

The assessment of functional adequacy in language performance

Folkert Kuiken and Ineke Vedder
University of Amsterdam

Assessment of functional adequacy: Rationale and objectives

Since the 1990s linguistic performance elicited by language tasks has often been operationalized in terms of complexity, accuracy, and fluency (CAF). As pointed out by e.g. Ortega (2003) and Pallotti (2009), less attention has been devoted to the functional dimension of language proficiency and to the appropriacy and efficacy of L2 performance. In more recent years, the importance of assessing functional adequacy (FA), in addition to CAF, as an essential component of L2 proficiency has been emphasized in a number of publications (De Jong et al., 2012; Hulstijn et al., 2012; Kuiken & Vedder, 2014, 2017, 2018; Révész et al., 2016).

In these studies FA is viewed as a multi-layered concept comprising different components. Over the years FA has been defined in terms of successful information transfer (Upshur & Turner, 1995), pragmatic appropriateness (McNamara & Roever, 2007) or text coherence and cohesion (Knoch, 2009). Within the framework of task-based language assessment (TBLA) and task-based language teaching (TBLT), Kuiken and Vedder (2017, 2018) consider FA as a task-related construct, in terms of successful task completion by the speaker/writer in conveying a message to the listener/reader. Inspired by Grice's (1975) conversational maxims of quantity, relation, manner and quality, Kuiken and Vedder (2017, 2018) presented a rating scale of FA for the assessment of oral and written performance. The scale comprises four dimensions: Task Requirements, Content, Comprehensibility and Coherence & Cohesion. In order to test the applicability of the FA rating scale, a number of experimental studies have been conducted, with various learners, tasks and settings (for a comprehensive account of the FA rating scale and an overview of studies in which the scale has been employed, see Kuiken and Vedder, this issue).

The goal of this special issue is to explore how FA of linguistic performance has been assessed for different source and target languages, types of learners, proficiency levels, task types and modalities. By taking a critical look at models of

FA and global and analytical performance measures, the assessment and development of FA, in relation to CAF, are addressed in a variety of learning contexts. Alongside implications for classroom practice and language pedagogy, key issues, perspectives, challenges, and directions for future research on FA are considered. The volume contains seven contributions: a position paper, four experimental papers and two commentaries. In these contributions the following questions are discussed:

1. How can FA be reliably assessed, both in L2 and L1?
2. How is FA related to CAF and to what extent do FA and CAF develop together?
3. What are the implications of research on FA for classroom practice?
4. Which perspectives and challenges can be identified for future research on FA?

The first question, regarding the applicability of the FA rating scale in L2 and/or L1, is investigated in the first five articles (Kuiken & Vedder, Ekiert et al., Strobl & Baten; Pallotti; Nuzzo & Bove, this issue). In these studies also the impact of proficiency level, task type and task modality on FA is discussed. The second question, concerning the relationship between FA and particular (sub)components of CAF, is explored by all authors, in particular by Ekiert et al., and Strobl and Baten. Pedagogical implications for classroom practice and assessment (question 3) are addressed by Kuiken and Vedder, Ekiert et al., and Nuzzo and Bove. Perspectives and challenges for future SLA research on FA (question 4) are discussed by all contributors, though especially by Loewen and González-Lloret in their commentaries.

Overview of the volume

In what follows we present a summary of the seven contributions included in the volume. In the position paper by Kuiken and Vedder the development of the six-point FA rating scale within the TBLA framework is sketched. This is followed by an overview of experimental studies in which the FA rating scale has been employed: for different source and target languages, task types (e.g. decision-making, instruction, narration), task modalities (oral vs. written), participants (children, university students) and proficiency levels (A2-C1 and native speakers). Next, future perspectives and pitfalls for research are discussed, such as (i) the relationship between FA and (sub)components of CAF, task type and language modality; (ii) the need for rater training and the possibility to use the FA rating scale both as a diagnostic instrument for teachers and a tool for learners' self-

assessment, and (iii) further research into the effect of task modality on FA and the use of the scale in interaction tasks.

The paper by Ekiert, Révész, Torgersen and Moss focuses on the relationship between breakdown fluency and FA in L2 speech. Specifically, the study examines the extent to which location of pausing and type of pausing may contribute to variation in FA in L2 oral performance and how task type moderates this relationship. Oral performances on two tasks by 40 L2 speakers of English (with Spanish as L1) were rated on an FA rating scale and analyzed for type of pausing (silent and filled at mid- or end-clause). The participants performed two oral tasks: making a complaint about a catering company and refusing an instructor's suggestion. The study showed that the location of filled pauses did not affect FA ratings, nor was the relationship between pause location and FA impacted by language proficiency. A significant relationship was, however, found between FA and the frequency of end-clause silent pauses: the fewer silent pauses participants produced between clauses, the more functionally adequate they were perceived to be. In other words, frequency of silent pauses appeared to be a strong predictor of FA ratings.

The study by Strobl and Baten focuses on advanced writing performance in L2 German. Texts were collected from 30 Dutch-speaking university students of L2 German who participated in a three to four month study abroad programme. The learners completed two narrative writing tasks regarding their experiences during an Erasmus stay, one before and one after their study trip. All texts were analyzed for lexical and syntactic complexity, accuracy, and fluency and evaluated by three expert raters by means of the FA rating scale. The results showed moderate to high intra-class correlations among the three raters, confirming the reliability of the FA rating scale for writing assessments based on a narrative task. It was found that syntactic complexity did not have an impact on the raters' judgments of FA, whereas lexical sophistication did. Another significant correlation was found between fluency and FA. This suggests that CAF and FA are interdependent constructs; although they develop together to a certain extent, they are different enough to be analyzed separately.

The study by Pallotti investigates the correlation between FA, measured holistically with the FA rating scale on the one hand, and by means of more analytic measures on the other. The data consisted of 227 texts written by primary school children in Italy, of whom 28% had Italian as a second or additional language. Texts were first assessed using three dimensions of the FA rating scale (Content, Comprehensibility, Coherence & Cohesion) and with an additional scale for coherence and cohesion from the CEFR (Council of Europe, 2001). Subsequently, the texts were analytically coded for a number of features directly bearing on FA and its subdimensions (e.g., main and secondary idea units and inappropriate

time shifts). With regard to the more analytic FA measures, some correlations were relatively clear, such as between Content and the number of main and secondary idea units, or the number of connecting devices and Coherence & Cohesion. Other correlations were less strong, yet were moving in the expected direction. The range of textual connectives, for example, was positively related to Coherence & Cohesion, while more ambiguous referential expressions were negatively correlated to Comprehensibility. Correlations were generally stronger for monolingual than for multilingual children, which raises the issue of language proficiency as a possible mediating factor between FA and its more immediate textual manifestations.

Nuzzo and Bove explored the potential use of the FA rating scale as a teaching tool for L₁ writing instruction. Participants in the study were 30 native Italian university students (MA level), who were randomly assigned to either an experimental or a control group. In the pre-test both groups were asked to complete an argumentative writing task (i.e., a motivation letter to apply as trainers for in-service secondary school teachers), related to the academic content of the course in which the experiment was taking place. Next, the students in the experimental condition received instruction, focusing on the use of the FA rating scale as a tool for self-assessment. In the post-test all participants performed a new argumentative task similar to the one in the pre-test. All texts produced in the pre- and post-test were assessed by five expert raters, using the FA rating scale. Results showed that the participants in the experimental condition benefited from the FA-based instructional treatment, although the gains they made in comparison to the control group were not statistically significant.

In his commentary on these five papers, Loewen focuses on how FA may be incorporated into the larger field of instructed second language acquisition (ISLA) and draws attention to a number of similarities and differences among the studies regarding their use of the FA rating scale. In some studies not all four of the scale dimensions were included (e.g., Pallotti) or they were altered (e.g., Strobl & Baten), which may hinder cross-study comparisons. In other studies a composite FA score was calculated (e.g., Ekiert et al., Pallotti, Nuzzo & Bove), which raises questions as to whether FA is fully captured in that way. As far as the use of the FA rating scale in both L₁ and L₂ is concerned, Loewen hypothesizes possible ceiling levels in speaking for L₁, but less so for writing: native speakers may also struggle in writing tasks, as shown by Nuzzo and Bove, who also demonstrate how the FA rating scale can be used as a pedagogical instrument for L₁ writers. Loewen considers this an important asset of the FA rating scale. Finally, as noted by Loewen, the study of FA rightly focuses our attention on the outcome of a language task in terms of its proposed goal: does the learner obtain what (s)he intends to achieve with his/her message, e.g., will the wish of the learner be granted by the addressee

or will his/her complaint be answered? Too often researchers and teachers look at what learners say or write, as well as how, whereas they should focus more on what their learners are able to achieve with their speech acts in terms of Grice's conversational maxims.

González-Lloret emphasizes in her commentary the impact FA may have on the field of TBLT and L2 pragmatics. With respect to TBLT she stresses the importance of pragmatic appropriacy for communication, especially in interactive settings. The author suggests that researchers should focus less on individual target tasks and more on creating target task types. In this way the generalizability problem (i.e., adaptation of the FA rating scale to any particular task) could be overcome. Viewed from the perspective of L2 pragmatics, the focus on FA may stimulate the use of tasks in which real interaction takes place. This implies that FA should take into account interactional features, cultural factors and the modality of communication. In order to assess interactional competence, González-Lloret proposes that three types of variables should be taken into consideration: (i) interactional variables that are culturally and situationally specific (such as flow of interaction, negotiation and politeness principles); (ii) participant variables (e.g., gender, degree of familiarity, power relations); (iii) artifacts that mediate the interaction (e.g., business cards, electronic messages, video conferencing). A final recommendation, raised by both Loewen and González-Lloret, is to test the use of the FA rating scale for computer-mediated modalities, especially those that are multimodal in nature with written and spoken components (e.g., a voice message, a note on the fridge, a text message) incorporating text and voice.

These suggestions and recommendations indicate that the assessment of FA in research and language teaching is both promising and challenging. A number of issues, clearly, need to be addressed in future studies. In the seven contributions, FA is explored from different angles, covering a wide range of target and source languages and learning contexts. The various studies presented here all use a cross-sectional design, which makes it difficult to assess how group-level differences in FA at different proficiency levels are connected to individual learner trajectories, as well as how they relate to CAF. Cross-sectional studies must therefore be complemented by longitudinal studies that focus on developmental patterns of FA in interaction with CAF, in both L2 and L1. Research in which a more in-depth investigation of the differential effects of task type and task modality on (sub)dimensions of FA is conducted, is also needed. Furthermore, it is important to explore more thoroughly the potential role of FA in classroom and assessment practice and to examine the impact of different instructional treatments on the development of FA in speaking and writing. Finally, as Loewen also suggests, other perspectives and challenges concern the standardization and adaptation of the FA rating scale, in relation to learning context, target language, task type, task

modality, together with the need to standardize methodology, rater training, and data analysis, in order to ensure comparability of studies.

References

- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- De Jong, N.H., Steinel, M.P., Florijn, A.F., Schoonen, R., & Hulstijn, J.H. (2012). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA* (pp. 121–142). John Benjamins. <https://doi.org/10.1075/llt.32.06jon>
- Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Speech acts* (pp. 41–58). Academic Press.
- Hulstijn, J.H., Schoonen, R., de Jong, N.H., Steinel, M.P., & Florijn, A. (2012). Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages (CEFR). *Language Testing*, 29(2), 203–221. <https://doi.org/10.1177/0265532211419826>
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275–304. <https://doi.org/10.1177/0265532208101008>
- Kuiken, F., & Vedder, I. (2014). Rating written performance: What do raters do and why? *Language Testing*, 31(3), 329–348. <https://doi.org/10.1177/0265532214526174>
- Kuiken, F., & Vedder, I. (2017). Functional adequacy in L2 writing. Towards a new rating scale. *Language Testing*, 34(3), 321–336. <https://doi.org/10.1177/0265532216663991>
- Kuiken, F., & Vedder, I. (2018). Assessing functional adequacy of L2 performance in a task-based approach. In N. Taguchi & Y-J. Kim (Eds.), *Task-based approaches to assessing pragmatics* (pp. 265–286). John Benjamins. <https://doi.org/10.1075/tblt.10.11kui>
- McNamara, T., & Roever, C. (2007). *Testing: The social dimension*. Blackwell. <https://doi.org/10.1111/j.1473-4192.2006.00117.x>
- Ortega, L. (2003). Syntactic complexity measures and their relationships to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518. <https://doi.org/10.1093/applin/24.4.492>
- Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590–601. <https://doi.org/10.1093/applin/amp045>
- Révész, A., Ekiert, M., & Torgersen, E. (2016). The effects of complexity, accuracy and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, 37(6), 828–848. <https://doi.org/10.1093/applin/amu069>
- Upshur, J.A., & Turner, C.E. (1995). Constructing rating scales for second language tests. *ELT Journal*, 49(1), 3–12. <https://doi.org/10.1093/elt/49.1.3>

Address for correspondence

Folkert Kuiken
Amsterdam Center for Language and Communication (ACLC)
University of Amsterdam
Postbus 1637
Amsterdam 1000 BP
The Netherlands
f.kuiken@uva.nl

Biographical notes

Folkert Kuiken is Professor Emeritus of Dutch as a Second Language and Multilingualism at the University of Amsterdam, The Netherlands, and Academic Director of the Institute for Dutch Language Education at that same university. His research interests include the effect of task complexity and interaction on SLA, Focus on Form, and the relationship between linguistic complexity and functional adequacy. He (co)authored and (co)edited various books and special issues, including “Dimensions of L2 performance and proficiency” (Housen, Kuiken & Vedder, 2012).

 <https://orcid.org/0000-0002-1667-2501>

Ineke Vedder is researcher at the University of Amsterdam, The Netherlands. Her research interests include Instructed Second Language Acquisition (particularly Italian), academic writing in L2 and L1, L2 pragmatics, and assessment of functional adequacy in L2 performance, in relation to linguistic complexity. Her publications have appeared in various edited books and journals, comprising two special issues, together with Housen, De Clercq and Kuiken, on syntactic complexity in SLA research (2019).

 <https://orcid.org/0000-0002-2677-0228>

Publication history

Date received: 26 February 2021

Date accepted: 7 September 2021

Published online: 20 June 2022