Theory and Practice of Second Language Acquisition
vol. 11 (1), 2025, pp. 1/22
https://doi.org/10.31261/TAPSLA.15946

Zdeňka Neumanová
https://orcid.org/0000-0002-1234-4885
Department of English and American Studies
Masaryk University, Brno, Czechia

An Investigation of Complexity, Accuracy, and Fluency in the Speech of EFL Learners

Abstract

This study investigates the relationship between several quantitative measures of L2 speech complexity, accuracy, and fluency (CAF) by comparing the oral productions of English L2 learners at different proficiency levels. Forty English as a Foreign Language (EFL) learners of varied proficiency levels performed a simple oral task. The performance of the L2 learners was analyzed regarding linguistic complexity, accuracy, and fluency to investigate the interplay of the CAF dimensions and how language proficiency levels (A2, Bl, and B2) relate to these dimensions in the speech of EFL learners.

This study confirms the prevailing view that the three CAF dimensions are interconnected and that the CAF triad is a useful and valid way to investigate and describe L2 performance development. The results indicate that all three dimensions of CAF significantly predict L2 learners' oral proficiency, that is, from the developmental perspective, L2 learners' output is produced with higher complexity, accuracy, and fluency as they progress to a higher proficiency level. Moreover, the current paper discusses ways to measure CAF. Specific measures of complexity (index of developmental levels), accuracy (syntactic error rate), and fluency (articulation rate, and dysfluency rate) were identified as effective discriminators between proficiency levels.

Keywords: accuracy, complexity, fluency, L2, proficiency

During the course of the current study, the focus will be on three proficiency dimensions: complexity, accuracy, and fluency (CAF), and their interplay at the group level. The purpose of this study is to investigate how complexity, accuracy, and fluency interact in English learners' second language (henceforth referred to as L2) development. Motivated by one of the principal questions in second language acquisition (SLA) research relating to the nature of linguistic changes taking place in the L2 system of the learners as they become more

proficient, the current study investigates the linguistic dimensions underlying complexity, accuracy, and fluency in the context of L2 speech produced by 40 Czech-speaking learners of English. It aims to resolve the multifaceted processes that underlie language proficiency, contributing to its broader understanding in the L2 context. L2 speech production samples were analyzed as manifestations of what Czech EFL learners are/are not capable of doing in their development, and provide a clear picture of the underlying L2 systems, which cannot be studied directly. By and large, empirical evidence gathered for the current study is relevant from a descriptive point of view and can contribute to the formulation of explanatory hypotheses regarding L2 development.

Moreover, being able to objectively measure progress, the study will examine the relationship between language proficiency levels and complexity, accuracy, and fluency and inform language teaching practices and curriculum development.

A methodological challenge in CAF research is, however, the appropriate selection of CAF measures. Michael (2017) suggests that researchers use some of the measures employed in previous studies to ensure comparability with previous findings and that these measures are supplemented by context-specific measures that take into account the specific characteristics of the research context. Therefore, the current study also aims to examine which specific measures of complexity, accuracy, and fluency are best able to discriminate between proficiency levels.

Furthermore, even though I accept the status of complexity, accuracy, and fluency as distinct dimensions of L2 performance and proficiency, I do not exclude the fact that they can be interconnected and that they may somehow interact in L2 production. Moreover, if we only examine the dimensions one by one, we miss their interaction and the fact that the way they interact changes with time as well (Larsen-Freeman, 2009, p. 582). For this reason, the present explanatory study also aims to broaden our understanding of the interplay of complexity, accuracy, and fluency in the speech of young adults studying at a university.

In this article, I first provide a theoretical and methodological overview of previous research and the CAF framework. This is followed by a description of research procedures and a presentation of the findings. Next, the results of the current research are discussed. The article concludes by highlighting the underlying relationship between the CAF dimensions and L2 proficiency and outlining future directions for research.

To address the aforementioned conceptual and methodological challenges, the present study addresses the following research questions:

RQ1: Can all three dimensions—complexity, accuracy, and fluency—significantly predict L2 learners' oral proficiency?

RQ2: Which measures are best able to discriminate between proficiency levels?

RQ3: What are the interactions between complexity, accuracy, and fluency?

Theoretical Part

In this section, I first present the individual aspects that have traditionally been described, albeit to varying degrees, in the literature. I then turn to the more recently uncovered aspects of the CAF triad.

Proficiency

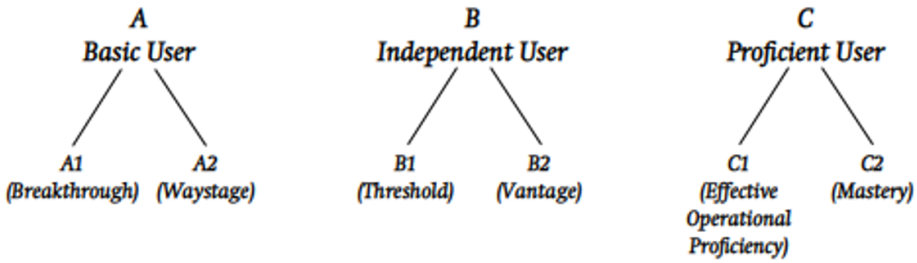
In the fields of SLA and foreign language learning (FLL), measuring L2 proficiency, defined as the ability to draw on and use competence in different tasks (Taylor, 1988, p. 166; R. Ellis, 2008, p. 976), is one of the key concerns, where the three dimensions of the CAF triad are widely recognized as the key components. Ortega (2014, p. 197) argues that "mastery and accuracy, that is, teleological arriving to isomorphic conformity with idealized native speaker norms," is the only thinkable way of defining linguistic and developmental success.

Proficiency has been interpreted in various ways in L2 research (for a discussion of how to define L2 proficiency, see Hulstijn, 2011). It is usually defined rather implicitly and its operationalisation differs across studies (Ortega, 2003). Studies have relied on institutional criteria which serve as a guarantee of EFL learners' proficiency (Gráf, 2015), different exam types, such as the Key English Test (KET) (De Felice & Pulman, 2009), or operationalised proficiency in terms of CEFR levels (Gyllstad et al., 2014).

The Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001a) is an established benchmark for language competence (Jones & Saville, 2009). It comprises six levels of competence (A1 to C2, as shown in Figure 1), which have become a common currency in language education, prevalent in curricula, syllabuses, textbooks, and teacher training courses (Anderson, 2007, p. 660).

Figure 1

The CEFR's Initial Division into Three Broad Levels—A, B and C (Council of Europe, 2001, p. 23)



The CEFR has established statements defining what is required for each stage within the framework and is based on performance-based "can-do statements," or "Reference Level Descriptors," evolved from the collective judgments of a body of experts (Van Ek & Trim, 1991a, 1991b, 2001), as depicted quite clearly in Figure 2.

Figure 2

An Illustrative Scale for Grammatical Accuracy (Council of Europe, 2001, p. 114)

	GRAMMATICAL ACCURACY
C2	*Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).*
C1	*Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot.*
B2	*Good grammatical control; occasional 'slips' or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.*
	Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding.
B1	*Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influence. Errors occur, but it is clear what he/she is trying to express.*
	Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.
A2	*Uses some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/she is trying to say.*
A1	*Shows only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.*

For empirical detail about levels of competence in learner English, see the English Profile (Harrison & Barker (2015), for example, introduce the EP program and discuss its latest findings). Moreover, its sub-project, The English Grammar Profile (EGP) resource, is a database of over 1,200 empirically derived statements and provides great detail on learner grammar competence.

CAF

Many researchers believe in the multi-componential nature of L2 performance, which can be captured by the notions of complexity, accuracy, and fluency (Ellis, 2003; Ellis & Barkhuizen, 2005). According to Skehan (2009), successful performance has been characterized as containing: (1) more advanced language, leading to complexity; (2) a concern to avoid errors, leading to higher accuracy if achieved; and (3) the capacity to produce speech at a normal rate and without interruptions, resulting in greater fluency (p. 510).

While the early working definitions of CAF are essentially used—complexity related to the size, richness, and diversity of L2 performance; accuracy as the measure for error-free language use; and fluency referring to the smooth production of speech with a limited number of hesitations and pauses (Michel, 2017)—many studies have featured CAF as variables that reflect the effect of other factors on language production (Housen et al., 2012, p. 2). These studies have included factors such as age (Mora, 2006), task design (Robinson et al., 2009), task complexity (Robinson, 2011), pre-task planning (Pang & Skehan, 2014), or the relationships among the CAF measures and learner oral proficiency (Miyamoto, 2019). Some studies have used CAF to examine longitudinal learner trajectories (Gunnarsson, 2012; Ferrari, 2012; Hokamura, 2018).

Complexity

Housen et al. (2012, p. 4) describe complexity as a palimpsest, a term that has acquired several closely related meanings in the context of SLA. For example, Ellis and Barkhuizen (2005, p. 139) define complexity as "the use of more challenging and difficult language," while Biber et al. (2011, p. 6) refer to "the more advanced grammatical structures that students exhibit as they progress in their language proficiencies." The second definition highlights the ability to employ a range of sophisticated structures and lexical items, defining complexity based on empirical observations of L2 production. Although Bulté and Housen (2012) propose a comprehensive typology of subdimensions of linguistic complexity, the current study adheres to an objective and quantitative definition of complexity. Manifestations of complexity in L2 production include syntactic structures and grammatical morphemes (for more details, see: Data and Methods).

Accuracy

Accuracy is often defined as the ability to produce error-free language (Ellis, 2008; Housen & Kuiken, 2009; Polio & Shea, 2014) and is largely associated with learner's linguistic knowledge representations. Palloti (2009, p. 592) mentions that "accuracy is perhaps the simplest and most internally coherent construct, referring to the degree of conformity to certain norms."

Although researchers often criticize the study of accuracy and mention that it is difficult to define an error (Gilquin & De Cock, 2011, p. 142), the existing definitions seem to concur in defining errors as deviations from a particular norm (Wolfe-Quintero et al., 1998). Straightforward though this general characterization seems, the difficulty seems to lie in selecting the criteria for evaluating accuracy and identifying errors, including the choice of the appropriate norm, that is, whether the criteria should be tuned to prescriptive standard norms or non-standard usages acceptable in some social contexts as well (Ellis, 2008; Polio, 1997).

The fact that the same language may have several normative standards or that raters might not always agree on what is accurate (Kuiken & Vedder, 2014) adds another layer to the discussion on the characterization and measures of accuracy. Moreover, "even if there was agreement regarding the norm, there remains the question of how 'far away' a deviation from this chosen norm is" (Michel, 2017, p. 9). All in all, "anyone who has worked on assessing accuracy in L2 data will know this only too well; some degree of personal judgment has to be invoked occasionally" (Foster & Wigglesworth, 2016, p. 112).

To avoid subjectivity, the current study focused on syntactic errors identified and tagged by two raters following Granger et al.'s (2022) error tagging manual and checked for inter-rater reliability (see: Data and Methods). Moreover, accuracy was gauged using a global measure (an error rate, i.e., number of errors per 100 words) for the purpose of this paper.

Having explored the intricate natures of language complexity and accuracy, which are both largely associated with learners' linguistic knowledge representations, we now turn to the third element of the triad: fluency, which is a performance phenomenon representing the outcome of psycholinguistic processing (Lennon, 1990).

Fluency

Language researchers have frequently analyzed oral production data to determine which linguistic phenomena contribute to L2 speech fluency (see, e.g., Kormos & Dénes, 2004), mainly because L2 speakers are often concerned with maintaining fluency since failure to do so can lead to a loss of a listener's attention and their face (Lennon, 2000).

Current research suggests that speech fluency is a multi-componential construct with various sub-dimensions, such as speed fluency, breakdown fluency, or repair fluency (Tavakoli & Skehan, 2005). According to Wood (2012, p. 9), fluency is often used as a synonym for the effective spoken use of a language, while Ellis and Barkhuizen (2005, p. 139) define fluency as "the production of language in real time without undue pausing or hesitation." Skehan (2009) adds that fluency is the capacity to produce speech at a normal rate and without interruption, and Segalowitz (2010) distinguishes between cognitive (the smoothness of the underlying processes), utterance (acoustically measurable aspects of performance), and perceived (the speaker's fluency impression on the hearer) facets of fluency. What these definitions emphasize is that we associate fluency not only with a quantifiable dimension for describing language performance but also with an impression made on the hearer. However, many conversational traits arise from the fact that conversation is typically spontaneous, that is, it is characterized by what has been called normal dysfluency, such as pauses, hesitators (er, um), and repeats. These impair a speaker's turn when the planning needs to catch up (Biber et al., 1999, p. 1048).

The current study examined some of the most typical characteristic features of spoken discourse (Biber et al., 1999; Wood, 2012), namely: silent pauses, filled pauses, repair, and repeats.

CAF at Different Proficiency Levels

The current article investigates complexity, accuracy, and fluency at different levels of CEFR. Larsen-Freeman (2009, p. 582) points out that examining these dimensions individually overlooks their interaction, which can change over time.

Some notable research focusing on the relationship between L2 proficiency and the CAF constructs sheds light on their interconnectedness and mutual influence. Some significant studies analyzed complexity and accuracy development (Polat & Kim, 2013; Spoelman & Verspoor, 2010, to name only a few), while other studies looked at the development of all three dimensions from a dynamic point of view. Larsen-Freeman (2006), for example, observed five Chinese adult learners over a six-month period. The average results showed progress in development in every dimension (although with great variation between subjects). Moreover, Iwashita et al. (2008) argue that features from grammatical accuracy and complexity, vocabulary, pronunciation, and fluency help distinguish overall levels of performance, with particular features of vocabulary and fluency having the strongest impact.

Finally, studies on information processing theory (Robinson, 2005; DeKeyser, 2007) and intra-individual variability (Spoelman & Verspoor, 2010) highlight

the non-linear nature of L2 development and the complex interaction between accuracy and complexity. Barrot and Gabinete (2021) suggest that complexity, accuracy, and fluency in writing are influenced not only by proficiency level but also by learners' L1 background, while Kowal (2018) found interplay between all three constructs.

Overall, these studies suggest correlations between and among the L2 proficiency and the components of complexity, accuracy, and fluency.

Data and Methods

The present study adopts a quasi-longitudinal design and focuses on the analysis of speech produced by Czech university learners of English.

The analyzed corpus consists of transcribed interviews with 40 first-year university learners of English (recorded over a period spanning from 2020 to 2021) who were approximately 20 years old on average (SD = 1.4).

L2 speech production samples analyzed in the current study cover a range of proficiency levels, falling within the A2, B1, and B2[1] levels, which were assessed using The Oxford Placement Test (OPT) before the individual interviews.

Table 1

Descriptive Statistics of the Transcribed Oral Performances

Category	Word count		
	A2	B1	B2
Minimum	467	416	746
Maximum	1091	1305	1571
Mean	724	800	1014
Median	723	777	922
SD	194	258	272

The interviews, each of which took approximately 15 minutes, consisted of three parts: a topic discussion, a picture description, and free conversation. First, the respondents were asked to choose a topic[2] and speak about it for three minutes without any interruptions. The first section of the oral interview was stimulated by the interviewer's question (*What topic have you chosen?*) when

[1] A2 (13 learners), B1 (15 learners), B2 (12 learners).

[2] Topic 1: A film you have seen/ a book you have read and think is particularly good/bad. Topic 2: A place or a country you have visited and liked.

the recording was started.[3] The learners were supposed to produce an extensive response. The second task was based on picture description, while the last task (free conversation) started with the students introducing themselves to enable them to talk about something familiar. Then, the interviewer posed topical scripted questions (e.g., *What can you tell me about your family?* or *Do you think English will be useful for you in the future?*), which were mostly concerned with familiar topics and with learning English. It should go without saying that some of the benefits of this task type can be re-constructed as weaknesses with lower-level learners. The process was rather free-flowing and indeterminate with talkative and accurate learners while less talkative (and less accurate) learners were often guided by prescripted questions.

In the present study, task performances were transcribed and analyzed, focusing on measures of complexity, accuracy, and fluency. To ensure comparability with previous research and address the research objectives, several general measures were assigned to each construct, taking into account their validity based on relevant studies (Bosker et al., 2013; Mostafa et al., 2020; Noris & Ortega, 2009; Polio & Shea, 2014; or Shungo & Kormos, 2020), and research objectives.

The study examined both lexical and syntactic complexity. Following Shungo and Kormos (2020), lexical complexity was measured as lexical density, which refers to the proportion of content words to the total words produced. This measure was computed using LexTutor (Cobb, 2011). In addition to lexical complexity, syntactic complexity was assessed using the Index of Developmental Levels (henceforth referred to as IDL), following the approach of Mostafa et al. (2020). They argue that IDL is a better predictor of L2 oral proficiency than widely-used structural complexity measures (mean length of AS-unit and subordination measure). By employing this measure of complexity, the study aims to examine the learners' ability to use advanced syntactic structures in their speech production. Table 2 likely presents the details of the IDL calculations for different morphemes and syntactic forms.

Following Mostafa et al. (2020), different weights were assigned to forms based on their developmental levels. Therefore, in the score sheets, 1 or 2 points (representing single or double occurrences of a form) were multiplied by the respective developmental levels of those forms. In other words, the EFL learners' total IDL scores increased as they produced more developmentally advanced target forms.

[3] Utilizing the built-in recording capabilities of a smartphone and a computer.

Table 2

Developmental Levels and Stages for the Acquisition of the English Morphemes, Negation, Questions, and Relative Clauses

Developmental levels for English morphemes	Developmental stages for the acquisition of English negation	Developmental levels for English questions	Acquisitional Order of Relative Clauses
Morpheme (level)	Item (level)	Item (level)	Relative clause type (level)
Ing (1) Plural -s (1) Be copula (1) Be auxiliary (2) a/the (2) Irregular past (3) Regular past -ed (4) Third person -s (4) Possessive -s (4)	Preverbal negation with no/not (1) Preverbal negation with don't (2) Postverbal negation in restricted contexts (3) Postverbal negation in all contexts (4)	Words and fragments with rising intonation (1) Canonical word order with rising intonation (2) Fronting of a questioning element (3) Inversion in two restricted contexts (4) Inversion expands to full range of target-like contexts (5) Negative questions/Question tags/ Questions in embedded clauses (6)	Subject (1) Direct object (2) Indirect object (3) Object of preposition (4) Genitive (5) Object of comparison (6)

Note. Separate IDL score sheets for morphemes, questions, negations, and relative clauses were developed for each participant. The participants received 1 point for producing a target form once, and they received 2 points for producing that form twice. For each type of form, the two examples had to be sufficiently dissimilar to be awarded the full points. They did not receive any additional point for producing a target form more than twice. In counting the IDL scores, grammatically acceptable forms were considered.

Regarding accuracy, a general measure known as the syntactic error rate (SER, i.e., the average number of syntactic errors per 100 words) was used. This measure was chosen for its comparability with previous research and its validity. A British English native speaker was trained to identify all syntactic errors (errors that contravene general rules of English grammar), which were categorized into sub-categories including determiners, articles, nouns, pronouns, adjectives, adverbs, verbs, word class, as described in Granger et al. (2022). To illustrate some of the syntactic errors observed, several examples were selected from an interview. Among other errors made by learner TT05, there were errors in the use of articles (GA), noun number (GNN), verb tense (GVT), verb number (GVN), personal pronouns (GPP), and errors involving independent prepositions (LSPR). Examples (1) to (4) demonstrate these errors.

1) ... and sight-seeing it's <GA corr="a very nice place"> very nice place </GA> with mountains and Alps it's <GA corr="a very good place"> very good place </GA> for young people maybe and old people... TT05
2) <B> Four <GNN corr="months"> month </GNN> in <name of a town> in <GA corr="a hotel"> hotel </GA> </B> TT05

3) ... I <GVT corr="visited"> visit </GVT> with my family and my family <GVN corr="don't"> doesn't </GVN> speak English and I must translate everything TT05
4) ...and <GPP corr="the people"> they people </GPP> are <LSPR corr="from other"> other </GPP> other <GNN corr="nations"> nation </GNN> I can see two girls and two boys... TT05

To assess the reliability of the accuracy measure, the author of this study examined a randomly selected 20% of the data. The inter-rater reliability, measured using Cohen's kappa, was found to be 0.92, indicating a high level of agreement (96.23%). This level of agreement can be considered excellent (Fleiss et al., 2013).

In the current study, no distinction was made between errors and mistakes (Ellis, 1997).[4] Any deviations from the norms of English grammar in terms of syntax were counted as errors. Additionally, omitted grammatical forms were counted as errors. However, self-corrections made by EFL learners after committing an error were not counted.

Finally, fluency, despite its multifaceted nature and a myriad of definitions, is defined in this study as "the production of language in real time without undue pausing or hesitation" (Ellis & Barkhuizen, 2005, p. 139). Drawing from Shungo and Kormos (2020), three subcomponents of utterance fluency were identified: speed, breakdown, and repair fluency. Unfilled pauses, defined by Bosker et al. (2013) as silence longer than 250 milliseconds, were manually coded by the researcher using Praat (Boersma & Weenink, 2012), marking the boundaries of clauses and pauses.

Speed fluency was assessed using the articulation rate (henceforth referred to as AR), calculated as the mean number of words per second, and divided by the total speech duration excluding pauses. Breakdown fluency was measured through the filled-pause ratio (FPR), which is the total number of filled pauses (e.g., ah, eh) divided by the total number of words. Repair fluency was determined by the dysfluency rate (DR), calculated as the mean number of dysfluencies (unfilled and filled pauses, repeats, and repairs) per minute, divided by the total speech duration (including pauses).

Table 3 provides a comprehensive overview of the specific measures of complexity, accuracy, and fluency.

[4] According to Ellis (1997), **errors** reflect gaps in the learners' linguistic knowledge (i.e., they do not know the correct form) while **mistakes** are occasional lapses in the learners' performance (which happens when the learners are not able to perform what they know).

Table 3

CAF Measurements

Measure	Calculation
Syntactic complexity	
Index of Developmental Levels (IDL)	separate IDLs
Lexical complexity	
Lexical density (LD)	content words/total words produced
Accuracy	
Syntactic error rate (SER)	syntactic errors/100words
Fluency	
Articulation rate (AR)	words per second/total speech duration
Filled pause ratio (FPR)	filled pauses/words
Dysfluency rate (DR)	dysfluencies per minute/total speech duration

To describe L2 speech production samples, descriptive statistics of linguistic measures were summarized in tables (see sections: Complexity in the Speech of EFL Learners; Accuracy in the Speech of EFL Learners; Fluency in the Speech of EFL Learners). The Shapiro-Wilk normality test suggested whether linguistic measures were normally distributed and where the Shapiro-Wilk normality test suggested that a linguistic measure was not normally distributed; nonparametric statistical tests were selected to correlate EFL learners' proficiency with specific linguistic measures. ANOVAs were performed to identify differences in the CAF measures among proficiency levels.

In the following section, I will provide an overview of the quantitative differences observed among the scrutinized proficiency levels.

Results

To address the first and second research questions, which investigate the predictive nature of complexity, accuracy, and fluency on L2 learners' oral proficiency, as well as the identification of the most discriminating measures between proficiency levels, the present study examined the relationship between these three dimensions and the learners' proficiency levels.

Initially, a comprehensive analysis was conducted, scrutinizing complexity, accuracy, and fluency individually in relation to the proficiency level of EFL learners. The Shapiro-Wilk normality tests indicated that three measures (IDL, FPR, and AR) were not normally distributed. As a result, nonparametric statistical tests were selected to assess the correlation between the learners' proficiency levels and the three measures. ANOVAs were performed to identify differences in the CAF measures among proficiency levels.

Furthermore, Pearson correlation analyses and Spearman correlation analyses were carried out between measures of complexity (LD and IDLs), accuracy (SER), and fluency (AR, FPR, and DR).

These analyses were conducted both overall and within each performance.

The following sections of the paper present descriptive and inferential statistics, as well as correlation analyses.

Complexity in the Speech of EFL Learners

As previously mentioned, lexical and syntactic complexity were assessed using lexical density (the ratio of content words to the total words produced, computed with LexTutor, Cobb, 2011) and an index based on the developmental levels of L2 morphological and syntactic forms (Index of Developmental Levels, IDL).

Before presenting the results, descriptive statistics for IDL and LD are provided in Table 4. Unlike IDL, which emerged as a significant positive predictor of L2 oral proficiency, lexical density did not prove to be a reliable predictor of proficiency. Specifically, B2-level learners exhibited significantly higher IDL scores compared to A2 and B1 levels, indicating that IDL scores increase as learners' proficiency levels improve.

Table 4

Descriptive Statistics—Complexity

Coefficient	CEFR level	Mean	SD
LD	A2	0.485	0.03
	B1	0.457	0.02
	B2	0.463	0.03
IDL	A2	26.31	11.10
	B1	46.13	18.52
	B2	60.25	12.02

Pairwise comparisons were conducted to detect differences between proficiency levels. The results indicated a significant overall effect on the proficiency level for IDL, as shown in Table 5.

Table 5

Pairwise Comparisons of Proficiency Levels

CEFR levels	IDL	
	Test	Sig.
A2–B1	–11.846	0.007
A2–B2	–21.346	0.000
B1–B2	–9.500	0.036

All in all, the results proved that L2 learners with higher proficiency levels used more difficult morphological and syntactic structures.

Accuracy in the Speech of EFL Learners

The corpus of 40 transcribed interviews was analyzed to identify errors that involved the violation of specific grammatical rules[5] (Granger et. al., 2022). Accuracy was measured using a general metric called syntactic error rate, which represents the mean number of syntactic errors per 100 words. For the corresponding data, refer to Table 6.

Table 6

Descriptive Statistics—Accuracy

Proficiency level	Mean	SD
A2	9.16	2.64
B1	6.91	2.31
B2	4.53	1.54
Total	6.93	2.85

The results of a detailed syntactic error analysis in the speech of Czech EFL learners revealed that by far the most frequent error type was errors in the use of articles, accounting for 30.8% of all errors. Verb tense was the second most frequent category (13.1%), followed by errors involving independent prepositions (10.4%). Interestingly, there did not appear to be significant differences in error types among the three proficiency levels, which shows that the use

[5] In their Error Tagging Manual, Granger et. al. (2022, p. 4) make an important distinction between errors (the breaking of a specific linguistic rule) and infelicities (instances of non-erroneous, but odd-sounding, language). Infelicities were not taken into account in the analysis.

of articles and verb tense is highly problematic even as the learners progress to a higher proficiency level.

Unfortunately, Granger et al. (2022) do not divide errors in the use of articles into subgroups. For this reason, it is rather difficult to determine which aspect of their use is the most problematic for Czech EFL learners. A more detailed analysis, however, revealed that omission of *a/an* in obligatory contexts accounts for 54.1% of all article errors, and that omission of *the* in obligatory contexts accounts for another 23.8% of all errors in the category of articles. Such a result seems to suggest that article omission in English might be the result of L1 transfer as Czech does not have articles.

As mentioned above, the analysis of the most frequent error types indicated similar frequencies across proficiency levels. Articles proved to be the most challenging grammatical feature for the majority of Czech EFL learners, including those in the A2–B2 proficiency groups. This was followed by verb tense and prepositions, which were found to be common sources of errors for Czech EFL learners across all three proficiency levels. For details on the types of errors made by advanced EFL learners, refer to Gráf (2015).

Moreover, to identify differences between proficiency levels, ANOVAs and pairwise comparisons were conducted. The results demonstrated a significant correlation between EFL learners' proficiency levels and syntactic error rate, indicating that B2 learners exhibited the highest level of accuracy, while A2 learners had the least accurate speech production. Refer to Table 7 for detailed information.

Table 7

Multiple Comparisons, LSD Test

Proficiency level		Mean difference	Sig.
A2	B1	2.2482	0.011
	B2	4.6282	0.000
B1	A2	–2.2482	0.011
	B2	2.3800	0.009
B2	A2	–4.6282	0.000
	B1	–2.3800	0.009

The current study proposes that syntactic accuracy on the part of EFL speakers could serve as a reliable predictor of language proficiency. The findings indicate a strong association between EFL learners' proficiency level and grammatical accuracy, which is significant for several reasons: (1) language assessors can use syntactic accuracy as a valuable indicator when evaluating the

proficiency of EFL learners, (2) educators may focus on designing instructional materials and activities that specifically target syntactic skills, (3) teachers can use this information to target their instruction to address syntactic challenges commonly faced by learners at different proficiency levels (for example, articles, or verb tense), (4) the finding encourages further investigation into the specific mechanisms through which syntactic skills impact overall proficiency.

Fluency in the Speech of EFL Learners

In this section of the paper, the analysis focused on examining the relationship between proficiency level and fluency in the speech of Czech EFL learners. Fluency was measured using three indicators: articulation rate (AR), filled pause ratio (FPR), and dysfluency rate (DR).

The statistical analysis revealed that both articulation rate ($p < 0.012$) and dysfluency rate ($p < 0.019$) exhibited a significant relationship with proficiency levels. Specifically, the results indicated that B2-level learners were statistically more fluent than A2 and B1 learners. Detailed information can be found in Table 8.

Table 8

Pairwise Comparisons

Proficiency levels	Test	Sig.
A2–B1	–4.503	0.307
A2–B2	–13.603	0.003
B1–B2	–9.100	0.043

Regarding the dysfluency rate, a significant difference was observed between A2 and B2 levels ($p < 0.005$). This indicates that B2-level learners exhibit a higher level of fluency compared to A2-level learners. For more information on fluency in the speech of EFL learners, see, for example, Huang and Gráf (2020), who compared speech rates and the frequency and location of unfilled pauses in the speech of native English speakers and learners of English. They found significant differences showing that between B2 and C1 levels the growth of proficiency is accompanied by an increase in speech rate and a decrease in the frequency of pausing, particularly within clauses and within constituents (Huang & Gráf, 2020, p. 57).

In summary, the findings indicate that all three dimensions of complexity, accuracy, and fluency (measured as SER, IDL, AR, and DR) significantly predict L2 learners' oral proficiency. Each dimension demonstrated at least

one coefficient that established a statistically significant relationship with EFL learners' proficiency levels. Therefore, the coefficients for syntactic error rate, index of developmental levels, articulation rate, and dysfluency rate can be considered as the measures that are most effective in discriminating between proficiency levels.

Interactions between Complexity, Accuracy, and Fluency

To address the third research question, which focused on the interactions between complexity, accuracy, and fluency, several correlation analyses were conducted separately for the measures of the three dimensions.

First, the correlation analysis between complexity (measured as IDL) and fluency (measured as AR) revealed a positive correlation between IDL and AR for the A2 proficiency level ($R = 0.574$, $p = 0.040$) and B1 proficiency level ($R = 0.651$, $p = 0.009$). This suggests that as learners used more complex morphological and syntactic structures, their oral production was also more fluent.

Second, negative correlations were observed between IDL and FPR for A2 ($R = –0.617$, $p = 0.025$) and B1 ($R = –0.527$, $p = 0.043$) proficiency levels, as well as between AR and FPR for the A2 ($R = –0.728$, $p = 0.005$) and B1 ($R = –0.692$, $p = 0.004$) proficiency levels. These findings indicate that as EFL learners produced more complex speech, their fluency increased while the filled pause ratio decreased.

In conclusion, there were no statistically significant correlations found between specific measures of complexity, accuracy, and fluency for B2-level learners, indicating that they, and more importantly all EFL learners across proficiency levels, can focus on all three dimensions simultaneously. Consequently, the findings confirm the interplay of complexity, accuracy, and fluency as interconnected systems. The development of these proficiency dimensions occurs concurrently.

Conclusions

The current research proposed measures for evaluating L2 learner oral performance and examined whether these predict L2 oral proficiency. It indicates that specific measures (for example, Index of Developmental Levels, based on developmental levels of L2 forms) better discriminate between learners of varied proficiency levels and provide empirical evidence that in oral performances, learners of higher proficiency use significantly more L2 forms that belong

to higher levels in developmental sequences than lower proficiency levels learners. As such, the current study provides empirical support "linking performance measures to the use of developmentally more advanced language" (Lambert & Kormos, 2014), which may have pedagogical significance. ESL teachers can, for example, gain insight into the current developmental levels of their learners for various L2 morpho-syntactic forms (by considering a measure such as IDL), or provide feedback on weak points of EFL learners, such as the use of articles and tenses, by considering syntactic error rate (SER) and examining the most frequent error types. These might be hard to detect for the speed with which learners produce speech, or might be wreathed in sophisticated avoidance strategies. Furthermore, the division of the respondents into proficiency levels helped identify the most problematic areas as those that occur throughout all of the proficiency bands.

Additionally, the analysis of accuracy highlights the practical significance of studying learner errors and has important implications for error correction, because understanding the nature of errors is crucial for effective error correction in language teaching (see, for example, Scrivener, 2005). Teachers need to have a thorough understanding of the types of errors made by learners to provide targeted and effective feedback. By identifying common error patterns and focusing on specific areas of difficulty, teachers can tailor their instruction and help learners improve their accuracy in language production.

Moreover, the current study contributes to the existing knowledge on the interplay between complexity, accuracy, and fluency in learner language by confirming the prevailing view that these three dimensions are interconnected. Furthermore, it demonstrates that complexity, accuracy, and fluency are influenced by proficiency level, aligning with the findings of previous research (Barrot & Gabinete, 2021). By and large, from the developmental perspective, it shows that L2 learners' output is produced with higher complexity, accuracy, and fluency when they progress to a higher proficiency level, that is, L2 learners with higher oral proficiency are more adept at using more advanced L2 forms, with fewer errors and dysfluencies. This suggests that as L2 learners develop proficiency, the structures they employ become more accurate, their speech becomes more fluent, and they tend to utilize L2 forms that belong to developmentally higher levels.

However, average group trajectories may not correspond to the developmental trajectory of a single individual subject (Larsen-Freeman, 2006). Thus, it is necessary to look at specific scores obtained by L2 learners individually and over more data collection points to reveal whether complexity, accuracy, and fluency follow separate developmental trajectories or not.

Finally, it is important to recognize that the components of the CAF triad are not isolated but interact with each other, as suggested by Larsen-Freeman (2006). The interplay between the three dimensions is multivariate and dynamic

(Spoelman & Verspoor, 2010, p. 547), meaning that changes in one dimension can affect the others. This highlights the complexity of language acquisition and the need for comprehensive research. To gain a deeper understanding of these relationships is necessary to conduct longitudinal and non-linear CAF research with a focus on difference and variation (Larsen-Freeman, 2009). For this reason, a more in-depth analysis of learner data, for example, a longitudinal one, would certainly reveal more interesting regularities.

References

Anderson, C. (2007). The CEFR and the need for more research. *Modern Language Journal*, *91*(4), 659–663.

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford University Press.

Barrot, J., & Gabinete, M. K. (2021). Complexity, accuracy, and fluency in the argumentative writing of ESL and EFL learners. *International Review of Applied Linguistics in Language Teaching, 59*(2), 209–232.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson.

Biber, D., & Conrad, S. (2001). Quantitative corpus-based research: Much more than bean counting. *TESOL Quarterly, 35*(2), 331–336.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5–35.

Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & de Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing, 30*, 159–175.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 23–46). John Benjamins.

De Felice, R., & Pulman, S. (2009). Automatic detection of preposition errors in learner writing. *CALICO Journal, 26*(3), 512–528.

DeKeyser, R. (2007). Skill acquisition theory. In B. Van Patten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (pp. 94–112). Lawrence Erlbaum.

Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.

Ellis, R. (1997). *Second Language Acquisition*. Oxford University Press.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ellis, R. (2008). *The Study of Second Language Acquisition (2nd edition*). Oxford University Press.

Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford University Press.

Ferrari, S. (2012). A longitudinal study of complexity, accuracy, and fluency variation in second language development. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy, and fluency in SLA* (pp. 277–298). John Benjamins Publishing Company.

Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions*. John Wiley & Sons.

Foster, P., & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a weighted clause ratio. *Annual Review of Applied Linguistics, 36,* 98–116.

Gilquin, G., & De Cock, S. (2011). Errors and disfluencies in spoken corpora: Setting the Scene. *International Journal of Corpus Linguistics, 16*(2), 141–172.

Gráf, T. (2015). *Accuracy and fluency in the speech of the advanced learner of English*. Univerzita Karlova v Praze.

Granger, S., Swallow, H., & Thewissen, J. (2022). *The Louvain Error Tagging Manual. Version 2.0.* CECL Papers 4. Centre for English Corpus Linguistics: Louvain-la-Neuve.

Gunnarsson, C. (2012). The development of complexity, accuracy, and fluency in the written production of L2 French. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy, and Fluency in SLA* (pp. 247–276). John Benjamins Publishing Company.

Gyllstad, H., Granfeldt, J., Bernardini, P., & Källkvist, M. (2014). Linguistic correlates to communicative proficiency levels of the CEFR. The case of syntactic complexity in written L2 English, L3 French and L4 Italian. *EUROSLA Yearbook, 14*, 1–30.

Harrison, J., & Barker, F. (2015). *English profile studies: English profile in practice*. Cambridge University Press.

Hokamura, M. (2018). The dynamics of complexity, accuracy, and fluency: A longitudinal case study of Japanese learners' English writing. *JALT Journal, 40*(1), 23–46.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in Second Language Acquisition. *Applied Linguistics*, *30*(4), 461–473.

Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy, and fluency in SLA*. John Benjamins Publishing Company.

Huang, L.-F., & Gráf, T. (2020). Speech rate and pausing in English: Comparing learners at different levels of proficiency with native speakers. *Taiwan Journal of TESOL, 17*(1), 57–86.

Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly, 8*(3), 229–249.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics, 29*(1), 24–49.

Jones, N., & Saville, N. (2009). European language policy: Assessment, learning, and the CEFR. *Annual Review of Applied Linguistics, 29*, 51–63.

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

Kowal, I. (2018). *The dynamics of complexity, accuracy, and fluency in second language development*. Jagiellonian University Press.

Kuiken, F., & Vedder, I. (2014). Rating written performance: What do raters do and why? *Language Testing, 31*(3), 329–348.

Lambert, C., & Kormos, J. (2014). Complexity, accuracy, and fluency in task based L2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics, 35*(5), 607–614.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of the Chinese learners of English. *Applied Linguistics, 27*(4), 590–619.

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 579–589.

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

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). University of Michigan Press.

Michel, M. C. (2017). Complexity, accuracy, and fluency in L2 production. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 50–68). Taylor & Francis.

Miyamoto, M. (2019). *Capturing L2 oral proficiency with CAF measures as predictors of the ACTFL OPI rating*. [Doctoral dissertation, Purdue University, School of Languages and Cultures, West Lafayette, Indiana, 2019].

Mora, J. C. (2006). Age Effect on Oral Fluency Development. In C. Muñoz (Ed.), *Age and the Rate of Foreign Language Learning* (pp. 65–88). Multilingual Matters.

Mostafa, T., Kim, Y., & Friginal, E. (2020). Examining a developmentally based measure of L2 oral performances: Does it predict L2 learners' oral proficiency? *System, 89*, 1–16.

Nation, I. S. P. (2009a). *Teaching ESL/EFL reading and writing*. Routledge.

Nation, I. S. P. (2009b). *Teaching ESL/EFL listening and speaking*. Routledge.

Nation, I. S. P. (2011). Research into practice: Vocabulary. *Language Teaching*, *44*(4), 1–11.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555–578.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*(4), 492–518.

Ortega, L. (2014). Trying out theories on interlanguage: Description and explanation over 40 years of L2 negation research. In Z. Han & E. Tarone (Eds.), *Interlanguage: Forty years later* (pp. 173–202). John Benjamins.

Palloti, G. (2009). CAF: Defining, refining, and differentiating constructs. *Applied Linguistics, 30*(4), 590–601.

Pang F., & Skehan, P. (2014). What do learners do when they plan? In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 95–128). John Benjamins.

Polat, B., & Kim, Y. (2013). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics, (35)*2, 184–207.

Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning, 47*, 101–143.

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing, 26*, 10–27.

Porcino, M. C., & Finardi, K. R. (2012). The tension between accuracy and fluency of L2 speech: Evidence from communicative tasks. *The ESPecialist, 33*(1), 25–44.

Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies, 21*(2), 45–107.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25*, 46–73.

Robinson, P. (2009). Syllabus design. In M. H. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 294–310). Blackwell Publishing.

Robinson, P. (2011). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–38). John Benjamins.

Rossiter, M. J., Derwing, T. M., Manimtim, L. G., & Thomson, R. I. (2010). Oral fluency: The neglected component in the communicative language classroom. *The Canadian Modern Language Review*, *66*(4), 583–606.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

Scrivener, J. (2005). *Learning teaching: A guidebook for English language teachers*. MacMillan Education.

Shungo, S., & Kormos, J. (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.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510–532.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in the development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics, 31*(4), 532–553.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). Benjamins.

Taylor, D. (1988). The meaning and use of the term 'competence' in linguistics and applied linguistics. *Applied Linguistics, 9*, 148–168.

Van Ek, J. A., & Trim, J. L. M. (1991a). *Waystage 1990*. Council of Europe/Cambridge University Press.

Van Ek, J. A., & Trim, J. L. M. (1991b). *Threshold 1990*. Council of Europe/Cambridge University Press.

Van Ek, J. A., & Trim, J. L. M. (2001). *Vantage*. Council of Europe/Cambridge University Press.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. University of Hawaii, Second Language Teaching and Curriculum Center.

Wood, D. (2012). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. Continuum.

Zhang, M. (2018). Collaborative writing in the EFL classroom: The effects of L1 and L2 use. *System, 76*, 1–12.