasia. *J Speech Lang Hear Res*, *59*(5), 1123–1132. https://doi.org/10.1044/2016_jslhr-l-15-0401

Gagnon, C., Olmand, M., Dupuy, E. G., Besnier, F., Vincent, T., Grégoire, C.-A., Lévesque, M., Payer, M., Bérubé, B., Breton, J., Lecchino, C., Bouabdallaoui, N., Iglesies-Grau, J., Gayda, M., Vitali, P., Nigam, A., Juneau, M., Hudon, C., & Bherer, L. (2022). Videoconference version of the Montreal Cognitive Assessment: Normative data for

Quebec-French people aged 50 years and older. *Aging Clinical and Experimental Research*, *34*(7), 1627–1633. https://doi.org/10.1007/s40520-022-02092-1

Goodglass, H., Kaplan, E., Barresi, B., Goodglass, H., Goodglass, H., Goodglass, H., Goodglass, H., Goodglass, H., Kaplan, E., & Kaplan, E. (2000). *The Boston Diagnostic Aphasia Examination (BDAE)*. Lippincott Williams.

Greenslade, K. J., Stuart, J. E. B., Richardson, J. D., Dalton, S. G. H., & Ramage, A. E. (2020). Macrostructural analyses of cinderella narratives in a large nonclinical sample. *American Journal of Speech-Language Pathology*, *29*(4), 1923–1936. https://doi.org/10.1044/2020_AJSLP-19-00151

Hameister, I., & Nickels, L. (2018). The cat in the tree – using picture descriptions to inform our understanding of conceptualisation in aphasia. *Language, Cognition and Neuroscience*, *33*(10), 1296–1314. https://doi.org/10.1080/23273798.2018.1497801

Jessen, F., Amariglio, R. E., Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., Dubois, B., Dufouil, C., Ellis, K. A., Flier, W. M., Glodzik, L., Harten, A. C., Leon, M. J., McHugh, P., Mielke, M. M., Molinuevo, J. L., Mosconi, L., Osorio, R. S., Perrotin, A., … Subjective Cognitive Decline Initiative (SCD-I) Working Group. (2014). A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimer's & Dementia*, *10*(6), 844–852. https://doi.org/10.1016/j.jalz.2014.01.001

Jessen, F., Amariglio, R. E., Buckley, R. F., Flier, W. M. van der, Han, Y., Molinuevo, J. L., Rabin, L., Rentz, D. M., Rodriguez-Gomez, O., Saykin, A. J., Sikkes, S. A. M., Smart, C. M., Wolfsgruber, S., & Wagner, M. (2020). The characterisation of subjective cognitive decline. *The Lancet Neurology*, *19*(3), 271–278. https://doi.org/10.1016/S1474-4422(19)30368-0

Kertesz, A. (2006). *Western Aphasia Battery- Revised*. Pearson.

Kong, A. P.-H. (2009). The use of main concept analysis to measure discourse production in Cantonese-speaking persons with aphasia: A preliminary report. *Journal of Communication Disorders*, *42*(6), 442–464. https://doi.org/10.1016/j.jcomdis.2009.06.002

Kong, A. P.-H. (2011). The main concept analysis in cantonese aphasic oral discourse: External validation and monitoring chronic aphasia. *Journal of Speech, Language, and Hearing Research*, *54*(1), 148–159. https://doi.org/10.1044/1092-4388(2010/09-0240)

Kong, A. P.-H. (2021). Dialectally-sensitive norms of the Spanish version of Main Concept Analysis (Span-MCA) for quantifying neurogenically impaired spoken discourse. *Revista de Investigación En Logopedia*, *11*(2), e69932. https://doi.org/10.5209/rlog.69932

Kong, A. P.-H., Whiteside, J., & Bargmann, P. (2016). The Main Concept Analysis: Validation and sensitivity in differentiating discourse produced by unimpaired English speakers from individuals with aphasia and dementia of Alzheimer type. *Logopedics Phoniatrics Vocology*, *41*(3), 129–141. https://doi.org/10.3109/14015439.2015.1041551

Kong, A. P.-H., & Yeh, C.-C. (2015). A Taiwanese Mandarin Main Concept Analysis (TM-MCA) for quantification of aphasic oral discourse: Taiwanese Mandarin Main Concept Analysis. *International Journal of Language & Communication Disorders*, *50*(5), 580–592. https://doi.org/10.1111/1460-6984.12157

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

Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., Roberts, C., Shoukri, M., & Streiner, D. L. (2011). Guidelines for Reporting Reliability and

Agreement Studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*, *64*(1), 96–106. https://doi.org/10.1016/j.jclinepi.2010.03.002

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). Aphasiabank: Methods for studying discourse. *Aphasiology*, *25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893

Marcotte, K., Lachance, A., Brisebois, A., Mazzocca, P., Désilets-Barnabé, M., Desjardins, N., & Brambati, S. M. (2022). Validation of videoconference administration of picture description from the Western Aphasia Battery- Revised in neurotypicals Canadian French speakers. *American Journal of Speech-Language Pathology*.

Mouton, A., Plonka, A., Fabre, R., Tran, T. M., Robert, P., Macoir, J., Manera, V., & Gros, A. (2022). The course of primary progressive aphasia diagnosis: A cross-sectional study. *Alzheimer's Research & Therapy*, *14*(1), 64. https://doi.org/10.1186/s13195-022-01007-6

Nicholas, L. E., & Brookshire, R. H. (1993a). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*, *36*(2), 338–350. https://doi.org/10.1044/jshr.3602.338

Nicholas, L. E., & Brookshire, R. H. (1993b). *A System for Scoring Main Concepts in the Discourse of Non-brain-damaged and Aphasic Speakers*. http://aphasiology.pitt.edu/1440/

Nicholas, L. E., & Brookshire, R. H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech and Hearing Research*, *38*(1), 145–156. https://doi.org/10.1044/jshr.3801.145

Richardson, J. D., & Dalton, S. G. H. (2016a). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology*, *30*(1), 45–73.

https://doi.org/10.1080/02687038.2015.1057891

Richardson, J. D., & Dalton, S. G. H. (2020). Main concepts for two picture description tasks: An addition to Richardson and Dalton, 2016. *Aphasiology*, *34*(1), 119–136. https://doi.org/10.1080/02687038.2018.1561417

Richardson, J. D., & Dalton, S. G. H. (2016b). *AphasiaBank | Main Concept Analysis*. https://aphasia.talkbank.org/discourse/MainConcepts/

Rivera, A., Hirst, J., & Edmonds, L. A. (2018). Evaluation of Language Predictors of Main Concept Production in Spanish/English Bilingual Discourse Using Nicholas and Brookshire Stimuli. *American Journal of Speech-Language Pathology*, *27*(1), 52–70. https://doi.org/10.1044/2017_AJSLP-15-0186

Sherratt, S. (2007). Multi-level discourse analysis: A feasible approach. *Aphasiology*, *21*(3–4), 375–393. https://doi.org/10.1080/02687030600911435

Sherratt, S., & Bryan, K. (2019). Textual cohesion in oral narrative and procedural discourse: The effects of ageing and cognitive skills. *International Journal of Language & Communication Disorders*, *54*(1), 95–109. https://doi.org/10.1111/1460-6984.12434

Slegers, A., Filiou, R.-P., Montembeault, M., & Brambati, S. M. (2018). Connected Speech Features from Picture Description in Alzheimer's Disease: A Systematic Review. *Journal of Alzheimer's Disease*, *65*(2), 519–542. https://doi.org/10.3233/JAD-170881

Stark, B. C. (2019). A comparison of three discourse elicitation methods in aphasia and age-matched adults: Implications for language assessment and outcome. *American Journal of Speech-Language Pathology*, *28*(3), 1067–1083. https://doi.org/10.1044/2019_AJSLP-18-0265

Stark, B. C., Alexander, J. M., Hittson, A., Doub, A., Igleheart, M., Streander, T., & Jewell, E.

(2022). *Reliability of spoken discourse in speakers with aphasia*. OSF Preprints. https://doi.org/10.31219/osf.io/b3s27

Stark, B. C., Bryant, L., Themistocleous, C., den Ouden, D. B., & Roberts, A. C. (2022). Best practice guidelines for reporting spoken discourse in aphasia and neurogenic communication disorders. *Aphasiology*, 1–24. https://doi.org/10.1080/02687038.2022.2039372

Stark, B. C., Dutta, M., Murray, L. L., Fromm, D., Bryant, L., Harmon, T. G., Ramage, A. E., & Roberts, A. C. (2021). Spoken discourse assessment and analysis in aphasia: An international survey of current practices. *Journal of Speech, Language, and Hearing Research*, *64*(11), 4366–4389. https://doi.org/10.1044/2021_JSLHR-20-00708

Stark, B. C., & Fukuyama, J. (2021). Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia. *Language, Cognition and Neuroscience*, *36*(5), 562–585. https://doi.org/10.1080/23273798.2020.1862258

Tompkins, C. A. (1995). *Right hemisphere communication disorders: Theory and management* / (Singular publishing group).

Yazu, H., Kong, A. P.-H., Yoshihata, H., & Okubo, K. (2022). Adaptation and validation of the main concept analysis of spoken discourse by native Japanese adults. *Clinical Linguistics & Phonetics*, *36*(1), 17–33. https://doi.org/10.1080/02699206.2021.1915385

Figure titles

Figure 1. Bland-Altman plots for Main Concepts variables of (a) MC Composite (b) AC and (c) AB.

Figure 2. Bland-Altman plot for the percentage of CIUs.

Supplementary material

Supplementary material 1. Best Practice Guidelines for Reporting Spoken Discourse in Aphasia and Neurogenic Communication Disorders.

Supplementary material 2
Table 1. CLAN commands used to extract discourse variables in the transcripts.
Table 2. Summary of inter-rater reliability results.

Supplementary material 3. Feuille de calcul excel des MC de Cendrillon (Cinderella's MC excel scoresheet).

Supplemental material 4. Canadian French adaptation of the Main concepts for the Cinderella retell task.

Table 1. Richardson and Dalton's (2016) Main Concept scoring system

Label	**Score for each MC**	**Definition**	**Examples** The target is MC #2: 'Cendrillon[1] vit[2] a belles soeurs[3]' [Cinderella lives with ste
Accurate and Complete (AC)	3 points	The statements contain all correct information.	'c'est une jeune fille[1] qui a perdu sa mèr et son père s'est remarié (...) donc le pèr belle-mère arrive à la maison avec les d [it's a young woman who's lost her moth dad and her dad got remarried (...) so he the stepmother]
Accurate and Incomplete (AI)	2 points	The statements contain correct pieces of information but fail to include one essential element.	'la jeune fille[1] vit[2] dans une maison' [the young woman lives in a house]
Inaccurate and Complete (IC)	2 points	The statements contain at least one incorrect piece of information but mention all essential elements.	'c'est une jeune fille[1] qui vit[2] avec sa tan [It's a young woman who lives with her
Inaccurate and Incomplete (II)	1 point	The statements contain at least one incorrect essential element and fail to include at least one essential element.	'elle[1] visite[2] une maison' [She visits a house]
Absent (AB)	0 points	The statements are absent.	

MC Composite (total composite score of all MCs) was computed according to Richardson and Dalton's (201 AC) + (2 x AI) + (2 x IC) + (1 x II)).

Table 2. Participants' characteristics.

Variable	
Age	
Mean (SD)	64.23 (6.54)
Median [Min - Max]	62 [55 - 79]
Gender	
Female	28 (65.12%)
Male	15 (34.88%)
Handedness	
Right	39 (90.70%)
Left	4 (9.30%)
Education	
Mean (SD)	16.44 (2.73)
Median [Min - Max]	16.0 [11 -25]
Time between sessions (days)	
Mean (SD)	241.77 (56.61)
Median [Min - Max]	253.0 [162 - 373]
Montreal Cognitive Assessment	
Mean (SD)	27.7 (1.64)
Median [Min - Max]	28.0 [24 - 30]
Naming score (TDQ30; Test de dénomination de Québec)	
Mean (SD)	28.88 (1.10)
Median [Min - Max]	29.00 [26 - 30]

SD = Standard Deviation; Min = Minimum; Max = Maximum

Table 3. Definition of the microstructural variables.

Measure	Definition
Duration	Duration of the sample in seconds
Tokens	Total number of words produced
Mean length of utterance (MLU)	Average number of words per utterance
Propositional density	Number of verbs, adjectives, adverbs, prepositions and conjunctions div by the total number of words
Words per minute	Total number of tokens divided by the duration (converted from seconds minute)
Verbs per utterance	Average number of verbs (verbs, copulas, auxiliaries followed by past o present participles) per utterance.
Open/closed class ratio	Ratio of open class words (all nouns, verbs, copulas, adjectives and adve divided by closed class words (all other words)
Noun/verb ratio	Ratio of nouns to verbs, excluding auxiliaries and modals
Moving Average Token-Type Ratio (MATTR)	Average of estimated Token-Type Ratios for successive nonoverlapping successive windows of flixed length
% Correct information units (CIUs)	Total number of words relevant to the stimulus and informative (CIUs) divided by the total number of words
CIU per minute	Total number of CIUs divided by the duration (converted from seconds minute)

Note. Data derived from the CLAN software (MacWhinney et al., 2010).

Table 4. Frequency for each Main Concept.

Main Concepts	Frequency			
	Test		Retest	
	n	%	n	%
1. Le père marie une femme avec deux filles. *Dad remarried a woman with two daughters.*	16	37,2	18	41,9
2. Cendrillon vit avec sa belle-mère/ses belles-sœurs. *Cinderella lives with stepmother/stepsisters.*	22	51,2	25	58,1
3. La belle-mère et les demi-sœurs étaient méchantes avec Cendrillon. *Stepmother/stepsisters were mean to Cinderella.*	32	74,4	28	65,1
4. Cendrillon était la servante de la belle-mère et des demi-sœurs. *Cinderella was a servant.*	30	65,1	23	53,5
5. Cendrillon doit faire le ménage. *Cinderella has to do the housework.*	30	69,8	31	72,1
6. Le roi pense que le prince devrait se marier. *The king thinks the prince should get married.*	24	55,8	25	58,1
7. Le roi annonce qu'il va y avoir un bal en l'honneur de son fils qui doit trouver une épouse. *King announces there is going to be a ball in honor of son who needs to find a wife.*	33	76,7	34	79,1
8. Elles ont eu une invitation au bal. *They got an invitation to the ball.*	22	51,2	22	51,2
***9. Elles sont excitées à l'idée d'aller au bal.** *They are excited about the ball.*	10	23,3	5	11,6
10. La belle-mère dit à Cendrillon qu'elle ne peut pas aller au bal à moins que/parce que *insérer la raison* *Cinderella is told by the stepmother she cannot go to the ball unless/because (insert reason).*	32	74,4	29	67,4
***11. Les demi-sœurs abîment la robe de Cendrillon.** *The stepsisters tore Cinderella's dress.*	13	30,2	12	27,9
***12. La belle-mère et les belles-sœurs sont allées au bal.** *Stepmother/stepsisters went to the ball.*	14	32,6	14	32,6
13. Cendrillon était triste. *Cinderella was upset.*	20	46,5	14	32,6
14. Une fée marraine est apparue à Cendrillon. *A fairy godmother appeared to Cinderella.*	29	67,4	30	69,8
15. La fée marraine fait en sorte que {éléments} se transforment en {éléments}. *The fairy godmother makes {item(s)} turn into {items}.*	29	67,4	30	69,8
16. La fée marraine fait de Cendrillon une belle princesse. *The fairy godmother makes Cinderella into a beautiful princess.*	38	88,4	38	88,4

17. Cendrillon est allée au bal en carrosse. *Cinderella went to the ball in the coach.*	36	83,7	37	86,0
18. Elle savait qu'elle devait être à la maison parce que tout va se retransformer à minuit. *She knew she had to be home by midnight because everything will turn back at midnight.*	39	90,7	39	90,7
19. Le prince et Cendrillon ont dansé dans la salle/toute la nuit/sans personne d'autre. *The prince and Cinderella danced around the room/all night/with no one else.*	28	65,1	31	72,1
20. Le prince tombe amoureux de Cendrillon. *Prince falls in love with Cinderella.*	20	46,5	17	39,5
21. Cendrillon a réalisé qu'il est minuit. *Cinderella realized it is midnight.*	34	79,1	34	79,1
22. Elle a descendu les escaliers. *She ran down the stairs.*	40	93,0	36	83,7
23. En courant dans les escaliers, elle a perdu une de ses pantoufles de verre. *As she was running down the stairs she lost one of her glass slippers.*	40	93,0	42	97,7
24. Le prince trouve la chaussure de Cendrillon. *The prince finds Cinderella's shoe.*	15	34,9	18	41,9
25. Tout retourne à sa forme originale. *Everything turns back to its original form.*	17	39,5	12	27,9
26. Elle est rentrée à la maison à temps. *She returned home in time.*	27	62,8	16	37,2
27. Le prince fait du porte à porte pour trouver Cendrillon *The prince searched door to door for Cinderella.*	41	95,3	40	90,7
28. Le prince vient à la maison de Cendrillon. *Prince comes to Cinderella's house.*	17	39,5	20	46,5
29. Les demi-sœurs essayent la pantoufle de verre. *The stepsisters try on the glass slipper.*	15	34,9	16	37,2
30. La pantoufle ne faisait pas aux demi-sœurs. *The slipper didn't fit the stepsisters.*	18	41,9	17	39,5
31. Il a mis la pantoufle au pied de Cendrillon. *He puts the slipper on Cinderella's foot.*	19	44,2	21	48,8
32. La pantoufle convient parfaitement à Cendrillon. *The slipper fits Cinderella perfectly.*	34	79,1	31	72,1
33. Cendrillon et le prince se sont mariés. *Cinderella and the prince were married.*	36	83,7	29	67,4
34. Cendrillon et le prince vécurent heureux pour toujours. *Cinderella and the prince lived happily ever after.*	30	69,8	28	65,1

*MCs 9, 11 and 12 were produced by less than 33% of the sample and were not considered in the statistical analyses.

Table 5. Descriptive statistics of the Main Concepts and microstructural variables. Statistical testing used Wilcoxon signed-rank test for paired samples ('V' = test statistic; p = p value) comparing test and retest and Spearman's correlation assessing the strength of association between test and retest.

Variable	**Test (n=43)**		**Retest (n=43)**		**Statistics**		**Interpretation**
	Mean (SD)	Median [Min - Max]	Mean (SD)	Median [Min - Max]	V (p value)	Spearman' rho (p value)	
Main Concepts codes							
MC Composite	55.4 (15.95)	57 [2 – 80]	53.2 (15.59)	55 [6 – 77]	355.0 (p=.154)	0.644 (p<.001)	No systematic difference, moderate relationship between sessions.
AC	15.7 (5.12)	16 [0 – 26]	14.8 (5.21)	16 [1 – 25]	316.5 (p=.090)	0.646 (p<.001)	No systematic difference, moderate relationship between sessions.
AI	1.2 (0.85)	1 [0 – 4]	1.2 (1.08)	1 [0 - 4]	214.0 (p=.790)	0.286 (p=.063)	No systematic difference, weak relationship between sessions.
IC	2.7 (1.91)	3 [0 - 8]	3.1 (2.24)	3 [0 - 12]	317.5 (p=.168)	0.535 (p<.001)	No systematic difference, moderate relationship between sessions.
II	0.4 (0.54)	0.0 [0 – 2]	0.1 (0.41)	0.0 [0 - 2]	42.0 (p=.007)	-0.070 (p=0.655)	Significant difference between sessions, weak relationship between session.
AB	10.9 (5.52)	10 [3 - 30]	11.7 (5.39)	11 [4 - 28]	490.5 (p=.159)	0.640 (p<.001)	No systematic difference, moderate relationship between sessions.

Variable	Test (n=43)		Retest (n=43)		Statistics		Interpretation
	Mean (SD)	Median [Min - Max]	Mean (SD)	Median [Min - Max]	V (p value)	Spearman' rho (p value)	
Microstructural variables							
Duration (seconds)	184.5 (74.86)	186 [21 - 423]	180.9 (63.33)	174 [50 - 395]	410.5 (*p* = .608)	0.722 (*p* < .001)	No systematic difference, strong relationship between sessions.
Tokens	758.2 (331.44)	688 [123 - 1937]	737.9 (301.99)	685 [43 - 1843]	398.0 (*p* = .365)	0.765 (*p* < .001)	No systematic difference, strong relationship between sessions.
MLU (words)	14.05 (1.86)	13.94 [10.38 - 19.65]	13.69 (2.98)	13.33 [8.5 - 21.25]	412.0 (*p* = .461)	0.105 (*p* = .504)	No systematic difference, very weak relationship between sessions.
Propositional Density	0.50 (0.03)	0.50 [0.42 - 0.54]	0.50 (0.03)	0.50 [0.43 - 0.56]	414.000 (*p* =.476)	0.722 (*p* < .001)	No systematic difference, strong relationship between sessions.
Words per minute	250.76 (45.77)	248.28 [116.13 - 351.43]	246.00 (46.83)	241.17 [124.09 - 392.40]	410.0 (*p* =.447)	0.722 (*p* <.001)	No systematic difference, strong relationship between sessions.
Verbs per utterance	2.31 (0.73)	2.33 [0.52 - 4.28]	2.27 (0.83)	2.38 [0.51 - 4.09]	465.0 (*p* =.923)	0.503 (*p* =.001)	No systematic difference, moderate relationship between sessions.
Open/closed class ratio	1.16 (0.10)	1.16 [0.89 - 1.37]	1.16 (0.09)	1.15 [1.02 - 1.39]	475.500 (*p* =.976)	0.165 (*p* =.289)	No systematic difference, very weak relationship between sessions.
Noun/verb ratio	1.76 (0.71)	1.56 [1.04 - 4.93]	1.79 (0.64)	1.61 [1.00 - 3.91]	488.500 (p=.644)	0.722 (p<.001)	No systematic difference, strong relationship between sessions.

Variable	Test (n=43)		Retest (n=43)		Statistics		Interpretation
	Mean (SD)	Median [Min - Max]	Mean (SD)	Median [Min - Max]	V (p value)	Spearman' rho (p value)	
MATTR	0.95 (0.01)	0.95 [0.93 - 0.98]	0.96 (0.01)	0.96 [0.94 - 0.98]	633.500 (p<.001)	0.446 (p=.003)	No systematic difference, weak relationship between sessions.
% CIU	58.99 (8.40)	57.24 [52.17 - 95.55]	59.20 (8.44)	57.45 [52.91 - 96.01]	485.000 (p=.885)	0.239 (p=.122)	No systematic difference, very weak relationship between sessions.
CIU per minute	145.74 (22.13)	141.99 [105.96 - 217.14]	143.80 (25.14)	141.11 [99.69 - 252.00]	426.000 (p=.570)	0.722 (p<.001)	No systematic difference, strong relationship between sessions.

SD = Standard Deviation; MC Composite = Main Concept total composite score; AC = Accurate and Complete; AI = Accurate and Incomplete; IC = Incorrect and Complete; II = Incorrect and Incomplete; AB = Absent; MLU = Mean Length of Utterances; CIU = Correct Information Units; MATTR = Moving-Average Type-Token Ratio.

Table 6. Summary of test-retest results.
Koo and Li (2016) gives the following suggestion for interpreting intraclass correlation coefficient (ICC). including confidence intervals: below 0.50 = poor; between 0.50 and 0.75 = moderate; between 0.75 and 0.90 = good; and above 0.90 = excellent.

Measure	**ICC**	**95% CI Low - High**	**Koo & Li (2016) ICC Quality (CI Quality)**	**Spearman' rho**		**Absolute Value Difference Between Test and Retest**		**MDC90**
				r	*p* value	M (SD)	Range	
Main Concepts								
MC Composite	0.775	0.622 - 0.871	Good (Moderate - Good)	0.644	< 0.001	8.77 (6.04)	1 - 23	17.40
AC	0.707	0.521 - 0.830	Good (Moderate - Good)	0.646	< 0.001	3.23 (2.34)	0 - 10	5.71
AI	0.213	-0.096 - 0.483	Poor	0.286	0.063	0.86 (0.86)	0 - 1	1.07
IC	0.563	0.323 - 0.736	Moderate (Poor - Moderate)	0.535	< 0.001	1.47 (1.32)	0 - 4	2.30
II	0.132	-0.127 - 0.391	Poor	-0.070	0.655	0.47 (1.32)	0 - 4	0.55
AB	0.790	0.644 - 0.880	Good (Moderate - Good)	0.640	< 0.001	2.86 (2.12)	0 - 8	6.02
Microstructural								
Duration (seconds)	0.806	0.670 - 0.890	Good (Moderate - Good)	0.722	0.001	35.19 (25.18)	0.00 - 102	76.32
Tokens	0.791	0.646 - 0.881	Good (Moderate - Good)	0.765	< 0.001	153.74 (116.97)	3 - 468	349.05
MLU (words)	0.147	-0.160 - 0.427	Poor (Poor)	0.105	0.504	2.69 (1.80)	0.07 - 7.07	2.74
Propositional Density	0.538	0.284 - 0.721	Moderate (Poor - Moderate)	0.722	< 0.001	0.02 (0.02)	0.00 - 0.10	0.03
Words per minute	0.747	0.579 - 0.854	Good	0.722	< 0.001	26.30 (20.07)	0.43 - 77.26	51.01

Measure	ICC	95% CI Low - High	Koo & Li (2016) ICC Quality (CI Quality)	Spearman' rho *r*	Spearman' rho *p* value	Absolute Value Difference Between Test and Retest M (SD)	Absolute Value Difference Between Test and Retest Range	MDC90
			(Moderate - Good)					
Verbs per utterance	0.566	0.322 - 0.740	Moderate (Poor - Moderate)	0.503	0.001	1.78 (0.81)	0.09 - 3.55	0.86
Open/closed ratio	0.165	-0.146 - 0.444	Poor (Poor)	0.165	0.289	0.10 (0.07)	0.01 - 0.31	0.10
Noun/verb ratio	0.675	0.472 - 0.810	Moderate (Poor - Good)	0.722	< 0.001	0.34 (0.43)	0.00 - 1.76	0.74
MATTR	0.343	0.043 - 0.585	Poor (Poor - Moderate)	0.446	0.003	0.01 (0.01)	0.00 - 0.03	0.01
% CIU	0.929	0.873 - 0.961	Excellent (Good - Excellent)	0.239	0.122	2.32 (2.17)	0.06 - 11.33	9.26
CIU per minute	0.742	0.571 - 0.851	Moderate (Moderate - Good)	0.722	< 0.001	13.19 (10.86)	0.28 - 35.57	26.08

SD = Standard Deviation; CI = Confidence Interval; MC_{total} = Main Concept total score; AC = Accurate and Complete; AI = Accurate and Incomplete; IC = Incorrect and Complete; II = Incorrect and Incomplete; AB = Absent; MLU = Mean Length of Utterances; CIU = Correct Information Units; MATTR = Moving-Average Type-Token Ratio; MDC90= Minimal Detectable Change at 90% confidence.

Figure 1.

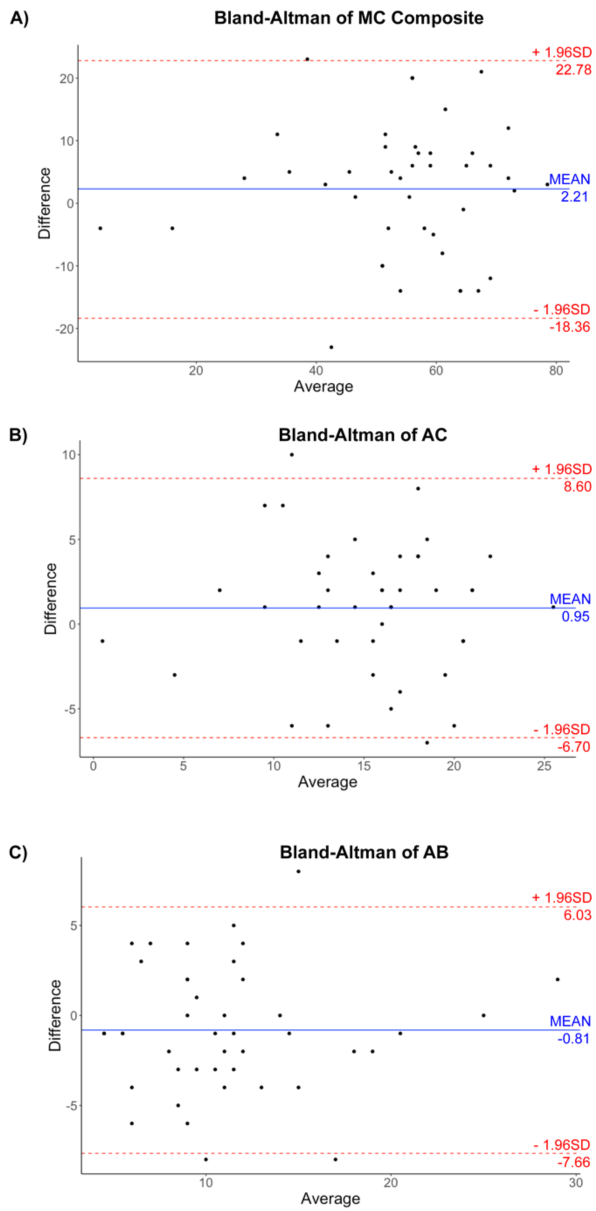

Figure 2.