Research Article

Pauliina Peltonen*

Connections between measured and assessed fluency in L2 peer interaction: a problem-solving perspective

https://doi.org/10.1515/iral-2020-0030
Received March 4, 2020; accepted January 19, 2021;
published online February 4, 2021

Abstract: This article examines the connections between ratings and measures of second language (L2) fluency based on dialogue data, extending previous studies on L2 individual fluency. The links between fluency and strategic competence are studied from a problem-solving perspective. Fifteen raters' assessments and comments of individual fluency, interactional fluency, and strategic competence based on six samples of L2 peer interaction were examined. Fluency measures corresponded to the rated dimensions. In a mixed-methods analysis, correlational analyses across the ratings and between the ratings and fluency measures were performed. The raters' comments were analyzed qualitatively. The results demonstrated correlations between the ratings and measures of fluency for individual and interactional fluency, but not for strategic competence. The mean length of turn pauses correlated strongly with interactional fluency ratings. The raters' comments revealed a multifaceted conceptualization of interactional fluency and strategic competence. The results have implications for L2 fluency measurement and assessment.

Keywords: assessment; communication strategies; fluency; interaction; second language speaking

1 Introduction

Second language (L2) speech fluency studies have established that objective, measurable aspects of temporal fluency, including speech rate and aspects of pausing, and raters' assessment of L2 fluency are connected (e.g., Bosker et al.

***Corresponding author: Pauliina Peltonen**, Department of English, University of Turku, Koskenniemenkatu 4, 20014 Turku, Finland, E-mail: paupelt@utu.fi

2013). However, the majority of studies examining the links between fluency measures and ratings has been based on monologic approaches to fluency; therefore, more research exploring these links based on L2 interactional data are needed. Due to collaboration among the participants being an integral aspect of interaction, the under-researched aspect of *interactional fluency*, referring to the co-construction of flow across turns (cf. McCarthy's 2010 *confluence*), deserves more attention in L2 speech fluency research (see also Van Os et al. 2020). Learners also use various resources, or problem-solving mechanisms, to maintain the flow of speech during speech production difficulties. Two main types of problem-solving mechanisms (Dörnyei and Kormos 1998), namely *stalling mechanisms* (e.g., fillers, such as *you know*, and repetitions) and *communication strategies* (CSs, e.g., paraphrases) were examined in the present study as potentially contributing to fluency. While some links between problem-solving mechanisms and fluency have been suggested in the literature (e.g., Segalowitz 2010), problem-solving mechanisms have rarely been in the focus of L2 speech fluency studies. Because how efficiently learners overcome problems related to gaps in vocabulary is linked to how well the flow of speech is preserved (see also Peltonen 2017a), more studies examining the links between the use of CSs and fluency are needed.

The present study thus aims to fill these two gaps in L2 speech fluency research by 1) exploring the notion of interactional fluency from the perspectives of L2 fluency measurement and assessment and by 2) adopting a problem-solving perspective on L2 fluency, which enables the study of strategic skills in maintaining fluency with measures for CSs and ratings of strategic competence along with temporal fluency. The main aim of the study is thus to examine the connections between measured fluency and raters' assessments of fluency based on samples of L2 dialogue. In the study, 15 raters assessed and commented on learners' individual fluency, interactional fluency, and strategic competence based on six samples of L2 peer interaction. Fluency measures corresponded to the assessed dimensions, including measures for individual temporal fluency, interactional fluency, and two types of problem-solving mechanisms (stalling mechanisms and CSs). As a whole, the study thus brings together aspects of individual fluency, interactional fluency, and strategic aspects of interaction, which have rarely been studied together. By incorporating individual fluency measures and ratings along with interactional fluency and strategic competence, the study also extends a well-established line of L2 fluency research to the context of L2 peer interaction. With a mixed-methods approach combining quantitative correlational analyses and a qualitative analysis of the raters' comments, the study presents a methodologically novel approach for L2 fluency analysis based on dialogue data. In addition to developing L2 fluency analysis, the study has implications for the teaching and assessment of L2 interaction.

2 Theoretical framework

Fluency is usually associated with the smoothness and effortlessness of (L2) speech (e.g., Chambers 1997; Lennon 1990) and considered an aspect of more general oral proficiency (along with accuracy and complexity; on the complexity, accuracy, and fluency [CAF] framework, see Housen et al. 2012). However, the division between *the broad sense of fluency/higher-order fluency*, which equates fluency with overall (oral) proficiency, and *the narrow sense of fluency/lower-order fluency*, which mainly regards fluency in terms of temporal aspects of speech (Lennon 1990, 2000), may not always be clear-cut in L2 assessment. In other words, while the focus in L2 fluency research has been on *utterance fluency*, i.e., examining fluency with objective, measurable aspects of temporal fluency (Skehan's 2009 three-fold division to *speed*, *breakdown*, and *repair fluency* dimensions forming an oft-cited basis for choosing fluency measures), in L2 assessment, *perceived fluency*, or listeners' interpretations of fluency indicators in speech (Segalowitz 2010[1]), may also be influenced by other, non-temporal aspects not included in the narrow sense of fluency (e.g., Freed 1995; Magne et al. 2019; Kormos and Dénes 2004; Rossiter 2009; see also Götz 2013).

The tension between the narrow operationalization of fluency from a measurement perspective and the variation in raters' understanding of fluency underlies one of the main strands of L2 speech fluency research: the connections between perceived and utterance fluency, which have chiefly been examined based on monologic data and/or from the perspective of an individual speaker's fluency (see also De Jong 2018). Studies linking oral proficiency ratings and fluency measures have demonstrated that overall, temporal fluency measures (such as speech rate) are moderately or strongly correlated with oral proficiency ratings (e.g., Brand and Götz 2011; Ginther et al. 2010; Iwashita et al. 2008; Kahng 2014). Fluency measures have also been found to predict communicative adequacy (task accomplishment) (Révész et al. 2016). Additionally, studies examining correlations between fluency ratings and fluency measures have demonstrated that temporal measures and fluency ratings are generally relatively strongly related even when raters have not explicitly been instructed to assess fluency according to the narrow sense (e.g., Cucchiarini et al. 2002; Freed 1995; Kormos and Dénes 2004; Suzuki and Kormos 2020). Other studies explicitly guiding the raters' assessment towards the narrow sense have obtained similar results of generally moderate to strong correlations, especially regarding speed of talk and pausing, while the results

1 Segalowitz (2010) also distinguishes a third dimension of fluency, *cognitive fluency*, which refers to the efficiency of cognitive processing underlying fluent speech; this dimension is left out of the scope of the present study.

related to repair are less straightforward (e.g., Bosker et al. 2013; Derwing et al. 2004; Magne et al. 2019; Préfontaine et al. 2016; Rossiter 2009).

While the substantial body of research on L2 monologue and individual fluency has established clear connections between utterance and perceived fluency, less is known about the connections in L2 dialogue data. Specifically, there is a gap in research on the interactional aspects of L2 fluency and how they are assessed; the few fluency studies on L2 dialogue data have generally compared performance in dialogue and monologue tasks and mainly analyzed the productions from the perspective of an individual's utterance fluency, not focusing on the joint construction of fluency (Tavakoli 2016; Witton-Davies 2014). However, in an early study of six NS–NNS conversations, Riggenbach (1991) examined interactional phenomena, such as backchannels, other-repetitions, and collaborative completions, in addition to individual fluency measures. More recently, Sato (2014) examined ratings of individual and interactional oral fluency based on monologue and peer interaction samples from 56 Japanese university students of English and related the ratings to measures of speech rate. In the study, turn-taking and scaffolding emerged as prominent themes in addition to temporal aspects in the raters' fluency assessment. Similarly, in a recent experimental study (Van Os et al. 2020) examining the effects of turn-taking behavior (delay steps, including gaps and overlaps) on perceived fluency based on manipulated speech, raters were found to combine cues from turn-taking with speech rate and to form their assessments based on the interaction between the two aspects. Thus, the three studies discussed here underscore the importance of considering the co-construction of fluency across turns in interactional settings more widely. This aspect of fluency is referred to as interactional fluency in the present study, being comparable to McCarthy's (2010) concept of confluence. Extending previous studies, the present study aims to analyze this under-researched dimension of fluency comprehensively from both utterance and perceived fluency perspectives.

While studying interactional fluency has been rare in the field of L2 speech fluency research, co-construction in L2 interaction has been of interest to researchers examining *interactional competence* from a conversation analytic perspective (CA-SLA; for a recent overview, see Salaberry and Kunitz 2019) and to language testing researchers exploring interaction in L2 assessment by analyzing the raters' comments (e.g., Borger 2019; Ducasse and Brown 2009; May 2009) or the participants' performance (e.g., Galaczi 2014). In these studies, various elements have been examined as indicators of interactional competence, relating especially to topic development, listener support, and turn-taking management (Galaczi 2014). Turn-taking management (turn management in Galaczi and Taylor 2018), notably cohesive turn-taking, comes closest to the concept of interactional fluency, which was operationalized with measures for cohesive devices (collaborative

completions and other-repetitions) and temporal interactional fluency (between-turn pause duration and frequency) in the present study (see also Peltonen 2017a, b). The cohesive devices entail acknowledgment of the interlocutor's speech, thus creating interactional flow by linking the participants' contributions across turns, while the temporal component captures how pause time between turns is jointly managed by the participants (for further discussion, see Peltonen 2020). Interactional fluency can thus be viewed as roughly comparable with the dimension of turn-taking management within the broader concept of L2 interactional competence and as an extension of the previously individually focused research on L2 fluency.

As research on L2 interactional fluency is scarce, a particular interest in the present study was to map how the raters conceptualized this construct in relation to individual fluency. While there are some studies exploring raters' comments of individual fluency assessment (Préfontaine and Kormos 2016), analyses of raters' comments based on interactional data assessment are rare in the context of L2 fluency (but see Sato 2014). Yet, studies examining assessment of interactional competence suggest that raters do acknowledge the participants' collaboration in maintaining (interactional) flow (e.g., Borger 2019; Ducasse and Brown 2009). Thus, while monologue-based L2 fluency studies have approached fluency mainly as an individual's characteristic, in paired tasks, fluency can be viewed, at least to some extent, as a result of collaborative effort (see also McCarthy 2010). The present study thus aims to explore the interplay between individual and interactional fluency based on ratings of individual fluency (individual score) and interactional fluency (joint score; see Section 4.2) as well as raters' comments of these aspects.

In addition to examining both individual and interactional dimensions of L2 fluency, the present study links fluency to a broader problem-solving perspective, including two resources for overcoming problems during speech production (Dörnyei and Kormos 1998) and maintaining fluency: stalling mechanisms and communication strategies (CSs). Stalling mechanisms (comparable to Götz's 2013 *speech management strategies*) include features that help with maintaining fluency and coping with processing time pressure, such as filled pauses and repetitions. CSs, such as paraphrases, facilitate lexical problem-solving. Different types of stalling mechanisms (Dörnyei and Kormos 1998; Götz 2013) and CSs (Dörnyei and Scott 1997) were analyzed from the interactions, and the raters assessed strategic competence[2] as its own dimension (see also Van Batenburg et al. 2018). The use of

2 While strategic competence has also been conceptualized broadly in terms of metacognitive strategies in models of communicative language ability employed in language testing (notably in Bachman and Palmer 1996), the present study employed a narrower approach, conceptualizing strategic competence in terms of CSs, in line with the problem-solving approach to fluency.

strategies is linked to fluency, since overcoming problems in word retrieval with the help of CSs is likely to result in more fluent production than "getting stuck" and staying silent (see Peltonen 2017a). Yet, the connections between strategic competence and fluency have not been explored from the listeners', or a perceived fluency, perspective until the present study.

3 Research questions

Based on gaps in previous research regarding the study of L2 interactional fluency and the links between fluency and strategic competence from a problem-solving perspective, the following research questions were formulated to examine individual fluency, interactional fluency, and strategic competence comprehensively in the present study:

1. Do analytic ratings of individual fluency, interactional fluency, and strategic competence correlate with each other, on the one hand, and with the holistic rating of oral proficiency, on the other hand?
2. Which fluency measures correlate most strongly with ratings of individual fluency, interactional fluency, and strategic competence?
3. Which features form the basis for raters' assessments of individual fluency, interactional fluency, and strategic competence?

4 Methodology

4.1 Raters

The raters were 15 Finnish-speaking advanced students of English at a Finnish university (13 female, 2 male). Most participants studied English as their major subject (n = 12). The raters were on average 26 years old (SD = 2.77) and reported normal hearing. The raters had studied English at university for 4.40 years on average (SD = 1.30) and at school for 10.03 years on average (SD = 1.32). The ratings were completed as a part of an MA-level course on language assessment (5 ECTS). In addition to a theoretical component (lectures), the course included practical assessment exercises in small groups. During the course, the participants had familiarized themselves with models of language ability and assessment frameworks, notably *the Common European Framework of Reference for Languages* (Council of Europe 2001).

While most raters reported having relatively limited experience especially in spoken language assessment before the course, 12 of the participants had

participated in teacher training as part of their degree and gained some experience in language assessment during their studies. Native speakers or teachers of the target language (e.g., Freed 1995; Lennon 1990; Préfontaine et al. 2016; Riggenbach 1991) or even professional examiners (Cucchiarini et al. 2002) have commonly rated L2 samples in fluency studies, but less experienced raters (e.g., Bosker et al. 2013; Derwing et al. 2004) and/or L2 speakers (e.g., Magne et al. 2019; Rossiter 2009) have also been included as raters of fluency and as raters of interactional strategies (Van Batenburg et al. 2018). While inexperienced raters may be expected to assess fluency differently from experienced raters, Rossiter's (2009) study demonstrated only minor differences in fluency assessment between three groups of raters (experienced teachers of English/experts, education students/novices, and advanced non-native speakers of English). In the present study, a somewhat heterogenous and less experienced group of raters was preferred over a more experienced group to capture a broad variety of impressions of the assessed dimensions for the purposes of the qualitative analysis, yet yielding sufficient inter-rater reliability for the quantitative analyses (see Section 5). The profile of the rater group was thus in line with the exploratory nature of the present study.[3]

4.2 Rated data and the rating procedure

A set of six samples of L2 English peer interaction (from 12 participants) was chosen to be rated from a larger data set (see Peltonen 2017a). The participants spoke Finnish as their L1 and represented two school levels: two pairs were ninth graders (final year of compulsory education; 15–16-year-olds) and four pairs were 2nd year upper secondary school students (17–18-year-olds). The samples were chosen according to the following criteria to ensure reliable assessment: 1) enough talk from both participants (over 100 syllables), 2) relatively symmetrical interaction, as judged by the author, and 3) the use of CSs during the interaction. It was also ensured that the samples represented a range of proficiency levels (see also e.g., Cucchiarini et al. 2002): the subjects' vocabulary test scores (LexTALE; Lemhöfer and Broersma 2012) suggested that there was sufficient variation in the sample regarding the participants' general proficiency in English. Two pairs (the ninth-grade pairs) represented the upper range of B1/lower range of B2 (scores 56.25–67.50%) in the CEFR (Council of Europe 2001), three pairs represented the B2

3 A similar approach based on untrained raters and minimal instructions has been used for ratings of other aspects of L2 proficiency as well, including lexical diversity (Jarvis 2017), to capture raters' intuitions about the concept under investigation.

level (scores 63.75–75.00%), and one pair represented levels B2 and C1–C2 (scores 72.50 and 96.50%).[4] Of potential interlocutor effects (see e.g., Taylor and Wigglesworth 2009), proficiency level, age, L1, and acquaintanceship were controlled in the present study.

The peer interaction samples (rated samples) involved the participants completing a problem-solving task in pairs. After 2 min of individual preparation, the pairs had 6 min for completing the task. The task was to discuss sixteen items in English and to rank them in the order of their potential usefulness for survival on a desert island. In addition to objects that the participants were likely to be able to name in English (e.g., umbrella), the task also included less familiar objects (e.g., flares) to encourage the use of CSs. The data were collected during regular school days for research purposes, and the participants were informed that their performance did not influence their English grades. Signed consent forms were obtained from the participants or from their parents in the case of underaged subjects.

The ratings were based on anonymized videos of the interaction samples. Transcriptions were available for the raters, e.g., for checking any unclear parts in the interactions. No background information about the participants except for their school level was given to the raters to secure the participants' anonymity. The instructions on the rating sheet (Appendix 1) informed the raters to assess all samples on 9-point scales for oral proficiency (holistic rating), followed by analytic ratings of (individual) fluency, interactional fluency, and strategic competence. Relatively detailed, 9- or 10-point scales have been used in previous fluency studies to ensure fine-grained distinctions in the participants' fluency (e.g., Bosker et al. 2013; Cucchiarini et al. 2002; Derwing et al. 2004; Götz 2013; Rossiter 2009). Despite the raters having familiarized themselves with the CEFR scales (Council of Europe 2001) for oral proficiency during the language assessment course, no descriptors were provided for them, as the study was exploratory in nature and the purpose was to capture the raters' perceptions of the concepts under investigation (see also Suzuki and Kormos 2020). The raters had 2 weeks to complete their ratings, and they could spend as much time on rating each sample as they wanted. As the raters completed the assessments on their own time, the researcher only had access to the final ratings.

For the analytic ratings, the raters were instructed to comment on and justify their numeric ratings by specifying the features to which they had paid attention when assigning the rating. For fluency, the raters were instructed to focus on fluency in the narrow sense and to take into account the main dimensions of fluency (speed, pausing, repair; Skehan 2009) to avoid them equating the concept

4 The variation was expected to be sufficient for the raters to detect differences across the samples based on the 9-point scales and thus to enable reliable assessment.

with general oral proficiency (fluency in the broad sense; following e.g., Derwing et al. 2004; Rossiter 2009). For interactional fluency, a joint score was given to the pair, but no further instructions or definitions were provided to capture the raters' understanding of the concept. Taking into account the co-constructed nature of interactional fluency, a joint score was viewed as more appropriate than an individual score for this dimension (see also May 2009; cf. discussion in Taylor and Wigglesworth 2009). Regarding strategic competence, the raters were instructed to focus on CSs, with paraphrases mentioned as an example, to guide the raters towards a relatively narrow construct of strategic competence that would match the problem-solving perspective on fluency adopted in the present study.

4.3 Fluency measures

For the purposes of fluency measurement, the interaction samples were transcribed and double-checked. Pauses (0.25 s or longer) were identified with a script (De Jong and Wempe 2009) in the speech analysis software Praat (Boersma and Weenink 2015) and adjusted manually if necessary. An upper limit for pauses was set at 3 s (Witton-Davies 2014), and the exceeding pause time was excluded from the total sample duration calculations. To answer RQ 2, a set of fluency measures based on well-established temporal (individual) fluency measures (for a summary, see De Jong 2018), capturing within-turn contributions from each participant, and exploratory measures of interactional fluency (e.g., Peltonen 2017a, b; Riggenbach 1991), capturing the collaborative, between-turn aspects, was used in the present study. Two types of problem-solving mechanisms, stalling mechanisms and CSs (Dörnyei and Kormos 1998; Dörnyei and Scott 1997; Götz 2013; Peltonen 2017a), were also examined. All frequency measures were standardized per minute of speaking time. Since the interaction samples were part of a larger data set, the group-level tendencies for the whole data set regarding the utterance fluency measures can be found in Peltonen (2017a).

The following temporal (individual) fluency measures were used:

1) speech rate (syllables per minute of total time excluding turn pauses),
2) articulation rate (syllables per minute of speaking time, excluding all pauses),
3) mean length of turn (the number of syllables divided by the number of turns; comparable to mean length of run used in monologue fluency studies),
4) the number of silent pauses (SPs; frequency of pauses lasting 0.25 s or longer),
5) the number of mid-clause SPs,
6) mean length of (within-turn) SPs (the total length of SPs divided by the total number of SPs, excluding TPs).

The following interactional fluency measures were used (the first two relate to temporal fluency, latter two to cohesive devices):
1) the number of (between-) turn pauses (TPs; divided equally between the participants),
2) mean length of TPs (the total length of TPs divided by the total number of TPs),
3) the number of other-repetitions (repetitions of words or longer utterances of interlocutor's speech), and
4) the number of collaborative completions (completions of an utterance the previous speaker has started).

All clause boundary pauses lasting over 2 s were included in TP calculations and considered "belonging" to both speakers (Witton-Davies 2014). Separate measures for frequencies or average durations for shorter clause boundary pauses were not included in the analyses, since mid-clause pauses are more strongly connected to L2 disfluency (e.g., Kahng 2014).

The following measures of stalling mechanisms were used: the number of
1) (non-lexicalized) filled pauses (e.g., *uh* and *um*),
2) drawls (vowel elongations of 0.30 s or longer; Witton-Davies 2014),
3) fillers (lexicalized pauses, e.g., *like* and *you know*; Götz 2013), and
4) (self-)repetitions.

The following measures of CSs were used: the number of
1) all CSs (measures 3–8),
2) paraphrase strategies (measures 3–6),
3) circumlocutions (describing properties of the target object),
4) approximations (alternative, related lexical items for objects),
5) all-purpose words ("empty" lexical items for an object; e.g., *thing*),
6) other paraphrase strategies,
7) transfer strategies (the use of the L1 for strategic purposes), and
8) appeals for assistance (requesting help from the interlocutor).

4.4 Analysis

Before the data were analyzed from the perspective of the research questions, inter-rater reliability was examined with intraclass correlation coefficients (ICCs; two-way random model, type: consistency). Average measures ICCs are reported. Spearman's rank-order correlations were calculated to answer the research question related to the connections between ratings of different aspects of fluency

(RQ 1) and to the connections between ratings of different aspects of fluency and measures of these aspects (RQ 2). Spearman's rank-order correlations were chosen instead of Pearson's correlations due to the Shapiro–Wilk's tests of normality indicating non-normal distributions for some of the variables. For the purpose of the correlational analyses, the pair scores for interactional fluency were treated as individual scores (the same score coded for both participants in a pair). The correlations were interpreted according to Plonsky and Oswald's (2014) guidelines, rs values close to 0.25 being small in effect, 0.40 being medium and 0.60 being large. SPSS version 25 was used for all statistical analyses. Regarding RQ 3 related to the raters' understanding of the rated dimensions, a qualitative content analysis (e.g., Dörnyei 2007) was conducted to identify central themes in the raters' comments.

5 Results

Before discussing the results related to the research questions 1–3, descriptive statistics about the ratings and inter-rater reliability statistics are provided to illustrate variation among the raters and the reliability of the ratings. The means and standard deviations of the ratings are provided in Table 1.

Overall, as can be seen from Table 1, pair 3 received the highest mean ratings for all of the rated aspects. The highest standard deviations were found for the strategic competence ratings, and the strategic competence ratings of subject G1-014 demonstrated the most variation among the raters (SD = 2.04). The inter-rater

Table 1: Descriptive statistics (M, SD) for all rated aspects (min. 1 – max. 9).

Pair #	Subject code	Oral proficiency rating M (SD)	Individual fluency rating M (SD)	Interactional fluency rating M (SD)	Strategic competence rating M (SD)
1	G2-004	7.00 (0.93)	7.13 (1.06)	7.73 (0.88)	7.60 (0.74)
1	G2-005	6.47 (1.30)	6.40 (1.30)		5.93 (1.58)
2	G1-011	5.87 (1.55)	5.93 (1.28)	5.47 (1.13)	5.40 (1.55)
2	G1-012	5.87 (1.41)	5.73 (1.28)		4.67 (1.76)
3	G2-001	8.20 (0.86)	8.07 (0.70)	8.53 (0.64)	8.21 (1.25)
3	G2-006	7.93 (0.80)	8.07 (0.83)		8.00 (1.57)
4	G2-206	5.53 (1.46)	5.13 (1.55)	5.93 (1.28)	6.20 (1.32)
4	G2-207	6.93 (1.16)	7.07 (1.22)		7.67 (0.98)
5	G1-013	6.80 (1.32)	6.87 (1.30)	6.57 (1.45)	7.00 (0.93)
5	G1-014	6.13 (1.77)	6.13 (1.60)		5.80 (2.04)
6	G2-306	6.36 (1.34)	6.13 (1.19)	6.21 (0.89)	6.13 (1.72)
6	G2-305	5.79 (1.31)	5.73 (1.44)		7.13 (1.06)

For interactional fluency, the pairs received a joint score.

reliability statistics showed that intraclass correlation coefficients were, nevertheless, quite high for all the rated dimensions (oral proficiency = 0.939, individual fluency = 0.925; interactional fluency = 0.981; strategic competence = 0.884). Yet, the ratings were slightly lower for strategic competence than for the other aspects.

5.1 Connections between ratings of different aspects of fluency and oral proficiency

The results related to RQ 1 addressing the connections between analytic ratings of different aspects of fluency as well as their connections to the holistic rating of oral proficiency are presented in Table 2.

As can be seen from Table 2, all aspects of fluency were statistically significantly correlated with oral proficiency, the Spearman's rank-order correlations between individual fluency ratings and oral proficiency ratings being the strongest (rs = 0.986). Furthermore, the different aspects of fluency correlated strongly and statistically significantly with each other.

Table 2: Correlations between mean ratings of different aspects of fluency and oral proficiency.

	Mean individual fluency rating	Mean interactional fluency rating	Mean strategic competence rating
Mean oral proficiency rating	0.986**	0.750**	0.697*
Mean individual fluency rating		0.773*	0.712*
Mean interactional fluency rating			0.622*

** = correlation significant at the 0.01 level. * = correlation significant at the 0.05 level.

5.2 Connections between fluency ratings and fluency measures

The results related to the connections between ratings of individual fluency and individual fluency measures (RQ 2) are compiled in Table 3.

As illustrated in Table 3, the individual fluency measures that demonstrated the strongest (and statistically significant) correlations with mean individual fluency ratings were SR (*rs* = 0.828), AR (*rs* = 0.681), and MLT (*rs* = 0.646). None of the silent pause measures were statistically significantly correlated with the individual fluency ratings and ranged from small to medium in effect (SPs/min.

Table 3: Correlations between central temporal fluency measures (speed fluency and pausing) and mean individual fluency ratings.

	Speech rate (SR)	Articulation rate (AR)	Mean length of turn (MLT)	Silent pauses (SPs)/min.	Mid-clause SPs/min.	Mean length of SPs
Mean individual fluency rating	0.828**	0.681*	0.646*	–0.495	–0.298	–0.428

** = correlation significant at the 0.01 level. * = correlation significant at the 0.05 level.

p = 0.102, mid-clause SPs/min. p = 0.346, mean length of SPs p = 0.165). Furthermore, the individual fluency ratings were weakly and not statistically significantly connected to individual stalling mechanisms (filled pauses/min. rs = 0.042, p = 0.897; drawls/min. rs = –0.232, p = 0.469; fillers/min. rs = 0.089, p = 0.783; repetitions/min. rs = 0.138, p = 0.669).

The results related to the connections between ratings of interactional fluency and the four interactional fluency measures are compiled in Table 4.

As can be seen from Table 4, the only statistically significant and very negative strong correlation was found between mean interactional fluency ratings and the mean length of turn pauses (rs = –0.943). This finding is illustrated in Figure 1: the higher the raters evaluated the pair's interactional fluency (mean ratings on the y-axis), the shorter the pauses were between turns in the interaction (mean durations in seconds on the x-axis). For the other variables, the correlations were not statistically significant (turn pauses/min. p = 0.102; other-repetitions/min. p = 0.222; collaborative completions/min. p = 0.457).

Finally, the results of the correlations between ratings and measures of strategic competence are compiled in Table 5.

As demonstrated in Table 5, there were no statistically significant correlations between the strategic competence ratings and communication strategy measures (total strategies p = 0.649; total paraphrase p = 0.633; circumlocutions p = 0.459;

Table 4: Correlations between interactional fluency measures and mean interactional fluency ratings.

	Turn pauses/min.	Other-repetitions/min.	Collaborative completions/min.	Mean length of turn pauses
Mean interactional fluency rating	–0.495	0.381	0.237	–0.943**

** = correlation significant at the 0.01 level.

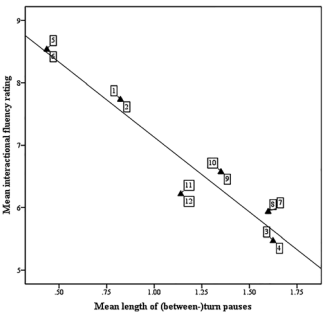

Figure 1: A scatter plot of correlations between mean interactional fluency ratings and the mean length of turn pauses.

Table 5: Correlations between communication strategies and mean strategic competence ratings.

	Total strategies	Total paraphrase	Circumlocutions	Approximations	All-purpose words	Other paraphrases	Transfer	Appeal for assistance
Mean strategic competence rating	−0.147	−0.154	0.218	−0.231	−0.232	0.431	−0.167	−0.033

approximations p = 0.471; all-purpose words p = 0.467; other paraphrases p = 0.161; transfer p = 0.605; appeal for assistance p = 0.918). Overall, the correlations were also relatively weak.

5.3 Raters' perceptions of different aspects of fluency

In this section, the results of the qualitative analysis (RQ 3) are presented. The themes identified from the raters' comments are presented in Tables 6–9, including descriptions for each theme, and the frequencies and percentages of comments in each thematic category. Altogether 1,110 comments were analyzed.

As illustrated in Table 6, 30.63% of the raters' comments related to individual fluency. Five main themes were identified: speech tempo and flow (9.10% of all comments), pausing (8.02%), hesitations and corrections (6.40%), amount of speech (3.69%), and stalling mechanisms (3.42%). The themes covered the aspects that the raters were instructed to focus on (speech rate, pausing, and corrections).

Table 6: Themes related to individual fluency with descriptions, comment frequencies, and percentages.

Theme	Description	Frequency of comments	Percentage of all comments
Individual fluency		340	30.63%
Speech tempo and flow	Rate of (individual) speech, naturalness and flow	101	9.10%
Amount of speech	Amount of (individual) speech	41	3.69%
Pausing	(Silent) interruptions to an individual's stream of speech	89	8.02%
Hesitations and corrections	Other interruptions and adjustments to speech besides silent pauses	71	6.40%
Stalling mechanisms	The use of fluency-related features to "buy time" and to avoid long pauses; e.g., filler words	38	3.42%

Table 7: Themes related to interactional fluency with descriptions, comment frequencies, and percentages.

Theme	Description	Frequency of comments	Percentage of all comments
Interactional fluency		343	30.90%
Flow of discussion	Pace and flow of the discussion	66	5.95%
Pausing during the discussion	(Silent) "shared" pauses in the discussion	34	3.06%
Turn-taking	Balance in participating in the discussion	63	5.68%
Acknowledging and encouraging interlocutor's contributions	Demonstrating agreement/understanding with different means and encouraging interlocutor to participate	100	9.01%
Task accomplishment	Taking task instructions into account (e.g., search for consensus)	26	2.34%
Mutual understanding	Participants seem to understand each other; indicators of understanding or the lack of it	34	3.06%
Other indicators of collaboration	Other references to collaboration (e.g., eye contact)	20	1.80%

Table 8: Themes related to strategic competence with descriptions, comment frequencies, and percentages.

Theme	Description	Frequency of comments	Percentage of all comments
Strategic competence		263	23.69%
Communication strategies, verbal	Compensation for gaps in vocabulary (e.g., paraphrases)	164	14.77%
Communication strategies, non-verbal	Compensation for gaps in vocabulary with non-verbal means (gestures)	48	4.32%
Collaborative problem-solving	Other means to ensure understanding during problem-solving, e.g., word searches	51	4.59%

Table 9: Other themes with descriptions, comment frequencies, and percentages.

Theme	Description	Frequency of comments	Percentage of all comments
Other		164	14.77%
Accuracy	Correctness of the production; errors	13	1.17%
Proficiency level	Level of target language knowledge	14	1.26%
Vocabulary	Vocabulary knowledge and size	51	4.59%
Content	Including clarity, coherence, and richness of the discussion	46	4.14%
Pronunciation	Including intonation	9	0.81%
The use of the L1	Mentions of (non-strategic) use of the L1	8	0.72%
Personality	Personal characteristics, e.g., natural speech rhythm	8	0.72%
Other	Aspects not included in other categories	15	1.35%

However, the raters also discussed the amount of speech, typically when the small amount of speech complicated fluency ratings.

Overall, the raters' comments reflected sensitivity to the different dimensions and features of individual fluency. For example, some of the raters distinguished pauses by their location (mid-clause vs. clause or turn boundary) and status (individual pauses vs. "shared" pauses) in their comments, and they even referred to

the reasons for pausing (due to content vs. language planning), as illustrated in comment 1 (comments originally in Finnish, translated by the researcher).

(1) Both participants had both silent and filled pauses in their speech, often in places where the speaker has to think about the content rather than language.
(Rater 9, sample 1, individual fluency)

Furthermore, while some raters discussed fluency-related features as means for buying time, all raters did not refer systematically to stalling mechanisms, which potentially explains the lack of statistically significant correlations between individual fluency ratings and stalling mechanisms.

As can be seen from Table 7, 30.90% of the raters' comments related to interactional fluency. Comments related to interactional fluency were grouped under seven main themes. The theme with the most comments related to the different ways of acknowledging (e.g., comments, indicators of active listening, collaborative completions) or encouraging (e.g., by asking questions) interlocutor's contributions (9.01% of all comments). Two categories related to the temporal aspects of interaction, the flow of discussion and pausing during the discussion, also covered 9.01% of all comments (5.95 and 3.06%, respectively). Other aspects related to the balance in turn-taking (5.68%), mutual understanding (3.06%), task accomplishment (2.34%), and other aspects of collaboration (1.80%).

While various interactional features were salient to the raters, categorizing them as interactional fluency was not always straightforward. In particular, comments related to the theme collaborative problem-solving were mentioned as part of interactional fluency almost as often (45.10% of the comments related to that theme) as they were as part of strategic competence (54.90%). Comment 2, categorized as belonging to the theme acknowledgment, illustrates how some aspects of interactional fluency could also be viewed as part of strategic competence by the raters. In fact, 8.75% of the comments related to interactional fluency were mentioned in conjunction with strategic competence.

(2) Both supported each other with their turns, "chiming in", so that the other participant could talk more based on that. I am not completely sure, however, whether this is strategic competence or simply the pair's fluent interaction.
(Rater 3, sample 3, strategic competence)

Furthermore, the distinction between individual and interactional fluency was not entirely clear-cut: 9.91% of the comments categorized as interactional fluency were mentioned in conjunction with individual fluency. While the pair received a joint score for interactional fluency, comment 3 illustrates how some aspects of

interactional fluency, such as turn-taking, could be approached from the individual's perspective.

(3) The person on the left had many silent pauses, and the partner had to occasionally take charge to ensure the continuation of the discussion. (Rater 13, sample 4, individual fluency)

The comment reflects how individual and interactional fluency can be viewed as intertwined, one aspect of an individual's (dis)fluency (pausing) affecting the interaction as a whole (balance in turn-taking).

As indicated in Table 8, 23.69% of the raters' comments related to strategic competence. For strategic competence, the raters were instructed to focus on communication strategies (for instance paraphrases) in their assessment, which was also reflected in the comments. As indicated in Table 8, the majority of the comments related to different types of verbal communication strategies (14.77% of all comments) and 4.32% to non-verbal strategies. The third theme, collaborative problem-solving, included comments (4.59%) related to other, collaborative means to ensure understanding during problem-solving, such as the negotiation of meaning. While the raters were able to identify strategies from the samples as instructed, the references to *collaborative* problem-solving in the raters' comments suggest that the measurement perspective, focusing on individual communication strategy frequencies, was not fully comparable with the raters' understanding of strategic competence.

In addition to the three main aspects of individual fluency, interactional fluency, and strategic competence, 14.77% of the raters' comments related to other themes, as summarized in Table 9. Vocabulary-related comments were relatively common in the data (4.59% of all comments), especially in conjunction with strategic competence. In particular, the raters' comments revealed challenges in assessing those participants' strategic competence who were perceived as not needing to use strategies due to sufficient vocabulary knowledge. Furthermore, different aspects related to content covered 4.14% of all comments. Other, less frequently mentioned aspects included accuracy (1.17%), general proficiency (1.26%), and pronunciation (0.81%). Some raters also referred to the use of Finnish (0.72%) or personality characteristics (0.72%).

6 Discussion

The purpose of the present study was to examine the connections between L2 fluency ratings and measures in the context of interactional data. Based on a mixed-methods analysis of comments about fluency ratings by 15 MA-level

university students of English and the connections between the ratings and fluency measures, the present study provided insights into the assessment of L2 peer interaction from the perspectives of individual fluency, interactional fluency, and strategic competence.

The correlations across ratings of different aspects of fluency (RQ 1) demonstrated that all of the aspects were connected. Furthermore, all aspects were correlated with oral proficiency. While there are no directly comparable studies examining correlations across the same dimensions, Iwashita et al. (2008) found that especially vocabulary and (temporal) fluency distinguished performances across proficiency levels based on holistic oral proficiency ratings, comparable to the correlations between oral proficiency and individual fluency found in the present study. The strong correlation could potentially be due to the raters understanding fluency in a broad sense (Lennon 1990), but the qualitative analysis does not fully support this conclusion, as references to other aspects of proficiency as part of individual fluency comments were relatively rare. It is more likely that the raters perceived the aspects as interrelated, as supported by the qualitative analysis demonstrating overlap and links across the rated dimensions. Furthermore, the raters could have conceptualized oral proficiency (holistic rating) in terms of the analytic categories (analytic ratings), leaving out aspects such as accuracy or complexity (see also Van Batenburg et al. 2018). While it is unclear whether such a halo effect influenced the ratings in the present study, in future studies, the holistic and analytic ratings could be provided by different rater groups to minimize potential halo effects, as suggested by Van Batenburg et al. (2018).

The correlations between fluency ratings and fluency measures (RQ 2) varied across different aspects of fluency. For individual fluency, the ratings were found to be most strongly correlated with key temporal fluency measures (speech rate, articulation rate, and mean length of turn). This finding is in line with previous studies on monologue L2 fluency that have demonstrated strong correlations between speed fluency measures and fluency ratings (e.g., Bosker et al. 2013; Cucchiarini et al. 2002; Kormos and Dénes 2004; Magne et al. 2019; Rossiter 2009; Suzuki and Kormos 2020). However, in contrast to fluency studies that have often found relatively strong and statistically significant correlations also between pausing (especially pause frequencies) and fluency ratings (e.g., Bosker et al. 2013; Rossiter 2009), (individual) pause frequencies or durations were only moderately and not statistically significantly correlated with raters' assessment in the present study (cf. Préfontaine et al. 2016). However, since speech rate contains information about within-turn pauses (included in total time), pauses seem to have been considered in the ratings as part of overall pace. The results thus complement previous studies that have demonstrated connections between utterance and perceived fluency based on monologue data by confirming a similar connection in

a dialogic context. This suggests that aspects of individual, temporal fluency can also be measured and assessed from L2 interactional data.

In contrast to the findings for individual fluency, strong correlations were found between the mean length of turn pauses and ratings of interactional fluency. The finding is in line with previous studies pointing to long between-turn silences differentiating learners across proficiency levels (Galaczi 2014; Peltonen 2017a) and being salient to raters in experimental settings (Van Os et al. 2020), but the present study is the first to demonstrate a strong correlation between an interactional fluency measure and the perceived degree of interactional fluency in the context of L2 peer interaction. The finding of the ratings not correlating with "individual" pauses but correlating strongly with "interactional" pauses also parallels Sato's (2014) findings of raters perceiving pauses differently across individual and interactional performances: in his study, pauses were linked to disfluencies in monologue, while in dialogue, pauses related to the interaction and turn-taking received more emphasis. The mean length of turn pauses could thus be used in future L2 interactional fluency studies for operationalizing an essential aspect of interactional competence, turn management (e.g., Galaczi and Taylor 2018). Furthermore, due to its strong correlation with interactional fluency ratings, the mean length of turn pauses could be incorporated as a criterion in assessment scales. For instance, while the CEFR (Council of Europe 2001: 28–29) qualitative aspects of spoken language use include a rating scale for "interaction", the relatively vague references to turn-taking and to keeping conversation going could be complemented with the more concrete feature of between-turn pause duration in conversation, which also seems to be salient to raters (see the discussion on the results of the qualitative analysis below).

Finally, for strategic competence, weak and non-significant correlations were found between the ratings and the measurements. Unlike in the frequency-based approach captured by the measures focusing on individual communication strategy use, the raters seemed to consider strategies in the context of collaborative problem-solving; how strategies are used in interaction might thus be more important than their frequency from the raters' perspective (see also Peltonen 2017a). The rating perspective, therefore, did not fully match the measurement approach to strategies.[5] Thus, compared to individual and interactional fluency, the assessment of strategic competence specifically in terms of communication

5 Despite the instructions guiding the raters towards a narrow operationalization of strategies, in line with the view of communication strategies as potentially fluency-enhancing resources from the measurement perspective, the raters' approach could have also been influenced by the broader notions of strategic competence in models of communicative language ability (e.g., Bachman and Palmer 1996).

strategies in a dialogue context may be less intuitive and could warrant the use of further instructions and/or specific descriptors. Furthermore, to better capture the collaborative problem-solving dimension of strategy use, which was quite prominently featured in the raters' comments, other interactional strategies, such as negotiation of meaning (e.g., Van Batenburg et al. 2018; see also Dörnyei and Kormos 1998), could also be incorporated and acknowledged in raters' instructions in future studies.

The qualitative analysis (RQ 3) revealed that the raters paid attention to a variety of features when assessing the samples, individual fluency comments matching the measured aspects more closely than the interactional fluency or strategy comments. Connections across the rated dimensions, between strategic competence and interactional fluency as well as individual fluency and interactional fluency, were also discovered (cf. RQ 1). The most salient themes related to individual fluency were comparable to aspects of individual fluency captured by the measures and themes identified in other studies analyzing raters' comments of (monologue) fluency assessment (Préfontaine and Kormos 2016).

The main themes identified for interactional fluency echo Sato's (2014) and Riggenbach's (1991) findings of displays of comprehension, scaffolding, and balance in turn-taking being key features of fluency in an interactional context (see also Borger 2019; Ducasse and Brown 2009). Yet, compared to the four interactional fluency measures used in the present study, the range of interactional phenomena attended to by the raters was broader; in addition to aspects that corresponded to the interactional fluency measures, the raters also mentioned different ways of encouraging the interlocutor to participate as well as indicators of mutual understanding, among others. While the raters could have understood interactional fluency broadly as interactional competence (e.g., Galaczi and Taylor 2018; Roever and Kasper 2018; Salaberry and Kunitz 2019), certain aspects usually associated with interactional competence, such as topic management (e.g., Galaczi 2014; Galaczi and Taylor 2018), were not prominently featured in the raters' comments in the present study. It thus seems possible to distinguish interactional fluency, with the co-construction of flow across turns as its essential element (McCarthy 2010), from interactional competence, which is a broader construct; much like it is possible to distinguish individual fluency from general oral proficiency. Yet it is also worth noting that based on the comments, individual and interactional fluency were to some extent interrelated; considering the interplay between individual contributions and joint management of interaction in dialogue contexts, it would perhaps be more surprising if these had been viewed as completely distinct aspects by the raters. In addition, interactional fluency and strategic competence were not completely distinct, as comments regarding collaborative problem-solving were mentioned in conjunction with both aspects.

Taking these findings together, in future studies, concrete features of interactional fluency (such as the duration of turn pauses) could be incorporated in the instructions to assist the raters in distinguishing interactional fluency from strategic competence, on the one hand, and from individual fluency, on the other hand. This could also help to ensure that interactional fluency is not confused with the broader notion of interactional competence.

For strategic competence, despite the fact that the score was awarded individually, the raters' comments related to the theme collaborative problem-solving demonstrated that this dimension was, to a certain extent, viewed in terms of mutual efforts (cf. "interactional effectiveness" in May 2009). A further challenge identified from the comments related to awarding scores to those participants who were perceived as having sufficient vocabulary knowledge and therefore no need for strategy use. In future studies, more specific instructions could be given to raters regarding the scoring of these types of performances. Finally, 14.77% of the comments related to other aspects besides individual and interactional fluency or strategic competence, including references to vocabulary, content or clarify, and accuracy. As the focus was on fluency in the narrow sense, this is a relatively high percentage. However, the result is not exceptional, as approximately 25% of the rater's negative impressions in Rossiter's (2009) study concerned non-temporal aspects despite being instructed to rate fluency in the narrow sense (see also Magne et al. 2019).

7 Conclusion

This study set out to examine the connections between measured and assessed L2 fluency based on interactional data. In particular, the study aimed to extend previous L2 fluency research by exploring the notion of interactional fluency and by analyzing fluency from a problem-solving perspective. The problem-solving perspective complemented previous studies, which have rarely explored the collaborative or strategic aspects of L2 fluency, by providing insights into the learners' strategic skills in maintaining fluency along with their temporal fluency. A mixed-methods approach, including quantitative correlational analyses and qualitative analyses of raters' comments, was employed to obtain complementary perspectives to the nature of individual fluency, interactional fluency, and strategic competence in the context of peer interaction. Overall, the results of the study demonstrated connections across the rated dimensions and across fluency ratings and measures. The results related to interactional fluency were particularly compelling, as the mean length of turn pauses correlated strongly with interactional fluency ratings. The findings thus attest to the feasibility of extending L2

fluency research to interactional contexts. Furthermore, the problem-solving perspective adopted in the present study revealed that despite the lack of correlations between the strategic competence ratings and the measures of communication strategies, the raters nevertheless attended to the problem-solving aspects of the interaction, including collaborative efforts to overcome interactional trouble. The dimensions of strategic competence and interactional fluency were also found to be somewhat interrelated, which warrants further study on the perceptions of collaborative strategies in L2 peer interaction.

Overall, the study has demonstrated the feasibility of the mixed-methods approach and provided a starting point for developing L2 interactional fluency measurements and assessment criteria further. The study also provided the basis for developing assessment criteria for rating strategic competence from individual and collaborative perspectives in the context of L2 peer interaction. However, as the raters of the present study were mainly preservice teachers and relatively inexperienced raters of L2 oral proficiency, future studies could conduct similar ratings with other rater populations (e.g., language teachers, native speakers) to confirm the findings. The relatively small number of raters (n = 15) and rated samples (n = 6, from 12 learners) can also be viewed as limitations of the study, influencing the generalizability of the findings; future research should confirm the findings provided here with other samples and larger data sets.

In line with the recent suggestions to expand the scope of L2 fluency to cover also interactional aspects of fluency (e.g., Lintunen et al. 2020; Tavakoli 2016), the present study has provided ideas for exploring this aspect further in the future. Based on the present study, other aspects of L2 speaking could also be examined more widely based on dialogue data and from an interactional perspective. While interactional data provide opportunities for analyzing additional aspects as part of L2 fluency or L2 speaking more broadly, it also brings with it new challenges. Especially the possibility for joint scoring, as it crucially relates to, and even challenges, the tradition of individual performance assessment in L2 testing (see also Roever and Kasper 2018), should be acknowledged more widely, including the context of L2 interactional fluency assessment. Applying Roever and Kasper's (2018: 332) notion that psycholinguistic–individualist and sociolinguistic–interactional approaches in L2 assessment "are not necessarily mutually exclusive" to L2 fluency assessment, individual aspects of fluency could still be assessed along with the more interactional aspects (interactional fluency and perhaps strategic competence; see also May 2009). Acknowledging both individual and interactional aspects of fluency is also important in L2 teaching: during speaking exercises, attention could be drawn to resources for maintaining fluency across turns to foster interactional skills along with individual speech fluency. Similarly, from the perspective of L2 fluency research examining interactional data, the results of the

present study suggest that measures capturing collaborative, between-turn aspects should be used along with measures for individual within-turn aspects to capture the co-constructed nature of fluency in interaction. In the future, interdisciplinary collaboration between researchers studying L2 fluency and L2 interactional competence, on the one hand, and between SLA and language testing researchers, on the other hand, is essential for developing the assessment and analysis practices of L2 speaking and L2 interactional competence further.

Appendix 1 Rating sheet

The instructions were translated by the author from the Finnish originals. Only the rating sheet for sample 1 is presented here; the sheets for other samples (2–6) were identical.

Instructions: Rate the six dialogue samples numerically according to the four criteria. Comment on your ratings for fluency, interactional fluency, and strategic competence in the boxes below the numeric scales. For oral proficiency, the numeric rating is enough.

Sample 1 (01-G-004-005)

I) Oral proficiency

Rate the oral proficiency of the speaker on the left on a scale from 1 to 9 (1 = extremely weak oral proficiency, 9 = extremely good oral proficiency). Mark the number in bold.

1 2 3 4 5 6 7 8 9

Rate the oral proficiency of the speaker on the right on a scale from 1 to 9 (1 = extremely weak oral proficiency, 9 = extremely good oral proficiency). Mark the number in bold.

1 2 3 4 5 6 7 8 9

II) Fluency

When rating fluency, pay attention to the temporal aspects of speech: speech rate, pausing (silent pauses and filled pauses, e.g. *um*) and repair (for instance reformulations, repetitions). Accuracy and complexity should not be a part of fluency ratings.

Rate the fluency of the speaker on the left on a scale from 1 to 9 (1 = extremely disfluent, 9 = extremely fluent). Mark the number in bold.

1 2 3 4 5 6 7 8 9

Rate the fluency of the speaker on the right on a scale from 1 to 9 (1 = extremely disfluent, 9 = extremely fluent). Mark the number in bold.

1 2 3 4 5 6 7 8 9

Which aspects did you pay attention to when rating the sample? Comment on and justify the ratings in the box below (note that you can enlarge the box if needed). The ratings should be based on concrete features of speech.

III) Interactional fluency

Rate the pair's interactional fluency on a scale from 1 to 9 (1 = extremely disfluent, 9 = extremely fluent). Mark the number in bold.

1 2 3 4 5 6 7 8 9

Which aspects did you pay attention to when rating the sample? Comment on and justify the rating in the box below. The rating should be based on concrete features of speech.

IV) Strategic competence

When rating strategic competence, pay particular attention to the use of communication strategies (for instance paraphrases).

Rate the strategic competence of the speaker on the left on a scale from 1 to 9 (1 = extremely weak strategic competence, 9 = extremely good strategic competence). Mark the number in bold.

1 2 3 4 5 6 7 8 9

Rate the strategic competence of the speaker on the right on a scale from 1 to 9 (1 = extremely weak strategic competence, 9 = extremely good strategic competence). Mark the number in bold.

1 2 3 4 5 6 7 8 9

Which aspects did you pay attention to when rating the sample? Comment on and justify the ratings in the box below. The ratings should be based on concrete features of speech.

References

Bachman, Lyle F. & Adrian S. Palmer. 1996. *Language testing in practice*. Oxford: Oxford University Press.

Boersma, Paul & David Weenink. 2015. Praat: Doing phonetics by computer. Version 6.0.07, http://www.praat.org/.

Borger, Linda. 2019. Assessing interactional skills in a paired speaking test: Raters' interpretation of the construct. *Apples – Journal of Applied Language Studies* 13(1). 151–174.

Bosker, Hans Rutger, Anne-France Pinget, Hugo Quené, Ted Sanders & Nivja H. De Jong. 2013. What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing* 30(2). 159–175.

Brand, Christiane & Sandra Götz. 2011. Fluency versus accuracy in advanced spoken learner language: A multi-method approach. *International Journal of Corpus Linguistics* 16(2). 255–275.

Chambers, Francine. 1997. What do we mean by fluency? *System* 25(4). 535–544.

Council of Europe. 2001. *Common European framework of reference for languages. Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Cucchiarini, Catia, Helmer Strik & Lou Boves. 2002. Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America* 111(6). 2862–2873.

De Jong, Nivja H. 2018. Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly* 15(3). 237–254.

De Jong, Nivja H. & Ton Wempe. 2009. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41(2). 385–390.

Derwing, Tracey M., Marian J. Rossiter, Murray J. Munro & Ronald I. Thomson. 2004. Second language fluency: Judgments on different tasks. *Language Learning* 54(4). 655–679.

Dörnyei, Zoltán & Judit Kormos. 1998. Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in Second Language Acquisition* 20(3). 349–385.

Dörnyei, Zoltán & Mary Lee Scott. 1997. Communication strategies in a second language: Definitions and taxonomies. *Language Learning* 47(1). 173–210.

Ducasse, Ana Maria & Annie Brown. 2009. Assessing paired orals: Raters' orientation to interaction. *Language Testing* 26(3). 423–443.

Freed, Barbara F. 1995. What makes us think that students who study abroad become fluent? In Barbara F. Freed (ed.), *Second language acquisition in a study abroad context*, 123–148. Amsterdam & Philadelphia: John Benjamins.

Galaczi, Evelina D. 2014. Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics* 35(5). 553–574.

Galaczi, Evelina D. & Lynda Taylor. 2018. Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly* 15(3). 219–236.

Ginther, April, Slobodanka Dimova & Rui Yang. 2010. Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing* 27(3). 379–399.

Götz, Sandra. 2013. *Fluency in native and nonnative English speech*. Amsterdam & Philadelphia: John Benjamins.

Housen, Alex, Folkert Kuiken & Ineke Vedder (eds.). 2012. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam & Philadelphia: John Benjamins.

Iwashita, Noriko, Annie Brown, Tim McNamara & Sally O'Hagan. 2008. Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics* 29(1). 24–49.

Jarvis, Scott. 2017. Grounding lexical diversity in human judgments. *Language Testing* 34(4). 537–553.

Kahng, Jimin. 2014. Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning* 64(4). 809–854.

Kormos, Judit & Mariann Dénes. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32(2). 145–164.

Lemhöfer, Kristin & Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods* 44. 325–343.

Lennon, Paul. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 40(3). 387–417.

Lennon, Paul. 2000. The lexical element in spoken second language fluency. In Heidi Riggenbach (ed.), *Perspectives on fluency*, 25–42. Ann Arbor, MI: University of Michigan.

Lintunen, Pekka, Maarit Mutta & Pauliina Peltonen (eds.). 2020. *Fluency in L2 learning and use*. Bristol: Multilingual Matters.

Magne, Viktoria, Shungo Suzuki, Yui Suzukida, Kazuya Saito, Meltem Ilkan & Mai Tran. 2019. Exploring the dynamic nature of second language listeners' perceived fluency: A mixed-methods approach. *TESOL Quarterly* 53(4). 1139–1150.

May, Lyn. 2009. Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing* 26 (3). 397–421.

McCarthy, Michael. 2010. Spoken fluency revisited. *English Profile Journal* 1(1). 1–15.

Peltonen, Pauliina. 2017a. Temporal fluency and problem-solving in interaction: An exploratory study of fluency resources in L2 dialogue. *System* 70. 1–13.

Peltonen, Pauliina. 2017b. L2 fluency in spoken interaction: A case study on the use of other-repetitions and collaborative completions. In Mikko Kuronen, Pekka Lintunen & Tommi Nieminen (eds.), *Näkökulmia toisen kielen puheeseen – Insights into second language speech*, 118–138. Jyväskylä: Finnish Association of Applied Linguistics AFinLA.

Peltonen, Pauliina. 2020. *Individual and interactional speech fluency in L2 English from a problem-solving perspective: A mixed-methods approach*. Turku: University of Turku.

Plonsky, Luke & Frederick L. Oswald. 2014. How big is "big"? Interpreting effect sizes in L2 research. *Language Learning* 64(4). 878–912.

Préfontaine, Yvonne & Judit Kormos. 2016. A qualitative analysis of perceptions of fluency in second language French. *International Review of Applied Linguistics in Language Teaching* 54(2). 151–169.

Préfontaine, Yvonne, Judit Kormos & Daniel Ezra Johnson. 2016. How do utterance measures predict raters' perceptions of fluency in French as a second language? *Language Testing* 33(1). 53–73.

Révész, Andrea, Monika Ekiert & Eivind Nessa Torgersen. 2016. The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics* 37(6). 828–848.

Riggenbach, Heidi. 1991. Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes* 14(4). 423–441.

Roever, Carsten & Gabriele Kasper. 2018. Speaking in turns and sequences: Interactional competence as a target construct in testing speaking. *Language Testing* 35(3). 331–355.
Rossiter, Marian J. 2009. Perceptions of L2 fluency by native and non-native speakers of English. *The Canadian Modern Language Review* 65(3). 395–412.
Salaberry, Rafael M. & Silvia Kunitz. 2019. Introduction. In Rafael M. Salaberry and Silvia Kunitz (eds.), *Teaching and testing L2 interactional competence: Bridging theory and practice*, 1–22. New York: Routledge.
Sato, Masatoshi. 2014. Exploring the construct of interactional oral fluency: Second language acquisition and language testing approaches. *System* 45. 79–91.
Segalowitz, Norman. 2010. *The cognitive bases of second language fluency*. New York: Routledge.
Skehan, Peter. 2009. Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics* 30(4). 510–532.
Suzuki, Shungo & Judit Kormos. 2020. Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition* 42(1). 143–167.
Tavakoli, Parvaneh. 2016. Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching* 54(2). 133–150.
Taylor, Lynda & Gillian Wigglesworth. 2009. Are two heads better than one? Pair work in L2 assessment contexts. *Language Testing* 26(3). 325–339.
Van Batenburg, Eline S. L., Ron J. Oostdam, Amos J. S. Van Gelderen & Nivja H. De Jong. 2018. Measuring L2 speakers' interactional ability using interactive speech tasks. *Language Testing* 35(1). 75–100.
Van Os, Marjolein, Nivja H. De Jong & Hans Rutger Bosker 2020. Fluency in dialogue: Turn-taking behavior shapes perceived fluency in native and non-native speech. *Language Learning* 70(4). 1183–1217.
Witton-Davies, Giles. 2014. *The study of fluency and its development in monologue and dialogue*. Lancaster: Lancaster University dissertation.