

STUDY

Linguistic complexity in second language writing: Insight from studies on task planning

Mark D. Johnson¹  and Mahmoud Abdi Tabari² 

¹East Carolina University, Greenville, USA and ²University of Nevada, Reno, USA

Corresponding author: Mark D. Johnson; Email: johnsonmark@ecu.edu

(Received 28 December 2022; revised 2 June 2024; accepted 16 June 2024)

Abstract

Task planning and its effect on the complexity of second language (L2) written production have been studied extensively. However, the results of these studies are inconclusive, and at times contradictory, potentially as a result of variation in metrics of linguistic complexity. This study is an extension of earlier research syntheses and quantitative meta-analyses on the effects of planning on oral and written L2 production. It examines the identification and selection of linguistic complexity metrics in previous research on planning and its subsequent effects on the linguistic complexity of written L2 production. This research-focused synthesis of studies surveys construct definitions and operational definitions of linguistic complexity in the research domain and provides an overview of rationales for metric selection in the included studies. Methodological implications for future research are discussed in light of the findings.

1. Introduction

As the most-frequently examined task complexity variable in task-based language teaching (TBLT)-informed second language (L2) writing research (Abdi Tabari, 2016, 2017, 2018; Ellis, 2021; Johnson, 2017, 2022), task planning has garnered considerable attention since Ellis (2005) introduced his model of task planning to the TBLT field. This construct has been appealing to TBLT scholars, evidenced by the large number of conceptual and empirical studies that have attempted to discover how task planning as an implementation variable, in tandem with task features as inherent variables in the task's topic or design, influence L2 writing performance and development. In particular, TBLT-informed studies have focused on the examination of task planning and its proposed effects on linguistic complexity, accuracy, and fluency (or the CAF triad), the three of which are multifaceted, multilayered, and multidimensional in nature (Housen et al., 2012; Norris & Ortega, 2009). The literature on TBLT highlights that findings from studies on task planning in L2 writing have shown inconsistencies, particularly concerning the impact of task planning on linguistic complexity.

These inconsistent results might stem from several factors. First, linguistic complexity has often not been fully defined at the level of the construct in previous task planning research. In other words, many researchers select metrics of linguistic complexity without fully defining what linguistic complexity is and how it might be recognized. Thus, operational definitions of linguistic complexity often take the place of construct definitions of complexity (Bulté & Housen, 2012). Second, there is a proliferation of diverse operationalizations of linguistic complexity within the TBLT domain (Long, 2014; Pallotti, 2022). Finally, L2 development and the response of the L2 writer to the demands of the genre of the writing task have often not been considered in the selection of linguistic complexity metrics (Abdi Tabari, 2023; Biber et al., 2022; Larsen-Freeman & Cameron, 2008; Lu, 2011; Norris &

Ortega, 2009). These salient factors have potentially obscured the effects of task planning on linguistic complexity, especially in the L2 writing domain, as no single stable finding has yet emerged.

To shed light on the nature of task planning and its effect on written L2 production, this research-focused synthesis (Chong & Plonsky, 2023) extends the work of earlier research syntheses and meta-analyses on the result of task planning on L2 written production (Ellis, 2021; Johnson, 2017; Johnson & Abdi Tabari, 2023) by examining the treatment of linguistic complexity in the domain.

2. Review of literature

2.1 Task complexity in L2 writing research

Given the centrality of linguistic complexity to TBLT-informed L2 writing research, this section situates planning within cognitive frameworks of task complexity and the theorized effect of task complexity on the linguistic complexity of L2 production. TBLT-informed L2 writing research is often situated in one of two models of cognitive task complexity – typically the Limited Attentional Capacity Model (Skehan, 1998, 2014; Skehan & Foster, 2001) or the Cognition Hypothesis (Robinson, 2001, 2003, 2005, 2011). However, both models have been criticized for prioritizing oral production (Manchón, 2014; Tavakoli, 2014). Thus, TBLT-informed L2 writing research often invokes Kellogg’s (1996) model of working memory in L1 writing to describe the role of cognition in L2 writing. The following provides a brief overview of these theoretical models.

2.1.1 The Limited Attentional Capacity Model

According to Skehan’s Limited Attentional Capacity Model (Skehan, 1998, 2014; Skehan & Foster, 2001), learners rely on two representations of the interlinguistic system when responding to a language task: (a) a rule-based representation, which is responsible for the generation of novel utterances, and (b) an exemplar-based representation, which fluently generates previously acquired automatized chunks of language. The Limited Attentional Capacity Model predicts that, when presented with a task that is novel or complex, the learner will consult the exemplar-based system as a method of buffering processing demands and facilitating the planning of the next utterance. Thus, fluency increases. However, this increase in fluency comes at the expense of complexity and accuracy. In contrast, tasks that are simple or familiar are thought to free attentional capacity so that the learner may consult the rule-based interlinguistic system to produce novel utterances, resulting in the production of language that is more complex or more accurate but less fluent.

2.1.2 The Cognition Hypothesis

According to the Cognition Hypothesis (Robinson, 2001, 2003, 2005, 2011), working memory/attention consists of multiple resource pools. A task can make demands on one or more attentional resource pools. But, so long as the complexity of a task does not place demands on the same attentional resource, language performance will not be affected. Robinson’s Triadic Componential Framework (2003, 2011) provides a means of describing the complexity of tasks on two axes (see Table 1): (a) a resource-directing

Table 1. Resource-directing and resource-dispersing features of cognitive task complexity (adapted from Robinson, 2011)

| Resource-directing features of cognitive task complexity | Resource-dispersing features of cognitive task complexity |
|--|---|
| ±Here-and-now | ±Planning time |
| ±Few elements | ±Prior knowledge |
| ±Spatial reasoning | ±Single task |
| ±Causal reasoning | ±Task structure |
| ±Intentional reasoning | ±Few steps |
| ±Perspective taking | ±Interdependency of steps |

axis of task complexity and (b) a resource-dispersing axis of task complexity. Resource-directing features of task complexity are thought to direct the learner's attention to the production of language that is more complex and more accurate – though less fluent – as a result of the cognitive demands of the task. Resource-dispersing features of task complexity are thought to distract the learner's attention from the production of language, resulting in decreased complexity, accuracy, and fluency.

One of the reasons, perhaps, that the Cognition Hypothesis has been accepted enthusiastically by L2 writing researchers is its conception of multiple attentional resources and the similarity of that conception to Kellogg's (1996) widely cited model of working memory and its involvement in the writing process. However, unlike Robinson's conception of attentional resources, Kellogg's model considers working memory to be a limited capacity system.

2.1.3 Working memory in first language (L1) writing

Kellogg's (1996) model of working memory in L1 writing links Baddeley's widely cited model of working memory (1986, 2007) to facets of the writing process. According to this model, writing systems and processes place varying demands on components of working memory (see Table 2). Of particular interest to a discussion of L2 writing is the formulation system, as this system comprises two further processes – planning and translating – which place considerable demands on all three components of working memory: (a) the visuospatial sketchpad, (b) the phonological loop, and (c) the central executive. Kellogg's early research (1987a, 1987b, 1990) has demonstrated that providing planning time prior to writing eases the demands of translation and results in increased fluency as well as increased ratings of language use and overall writing quality.

Given the high demands placed on working memory capacity by the formulation system, it is, perhaps, unsurprising that the majority of TBLT-informed L2 writing research has examined planning as a resource-dispersing feature of task complexity (Johnson, 2017; Johnson & Abdi Tabari, 2023).

2.2 Research on planning and L2 writing

As noted previously, much of the research on planning and its effect on the linguistic complexity of L2 written production has been informed by task-based theories of L2 learning and teaching, often in conjunction with Kellogg's (1996) model of working memory in L1 writing. Typically, such studies examine the effect of planning on the linguistic complexity of subsequent L2 production. However, the results of primary research on the effect(s) of planning on the linguistic complexity of L2 written production differ considerably. For example, a widely cited study by Ellis and Yuan (2004) reports a medium effect of pre-task planning on the syntactic complexity of L2 written production ($d = 0.73$) and a very large effect of pre-task planning on the syntactic variety of L2 written production ($d = 4.42$) among 42 learners of English as a foreign language. In terms of online planning – that is, planning that takes place during the composing process – Ellis and Yuan (2004) report small effects of online planning on the syntactic complexity of L2 written production ($d = 0.56$) and a medium effect of online planning on the syntactic variety of written L2 production ($d = 0.73$). In contrast, a

Table 2. Kellogg's (1996) model of working memory in L1 writing (adapted from Kellogg et al., 2013)

| Writing system | Writing process | Visuo-Spatial sketchpad | Central executive | Phonological loop |
|----------------|-----------------|-------------------------|-------------------|-------------------|
| Formulation | Planning | ✓ | ✓ | |
| | Translating | | ✓ | ✓ |
| Execution | Programming | | ✓ | |
| | Executing | | | |
| Monitoring | Reading | | ✓ | ✓ |
| | Editing | | ✓ | |

large-scale study by Johnson et al. (2012) found no effect of pre-task planning on the linguistic complexity of L2 written production among 918 learners of English as a foreign language. One potential explanation for the disparity in research findings is differences in the operational definition of linguistic complexity in each study. Ellis and Yuan (2004) examined the number of different grammatical verb forms and the ratio of clauses to T-units, while Johnson et al. (2012) examined Flesch reading ease and mean number of words preceding the main verb of clauses.

While the results of individual primary studies on the effect of planning on the complexity of L2 written production run the gamut, some trends are evident in secondary research on planning and its effect on the complexity of L2 written production (Ellis, 2021; Johnson, 2017; Johnson & Abdi Tabari, 2023). Johnson (2017) focused more broadly on the pooled effect(s) of task complexity features on the complexity of L2 written production. However, Johnson (2017) further combs apart comparisons according to task complexity features identified in Robinson (2011). In reporting the pooled effects of pre-task planning on the complexity of L2 written production, Johnson (2017) notes a small ($d = 0.42$) effect of pre-task planning on L2 written syntactic complexity among independent-samples comparisons and a small to medium effect of pre-task planning on the lexical complexity of L2 written production ($d = 0.65$) among dependent-samples comparisons. The author catalogues metrics of linguistic complexity used in the individual studies and provides some discussion of metric selection in the included studies, but the discussion is somewhat limited.

In a recent systematic review, Ellis (2021) provides a detailed review of studies examining the effect of planning on L2 written production, reporting individual effect sizes calculated from the results of the included studies. However, these effects are not aggregated. Thus, the reader is, again, left with highly variable results. Further, discussion of the various complexity metrics in the included studies is limited.

A more recent research synthesis and meta-analysis (Johnson & Abdi Tabari, 2023) pooled effects across included studies and reported consistent effects of pre-task planning on the syntactic complexity of written L2 production ($d = 0.60$). Consistent with previous syntheses, the authors catalogued metrics of linguistic complexity in the included studies, but they did not provide a critical analysis of those metrics, the rationale for their inclusion, and their ability to capture linguistic complexity at multiple levels.

Taken together, the results of the meta-analyses reviewed above suggest consistent effects of pre-task planning on the linguistic complexity of L2 written production. However, a more basic question seems to be unaddressed in the research domain: how are metrics of linguistic complexity selected by researchers in the domain, and what rationale is presented for their selection? In other words, what is linguistic complexity and how should it be measured?

2.3 Linguistic complexity

Bulté and Housen (2012, 2014) advocate for a narrow linguistic definition of complexity, emphasizing quantifiable linguistic features and their interplay. The authors argue against definitions of linguistic complexity that make reference to constructs such as development and proficiency. Rather, the authors argue that features of linguistic complexity, which may or may not be associated with development and/or proficiency, should be determined empirically.

In an early discussion of linguistic complexity and its operationalization in second language acquisition research, Norris and Ortega (2009) present an argument for the measurement of syntactic complexity as a multi-componential construct. Basing their argument on systemic functional linguistics, the authors argue that language development occurs in three distinct stages:

1. Expression of ideas through parataxis (i.e., coordination).
2. Expansion of hypotaxis through subordination (among other means).
3. Emergence of/reliance on grammatical metaphor expressed through increased nominalization.

Consistent with Bulté and Housen (2012, 2014), Norris and Ortega (2009) argue for the use of quantifiable metrics that capture the number and range of forms and the interaction of those

forms. Norris and Ortega (2009) further argue that metrics of syntactic complexity must be selected to capture development at each of the above stages, recommending metrics of syntactic complexity at four levels: (a) global, (b) variety of forms, (c) clausal, and (d) phrasal. More importantly, the authors argue that metrics of linguistic complexity cannot/should not be considered in isolation as development in the L2 is dynamic and does not necessarily follow a linear path. Thus, the authors argue for the selection of multiple, non-redundant metrics selected to capture the performance of L2 writers at the stages of development outlined above. However, a more recent synthesis of corpus analytic research (Biber et al., 2022) suggests that the developmental sequence proposed in systemic functional linguistics may confound general development in the L2 with L2 writing development and increased familiarity with academic genres.

In a recent volume on register-functional approaches to assessing grammatical complexity, Biber et al. (2022) suggest that the movement from clausal complexity toward more nominal complexity, as proposed by systemic functional linguistics (Norris & Ortega, 2009), may not be associated with general L2 development but rather with the development of familiarity and proficiency with academic prose as a register/genre. Based on extensive corpus-analytic research (Biber & Gray, 2010; Biber et al., 2011, 2013; Staples et al., 2016), the authors outline five stages of academic writing development, the earliest of which is marked by extensive use of finite complement clauses, the latest of which is marked by extensive phrasal embedding. Importantly, Biber and his colleagues link the use of these grammatical features not only to the development of academic writing proficiency but also to the functional demands of academic genres, arguing that academic genres necessitate the dense packaging of information, which is facilitated by increasing phrasal elaboration. Thus, Biber et al. (2022) argue that linguistic complexity is mediated by register, and they eschew omnibus metrics of linguistic complexity in favor of fine-grained metrics that are specific to the register under examination and that can be interpreted in terms of their function.

Common to Norris and Ortega's (2009) and Biber et al.'s (2022) conception of linguistic complexity is a call for researchers to examine multiple metrics of linguistic complexity to capture features of L2 development, academic writing development, and/or familiarity with the genre of the writing task. However, many TBLT-informed studies of planning and its effect on L2 writing examine only a handful of metrics of syntactic and/or lexical complexity. Thus, it begs the question of whether the inconsistent results of prior studies are due to differences in operational definitions of linguistic complexity. This systematic review examines such definitions in the research domain. In particular, it is driven by the following questions:

1. What construct definitions of linguistic complexity are provided in the included studies?
2. What are the common features of construct definitions in the included studies?
3. How is linguistic complexity operationally defined in the included studies?
 - a. Are metrics selected for each level of syntactic complexity identified in Norris and Ortega (2009)?
 - b. Are metrics selected to identify sub-types of lexical complexity?
 - c. Are syntactic and lexical complexity metrics examined in concert with one another?
4. What criteria/rationales are presented when selecting operational definitions of linguistic complexity?

3. Method

3.1 Literature search and inclusion criteria

This systematic review is part of a larger meta-analytic study on the effects of planning on the syntactic complexity, accuracy, lexical complexity, and fluency of L2 production. Two previous research syntheses and quantitative meta-analyses found consistent, significant effects of pre-task planning on oral L2 production (Johnson & Abdi Tabari, 2022) and written L2 production (Johnson & Abdi Tabari, 2023). In the study selection process for each of the previous meta-analyses, the authors took every measure

Table 3. Search terms used

| Planning terms | L2 production terms | TBLT terms |
|----------------------|------------------------|------------------------------------|
| Strategic planning | Syntactic complexity | TBLT |
| Pre-task planning | Accuracy | Cognition Hypothesis |
| Online planning | Lexical complexity | Trade-off Hypothesis |
| Within-task planning | Lexical richness | Limited Attentional Capacity Model |
| Task repetition | Lexical diversity | |
| Task rehearsal | Lexical density | |
| | Lexical sophistication | |
| | Fluency | |
| | L2 writing | |

to be as inclusive as possible. To identify studies for inclusion, the authors worked independently to search the following databases using the search terms listed in Table 3:

- Linguistics and Language Behavior Abstracts
- Educational Resources Information Center
- ProQuest Linguistics
- ProQuest Dissertations and Theses Global
- ResearchGate
- Google Scholar.

Furthermore, the authors conducted searches in the tables of contents of the top 50 journals in applied linguistics, as determined by Scimago Journal and Country Rank. Additionally, they reviewed the following journals, which either featured special issues on linguistic complexity or published a substantial number of studies related to linguistic complexity:

- *International Journal of Applied Linguistics*
- *International Review of Applied Linguistics*
- *The Canadian Modern Language Review*.

To secure as many studies of written L2 production as possible, the authors also searched the annotated bibliographies regularly published in the *Journal of Second Language Writing*.

The authors' searches yielded well over 2,000 studies for potential inclusion. A preliminary review of the abstracts of these studies indicated that the vast majority (a) did not examine L2 production, (b) examined the language production of participants with language disorders, or (c) did not examine the effects of planning on L2 production. These studies were eliminated, leaving 126 studies for further review. The first author reviewed the method of each study to determine its research focus. Studies that did not examine the effect(s) of planning on the complexity of L2 production were further eliminated. Extra attention was devoted to the method of dissertation studies to ensure that samples were not duplicated in later published studies. Together, these measures brought the total number of studies for inclusion to 103. Because this systematic review focuses on the linguistic complexity of L2 written production, studies of oral L2 production were eliminated from further consideration. For the purposes of this systematic review, computer-mediated chat (CMC) and collaborative writing tasks were retained. Kim (2018) was the only study that examined both oral and written modes of production. Data reported on the written mode of production was retained for further analysis. These screening measures yielded a total of 44 studies for further review.

The first author worked independently to code the remaining 44 studies for the following information:

- target language
- context and participants
- L2 proficiency of the participants
- method of ascertaining L2 proficiency
- construct definition of linguistic complexity
- metrics of syntactic complexity
- metrics of lexical complexity
- the rationale for metrics selected.

To establish reliability, the second author coded a subset of approximately 10% of the included studies. Inter-rater reliability was assessed as the percentage of agreement and was considered acceptable (96%). Only three disagreements were noted between the first author and the second author. They were easily resolved through discussion, bringing the agreement to 100%.

4. Results and discussion

The following sections present the results of this systematic analysis and examine their implications for TBLT-informed L2 writing research. Readers will note that, throughout, the parenthetical frequencies may exceed the number of included studies. This is because many of the included studies examined multiple metrics of linguistic complexity.

4.1 Construct definitions and their features

The results of this systematic review with regard to the first research question clearly illustrate one of the criticisms leveled by Bulté and Housen (2014). In many instances, operational definitions of linguistic complexity take the place of construct definitions of linguistic complexity. In fact, of the 44 studies included in this review, only eight provided construct definitions of linguistic complexity (see Supplementary Appendix A). The most frequent ($n = 5$) theme of the included construct definitions was elaboration of language, with authors citing Bulté and Housen (2012), Ellis and Barkhuizen (2005), and/or Skehan (1996) to support their focus on elaboration. The second most frequent theme ($n = 4$) referred to latent psychological constructs such as eagerness, ambition, and/or willingness to experiment. Finally, the third most frequent theme ($n = 3$) addressed the use of “advanced” language with no clear indication of what forms constituted “advanced” language.

In the majority of the eight studies, the authors cite Skehan (1996) and Skehan and Foster (2001) for construct definitions of linguistic complexity, referring to linguistic complexity as “the capacity to use more advanced language, with the possibility that such language may not be controlled so effectively” (Skehan & Foster, 1999, p. 96) or as an indication of “the stage and elaboration of the underlying interlanguage system” (Skehan, 1996, p. 46). While the first of these definitions is widely cited – and widely recognized as valid – it is problematic when applied to written L2 production for two reasons: (a) it is not clear what is considered to be advanced language, and (b) it is not clear whether such “advanced language” features are appropriate to a given writing task. When examined in light of the criticisms of Bulté and Housen (2012), such definitions are further problematic as they incorporate reference to constructs outside the linguistic system, its components, and their articulation. Further, if, as Larsen-Freeman and Cameron (2008) argue, language use is an adaptive response to the context of its use, the use of “advanced language” – whatever that may be – may not be the most appropriate response to a given context if L2 written production is conceived of as the L2 writer’s response to the rhetorical context.

For example, one of the included studies illustrates the fundamental mismatch between conceptions of linguistic complexity at the level of the construct and operational definitions of linguistic

complexity. Citing the seminal work of Larsen-Freeman and Cameron (2008), the study provides a sound rationale for considering linguistic complexity as a dynamic, adaptive, multi-componential construct. However, the study examined only two metrics of syntactic complexity – both of which were global metrics – together with a single lexical complexity metric. Thus, it would appear that many TBLT-informed L2 writing studies of task planning have fallen prey to the circularity of reasoning that Bulté and Housen (2014) caution against by either conflating construct definitions with operational definitions of linguistic complexity or by fundamentally ignoring the articulation between linguistic complexity as a construct and operational definitions adopted as measurements of that construct.

4.1.1 Metrics of syntactic complexity and their levels (Norris & Ortega, 2009)

Of the 44 studies included in this review, 42 examined metrics of syntactic complexity. Metrics of syntactic complexity in the included studies are presented by study (in reverse chronological order) in Supplementary Appendix B and categorized at four levels using the guidelines proposed by Norris and Ortega (2009) and Johnson (2017): (a) metrics of global complexity, (b) metrics of variety of forms, (c) metrics of phrasal elaboration, and (d) metrics of clausal elaboration.

Twenty-nine different metrics of syntactic complexity were employed in the included studies, the most frequently employed of which were mean length of T-unit ($n = 11$), dependent clauses per T-unit ($n = 10$), clauses per T-unit ($n = 10$), mean length of clause ($n = 9$) and dependent clauses per clause ($n = 7$). When categorized using Norris and Ortega's framework, the five most frequently used metrics of syntactic complexity indicate a preference for global metrics of syntactic complexity (i.e., mean length of T-unit) and for metrics of clausal elaboration commonly associated with oral language production (Biber et al. 2013). In fact, 16 of the included studies examined metrics from a single level of syntactic complexity, typically clausal elaboration. Eighteen of the included studies examined metrics from two levels of syntactic complexity. Five of the included studies included metrics from three levels of complexity. Only two studies (Hsu, 2012a, 2017) included metrics from all four levels of syntactic complexity. These results further illustrate a mismatch between conceptions of linguistic complexity as a multidimensional construct and operational definitions selected to reflect this construct. If, as Larsen-Freeman and Cameron (2008) argue, language use is a dynamic response to the context of language use – a view consistent with the functional view of linguistic complexity advocated by Biber et al. (2020) as well as the views of systemic functional linguistics (Byrnes, 2014) – then the operational definition of syntactic complexity at a single level (Norris & Ortega, 2009) seems at odds with such a view. Further still, if, as Bulté and Housen (2012) argue, measuring linguistic complexity involves iterating the number of linguistic components and their articulation as well as the use of diverse, elaborate, sophisticated linguistic forms, then the selection of a handful of metrics – in the case of 26 of the included studies, the selection of only one ($n = 15$) or two metrics ($n = 11$) – as operational definitions of syntactic complexity falls short of capturing the syntactic complexity of written L2 production.

Of the 16 studies that included more than two metrics of syntactic complexity, four included metrics of syntactic complexity at three levels. In all cases, metrics were combinations of global complexity, phrasal elaboration, and clausal elaboration.

4.1.2 Metrics of lexical complexity and sub-types of lexical complexity

Metrics of lexical complexity in the included studies are presented in Supplementary Appendix C and are categorized by sub-type of lexical complexity as metrics of (a) lexical diversity, (b) lexical sophistication, or (c) lexical density. As discussed in the previous section, studies of task planning and its effect on lexical complexity appear to suffer from a mismatch between metric selection and conceptions of linguistic complexity – in this case, lexical complexity – as a multidimensional construct. Of the included studies, 32 examined the effect of task planning on lexical complexity. The majority of those studies ($n = 19$) examined only lexical diversity metrics. This was typically a single metric of lexical diversity, most frequently mean segmental type-token ratio ($n = 13$). Nine studies examined lexical diversity in addition to lexical sophistication. Metrics of lexical sophistication were quite varied, but

the most frequently used metric was CELEX frequency rating or its log transformation. Only three studies examined all three sub-sub-types of lexical complexity.

4.1.3 Syntax and lexis examined in concert

Of the 30 studies that examined both syntactic complexity and lexical complexity, five provide some discussion of the interplay of the two, either in the literature review, the method, and/or the results and discussion. In most cases, this discussion was limited to (a) the simple observation that either syntactic complexity or lexical complexity was “improved” through the provision of planning time, and the other was not; (b) some suggestion that planning time allowed participants to prioritize lexical retrieval over the retrieval of syntactic forms; or (c) some mention that lexical complexity and syntactic complexity “complement” one another. Missing from these studies was a discussion of potential tradeoffs between syntactic and lexical complexity as has been observed in Dynamic Systems Theory studies of L2 writing (Spoelman et al., 2010; Verspoor et al., 2008, 2012) as well as corpus-driven studies of academic prose, which found increased nominalization associated with decreased lexical diversity (Biber, 1988). Thus, metrics of linguistic complexity are observed in isolation despite extensive calls to address linguistic complexity as a dynamic, multi-componential system (Larsen-Freeman & Cameron, 2008) as well as corpus-based calls to address multidimensional metrics of linguistic complexity (Biber et al., 2022).

4.2 The rationale for metric selection

Of the included studies, 35 provide some discussion of the rationale for metric selection. Metric selection was primarily justified through reference to prior research, either to justify the selection of specific metrics or to facilitate direct comparison with prior research. Seven of these studies provided more robust discussions of the metrics through reference to developmental sequences of L2 writing (Abrams & Byrd, 2017) or syntactic features associated with a specific register and/or genre (Abdi Tabari, 2021, 2022; Abdi Tabari & Wang, 2022; Farshi et al., 2013; Hsu, 2012b; Johnson et al., 2012). Of particular concern in metric selection in the included studies is the potential for redundant or collinear metrics. Few of the included studies report collinearity screening practices to determine whether the included metrics are, in fact, distinct from one another. Thus, a number of studies included multiple metrics of lexical diversity (e.g., type-token ratio, mean segmental type-token ratio, and corrected type-token ratio) which are likely redundant.

5. Conclusion, future directions, and limitations

TBLT-informed L2 writing research on the effects of planning on L2 production has largely taken a neutral position on the definition and nature of linguistic complexity of L2 production, likely as a methodological expedient to focus on demonstrating the effect of planning on L2 written production. The results of this systematic review of studies suggest that when construct definitions of linguistic complexity are provided, they are often framed in terms of elaboration of language, a definition consistent with the recommendations of Bulté and Housen (2012). However, in a number of instances, construct definitions are framed in terms unrelated to the linguistic system such as *EAGERNESS*, *AMBITION*, or *WILLINGNESS TO EXPERIMENT*. Further, much of the rationale in metric selection is one of direct comparison with previous studies – studies that may or may not have defined linguistic complexity. Thus, in an ideal scenario, future research on the effects of task planning on linguistic complexity would provide construct definitions of linguistic complexity that conform to the following recommendations:

1. Definitions of linguistic complexity should address the use of discrete linguistic units and their interconnection (Bulté & Housen, 2014).
2. Selection of linguistic complexity metrics – particularly in studies of L2 written production – should address register/genre as a mediating variable (Biber et al., 2020).

3. Selection of linguistic complexity metrics should also address the role of L2 development (Norris & Ortega, 2009).

Metric selection would, then, be supported through specific reference to the criteria established by such construct definitions. However, given the current state of the domain, only the first point is tenable as a comprehensive set of linguistic features associated with each register/genre and L2 development is unavailable. Thus, a multidimensional approach to the measurement of linguistic complexity is needed to determine complex linguistic features associated with L2 development (Norris & Ortega, 2009) as well as the L2 writer's response to the functional demands of the writing task (Biber et al., 2022).

Though this systematic review of studies illustrates several trends in past research as well as opportunities for future research, these trends and opportunities should be noted in light of several limitations. First, while the authors conducted an exhaustive literature search to capture an accurate picture of the research domain, no efforts were made to locate unpublished literature other than unpublished doctoral dissertations. Second, although the study of planning and its effect on the linguistic complexity of L2 written production dates back to 1998, the field is still comparatively young. Thus, further research – in particular, replication research using carefully selected metrics of linguistic complexity – is needed to fully understand the effects of task planning on the linguistic complexity of L2 production.

Supplementary material. To view supplementary material for this article, please visit: <https://doi.org/10.1017/S0261444824000284>.

Conflict of interest. The authors have no known conflict of interest to disclose.

References

Note: Studies marked with an asterisk were included in the systematic review.

- *Abdi Tabari, M. (2016). The effects of planning time on complexity, accuracy, fluency, and lexical variety in L2 descriptive writing. *Asian-Pacific Journal of Second and Foreign Language Education*, 1(1), Article 10. doi:10.1186/s40862-016-0015-6
- *Abdi Tabari, M. (2017). Investigating the effects of planning time on the complexity of L2 argumentative writing. *TESL-EJ*, 21(1), 1–24.
- *Abdi Tabari, M. (2018). *The effect of planning time conditions on complexity, accuracy, and fluency in second language written narratives* [Unpublished Ph.D. dissertation]. Oklahoma State University.
- *Abdi Tabari, M. (2020). Differential effects of strategic planning and task structure on L2 writing outcomes. *Reading and Writing Quarterly*, 36(4), 320–338. doi:10.1080/10573569.2019.1637310
- *Abdi Tabari, M. (2021). Task preparedness and L2 written production: Investigating effects of planning modes on L2 learners' focus of attention and output. *Journal of Second Language Writing*, 52, Article 100814. doi:10.1016/j.jslw.2021.100814
- *Abdi Tabari, M. (2022). Investigating the interactions between L2 writing processes and products under different task planning time conditions. *Journal of Second Language Writing*, 55, Article 100871. doi:10.1016/j.jslw.2022.100871
- *Abdi Tabari, M. (2023). Unpacking the effects of different lengths of pre-task planning time: L2 writing outcomes and learners' perceptions. *The Language Learning Journal*. Advance online publication. doi:10.1080/09571736.2023.2213237
- *Abdi Tabari, M., & Wang, Y. (2022). Assessing linguistic complexity features in L2 writing: Understanding effects of topic familiarity and strategic planning within the realm of task readiness. *Assessing Writing*, 52, Article 100605. doi:10.1016/j.asw.2022.100605
- *Abrams, Z. I., & Byrd, D. R. (2016). The effects of pre-task planning on L2 writing: Mind-mapping and chronological sequencing in a 1st-year German class. *System*, 63, 1–12. doi:10.1016/j.system.2016.08.011
- *Abrams, Z., & Byrd, D. R. (2017). The effects of meaning-focused pre-tasks on beginning-level L2 writing in German: An exploratory study. *Language Teaching Research*, 21, 434–453. doi:10.1177/1362168815627383
- *Aref, G. S., & Mojavezi, A. (2019). The impact of pre-task instruction and task rehearsal on fluency, accuracy, and complexity of Iranian EFL learners' writing. *Journal of Language Teaching and Research*, 10(2), 281–288. doi:10.17507/jltr.1002.09
- Baddeley, A. (1986). *Working memory*. Oxford University Press.
- Baddeley, A. (2007). *Working memory, thought, and action*. Oxford University Press.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2–20. doi:10.1016/j.jeap.2010.01.001
- 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. doi:10.5054/tq.2011.244483

- Biber, D., Gray, B., & Poonpon, K. (2013). Pay attention to the phrasal structures: Going beyond T-units—A response to WeiWei Yang. *TESOL Quarterly*, 47(1), 192–201. doi:10.1002/tesq.84
- Biber, D., Gray, B., Staples, S., & Egbert, J. (2020). Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. *Journal of English for Academic Purposes*, 46, 1–15. doi:10.1016/j.jeap.2020.100869
- Biber, D., Gray, B., Staples, S., & Egbert, J. (2022). Rationale and hypothesis for the study of writing development from the RF perspective. In D. Biber, B. Gray, S. Staples, & J. Egbert (Eds.), *The register-functional approach to grammatical complexity: Theoretical foundations, descriptive research findings, application* (pp. 276–290). Taylor Francis.
- *Biria, R., & Karimi, Z. (2015). Contributory role of pre-task planning in improving Iranian EFL learners' writing of argumentative essays: The case of accuracy and complexity. *Theory and Practice in Language Studies*, 5(5), 1026–1036. doi:10.17507/tpls.0505.18
- Bulté, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy, and fluency in SLA* (pp. 21–46). John Benjamins.
- Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42–65. doi:10.1016/j.jslw.2014.09.005
- Byrnes, H. (2014). Conceptualizing FL writing in collegiate settings: A genre-based systemic functional linguistics approach. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 191–219). De Gruyter Mouton.
- Chong, S., & Plonsky, L. (2023). A typology of secondary research in Applied Linguistics. *Applied Linguistics Review*, 15(4), 1569–1594. doi:10.1515/applrev-2022-0189
- Ellis, R. (2005). Planning and task-based performance: Theory and research. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 3–34). John Benjamins.
- Ellis, R. (2021). The effects of pre-task planning on second language writing: A systematic review of experimental studies. *Chinese Journal of Applied Linguistics*, 44(2), 131–165. doi:10.1515/CJAL-2021-0009
- Ellis, R., & Barkhuizen, G. (2005). *Analyzing learner language*. Oxford University Press.
- *Ellis, R., & Yuan, F. (2004). The effect of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26(1), 59–84. doi:10.1017/S0272263104026130
- *Farahani, A. A. K., & Meraji, S. R. (2011). Cognitive task complexity and L2 narrative writing performance. *Journal of Language Teaching and Research*, 2(2), 445–456. doi:10.4304/jltr.2.2.445-456
- *Farshi, N., Tavakoli, M., & Ketabi, S. (2013). The effect of different online planning conditions on EFL learners' writing test performance. *The Iranian EFL Journal*, 9(4), 325–345.
- *Fazilatfar, A. M., Kasiri, F., & Nowbakht, M. (2020). The comparative effects of planning time and task conditions on the complexity, accuracy, and fluency of L2 writing by EFL learners. *Iranian Journal of Language Teaching Research*, 8(1), 93–110.
- *Fei, F. (2015). Formulaic language use in L2 Chinese: The role of pre-writing planning [Unpublished Doctoral dissertation]. Michigan State University.
- *Gauthier, M. (2007). *The effects of structured pre-task writing plans in L1 and L2 on high school learners' FSL writing performance* [Unpublished Master's thesis]. The University of Toronto.
- *Golparvar, S. E., & Azizsahra, M. (2023). The effect of graph complexity and planning on graph writing performance and descriptive strategies. *Foreign Language Annals*, 56, 117–143. doi:10.1111/flan.12676
- Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy, and fluency in SLA*. John Benjamins.
- *Hsu, H. (2012a). Planning and language development in task-based synchronous computer-mediated communication [Unpublished Doctoral dissertation]. The University of Texas at San Antonio.
- *Hsu, H. (2012b). Investigating the effects of planning on L2 text chat performance. *CALICO Journal*, 29(4), 619–638. doi:10.11139/cj.29.4.619-638
- *Hsu, H. (2017). The effect of task planning on L2 performance and L2 development in text-based synchronous computer-mediated communication. *Applied Linguistics*, 38(3), 359–385. doi:10.1093/applin/amv032
- Johnson, M. D. (2017). Cognitive task complexity and L2 written syntactic complexity, lexical complexity, accuracy, and fluency: A research synthesis and meta-analysis. *Journal of Second Language Writing*, 37, 13–38. doi:10.1016/j.jslw.2017.06.001
- Johnson, M. D. (2022). Task complexity studies. In R. M. Manchón & C. Polio (Eds.), *Handbook of second language acquisition and writing* (pp. 52–64). Routledge.
- Johnson, M. D., & Abdi Tabari, M. (2022). Task planning and oral L2 production: A research synthesis and meta-analysis. *Applied Linguistics*, 43(6), 1143–1164. doi:10.1093/applin/amac026
- Johnson, M. D., & Abdi Tabari, M. (2023). Planning in L2 writing: A research synthesis and meta-analysis. *System*, 118, 103152. doi:10.1016/j.system.2023.103152
- *Johnson, M. D., Mercado, L., & Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing*, 21(3), 264–282. doi:10.1016/j.jslw.2012.05.011
- *Kang, S., & Lee, J. (2019). Are two heads always better than one? The effects of collaborative planning on L2 writing in relation to task complexity. *Journal of Second Language Writing*, 45, 61–72. doi:10.1016/j.jslw.2019.08.001

- Kellogg, R. T. (1987a). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory and Cognition*, 15(3), 256–266. doi:10.3758/bf03197724
- Kellogg, R. T. (1987b). Writing performance: Effects of cognitive strategies. *Written Communication*, 4(3), 269–298. doi:10.1177/0741088387004003003
- Kellogg, R. T. (1990). Effectiveness of prewriting strategies as a function of task demands. *American Journal of Psychology*, 103(3), 327–342. doi:10.2307/1423213
- Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 57–71). Lawrence Erlbaum Associates.
- Kellogg, R. T., Whiteford, A. P., Turner, C. E., Cahill, M., & Mertens, A. (2013). Working memory in written composition: An evaluation of the 1996 model. *Journal of Writing Research*, 5(2), 159–190. doi:10.17239/jowr-2013.05.02.1
- *Kherzlou, S. (2020). Training planning in second language narrative writing. *ELT Journal*, 74(1), 49–62. doi:10.1093/elt/ccz050
- *Khorasani, R., Pandian, A., & Ismail, S. A. M. M. (2012). The effect of strategic planning on accuracy, fluency, and complexity of written narrative task production. *Language in India*, 12(7), 422–446.
- *Kim, N. (2018). The effects of online planning on CAF in L2 spoken and written performance. *English Teaching*, 73(3), 3–28. doi:10.15858/engtea.73.3.201809.3
- *Kim, Y., Choi, B., Yun, H., Kim, B., & Choi, S. (2020). Task repetition, synchronous written corrective feedback and the learning of Korean grammar: A classroom-based study. *Language Teaching Research*, 26(6), 1106–1132. doi:10.1177/1362168820912354
- Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford University Press.
- *Liang, M., & Xie, Y. (2023). Exploring the effects of planning conditions and peer familiarity on Chinese EFL learners' argumentative writing. *System*, 115, 103057. doi:10.1016/j.system.2023.103057
- *Liao, J. (2018). The impact of face-to-face oral discussion and online text-chat on L2 Chinese writing. *Journal of Second Language Writing*, 41, 27–40. doi:10.1016/j.jslw.2018.06.005
- *Lin, Y. (2013). The effects of task planning on L2 writing [Unpublished Doctoral dissertation]. The University of Auckland.
- Long, M. (2014). *Second language acquisition and task-based language teaching*. Wiley.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36–62. doi:10.5054/tq.2011.240859
- *Luo, X. (2022). The effect of manipulating task complexity along resource-dispersing dimension on L2 written performance from the perspective of complexity theory. *English Language Teaching*, 15(9), 151–159. doi:10.5539/elt.v15n9p151
- Manchón, R. M. (2014). The internal dimension of tasks: The interaction between task factors and learner factors in bringing about learning through writing. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 27–52). John Benjamins.
- *McDonough, K., & De Vleeschauwer, J. (2019). Comparing the effect of collaborative and individual prewriting on EFL learners' writing development. *Journal of Second Language Writing*, 44, 123–130. doi:10.1016/j.jslw.2019.04.003
- *Meraji, S. R. (2011). Planning time, strategy use, and written task production in a pedagogic vs. a testing context. *Journal of Language Teaching and Research*, 2(2), 338–352. doi:10.4304/jltr.2.2.338-352
- *Nitta, R., & Baba, K. (2014). Task repetition and L2 writing development: A longitudinal study from a dynamic systems perspective. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 107–136). John Benjamins.
- 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. doi:10.1093/applin/amp044
- *Ong, J., & Zhang, L. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, 19(4), 218–233. doi:10.1016/j.jslw.2010.10.003
- Pallotti, G. (2022). Commentary on “understanding variability in second language acquisition, Bilingualism, and cognition: A second language acquisition perspective”. In K. Kersten & A. Winsler (Eds.), *Understanding variability in second language acquisition, bilingualism, and cognition: A second language acquisition perspective* (pp. 323–336). Routledge.
- *Piri, F., Barati, H., & Ketabi, S. (2012). The effects of pre-task, on-line, and both pre-task and on-line planning on fluency, complexity, and accuracy: The case of Iranian EFL learners' written production. *English Language Teaching*, 5(6), 158–167. doi:10.5539/elt.v5n6p158
- *Rahimi, M., & Zhang, L. J. (2018). Effects of task complexity and planning conditions on L2 argumentative writing production. *Discourse Processes*, 55(8), 726–742. doi:10.1080/0163853x.2017.1336042
- *Rahimpour, M., & Safarie, M. (2011). The effects of on-line and pre-task planning on descriptive writing of Iranian EFL learners. *International Journal of English Linguistics*, 1(2), 274–280. doi:10.5539/ijel.v1n2p274
- Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57. doi:10.1093/applin/22.1.27
- Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies*, 21(2), 45–105.
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *IRAL-International Review of Applied Linguistics in Language Teaching*, 43(1), 1–33. doi:10.1515/iral.2005.43.1.1

- Robinson, P. (Ed.). (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. John Benjamins.
- *Rostamian, M., Fazilatfar, A. M., & Jabbari, A. A. (2018). The effect of planning time on cognitive processes, monitoring behavior, and quality of L2 writing. *Language Teaching Research*, 22(4), 418–438. doi:10.1177/1362168817699239
- *Salimi, A., Alavinia, P., Hosseini, P., & Shafaei, A. (2012). The impact of task complexity and strategic planning time on EFL learners' accuracy and fluency in written task production. *International Journal of Applied Linguistics & English Literature*, 1(7), 104–116. doi:10.7575/ijalel.v1n.7p.104
- *Salimi, A., & Fatollahnejad, S. (2012). The effects of strategic planning and topic familiarity on Iranian intermediate EFL learners' written performance in TBLT. *Theory and Practice in Language Studies*, 2(11), 2308–2315. doi:10.4304/tpls.2.11.2308-2315
- *Sattarpour, S., & Farrokhi, F. (2017). Exploring the interplay of planning time, reasoning demands, and language learning aptitude in Iranian EFL learners' written production. *The Journal of Asia TEFL*, 14(4), 736–754. doi:10.18823/asiatefl.2017.14.4.10.736
- *Seyyedi, K., Ismail, S. A. M. M., Orang, M., & Nejad, M. S. (2013). The effect of pre-task planning time on L2 learners' narrative writing performance. *English Language Teaching*, 6(12), 1–10. doi:10.17509/ijal.v7i3.9803
- Skehan, P. (1996). Second language acquisition research and task-based instruction. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 17–30). Heinemann.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Skehan, P. (2014). *Processing perspectives on task performance*. John Benjamins.
- Skehan, P., & Foster, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research*, 3(3), 215–247. doi:10.1016/j.sbspro.2011.11.445
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge University Press.
- Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics*, 31(4), 532–553. doi:10.1093/applin/amq001
- Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, 33(2), 149–183. doi:10.1177/0741088316631527
- Tavakoli, P. (2014). Storyline complexity and syntactic complexity in writing and speaking tasks. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 217–236). John Benjamins.
- Verspoor, M., Lowie, W., & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *Modern Language Journal*, 92(2), 214–231. doi:10.1111/j.1540-4781.2008.00715.x
- Verspoor, M., Schmid, M. S., & Xu, X. (2012). A dynamic usage-based perspective on L2 writing. *Journal of Second Language Writing*, 21(3), 239–263. doi:10.1016/j.jslw.2012.03.007
- *Xie, Y., & Zhu, D. (2023). Effects of participatory structure of pre-task planning on EFL learners' linguistic performance and alignment in the continuation writing task. *System*, 114, 103008. doi:10.1016/j.system.2023.103008
- *Ziegler, N. (2018). Pre-task planning in L2 text-chat: Examining learners' process and performance. *Language Learning & Technology*, 22(3), 193–213. <http://hdl.handle.net/10125/44664>

Mark D. Johnson is Associate Professor of TESOL and Applied Linguistics at East Carolina University, where he teaches classes on linguistics and teaching English to speakers of other languages. His research focuses on cognition in second language writing, particularly on features of writing tasks and the demands they make on language production. His publications have appeared in *TESOL Journal*, *Writing and Pedagogy*, *Applied Linguistics*, and the *Journal of Second Language Writing*. His article, Cognitive task complexity and L2 written syntactic complexity, lexical complexity, accuracy, and fluency: a research synthesis and meta-analysis, received an honorable mention from the *Journal of Second Language Writing* in their competition for the best paper of 2017.

Mahmoud Abdi Tabari is a Teaching Assistant Professor of TESOL and Writing Studies at the University of Nevada, Reno (UNR). His research focuses on second language acquisition, second language writing, and task-based language teaching. He has published his work in various top-tier academic journals, including *Applied Linguistics*, *Modern Language Journal*, *Journal of Second Language Writing*, *Language Teaching Research*, *System*, *Reading and Writing*, *Assessing Writing*, and *Journal of English for Academic Purposes*.