# Choosing Discourse Outcome Measures to Assess Clinical Change

**Mary Boyle, Ph.D., CCC-SLP, BC-ANCDS[1]**

## ABSTRACT

Surveys of speech-language pathologists who work with people with aphasia indicate that they view the large number of existing measures to be a barrier to using discourse analysis in their practice. This article provides a process that can help determine whether a particular discourse outcome measure might be useful with a particular client. The process involves answering questions about the client, the treatment, the work setting, and the psychometric properties of the discourse outcome measure in question. By following this systematic process, clinicians can eliminate outcome measures that are not likely to provide useful data and can focus on those that can help them demonstrate treatment-related change.

**KEYWORDS:** aphasia, discourse, outcomes

**Learning Outcomes:** As a result of this activity, the reader will be able to (1) discuss practical factors to consider when choosing a discourse outcome measure to use in treatment; (2) discuss psychometric properties to consider when choosing a discourse outcome measure to use in treatment; (3) and explain why it is important to know about an outcome measure's stability before using it to measure treatment-related change.

The desired outcome of therapy for people with stroke-induced aphasia is an improvement in communication ability. Ideally, this improvement should have a noticeable impact on a person's everyday communication activities. Because discourse activities are the most common communication activities for adults with aphasia,[1] it is not surprising that discourse is increasingly a target of clinical treatment and research.[2–6] There have been calls by some researchers to establish a core set of discourse outcome measures to be used across studies,[5] but to date no consensus about a group of measurements has been reached. Because there is no agreed-upon core set of discourse outcome measures, researchers often develop new

[1]Department of Communication Sciences and Disorders, Montclair State University, Bloomfield, New Jersey.

Address for correspondence: Mary Boyle, Ph.D., CCC-SLP, BC-ANCDS, Department of Communication Sciences and Disorders, Montclair State University, 1515 Broad Street, Bloomfield, NJ 07003 (e-mail: boylem@montclair.edu).

measures tailored to the aims of their particular study.[5] One result has been the proliferation of discourse outcome measures. Bryant et al identified 536 different discourse outcome measures in the research literature, and Pritchard and her colleagues identified 58 measures that focused solely on the information content of discourse.[2,6] Bryant et al speculated that the existence of such a large number of measures might cause confusion regarding selection of appropriate measures for individual clients.[3] The results of their survey confirmed that, indeed, clinicians viewed the selection of appropriate discourse outcome measures for specific clients as a barrier to using discourse analysis in their practice.

It is beyond the scope of this article to offer guidance about each of the hundreds of discourse outcome measures that have been reported.[2,6] Speech-language pathologists can learn about discourse outcome measures by reviewing the professional literature, participating in professional continuing education offerings, and reviewing materials available from publishers of assessment and treatment materials. This article aims to provide a general process that clinicians can use to determine whether a particular discourse outcome measure might be a reasonable choice for a particular client. The process involves principles of evidence-based practice, the model of health and disability provided by the International Classification of Functioning, Disability, and Health (ICF),[7] and psychometric properties of the measure under consideration.

Let us start with the assumption that the clinician has followed the principles of evidence-based practice to integrate the needs and perspectives of the client and/or other recipient of the treatment (such as an important communication partner) with assessment data concerning the client's aphasia and its effect on communication. Using this information, the clinician has developed a treatment plan to achieve client-centered goals, choosing a treatment approach that is within his or her clinical expertise to carry out and that is acceptable to the client and other treatment participants. Let us also assume that the clinician is aware of the external scientific evidence associated with each treatment option and has chosen the one with the highest level of evidence that is also com-

patible with the client's preferences and the clinician's own expertise.[8]

## A PROCESS FOR CHOOSING A DISCOURSE OUTCOME MEASURE

Once a treatment plan to achieve the client's goals has been formulated, the clinician can consider a series of questions (see Table 1) to determine which discourse outcome measure or measures might be best for a particular client.

### Questions Related to the Client, Treatment, and Work Setting

1. *What aspect or level of discourse are you expecting the treatment to improve?* Because the primary purpose of an outcome measure is to demonstrate improvement following treatment, the outcome measure should be aligned with the focus of treatment. Some treatments concentrate on improving the microstructural level of discourse, such as words, phrases, clauses, or sentences, thus emphasizing the lexical, semantic, or syntactic aspects of language.[6,9] So, for example, if the treatment is focused on improving word-retrieval ability, an outcome measure that would show improvement in that area, such as an increase in the number of words produced per minute,[10] a reduction in the occurrence of word-finding behaviors,[11] or an improvement in the percentage of words that convey accurate, relevant information[10] might be a logical outcome measure to choose. If the treatment is focused on improving utterance production, then an outcome measure that could demonstrate an increase in the percentage of complete utterances[12] or in the number of embedded clauses per utterance[13] might be considered, depending on the exact goal(s) of the treatment.

The macrostructural level of discourse is concerned with its overall meaning, with the way meaning is organized within the discourse, and with its social or interpersonal purpose.[6,9] If treatment is expected to improve discourse production at the macrostructural level, then outcome measures focused on the adequacy of cohesive ties between utterances,[14] elements of story grammar,[15,16] or turn-taking interchanges[17] might be expected to reveal improvement depending on the specific focus of the treatment.

**Table 1    A Process for Choosing a Discourse Outcome Measure**

**Questions related to the client, treatment, and work setting**

1 What aspect or level of discourse are you expecting the treatment to improve?

    a. Microstructure

    b. Macrostructure

    c. Does the discourse genre of the outcome measure match the genre that you plan to use in treatment?

2 Do you expect that improving discourse might result in changes in activity, participation, or quality of life?

    a. Consider aphasia-specific patient-reported outcome measures

    b. Does your client have family/social support crucial for change at this level?

3 Can you implement the outcome measure in your workplace?

    a. Do you have access to the materials necessary to administer the outcome measure?

    b. Do you have time in your workday to analyze the discourse according to the outcome measure's
protocol?

4 Is there evidence that the discourse outcome measure is relevant for people similar to your client?

    a. Has the measure been used with people who have aphasia?

    b. Has the measure been used with people whose aphasia is similar in severity and type to your client's
aphasia?

**Questions related to the psychometric properties of the discourse outcome measure**

1 Is there evidence concerning the scoring reliability of the outcome measure?

    a. Are there reports that intra-rater and inter-rater reliability with the outcome measure is 0.70 or better?

2 Is there evidence concerning the stability (test–retest reliability) of the outcome measure?

    a. Is the test–retest reliability coefficient at least 0.90?

    b. Is the standard error of measurement (SEM) reported?

     i. If so, add and subtract the SEM from your client's score. If the resulting range does not include your
client's pretreatment score, it is likely that treatment has truly changed your client's performance.

    c. Is the minimal detectable change (MDC) value reported?

     i. If so, compare how much your client changed on the measure pre- to posttreatment. If this value is
equal to or larger than the MDC value, it is likely that treatment has truly changed your client's performance.

3 Is there evidence that the outcome measure is responsive to change?

    a. If an effect size or an MDC is reported, you can use this to gauge whether your client's change on the
outcome measure is likely to reflect treatment-related change.

Another aspect that will influence the choice of an outcome measure is the genre of discourse that is targeted in treatment. Genre refers to different ways of using language for a particular purpose that are shared within a culture. Different discourse genres are marked by different words and structures.[9] Some discourse outcome measures are specific to a particular genre of discourse (e.g., story grammar is used to analyze narrative discourse), whereas other outcome measures might be used across discourse genres (e.g., adequacy of cohesive ties; the number of words produced per minute). Although some outcome measures can be used across genres, you cannot necessarily make valid comparisons about performance on one of those measures across genres because of different requirements from one genre to another. For example, although increasing the number of words produced per minute might be a desired outcome in a story retell (because such an increase could provide more information more efficiently), increasing the number of words produced per minute in a procedural discourse might result in unnecessarily long, complicated, or rapid instructions, which would not be positive changes.

Finally, some discourse outcome measures were specifically designed to be used with particular elicitation stimuli and cannot be applied when discourses are elicited with other stimuli. For example, main concepts analysis, which analyzes both microstructural (specific words) and macrostructural (concepts essential

to convey the gist of a story) discourse elements, was developed by Nicholas and Brookshire to be used with a specific set of elicitation stimuli.[18] Richardson and Dalton developed a different list of main concepts that can be used with a different set of elicitation stimuli.[19,20] Because the lists of the main concepts in these two analysis schemes contain vocabulary that is specific to each different stimulus set, they cannot be used with any other stimuli than those for which they were developed. Obviously, in these cases, it would be important for the clinician to have access to the elicitation stimuli that were used to develop the main concepts list.

2. *Do you expect that improving discourse might result in changes at the level of activity, participation, or quality of life?* Thus far, the emphasis has been on outcome measures that might demonstrate change at the body function, or impairment level.[7] Barak and Duncan stated that "measurement of recovery at just one level gives only a partial picture of the recovery process."[21] Kagan and colleagues pointed out that we should also try to capture the ways that changes in the impairment lead to changes in participation, confidence, or quality of life.[22] They developed a framework, based on the ICF model, to illustrate this idea in aphasia. The overlapping circles of the

*Living with Aphasia: Framework for Outcome Measurement* (A-FROM; Fig. 1) indicate that working in one domain of the model is likely to impact other domains of the model. For example, improving a client's ability to convey information in discourse could increase participation in conversations with family members and friends. It might also increase the client's confidence during conversations.

Outcome measures that can capture changes in activities, participation, attitudes, and quality of life are usually patient-reported outcome measures (PROMs), meaning that the client (or patient) completes the outcome tool. There are a variety of PROMs that have been developed specifically for people with aphasia, including the Aphasia Communication Outcome Measure (ACOM),[23] the Assessment for Living with Aphasia—second edition (ALA-2),[24] the Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39),[25] and the Communication Confidence Rating Scale for Aphasia (CCRSA).[26] Including an outcome measure that assesses change at the activity and participation level can highlight the way that treatment has an impact on real-life activities and situations. However, it is important to consider a client's individual situation when deciding
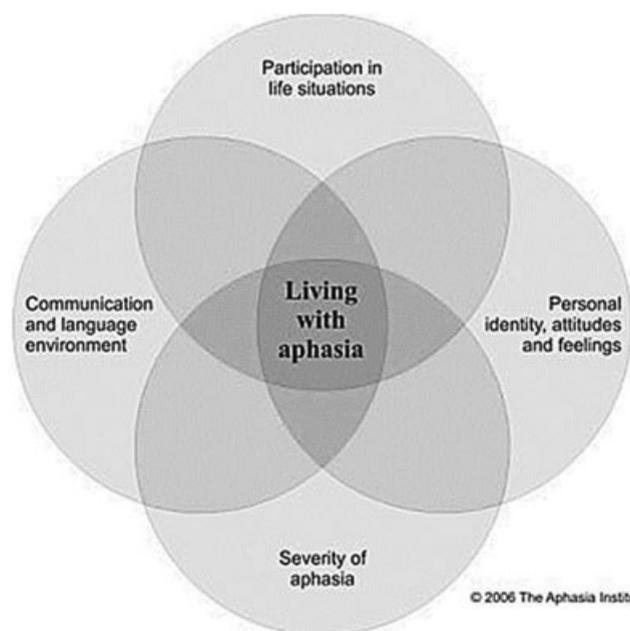


**Figure 1** Living with aphasia: framework for outcome measurement (A-FROM). Reprinted with permission from the Aphasia Institute.

whether a particular PROM might capture treatment-related changes. For example, the presence of family or social support can influence whether impairment-level changes translate into changes at the activity or participation level. That is, a client's discourse production might improve in treatment, but without someone at home or in the community to talk to daily, that improvement might not result in improved activities or participation for that client.[21]

3. *Can you implement the outcome measure in your work setting?* Some practical factors should be considered as you choose a discourse outcome measure. These include ability to access the materials necessary to complete the measure and availability of time to elicit and analyze the discourse sample. In terms of accessing the materials, you should determine whether you or your employer has the funds to purchase the necessary materials if the measure is not freely available. Likewise, think about whether you can complete the procedures to elicit and analyze the discourse in the time allowed for a diagnostic session in your work setting. Some discourse outcome measures require that the discourse be recorded and transcribed before it can be analyzed, and this may take more time than most clinicians have in their daily schedules. For other measures, the analysis can be done during the elicitation of the discourse or while listening to a recording, so that transcribing the discourse is not required. For example, Hula and colleagues reported that there was good reliability between transcription-based scoring of the Story Retell Procedure and scoring from an audiorecording only, without transcription.[27] In this issue, Dalton et al review other non–transcription-based discourse analysis methods.[28]

4. *Is there evidence that the discourse outcome measure is relevant for people similar to your client?* There are several aspects of this question to consider. First, are there reports demonstrating that the outcome measure has been used successfully with people who have aphasia? Some discourse outcome measures might have been developed to assess discourse in people who sustained traumatic brain injuries or right-hemisphere cerebrovascular accidents. If that is the case, are there also studies that used the measure with people who have aphasia? Al-though all three groups have discourse impairments, the way that discourse is impaired differs markedly among them.[29] Choosing a measure that has been used to assess the discourse of people with aphasia improves the likelihood that the measure will be relevant for your client.

A second aspect of this question pertains to outcome measures that assess quality of life. Often, people with aphasia have been excluded from participating in research to assess quality of life following stroke.[25] This means that measures developed to assess how stroke affects quality of life may not include questions pertinent to aphasia and may be too linguistically difficult for people with aphasia to complete. Choosing a quality-of-life measure that was developed specifically for people with aphasia, like those mentioned earlier in this article, improves the chances that the measure will allow you to show change related to aphasia treatment.

Finally, if the discourse outcome measure has been used with people who have aphasia, how similar are those people to your client? If your client has severe aphasia but the participants in studies using the measure all had mild aphasia, it might not be the best choice for your client. If your client has Wernicke's aphasia but all of the participants in studies using the measure had agrammatic Broca's aphasia, you should think critically about whether it would provide you with useful information about a change in your client's fluent discourse production. It is the responsibility of the clinician rather than of the developers of the outcome measure to make judgments about the suitability of the measure for his or her own purpose or clients.[30]

## Questions Related to the Psychometric Properties of the Discourse Outcome Measure

Clinicians need to be as concerned as researchers about the psychometric robustness of an outcome measure. Clinicians use outcome measures to establish pretreatment performance and to assess treatment-related change. We want to be confident that an increased score on an outcome measure reflects real change and is not a result of error in the measure. Unless an outcome measure's psychometric properties have been established, we cannot be confident about this.[2,4,6,21,30]

The development of a psychometrically sound measure is a long process and may involve more than one study.[30] For example, the CCRSA was introduced in a paper published in 2010 and its psychometric properties were established in two follow-up papers.[26,31,32] The Story Retell Procedure was introduced in a paper published in 1998 and various aspects of its psychometric properties were established in five subsequent publications.[27,33–37] It is important to be aware, however, that sometimes discourse outcome measures appear in refereed journals despite the fact that there is little or no information about their psychometric properties, either in the original paper describing the outcome measure or in subsequent papers.[6,38] For these reasons, clinicians should be aware of the existence or absence of information about the psychometric properties of an outcome measure.[30] The psychometric properties of an outcome measure most often identified as being important for deciding on its clinical use are reliability and responsiveness to change.[6,21,30]

1. *Is there evidence concerning the scoring reliability of the outcome measure?* Scoring reliability refers to whether an outcome measure can be scored consistently across scoring attempts and across scorers. It includes intra-rater and inter-rater reliability. Intra-rater reliability is the ability of one rater to score an outcome measure from an individual in the same way on two different occasions without referring back to the earlier scoring outcome. Inter-rater reliability is the ability of two raters to independently score a person's results in the same way. Intra-rater and inter-rater reliability is reported as reliability coefficients that range from 0 to 1. Generally, reliability coefficients less than 0.40 are considered weak, between 0.40 and 0.70 are considered moderate, and above 0.70 are considered strong.[39] Good scoring reliability (i.e., a reliability coefficient of 0.70 or better) is generally a reflection of a well-described, clear, and detailed protocol for administering and scoring the outcome measure. This contributes to confidence that the measure will be administered and scored consistently, thus minimizing measurement error.

2. *Is there evidence concerning the stability of the outcome measure?* The stability, also called "test–retest reliability" or "session-to-session stability," of an outcome measure refers to whether it produces the same result on repeated applications when the person being assessed has not changed on the domain or behavior being measured.[40] It is important to establish the stability of an outcome measure to provide confidence that changes on the measure are related to treatment rather than to spurious, day-to-day variability inherent either in the measure itself or in the behavior that it is measuring, which is frequently more variable in clinical populations than in neurologically healthy individuals.[41,42] For example, the ability to retrieve words to produce discourse may vary from one day to the next because of variations in the speaker's physiologic and cognitive states—he or she may be more tired or more distracted on one day than another—and this variability, leading to a change in score, might be misinterpreted as change due to treatment. If we know the amount by which an outcome measure varies when there has been no change in the behavior being measured, then we know that a larger change is necessary to be confident that it is a true change that resulted from treatment.

The stability is reported as a reliability coefficient that ranges from 0 to 1. The statistic used to calculate the test–retest reliability coefficient will depend on the kind of data the study is analyzing. For example, the weighted Kappa statistic might be used for categorical data and the intraclass correlation coefficient might be used for continuous data. Generally, values less than 0.50 indicate poor stability, values between 0.50 and 0.75 indicate moderate stability, values between 0.75 and 0.90 indicate good stability, and values above 0.90 indicate excellent stability.[43] Fitzpatrick and colleagues recommended that an outcome measure should have a minimum test–retest reliability coefficient of 0.90 if it will be used to make clinical decisions about changes in an individual's performance, because confidence intervals around an individual's score are wide at reliability levels less than 0.90.[40] Confidence intervals indicate the range of values within which an individual's true score lies at a particular level of confidence (e.g., confidence 90% of the time). Smaller confidence intervals indicate higher precision and less error.[30]

Two other values are very useful when making decisions about whether a client has truly changed on an outcome measure. Donoghue and Stokes[44] suggested that it is better to use the standard error of measurement (SEM), which indicates how much a score varies randomly on repeated measurements, for clinical decisions rather than the value of the reliability coefficient. The SEM can be used to calculate the minimal detectable change (MDC) value. The MDC estimates the change in score on an outcome measure necessary to be confident that the change is a real one and not simply a reflection of measurement error.[45] Studies that report the MDC provide clinicians with a means to objectively determine whether a client's change in score is large enough to be considered a true improvement rather than random day-to-day variability of the behavior. If a client's score changes by an amount equal to or greater than the MDC reported for the outcome measure, a clinician can be fairly confident that treatment, rather than random variability, is responsible for the change.

Occasionally, a study will report a value called the "minimally important difference" (MID) or the minimally clinically important difference (MCID). This represents the smallest change on an outcome measure that would be considered important by a client or a clinician. Although this sounds like it would be a useful value, there is no standard method to derive an MCID, and this has led to problems in interpreting such values.[21,46] Establishing agreed-upon methods for deriving and interpreting the MCID is an area of current research in stroke outcomes research.[21]

In summary, when reviewing an outcome measure that has been reported as part of a group study, a clinician can use the MDC value to assess whether a client's change in score represents real change. If the MDC value is not reported but SEMs are reported, a clinician can add and subtract the SEM from a client's score to get a range that includes the client's true performance. If the range of scores obtained in this way does not include the client's pretreatment score on the outcome measure, then it is likely that the client has truly changed in the behavior in question. If neither the MDC nor the SEM is reported, but the study reports that the test–retest reliability coefficient for the outcome measure is greater than 0.90, the clinician can assume that the measurement error is probably small, but the clinician still has no way of knowing objectively whether the client's change in score represents true change.

3. *Is there evidence that the outcome measure is responsive to change?* Responsiveness is the sensitivity of the measure to change over time, and so may indicate the effects of treatment.[21] Responsiveness is a component of an outcome measure's validity, and it is important if the measure is going to be used to evaluate whether treatment caused a change in score on the measure.[30] The responsiveness of an outcome measure quantifies the magnitude of change, which is often reported as an effect size.[21,30] Effect size is a statistical calculation that provides information about the magnitude (e.g., small, medium, or large) of an effect. Calculated effect sizes can transform different scales of measurement to a common scale, so that they can then be compared with each other. An outcome measure that has a large effect size would be considered more responsive to change than an outcome measure that has a medium or small effect size. The MDC, discussed earlier, could also be considered an indicator of an outcome measure's responsiveness.

## SUMMARY AND CONCLUSIONS

The large number of discourse outcome measures reported in the literature can be overwhelming to a clinician who is trying to choose a measure to use with his or her client. However, this article provides a series of questions that clinicians can ask about an outcome measure to determine whether it might be useful. The questions include practical considerations about what aspect of discourse production the chosen treatment is expected to change and whether the time and resources necessary to apply the discourse outcome measure are available in the clinician's workplace. Since the outcome measure will be used to assess whether treatment resulted in a change in discourse production, psychometric properties of the outcome measure, such as its reliability and responsiveness, should also be considered. As with most things related to clinical practice, staying current with

the discourse outcome measures reported in the literature, whether by reading professional journals or by pursuing continuing education activities, can improve a clinician's ability to make sound choices. Finally, it is always important to consider how closely the client resembles the participants in the study reporting the outcome measure. The more closely the client resembles the participants, the more likely it is that the measure will be a useful one for that client.

## REFERENCES

1. Davidson B, Worrall L, Hickson L. Identifying the communication activities of older people with aphasia: evidence from naturalistic observation. Aphasiology 2003;17:243–264

2. Bryant L, Ferguson A, Spencer E. Linguistic analysis of discourse in aphasia: a review of the literature. Clin Linguist Phon 2016;30(07):489–518

3. Bryant L, Spencer E, Ferguson A. Clinical use of linguistic discourse analysis for the assessment of language in aphasia. Aphasiology 2017;31:1105–1126

4. Linnik A, Bastiaanse R, Höhle B. Discourse production in aphasia: a current review of theoretical and methodological challenges. Aphasiology 2016;30:765–800

5. Dietz A, Boyle M. Discourse measurement in aphasia research: have we reached the tipping point? Aphasiology 2018;32:459–464

6. Pritchard M, Hilari K, Cocks N, Dipper L. Reviewing the quality of discourse information measures in aphasia. Int J Lang Commun Disord 2017;52(06):689–732

7. International Classification of Functioning, Disability, and Health: ICF. Geneva: World Health Organization; 2001

8. Evidence-based practice. American Speech-Language-Hearing Association. Available at: https://www.asha.org/research/ebp/evidence-based-practice/. Accessed August 10, 2019

9. Armstrong E. Aphasic discourse analysis: the story so far. Aphasiology 2000;14:875–892

10. Nicholas LE, Brookshire RH. A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. J Speech Hear Res 1993;36(02):338–350

11. Boyle M. Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. Am J Speech Lang Pathol 2004;13(03):236–249

12. Edmonds LA, Babb M. Effect of verb network strengthening treatment in moderate-to-severe aphasia. Am J Speech Lang Pathol 2011;20(02):131–145

13. Rochon E, Saffran EM, Berndt RS, Schwartz MF. Quantitative analysis of aphasic sentence production: further development and new data. Brain Lang 2000;72(03):193–218

14. Andreetta S, Marini A. The effect of lexical deficits on narrative disturbances in fluent aphasia. Aphasiology 2015;29:705–723

15. Coelho CA, Liles B, Duffy R, Clarkson J, Elia D. Longitudinal assessment of narrative discourse in a mildly aphasic adult. Clin Aphasiol 1994;17:145–155

16. Whitworth A. Using narrative as a bridge: linking language processing models with real-life communication. Semin Speech Lang 2010;31(01):64–75

17. Savage MC, Donovan NJ, Hoffman PR. Preliminary results from conversation therapy in two cases of aphasia. Aphasiology 2014;28:616–636

18. Nicholas LE, Brookshire RH. Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. J Speech Hear Res 1995;38(01):145–156

19. Richardson JD, Dalton SG. Main concepts for three different discourse tasks in a large non-clinical sample. Aphasiology 2016;30(01):45–73

20. Richardson JD, Dalton SG. Main concepts for two picture description tasks: an addition to Richardson and Dalton, 2016. Aphasiology 2019. Doi: 10.1080/02687038.2018.1561417

21. Barak S, Duncan PW. Issues in selecting outcome measures to assess functional recovery after stroke. NeuroRx 2006;3(04):505–524

22. Kagan A, Simmons-Mackie N, Rowland A, et al. Counting what counts: a framework for capturing real-life outcomes of aphasia intervention. Aphasiology 2008;22:258–280

23. Hula WD, Doyle PJ, Stone CA, et al. The aphasia communication outcome measure (ACOM): dimensionality, item bank, calibration, and initial validation. J Speech Lang Hear Res 2015;58(03):906–919

24. Kagan A, Simmons-Mackie N, Victor J, et al. Assessment for Living with Aphasia (ALA), —2nd ed. Toronto, ON: Aphasia Institute; 2013

25. Hilari K, Byng S, Lamping DL, Smith SC. Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39): evaluation of acceptability, reliability, and validity. Stroke 2003;34(08):1944–1950

26. Babbitt EM, Cherney LR. Communication confidence in persons with aphasia. Top Stroke Rehabil 2010;17(03):214–223

27. Hula WD, McNeil MR, Doyle PJ, Rubinsky HJ, Fossett TRD. The inter-rater reliability of the story retell procedure. Aphasiology 2003;17:523–528

28. Dalton SG, Hubbard HI, Richardson JD. Moving toward non-transcription based discourse analysis in stable and progressive aphasia. Seminars in Speech and Language 2020;41:32–44

29. Davis GA. Aphasia and Related Cognitive-Communicative Disorders. Boston, MA: Pearson; 2014

30. Jerosch-Herold C. An evidence-based approach to choosing outcome measures: a checklist for the critical appraisal of validity, reliability, and responsiveness studies. Br J Occup Ther 2005;68:347–353

31. Cherney LR, Babbitt EM, Semik P, Heinemann AW. Psychometric properties of the communication Confidence Rating Scale for Aphasia (CCRSA): phase 1. Top Stroke Rehabil 2011;18 (04):352–360

32. Babbitt EM, Heinemann AW, Semik P, Cherney LR. Psychometric properties of the Communication Confidence Rating Scale for Aphasia (CCRSA): Phase 2. Aphasiology 2011;25:727–725

33. Doyle PJ, McNeil MR, Spencer K, Goda AJ, Cotrell K, Lustig AP. The effects of concurrent picture presentations on retelling of orally presented stories by adults with aphasia. Aphasiology 1998;12:561–574

34. Doyle PJ, McNeil MR, Park G, et al. Linguistic validation of four parallel forms of a story retelling procedure. Aphasiology 2000;14:537–549

35. McNeil MR, Doyle PJ, Fossett TRD, Park GH, Goda AJ. Reliability and concurrent validity of the information unit scoring metric for the story retelling procedure. Aphasiology 2001;15:991–1006

36. McNeil MR, Doyle PJ, Park GH, Fossett TRD, Brodsky MB. Increasing the sensitivity of the Story Retell Procedure for the discrimination of normal elderly subjects from persons with aphasia. Aphasiology 2002;16:815–822

37. McNeil MR, Sung JE, Yang D, et al. Comparing connected language elicitation procedures in persons with aphasia: concurrent validation of the Story Retell Procedure. Aphasiology 2007;21:775–790

38. Dietz A, Boyle M. Discourse measurement in aphasia: consensus and caveats. Aphasiology 2018;32:487–492

39. Johnston I. I'll Give You a Definite Maybe: An Introductory Handbook on Probability, Statistics, and Excel. Available at: https://johnstoi.web.viu.ca/maybe/maybe4.htm. Accessed January 11, 2010

40. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assess 1998;2 (14):i–iv, 1–74

41. Brookshire RH, Nicholas LE. Speech sample size and test-retest stability of connected speech measures for adults with aphasia. J Speech Hear Res 1994;37(02):399–407

42. Herbert R, Hickin J, Howard D, Osborne F, Best W. Do picture-naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? Aphasiology 2008;22:184–203

43. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(02):155–163

44. Donoghue D, Stokes EK; Physiotherapy Research and Older People (PROP) group. How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. J Rehabil Med 2009;41(05):343–346

45. Stratford PW. Getting more from the literature: estimating the standard error of measurement from reliability studies. Physiother Can 2004;56:27–30

46. Cook CE. The minimal clinically important change score (MCID): a necessary pretense? J Man Manip Ther 2008;16:E82–E83