



Discourse measurement in aphasia: consensus and caveats

Aimee Dietz & Mary Boyle

To cite this article: Aimee Dietz & Mary Boyle (2018) Discourse measurement in aphasia: consensus and caveats, *Aphasiology*, 32:4, 487-492, DOI: [10.1080/02687038.2017.1398814](https://doi.org/10.1080/02687038.2017.1398814)

To link to this article: <https://doi.org/10.1080/02687038.2017.1398814>



Published online: 13 Nov 2017.



[Submit your article to this journal](#)



Article views: 427



[View related articles](#)



[View Crossmark data](#)



Citing articles: 5 [View citing articles](#)

REPLY



Discourse measurement in aphasia: consensus and caveats

Aimee Dietz^a and Mary Boyle^b

^aDepartment of Communication Sciences & Disorders, University of Cincinnati, Cincinnati, OH, USA;



^bDepartment of Communication Sciences & Disorders, Montclair State University, Bloomfield, NJ, USA

Our intention was to use this clinical forum to start a dialogue regarding the multitude of discourse outcome measures currently used in studies of aphasia. Based on the commentary papers, it is obvious that we are not alone in wondering whether it is time to consider a more organized process. Kintz and Wright (this issue) agree that we have reached a tipping point and that a more systematic approach is warranted. Several of the authors point to the benefits of collecting common data across studies. These potential benefits include (1) basing evidence for treatments and support for hypotheses regarding brain and communicative functions on larger amounts of data, thus yielding more robust conclusions than is currently the norm (Armstrong, [this issue](#); de Riesthal & Diehl, [this issue](#)); (2) elucidating best practices to improve communication (Kintz & Harris Wright, [this issue](#)); (3) allowing meaningful comparison of outcomes among treatment studies (Kurland & Stokes, [this issue](#)); and (4) achieving aims of people affected by aphasia, clinicians, and researchers to report the effects of a treatment on both language and communication (Wallace, Worrall, & Rose, [this issue](#)). Additionally, several of the authors cite timely and informative reviews (Bryant, Ferguson, & Spencer, 2016; Linnik, Bastiaanse, & Hohle, 2016; Pritchard, Hilari, Cocks, & Dipper, 2017) that were published shortly after we submitted our initial paper for this forum, which substantiate the proliferation of discourse outcome measurements that we described.

Despite this general consensus, each commentary paper raises concerns and caveats. The primary concerns seem to pertain to the issues of psychometric properties of the outcome measurements and the complexity of discourse, as well as elicitation, transcription, and coding methodology. We largely concur with these concerns. In this response, we will add our thoughts to those provided by the authors of the commentaries and conclude with suggestions for moving forward.

Psychometric properties and complexity of discourse

Armstrong ([this issue](#)) points out that few discourse measures have demonstrated high stability, reliability, and validity. Perhaps this is because, as De Riesthal and Diehl ([this issue](#)) noted, few researchers have ascertained and reported on the psychometric properties of the measures they have used. These are serious issues that require attention. Based on the large number of discourse measures reported in recent reviews (Bryant et al., 2016; Linnik et al., 2016; Pritchard et al., 2017) and

CONTACT Aimee Dietz  dietzae@ucmail.uc.edu  Department of Communication Sciences & Disorders, University of Cincinnati, PO Box 670379, 3202 Eden Avenue, Cincinnati, OH 45267-0379, USA

© 2017 Informa UK Limited, trading as Taylor & Francis Group

the relatively low levels of reliability and validity reporting cited therein (Pritchard et al., 2017), it seems that journal reviewers and editors are not questioning the use of these new, untested measures when making publication decisions. Consideration of a measure's reliability and validity seems especially crucial when it is used to report change as part of a treatment study. We are not opposed to the development of new discourse measures. Rather, we call on researchers to consider what they know about a measure's psychometric properties before stating that changes on the measure reflect true changes in discourse production. Likewise, we call on reviewers and editors to challenge the use of untested measures. One approach for a researcher who is interested in a new discourse measure might be to examine its psychometric properties using large discourse databases such as AphasiaBank (www.aphasia.talkbank.org) (MacWhinney, Fromm, Forbes, & Holland, 2011), assuming the appropriate genres are available for analysis. Another approach might be to do a pilot study using the measure to ascertain its stability, reliability, and validity before applying it in a treatment study. We acknowledge the additional burden this may impose on those managing the already complex task of developing and testing novel treatments. However, we cannot forget the fundamental importance of establishing the stability of a measure before using it to assess treatment-related change (Herbert, Hickin, Howard, Osborne, & Best, 2008). Failing to do so risks the misinterpretation of spurious changes related to a measure's instability as genuine changes caused by treatment (Brookshire & Nicholas, 1994). Wallace and colleagues ([this issue](#)) note that in order to be included in a core outcome set, a measure must have strong psychometric properties and they recommend that efforts be directed at standardizing and validating discourse outcome measures. We strongly agree with this recommendation.

Some of the commentary authors raise an even stronger cautionary note about measuring discourse, suggesting that the complexity of discourse creates uncertainty regarding measurement. For example, Armstrong questions whether something as innately variable and complex as discourse can, in fact, be quantified, while Whitworth ([this issue](#)) wonders whether discourse behavior can be reliably sampled and whether we can expect discourse to be stable over time. Several of the authors (Kintz & Wright; de Riesthal & Diehl; and Whitworth) remind us that we need a better understanding of the relationship between microlinguistic and macrolinguistic skills, and how changes in each contribute to communicative success. We agree that much work is needed to tease apart the factors associated with the multi-level approach to discourse analysis (Henderson, Kim, Kintz, Frisco, & Wright, 2017; Marini, Andreetta, del Tin, & Carlomagno, 2011; Marini, Boewe, Caltagirone, & Carlomagno, 2005). However, we argue that we won't be able to answer these questions unless and until we know more about the psychometric properties of the quantitative measures that are being used in studies of discourse. If psychometrically acceptable quantitative measures that provide meaningful information about discourse cannot be found or developed because discourse is too complex and variable to be measured in such a way, then researchers might need to consider using qualitative methodology. At the moment, we do not have enough information to answer these serious questions.

Elicitation, transcription, and coding methodology

The failure of some researchers to clearly explain elicitation, transcription, and coding protocols was another concern raised by some commentators. Kintz and Wright emphasize the fact that differences organizing discourse samples into utterances for subsequent analyses can result in different outcomes – even when the same error analysis is applied. They also remind us of the impact that sample length has on discourse outcome measures such as lexical diversity. These issues underscore the importance of establishing clear, consistent protocols for eliciting, transcribing, and coding discourse samples. Therefore, we urge editors and reviewers to require that authors include transparent descriptions of these protocols in their publications.

There is little argument regarding the necessity of making discourse analyses clinically feasible – especially the transcription and coding aspects of the task. Kurland and Stokes (as well as Whitworth and Kintz and Wright) rightly pointed out our seeming dismissal of the challenges involved in applying research-based discourse measures to clinical settings. We are acutely sensitive to this need; however, in our Target paper, we were attempting to focus the discussion on the need to improve the state of the science in discourse measurement and not dwell on the barriers. Bryant and her colleagues' recent report (2016) underscores the importance of offering clinically feasible solutions to transcription and analysis roadblocks for clinicians. Armstrong suggests that our use of computer technology will be key to examining discourse beyond word level informational units (e.g., correct information units) or efficiency measures such as words per minute. There are several available software programs for discourse analysis such as the Computerized Propositional Idea Density Rater (CPIDR) (Brown, Snodgrass, Kemper, Herman, & Covington, 2008), Computerized Language Analysis (CLAN) (MacWhinney, 2014; MacWhinney et al., 2011), and the Systematic Analysis of Language Transcripts (SALT) (Miller, Andriacchi, & Knockerts, 2011); however, much work is needed to make these feasible for clinical use. All require manual transcription and some degree of manual coding of the discourse sample. It is these time-intensive requirements that make them impractical in busy clinical settings. We agree that we need to embrace the power of computers to the extent possible; however, at this point, the amount of preparatory transcription and coding time needed to use their analytic power poses the biggest barrier to everyday clinical use of computer analysis programs. Another option is to develop discourse measurement systems that do not require transcription. As de Riesthal and Diehl remind us, Olness and her colleagues (Williamson et al., 2012) developed a qualitative, criterion referenced tool for assessing narrative discourse in the clinical setting. Others (Doyle et al., 2000; Doyle, Tsironas, Goda, & Kalinyak, 1996; Hula, McNeil, Doyle, Rubinsky, & Fossett, 2003; Richardson & Dalton, 2016) have measured aspects of discourse using checklists of main concepts that were developed from the discourses of neurologically healthy adults. These methods can be used without transcribing the discourse. Creative solutions for barriers to the clinical implementation of discourse analysis seem especially necessary because there is some evidence that changes in discourse may be more sensitive to change/diagnosis than changes on standardized tests (Henderson et al., 2017; Kintz & Wright).

Suggestions for moving forward

As summarized in our Target paper (Dietz & Boyle, [this issue](#)), we applaud the efforts by Wallace and her colleagues to spearhead the effort to establish a COS for aphasiology (Wallace, Worrall, Rose, & Le Dorze, [2014a](#), [2014b](#)); moreover, we appreciate their willingness to consider discourse measures as a part of the ROMA COS (Wallace et al., [2014b](#)). Participation in this discussion forum has helped us to answer our original question: “Do we need a core outcome subset on discourse?” We agree with Wallace and her colleagues that psychometrically sound discourse measures with clear elicitation protocols, that are also deemed valuable by stakeholders, must first emerge. Kintz and Wright suggested that Wallace et al. ([2014a](#), [2014b](#)) procedures be followed to work toward identification of discourse measures that meet these criteria. Generally speaking, we concur. We would be remiss, though, if we did not reiterate the concerns Hula, Fergadiotis, and Doyle ([2014](#)) summarized regarding these procedures when Wallace and her colleagues first introduced the ROMA COS. Hula and colleagues acknowledged the merits of a systematic review of outcome measures, but warned against the assumption that appropriate instruments already exist. To this point, Armstrong and Whitworth both expressed concerns regarding the potential for a discourse core outcome set to hamper creativity in the research process. In discourse, especially, one could argue that different measures may be required for spoken and written discourse, various genres, as well as for microlinguistic or macrolinguistic levels of discourse analysis. As such, we urge continued international conversations, via a variety of stakeholder consensus processes (Williamson et al., [2012](#)), to determine the best way to manage the complexity of determining what is core to the measurement of discourse.

The reviews (Bryant et al., [2016](#); Pritchard et al., [2017](#)) that emerged in the short time frame between the submission of the original Target paper (Dietz & Boyle) and the commentary papers submitted in response offer evidence that there has been movement toward refining the field’s approach to the use of discourse in aphasiology. We are excited by these efforts and believe that this bodes well for establishing discourse measures with psychometric properties acceptable for inclusion in the ROMA COS, spearheaded by Wallace and her colleagues ([2014a](#)). Further, we are optimistic that, as the reviews continue and awareness increases regarding the key issues discussed in this forum, researchers will be forthcoming about their discourse elicitation, transcription, and coding protocols, with reviewers and editors acting as the gatekeepers in this regard. We also hope that this discussion forum has inspired readers to consider the psychometric properties of measurements used when analyzing the discourse of people with aphasia. Finally, we hope this discussion prompts additional work to improve technology and discourse analysis protocols so that transcription and coding become less burdensome. Such improvements might allow discourse analysis and treatment to become routine aspects of clinical care.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Armstrong, E. ([this issue](#)). The challenges of consensus and validity in establishing core outcome sets. *Aphasiology*.
- Brookshire, R. H., & Nicholas, L. E. (1994). Test-retest reliability of measures of connected speech in aphasia. *Clinical Aphasiology*, 22, 119–133.
- Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., & Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods*, 40, 540–545. doi:[10.3758/BRM.40.2.540](https://doi.org/10.3758/BRM.40.2.540)
- Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics*, 30, 489–518. doi:[10.3109/02699206.2016.1145740](https://doi.org/10.3109/02699206.2016.1145740)
- de Riesthal, M., & Diehl, S. K. ([this issue](#)). Conceptual, methodological, and clinical considerations for a core outcome set for discourse. *Aphasiology*.
- Dietz, A., & Boyle, M. ([this issue](#)). Discourse measurement in aphasia: Have we reached the tipping point? *Aphasiology*.
- Doyle, P. J., McNeil, M. R., Park, G., Goda, A., Rubenstein, E., Spencer, K., ... Szwarc, L. (2000). Linguistic validation of four parallel forms of a story retelling procedure. *Aphasiology*, 14, 537–549. doi:[10.1080/026870300401306](https://doi.org/10.1080/026870300401306)
- Doyle, P. J., Tsironas, D., Goda, A. J., & Kalinyak, M. (1996). The relationship between objective measures and listener's judgments of the communicative informativeness of the connected discourse of adults with aphasia. *American Journal of Speech-Language Pathology*, 5, 53–60. doi:[10.1044/1058-0360.0503.53](https://doi.org/10.1044/1058-0360.0503.53)
- Henderson, A., Kim, H., Kintz, S., Frisco, N., & Wright, H. H. (2017). Working memory in aphasia: Considering discourse processing and treatment implications. *Semin Speech Lang*, 38, 040–051. doi:[10.1055/s-0036-1597257](https://doi.org/10.1055/s-0036-1597257)
- Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2008). Do picture naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology*, 22, 184–203. doi:[10.1080/02687030701262613](https://doi.org/10.1080/02687030701262613)
- Hula, W. D., Fergadiotis, G., & Doyle, P. J. (2014). A core outcome set for aphasia treatment research: Obstacles, risks, and benefits. *Aphasiology*, 28, 1396–1399. doi:[10.1080/02687038.2014.930264](https://doi.org/10.1080/02687038.2014.930264)
- Hula, W. D., McNeil, M. R., Doyle, P. J., Rubinsky, H. J., & Fossett, T. R. D. (2003). The inter-rater reliability of the story retell procedure. *Aphasiology*, 17, 523–528. doi:[10.1080/02687030344000139](https://doi.org/10.1080/02687030344000139)
- Kintz, S., & Harris Wright, H. ([this issue](#)). Response: Discourse measurement in aphasia research. *Aphasiology*.
- Kurland, J., & Stokes, P. ([this issue](#)). Let's talk real talk: An argument to include conversation in a D-COS for aphasia research with an acknowledgement of the challenges ahead. *Aphasiology*.
- Linnik, A., Bastiaanse, R., & Hohle, B. (2016). Discourse production in aphasia: A current review of theoretical and methodological challenges. *Aphasiology*, 30, 765–800. doi:[10.1080/02687038.2015.1113489](https://doi.org/10.1080/02687038.2015.1113489)
- MacWhinney, B. (2014). Challenges facing COS development for aphasia. *Aphasiology*, 28, 1393–1395. doi:[10.1080/02687038.2014.930263](https://doi.org/10.1080/02687038.2014.930263)
- MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, 25, 1286–1307. doi:[10.1080/02687038.2011.589893](https://doi.org/10.1080/02687038.2011.589893)
- Marini, A., Andreetta, S., del Tin, S., & Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology*, 25, 1372–1392. doi:[10.1080/02687038.2011.584690](https://doi.org/10.1080/02687038.2011.584690)
- Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related differences in the production of textual descriptions. *Journal of Psycholinguistic Research*, 34, 439–463. doi:[10.1007/s10936-005-6203-z](https://doi.org/10.1007/s10936-005-6203-z)
- Miller, J. F., Andriacchi, K., & Knockerts, A. (2011). *Assessing language production using salt software: A clinician's guide to language sample analysis*. Middletown, WI: SALT Software LLC.

- Pritchard, M., Hilari, K., Cocks, N., & Dipper, L. (2017). Reviewing the quality of discourse information measures in aphasia. *International Journal of Language and Communication Disorders*. doi:10.1111/1460-6984.12318
- Richardson, J. D., & Dalton, S. G. (2016). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology*, 30, 45–73. doi:10.1080/02687038.2015.1057891
- Wallace, S. J., Worrall, L., & Rose, T. (this issue). A core outcome set...or greater standardization of discourse measures?. *Aphasiology*.
- Wallace, S. J., Worrall, L., Rose, T., & Le Dorze, G. (2014a). A good outcome for aphasia. *Aphasiology*, 28, 1400–1404. doi:10.1080/02687038.2014.935119
- Wallace, S. J., Worrall, L., Rose, T., & Le Dorze, G. (2014b). Measuring outcomes in aphasia research: A review of current practice and an agenda for standardisation. *Aphasiology*, 28, 1364–1384. doi:10.1080/02687038.2014.930262
- Whitworth, A. (this issue). The tipping point: Are we nearly there yet?. *Aphasiology*.
- Williamson, P. R., Altman, D. G., Blazeby, J. M., Clarke, M., Devane, D., Gargon, E., & Tugwell, P. (2012). Developing core outcome sets for clinical trials: Issues to consider. *Trials*, 13, 132. doi:10.1186/1745-6215-13-132