



Research Article

Shungo Suzuki* and Judit Kormos

L2 fluency across tasks: disentangling demands on conceptualisation and formulation in speech production

<https://doi.org/10.1515/iral-2024-0185>

Received September 11, 2024; accepted June 2, 2025; published online June 20, 2025

Abstract: Second language (L2) speaking research has highlighted intra-speaker variability of fluency performance across tasks. To better understand such task effects on fluency, the framework of speech processing demands has been proposed as a systematic approach to relating task characteristics to L2 speech production mechanisms and the limited capacity of attentional resources (Skehan 2009. *Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis*. *Applied Linguistics* 30(4). 510–532, 2014. Limited attentional capacity, second language performance, and task-based pedagogy. In Peter Skehan (ed.), *Processing perspectives on task performance*, 211–260. Amsterdam: John Benjamins). However, the framework has been tested on a limited range of task types, using carefully designed experimental tasks. For the sake of the ecological validity of findings, the current study thus further explores how L2 learners' fluency varies across four spontaneous speaking tasks differing in their processing demands. A total of 128 Japanese learners of English completed four speaking tasks: Argumentative task, Picture narrative task, Reading-to-Speaking task, and Reading-while-listening-to-speaking task. Their speech was analysed in terms of speed, breakdown, and repair fluency and was compared across tasks. The results of Generalised Linear Mixed-effects Modelling showed that conceptualising demands were reflected in the frequency of filled pauses, while formulation demands were associated with the articulation rate, mid-clause pause ratio, and mid-clause pause duration. These findings unveil the interrelationship between task characteristics, fluency measures, and how learners approach tasks.

***Corresponding author: Shungo Suzuki**, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8601, Japan; and Lancaster University, Lancaster, UK, E-mail: suzuki.shungo.r5@f.mail.nagoya-u.ac.jp. <https://orcid.org/0000-0002-6327-3298>

Judit Kormos, Lancaster University, Lancaster, UK, E-mail: j.kormos@lancaster.ac.uk. <https://orcid.org/0000-0002-2643-7222>

Keywords: dual-process theory; fluency; limited attentional capacity model; speech production; speech processing demands; task effects

1 Introduction

Among various aspects of second language (L2) oral skills, fluency has been regarded as one of the important constructs, considering its reliability as an indicator of L2 proficiency (Tavakoli et al. 2020). Using a variety of speaking tasks, prior research has demonstrated that L2 fluency varies as a function of the characteristics of speech elicitation tasks (e.g., Préfontaine and Kormos 2015; Tavakoli and Skehan 2005). For a reliable assessment of L2 fluency, the knowledge of fluency performance variability in accordance with task characteristics is thus essential particularly when comparing learners' performance across different tasks. However, the findings regarding task effects on L2 fluency are mixed and often contradictory (Segalowitz 2010; Skehan 2009). To better understand task effects on fluency, scholars have proposed different cognitive approaches to systematically manipulate and interpret task characteristics, shedding light on how task characteristics affect speech production processes (Robinson 2011; Segalowitz 2010; Skehan 2014). Among various theoretical frameworks, the current study follows the framework of speech processing demands (Skehan 2009; Skehan et al. 2012), which has been extended from Skehan's (1996; updated in 2014) Trade-Off Hypothesis and Limited Attentional Capacity Model. One advantage of this framework is that it attempts to relate task characteristics to relevant speech production processes (cf. de Bot 1992; Kormos 2006; Levelt 1989). As such, the framework of speech processing demands may offer insights into the construct validity of speaking performance measures in terms of their underlying cognitive processes as well as their sensitivity to variations in task characteristics.

Previous research on speech processing demands has carefully designed experimental tasks for the sake of causal inferences (e.g., Felker et al. 2019). To extend the framework of speech processing demands, the current study rather adopts spontaneous speaking tasks that are widely used in pedagogical and assessment contexts for the sake of ecological validity of findings. More specifically, the current exploratory study compared L2 learners' fluency across four speaking tasks differing in their speech processing demands on conceptualisation and formulation – argumentative task, picture narrative task, reading-to-speaking task, and reading-while-listening-to-speaking task. Conceptualisation entails language-general processes such as content planning (for details, see Section 2.2). The variability of fluency performance caused by conceptualisation demands should thus be regarded as construct irrelevant variance in L2 assessment contexts (Suzuki et al. 2022). Meanwhile, formulation consists of a range of manipulation of linguistic knowledge and thus is reflective of the degree of

automatisation of L2 knowledge (Suzuki and Révész 2023). Due to the exploratory nature of the study, learners' perceptions of the intended processing demands of the tasks were also qualitatively examined.

2 Literature review

2.1 Two competing approaches to task characteristics

Previous research has proposed different frameworks to theorise how different task characteristics can explain intra-speaker variability in speaking performance. Among them, two major frameworks have emerged in task-based performance research: Cognition Hypothesis (Robinson 2011) and Limited Attentional Capacity Model (Skehan 1996, 2014), along with its related notion of speech processing demands (Skehan 2009). One of the differences between these frameworks lies in how task effects are operationalised and interpreted. The former manipulates the cognitive demands of tasks by manipulating task characteristics, such as the number of steps and elements involved in the task (Robinson 2011). Meanwhile, the latter sheds light on how different task characteristics affect speakers' speech production processes, including content planning and lexical retrieval, in relation to the limited amount of attentional resources (Skehan 2009, 2014).

The debate about the two aforementioned frameworks has continued in L2 research (Jackson and Suethanapornkul 2013; Pallotti 2020). However, it is possible to revisit the cognitive demands of tasks from the perspective of two competing theories in the field of cognitive psychology. On the one hand, the intrinsic cognitive load theory (Chandler and Sweller 1991) hypothesises that tasks have an inherent level of difficulty which is determined by various factors, such as whether the task can be solved in one step or requires several steps, or whether they can be accomplished by retrieving ready-made chunks or units from long-term memory. In this respect, the intrinsic cognitive load theory shares a number of features with Robinson's (2011) Cognition Hypothesis, which postulates that a set of task characteristics can predict the cognitive demands of language learning tasks. On the other hand, the dual-process theory of higher cognition (Evans and Stanovich 2013) argues that it is not task complexity but the nature of specific processing demands on attentional and working memory resources that explains how efficiently one can perform a task. The dual-process theory, which gained strong support in a recent meta-analysis of the relationship between reading task type, first language reading development and working memory conducted by Peng et al. (2018), is compatible with Skehan's (2009) speech processing framework. Therefore, the current study adopts Skehan's (2009) framework and draws on Evans and Stanovich's (2013) dual-process theory in the discussion.

2.2 Speech processing demands: applying speech production mechanisms to task effects

Given that the framework of speech processing demands is theoretically underpinned by the mechanisms of speech production, it is essential to understand how speech production proceeds. Following previous work on speech processing demands (Skehan 2009, 2014), the current study adopts Kormos' (2006) L2 speech production model, which was developed based on L1 speech production models (e.g., Levelt 1989, 1999). As with other L2 speech production models (de Bot 1992; Segalowitz 2010), Kormos' model assumes that speech production proceeds in three major phases: *Conceptualisation*, *Formulation*, and *Articulation*. Conceptualisation is responsible for planning the content of speech, including the organisation of content of the message. A planned message (i.e., so-called *preverbal message*) is then linguistically encoded in the formulation stage, drawing on the speaker's lexical, grammatical and phonological resources. The outcome of formulation is pronounced as a stream of sounds in the phase of articulation via the utilisation of speech organs. In addition to these three phases, successful speech production is achieved with the assistance of *self-monitoring* processes, which inspect the interim content and eventual outcome of the three phases of speech production for informational precision and linguistic accuracy.

From the perspective of the current study, the subprocesses of conceptualisation and formulation need further review (for a more detailed discussion, see Suzuki 2021). Conceptualisation consists of two sequential processes: *macroplanning* and *microplanning*. During macroplanning, the speaker decides what speech acts and informational content to express and the order of presenting such information (Levelt 1989). The outcome of macroplanning is an ordered sequence of speech acts and information. During microplanning, the speaker further specifies the outcome of macroplanning, by adding an informational perspective, such as semantic representations of the message and the given versus new status of information. Consequently, the message entails all necessary conceptual specifications for the message to be translated into a corresponding linguistic form at formulation. Meanwhile, linguistic encoding processes involved in formulation are commonly conducted by retrieving target linguistic resources according to the input from a previous process. In the case of lexical retrieval, for instance, semantic information in the preverbal message serves as input and sends activation to the store of lexical resources (i.e., the mental lexicon). Activated lexical entries are then retrieved for subsequent linguistic encoding processes, such as grammatical and phonological encoding. Notably, following Levelt's model (1989, 1999), Kormos (2006) assumes that linguistic encoding processes are achieved by the mechanisms of activation spreading (Bock 1982). The

efficiency and effectiveness of linguistic retrieval is thus affected by the level of activation; the more strongly a target item is activated, the more likely it is to be selected.

From a theoretical perspective, speech production entails a range of information processing, such as lexical retrieval, and thus proceeds by consuming attentional resources (Baddeley 2003). Moreover, the attentional resources available for spontaneous L2 speech production are considered to be limited due to the partially automatised status of L2 skills (de Bot 1992; Kormos 2006; Skehan 2014). Accordingly, L2 speaking performance can be affected by how learners distribute their limited attentional resources to different phases of speech production (cf. Limited Attentional Capacity Model, Skehan 2009, 2014; dual-process theory; Evans and Stanovich 2013). In addition, L2 learners vary in their efficiency in linguistic encoding processes, that is, L2-specific automaticity (Segalowitz 2010, 2016), meaning that depending on their degree of L2 automatisation, the amount of attentional resources needed for a certain encoding process can also vary. In accordance with the dual-process theory (Evans and Stanovich 2013), learners' L2 automaticity might modulate the effects of speech processing demands on their speaking performance (see also Skehan 2009). Building on these theoretical assumptions, the framework of speech processing demands has attempted to explain intra-individual variability in speaking performance across tasks. While task characteristics can generally affect the assignment of attentional resources available to different aspects of speech processing, learners' individual differences in how they approach a given task and their L2 automaticity moderate the effects of task characteristics. The mechanism of task effects on L2 fluency is visualised in Figure 1.

2.3 Theoretical perspective on fluency

The construct definition of fluency can vary depending on the scope with which it is approached. These scopes may range from a very broad, holistic view (i.e., overall speaking ability) to more narrowly defined ones (e.g., temporal features of speech). While practitioners often hold a broader conceptualisation of fluency, L2 researchers has consistently followed a very narrower definition, focusing on speed, pausing and repair features of speech (see Tavakoli and Hunter 2018). Besides, Segalowitz's (2010) triadic model offers three different perspectives on fluency: cognitive fluency (i.e., efficiency of underlying speech production processes), utterance fluency (i.e., observable temporal characteristics of speech), and perceived fluency (i.e., judgements of fluency made by listeners). One of the notable contributions of this model is its theorisation of cognitive fluency by associating possible causes of fluency disruptions with Levelt's L1 speech production model (i.e., fluency

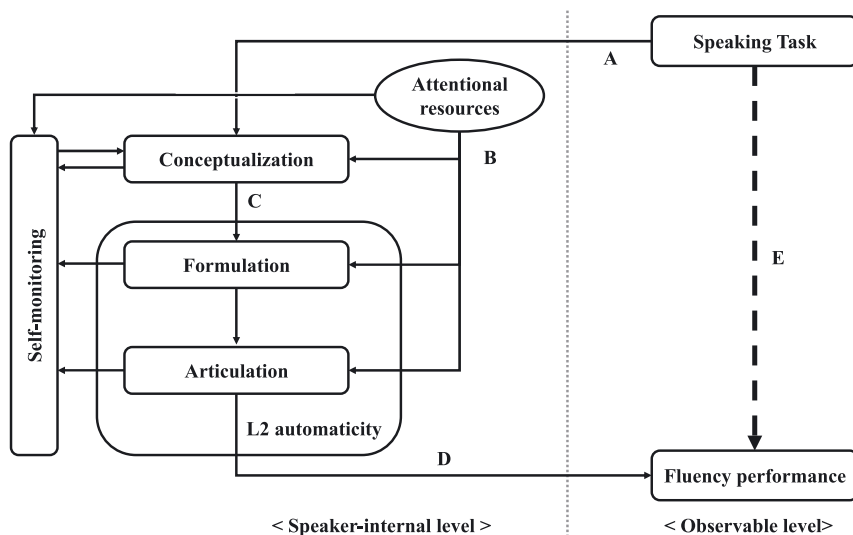


Figure 1: A visual representation of task effects on L2 fluency (on the basis of Evans and Stanovich 2013; Kormos 2006; Segalowitz 2010; Skehan 2009, 2014; Skehan et al. 2012). *Note.* From the observable level (right-hand side) to the speaker-internal level (left-hand side), a speaking task first imposes different processing demands in each stage of speech production (arrow A). In response to a given task, speakers use their attentional resources for their conscious and controlled processing in each stage of speech production in a serial manner (arrow B). Depending on the demands on conceptualisation, the amount of attentional resources available for subsequent processes (i.e., formulation and articulation) varies (arrow C). Finally, speakers' automaticity as well as the amount of attentional resources available can determine the overall smoothness of speech processing (arrow D), which may be reflected as fluency of the speech produced (arrow E).

vulnerability points). These fluency vulnerability points incorporate how each process of speech production, such as microplanning and lexical retrieval, may be disrupted and how such disruptions manifest at the level of overt speech. Researchers have examined the interrelationship among the three fluency constructs (for meta-analytic and systematic reviews, see Suzuki et al. 2021; Suzuki and Révész 2023). In the context of task-based performance research, previous studies on task effects have focused primarily on utterance fluency, often within the framework of complexity, accuracy, and fluency (CAF; Housen et al. 2012; Lambert and Kormos 2014).

Utterance fluency itself is multifaceted. Tavakoli and Skehan (2005) have empirically demonstrated that it comprises three distinct subconstructs: speed fluency (i.e., speed of delivery), breakdown fluency (i.e., pausing behaviour), and repair fluency (i.e., verbatim repetitions and corrections). This three-dimensional

model has been further validated across different speaking tasks and in relation to underlying cognitive processes, that is, cognitive fluency (Suzuki and Kormos 2023).

2.4 Effects of conceptualising demands on fluency

Previous studies have manipulated conceptualising demands using one of three major approaches – tightness of task structure, requirement for content generation and the number of options provided (see also Suzuki 2021). The first type of manipulation of conceptualising demands is the adjustment of tightness of task structure (Foster and Tavakoli 2009; Tavakoli and Foster 2008; Tavakoli and Skehan 2005). Task structure refers to the extent to which task design can help speakers plan the expected macrostructure of their speech to fulfil the given task requirements. Tightly structured tasks (e.g., conventional storyline in picture narrative tasks; Tavakoli and Skehan 2005) enable students to accomplish conceptualisation processes with a small amount of attentional resources, subsequently saving their remaining attentional resources for formulation and articulation processes. Fluency in tightly structured picture narrative tasks is commonly characterised by a reduction in silent pauses (Tavakoli and Foster 2008; Tavakoli and Skehan 2005) and false starts (Tavakoli and Foster 2008), compared to loosely structured ones.

The second approach to altering conceptualising demands can be done in terms of how necessary it is for speakers to plan the content of speech (i.e., content generation). Préfontaine and Kormos (2015) examined the effects of conceptualisation demands on L2 fluency by comparing unrelated and related picture narrative tasks. In the unrelated picture narrative task, L2 speakers were asked to create a storyline from six unrelated pictures, whereas in the related picture narrative task, they narrated a predefined sequence of events with an 11-frame cartoon. The results showed that in the unrelated picture narrative task, learners' speech was characterised by a lower articulation rate and more silent pauses, compared to the related picture narrative task. Their findings confirmed that the enhanced demands on conceptualisation can negatively affect speed fluency and breakdown fluency.

The final approach manipulates the demands on conceptualisation by presenting alternative choices. Although this approach is specific to one particular type of task, that is, network description task, it can control for the influence on formulation demands (Felker et al. 2019). The network description task is a type of picture naming task where participants are presented with a network of objects linked via paths and are asked to describe the route of the paths connecting the highlighted objects (Felker et al. 2019). Felker et al. (2019) prepared an easy and a difficult condition of network description tasks based on the number of potential alternative choices participants are presented with (i.e., distractors). In addition, they changed the target picture and

its path immediately after participants fixated on some picture stimuli for 500 ms. Such online changes to target paths may force participants to revise their plan of the preverbal message, consuming a certain amount of attentional resources. The results revealed that online changes to picture stimuli and paths clearly impeded fluency of performance in terms of filled and silent pause frequency, syllable lengthening, and speech rate. However, Felker et al. (2019) acknowledged that the output of the task was neither lexically nor syntactically complex, as only familiar lexical items were used as picture stimuli and the same syntactic structures could be recycled across paths. Therefore, despite the careful controlling of factors other than conceptualising demands, the ecological validity of Felker et al.'s findings for spontaneous L2 speech might need further examination.

To operationalise conceptualisation demands with an adequate level of ecological validity, it may be plausible to manipulate characteristics of spontaneous speaking tasks in relation to the required outcome of the tasks. As reviewed earlier, conceptualisation is responsible for planning how a given communicative intention is achieved in speech production. In the light of how target task characteristics are essential for achieving a given task, the validity of manipulating task characteristics in picture narrative tasks – one of the most widely used task types in previous studies – for conceptualisation demands is worth revisiting. The task requirement of narrative tasks typically includes identifying characters and objects and referring to them consistently, specifying the main events of the story and narrating the events in a coherent manner (Luoma 2004). With the support of visual prompts, the identification of characters and objects can be considered to be relatively easy in picture narrative tasks (see Tavakoli and Foster 2008). Even with some modifications to task structure, picture narrative tasks might thus inherently impose low conceptualisation demands. Among other task types, the task requirement of argumentative tasks involves expressing an opinion on a given issue, contrasting it with other opinions, and discussing rationales and supporting information for that opinion (Luoma 2004). This task type thus requires learners to engage with the selection of information from among possible alternatives and the organisation of discourse (i.e., conceptualisation).

2.5 Manipulation of formulation demands

Scholars have been interested in how limitations on attentional resources in formulation might affect fluency, because this line of research could yield information on which temporal features are reflective of L2-specific competence or cognitive fluency (Segalowitz 2010, 2016). Due to the serial nature of speech production (Kormos 2006; Levelt 1989), previous studies have attempted to operationalise the demands on formulation through manipulating conceptualising demands,

assuming that the high consumption of attention resources in conceptualisation can reduce the attentional resources available in subsequent formulation processes (Skehan 2009). However, enhanced conceptualising demands do not necessarily enhance the demands on formulation processes (cf. Felker et al. 2019). Linguistic formulation demands might thus need to be manipulated in a relatively direct manner. However, to the best of our knowledge, few studies have directly manipulated formulation demands within the speech processing framework (cf. Mirdamadi and De Jong 2015). Therefore, the current study adopted an exploratory approach to the operationalisation of formulation demands based on the mechanisms of speech production.

As reviewed earlier, one of the fundamental characteristics of formulation is activation spreading (Bock 1982; Levelt 1989, 1999), which assumes that the stronger the activation of an item, the greater the probability that the item is selected. This principle of activation spreading has been applied to priming research methods, in which speed and accuracy of linguistic retrieval has been found to be facilitated when the resting level of activation is enhanced by primes (Indefrey and Levelt 2004; Sprenger et al. 2006). Among spontaneous speaking tasks, reading-to-speaking (RtoS) tasks can be regarded as one such type of speaking task, as the source text activates in-text linguistic items prior to speaking performance. One advantage of using RtoS tasks is that the level of activation can be manipulated by the modality of the source text. According to psycholinguistic research, a bimodal input (i.e., reading-while-listening, in this case) has been found to enhance the activation level of linguistic representations, compared to a unimodal input (Ferrand and Grainger 1993). Therefore, the current study operationalised formulation demands by using RtoS tasks with the conditions of reading-only and reading-while-listening, with the assumption that higher activation level in the reading-while-listening condition should facilitate speech production processes and thus enhance fluency performance (see also Section 4.2).

3 The current study

The above review of the literature on speech processing demands suggests that for the sake of the ecological validity of findings in pedagogical contexts, it is worth exploring the applicability of the framework of speech processing demands using spontaneous speaking tasks. Prior work also reveals that the validity of operationalising the demands on conceptualisation and formulation might be enhanced by manipulating task characteristics relevant to task requirements and by activating task-relevant linguistic items prior to speech production. Therefore, the current study investigated how L2 learners' fluency varied across four speaking tasks potentially differing in their demands on content generation, including discourse

organisation, and linguistic retrieval. Due to the exploratory nature of the study, students' perceptions of speech processing demands were also examined to cross-validate the current operationalisation of speech processing demands. The study was guided by the following research questions:

How does L2 fluency vary between tasks that differ in their speech processing demands on

RQ1: *Content generation including discourse organisation?*

RQ2: *The activation of lexical representations?*

RQ3: *The activation of phonological representations?*

As part of a larger project, the current study used the same dataset from our precursor studies (Kormos et al. 2022; Suzuki and Kormos 2023). Note that although the precursor studies analysed participants' utterance fluency in four speaking tasks, the comparison of their fluency measures across tasks and their perceptions of processing demands collected through the post-speaking questionnaire are exclusively used in the study reported in this article.

4 Methods

4.1 Participants

One hundred and twenty-eight Japanese learners of English were recruited at a private university in Japan (females = 73, males = 55). The ages ranged from 18 to 27 years ($M_{\text{age}} = 20.43$, $SD_{\text{age}} = 1.81$). Their self-reported university placement test scores suggested that their proficiency levels were mostly B1–B2 on the Common European Framework of Reference for Languages (CEFR) scale (Council of Europe 2001).

4.2 Speaking tasks

The current study focuses on the speech processing demands on conceptualisation and formulation. The former was manipulated by content generation and discourse

Table 1: Summary of the contrasts of speaking tasks in relation to different speech processing demands.

Target components of speech production	Relevant task characteristics	Contrast
Conceptualisation	Content generation and discourse organisation (RQ1)	Arg PicN
Formulation	The activation level of lexical representations (RQ2)	PicN RtoS
	The activation level of phonological representations (RQ3)	RtoS RwLtoS

Note. Arg = Argumentative task; PicN = Picture narrative task; RtoS = Reading-to-speaking task; RwLtoS = Reading-while-Listening to speaking task.

organisation, and the latter by the activation level of lexical and phonological representations of task-relevant linguistic items. The effects of manipulating these task characteristics were examined by comparing fluency across three contrasts via four speaking tasks (see Table 1).

In the argumentative task, adopted from Suzuki and Kormos (2020), students were instructed to argue how far they agree with a statement – *The Tokyo Olympics in 2020 will bring economic growth to Japan*.¹ For the picture narrative (PicN) task, Préfontaine and Kormos’ (2015) 11-frame picture cartoon describing the story of a successful businessman in a farming business was adopted. The argumentative task was regarded as the condition representing increased conceptualising demands, as the task requires students to generate their own ideas and connect different pieces of information logically to construct a coherent argument. In contrast, the conceptualising demands in the PicN task were considered relatively low, as the information that needs to be communicated and its order of presentation are visually described in a clear time sequence in the prompt.

Adopting the priming method (McDonough and Trofimovich 2008) and the theory of activation spreading (Bock 1982), the activation level of lexical representations was enhanced by presenting task-relevant linguistic items prior to speech by means of the source text of RtoS (reading-to-speaking) tasks. In the RtoS task, students first read a 300-word expository text written in L2 English, and then orally summarised the text. In order to prevent speakers simply reading aloud the source text in their oral summary, they were not allowed to refer to the text while speaking. While both the PicN task and the RtoS task largely predefined the content of speech, the

¹ All the speech data were collected before the outbreak of Covid-19 in 2019; thus, the topic of the argumentative task was meaningful to the participants at the time of data collection.

linguistic input was only processed in the RtoS task through reading the source text. Hence, the activation of task-relevant linguistic items can be assumed to be higher in the RtoS task than in the PicN task. Similarly, the activation level of phonological representations was enhanced by presenting a bimodal source text, as evidenced in L2 reading research (Kořak-Babuder et al. 2019; Liu and Todd 2014). To this end, we designed another RtoS task, where students read the text while simultaneously listening to an aural recording of the text (henceforth, a reading-while-listening-to-speaking [RwLtoS] task).

To minimise the effects of source texts on students' speaking performance, two different expository texts were adapted from Millington (2019). The combination of source texts and input modes, as well as the order of the input modes, was counterbalanced across participants. One text (Text A) was about the history of the national flag of the United States, while the other text (Text B) illustrated the history of the flag of the International Committee of the Red Cross. Both texts were modified for lexical complexity to ensure that the participants could comprehend the texts. The frequency levels of words in the texts were examined according to the JACET8000 wordlist which is specifically tailored for Japanese learners of English (JACET 2003). Since our target population of participants was university students, low-frequency vocabulary items at Level 5 (the upper-intermediate level of university) or above in the list were replaced with synonyms at Level 4 or below (the beginning level of university). Collocational accuracy was checked by a native speaker of English. The readability of the texts was evaluated using Flesh-Kincaid Reading Ease values which were calculated by Coh-Metrix software (McNamara et al. 2014). An overall linguistic complexity score was also derived using TextEvaluator[®] software (Educational Testing Service 2013). The audio input of the RwLtoS task was recorded by an L1 Canadian English speaker with 15 years of teaching experience of English at universities in Japan. Following Kořak-Babuder et al. (2019), a delivery rate of 120 words per minute was adopted. The characteristics of the source texts are summarised in Table 2 (for details, see Suzuki 2021).

Table 2: The textual characteristics of the source texts.

	Text A	Text B
Topic	US flag	Red cross
Flesh-Kincaid value	71.21	64.79
TextEvaluator [®] score	380	660
Text length (words)	324	303
Speed of delivery (words/min)	116.4	119.6

Note. Adopted from Suzuki (2021; see also Kormos et al. 2022).

4.3 Post-speaking questionnaire

Considering the exploratory nature of the current study, a post-speaking questionnaire was administered to 40 participants to cross-validate the operationalisation of different speech processing demands from the students' perspective. The questionnaire asked participants to briefly report what they found difficult or challenging when they were speaking. A deductive approach to analysing their qualitative responses was adopted, given the predetermined theoretical framework (speech production mechanisms, speech processing demands) as well as the explanatory purpose of the data in the study.

4.4 Procedure

To control for the effects of task implementation factors on fluency, three minutes were provided for pre-task planning in all four tasks. Note-taking during the pre-task planning time was not allowed for any of the tasks, while no time limit was provided during speaking performance. Regarding the order of tasks, care was taken to minimise the influence of other tasks on speaking performance in subsequent tasks. Tasks without linguistic input (argumentative task, PicN task) were administered as the first block, while tasks with linguistic input (RtoS task, RwLtoS task) comprised the second block. The order of tasks within each block was counterbalanced across participants (see Table 3).² Finally, students ($n = 40$ out of 128) completed the post-speaking questionnaire immediately after they finished each task.

4.5 Fluency analysis

Following prior research on L2 fluency (De Jong 2016; Suzuki et al. 2021; Tavakoli et al. 2020), a comprehensive set of utterance fluency measures were adopted, covering three major aspects of utterance fluency – speed fluency, breakdown fluency, and repair fluency (Tavakoli and Skehan 2005). The selected fluency measures are listed below:

² As the current study was part of a large project to examine the interrelationship between L2 speaking performance, linguistic knowledge, and L1 speaking style, some participants ($n = 104$) were provided with two argumentative tasks. These two argumentative tasks were regarded as a sub-block of the first block, and the students performed two argumentative tasks in a row with the order of prompts counterbalanced across participants.

Table 3: The assignment of participants for each group of the task set.

Group	N	First block		Second block			
		Task 1	Task 2	Task 3	Text	Task 4	Text
1	16	PicN	Arg	RtoS	B	RwLtoS	A
2	16	PicN	Arg	RwLtoS	B	RtoS	A
3	16	PicN	Arg	RtoS	A	RwLtoS	B
4	16	PicN	Arg	RwLtoS	A	RtoS	B
5	16	Arg	PicN	RtoS	B	RwLtoS	A
6	16	Arg	PicN	RwLtoS	B	RtoS	A
7	16	Arg	PicN	RtoS	A	RwLtoS	B
8	16	Arg	PicN	RwLtoS	A	RtoS	B

Note. The total number of participants = 128; Arg = Argumentative task; PicN = Picture narrative task; RtoS = Reading-to-speaking task; RwLtoS = Reading-while-Listening to speaking task.

Speed fluency

1. *Articulation rate.* The mean number of syllables produced per second, divided by total speech duration excluding pauses.

Breakdown fluency

2. *Mid-clause pause ratio.* The mean number of silent pauses within clauses, divided by the total number of syllables produced.
3. *End-clause pause ratio.* The mean number of silent pauses between clauses, divided by the total number of syllables produced.
4. *Filled pause ratio.* The mean number of filled pauses, divided by the total number of syllables produced.
5. *Mid-clause pause duration.* Mean duration of pauses within clauses.
6. *End-clause pause duration.* Mean duration of pauses between clauses.

Repair fluency

7. *Self-correction ratio.* The mean number of self-correction behaviours, divided by the total number of syllables produced.
8. *False start ratio.* The mean number of false starts/reformulations, divided by the total number of syllables produced.
9. *Self-repetition ratio.* The mean number of self-repetitions, divided by the total number of syllables produced.

The speech data were transcribed and then annotated for clause boundaries. The number of syllables was calculated based on written pruned transcripts using the *syllcount* package (Schmidt 2020) in R statistical software 4.0.2 (R development Core Team 2020). Relevant temporal features were annotated using Praat software (Boersma and Weenink 2012). Silent pauses were defined as silences longer than 250 ms (Bosker et al. 2013; De Jong and Bosker 2013). Using Praat's automatic detection of silences, silent pauses were first roughly identified, and the boundaries of silent pauses were then manually modified. Subsequently, information about clause boundaries and pause location was added to TextGrid files in Praat, following Foster et al.'s (2000) definition of clauses.

4.6 Statistical analysis

We first calculated descriptive statistics as a preliminary analysis and performed Shapiro-Wilk tests to examine the distributions of all the fluency measures (for statistics, see Suzuki and Kormos 2023). Most of the fluency measures were found to be non-normally distributed, while articulation rate seemed to be normally distributed across tasks. The density plots for each measure indicated that the distributions of all the fluency measures were positively skewed, except for articulation rate. Given the observed non-normal distribution of fluency measures, we thus decided to employ a Generalised Linear Mixed-effects Model (GLMM) with a log link function. A Gaussian distribution (i.e., normal distribution) was applied to the GLMM of articulation rate, while the GLMMs of the remaining fluency measures adopted a Gamma distribution, which is one of the continuous probability distributions where a possible range of values is from zero to $+\infty$ (Coupé 2018). Non-positive values (in the current data set, basically 0 values) may prevent the estimation of statistical models based on a Gamma distribution. Accordingly, when building GLMMs with a non-normal distribution, 0 values were replaced with $-3SD$ values for the theoretical distributions of the variables, estimated by Maximum Likelihood (ML). All GLMMs were estimated using the *glmer* function in the *lme4* package (Bates et al. 2015), using R statistical software 4.0.2 (R development Core Team 2020). We constructed GLMMs for each fluency measure (i.e., outcome variable), using task type as a categorical fixed-effect predictor variable and individual participants as a random-effects predictor. Since task type was a within-subject variable, the random slope of participants may not be distinguishable from random error variance (Barr 2013). We thus only included random intercepts of participants. Regarding the predictor variable of task type, to minimize the rate of type I errors, we decided to take a confirmatory approach to comparing outcome variables between three predetermined contrasts (see Table 1). To this end, we adopted forward difference contrast coding for the

categorical variable of task type. The R code and anonymised dataset will be made available on the IRIS database (<https://www.iris-database.org/iris/app/home/index>). For a full description of statistical analyses, see Suzuki (2021).

5 Results

5.1 Task effects on fluency

To investigate whether the participants' fluency measures differed across four speaking tasks (Argumentative task, PicN task, RtoS task, RwLtoS task), a series of GLMMs were constructed. For each GLMM, fluency measures were compared across four task types (i.e., fixed-effect predictor variable) with random intercepts of individual participants. Table 4 summarises the effects of three predetermined contrasts of speaking tasks on fluency (for further details, see Supplementary Information). The results demonstrated significant effects of tasks on all of the fluency measures.

Learners' articulation rate was faster in the argumentative task than in the PicN task. The measure was also higher in the PicN task than in the RtoS task. Meanwhile, there were no significant differences in the speed fluency measure between the RtoS task and the RwLtoS task.

With regard to the frequency aspect of breakdown fluency, students produced fewer silent pauses in the argumentative task than in the PicN task in both pause locations. In contrast, the frequency of filled pauses was higher in the argumentative task than in the PicN task. Meanwhile, they produced fewer mid-clause pauses but more end-clause pauses in the PicN task than in the RtoS task. As with mid-clause pauses, the frequency of filled pauses was found to be lower in the PicN task than in the RtoS task. Finally, there were no significant differences in pause ratio measures between the two text summary tasks (i.e., RtoS task vs. the RwLtoS task). In terms of the duration aspect of breakdown fluency, there were no significant differences between the argumentative task and the PicN task in the duration of both mid-clause and end-clause pauses. In contrast, end-clause pauses tended to be shorter in the PicN task than in the RtoS task. Meanwhile, longer mid-clause pauses were observed in the RwLtoS task, compared to the RtoS task.

Finally, the results of repair fluency measures showed a nuanced picture of task effects across types of disfluency phenomena. First, students produced fewer self-repetitions and self-corrections in the argumentative task than in the PicN task. Second, the frequency of self-repetitions was found to be higher in the PicN task than in the RtoS task. Third, students produced fewer false starts in the PicN task than in the RtoS task.

Table 4: Summary of the effects of three predetermined contrasts of speaking tasks on fluency.

Fixed-effects	Estimate	SE	z-Value	p	Contrast
Articulation rate					
Intercept	2.831	0.041	68.806	<0.001	
Arg – PicN	0.316	0.030	10.537	<0.001	Arg > PicN
PicN – RtoS	0.160	0.030	5.329	<0.001	PicN > RtoS
RtoS – RwltoS	–0.032	0.030	–1.077	0.282	n.s.
Mid-clause pause ratio					
Intercept	–1.507	0.048	–31.407	<0.001	
Arg – PicN	–0.117	0.023	–4.972	<0.001	Arg < PicN
PicN – RtoS	–0.072	0.023	–3.068	0.002	PicN < RtoS
RtoS – RwltoS	–0.018	0.023	–0.754	0.451	n.s.
End-clause pause ratio					
Intercept	–2.640	0.028	–94.164	<0.001	
Arg – PicN	–0.381	0.024	–15.821	<0.001	Arg < PicN
PicN – RtoS	0.117	0.024	4.881	<0.001	PicN > RtoS
RtoS – RwltoS	0.005	0.024	0.210	0.834	n.s.
Filled pause ratio					
Intercept	–2.541	0.087	–29.274	<0.001	
Arg – PicN	0.191	0.050	3.819	<0.001	Arg > PicN
PicN – RtoS	–0.208	0.049	–4.205	<0.001	PicN < RtoS
RtoS – RwltoS	–0.065	0.049	–1.312	0.190	n.s.
Mid-clause pause duration					
Intercept	0.065	0.038	1.711	0.087	
Arg – PicN	0.003	0.023	0.112	0.911	n.s.
PicN – RtoS	–0.027	0.023	–1.161	0.246	n.s.
RtoS – RwltoS	–0.054	0.023	–2.315	0.021	RtoS < RwltoS
End-clause pause duration					
Intercept	0.201	0.056	3.606	<0.001	
Arg – PicN	0.003	0.032	0.081	0.935	n.s.
PicN – RtoS	–0.149	0.032	–4.735	<0.001	PicN < RtoS
RtoS – RwltoS	0.024	0.032	0.750	0.454	n.s.
Self-repetition ratio					
Intercept	–2.679	0.075	–35.810	<0.001	
Arg – PicN	–0.415	0.064	–6.440	<0.001	Arg < PicN
PicN – RtoS	0.183	0.064	2.846	0.004	PicN > RtoS
RtoS – RwltoS	–0.110	0.064	–1.722	0.085	n.s.

Table 4: (continued)

Fixed-effects	Estimate	SE	z-Value	p	Contrast
Self-correction ratio					
Intercept	−3.834	0.049	−77.518	<0.001	
Arg – PicN	−0.164	0.077	−2.133	0.033	Arg < PicN
PicN – RtoS	0.039	0.077	0.505	0.614	n.s.
RtoS – RwLtoS	−0.007	0.077	−0.086	0.931	n.s.
False start ratio					
Intercept	−4.832	0.064	−75.150	<0.001	
Arg – PicN	0.062	0.115	0.538	0.591	n.s.
PicN – RtoS	−0.586	0.115	−5.088	<0.001	PicN < RtoS
RtoS – RwLtoS	0.131	0.115	1.142	0.254	n.s.

Note. The models of all fluency measures but articulation rate adopted a log link function; Arg = Argumentative task; PicN = Picture narrative task; RtoS = Reading-to-speaking task; RwLtoS = Reading-while-Listening to speaking task; n.s. indicates the lack of statistical difference.

5.2 Students’ perceived demands of tasks

To gain insights from students’ perspective, their qualitative responses to what they found difficult were coded for different speech production processes. The codes were then grouped into three major processes of speech production, that is, *Conceptualisation*, *Formulation*, and *Articulation*. Table 5 summarises the raw frequency of participants who mentioned each coding label. Overall, participants were rarely conscious of the processing demands of articulation ($n = 6$ out of 111 responses). However, the relative frequency of speakers reporting the demands of conceptualisation and formulation seemed to vary across tasks. Specifically, in the argumentative task, almost equal numbers of students reported perceived demands on conceptualisation and formulation, while in the PicN task, they were more likely to be aware of demands on formulation than on conceptualisation. Furthermore, in both conditions of the text summary task, conceptualisation processes were commonly perceived as more demanding than formulation processes. In the discussion section, individual excerpts are introduced to shed further light on the GLMM results.

6 Discussion

6.1 Speech processing demands on conceptualisation

Comparing the argumentative and PicN tasks, the GLMMs suggested that the enhanced conceptualising demands resulted in faster articulation rate and fewer

Table 5: Descriptive summary of students’ perceived demands while speaking.

Category	Arg	PicN	RtoS	RwLtoS	Total
Conceptualisation	12	8	16	24	59
Formulation	11	22	5	8	46
Articulation	2	3	0	1	6
Total	25	32	21	33	111

Note. The total number of respondents = 40; Arg = Argumentative task; PicN = Picture narrative task; RtoS = Reading-to-speaking task; RwLtoS = Reading-while-Listening to speaking task.

silent pauses (both mid- and end-clause pauses), but more filled pauses, and fewer self-repetitions and self-corrections. From a theoretical perspective, the necessity for content generation and organizing different ideas, including opinions and supporting information, in a coherent manner in the argumentative task may elevate the conceptualising demands, subsequently lowering fluency (Préfontaine and Kormos 2015). However, learners’ speech was more fluent in all the subdimensions of fluency (speed, breakdown, repair fluency) in the argumentative task than in the PicN task.

To elucidate the overall fluency advantage in the argumentative task, the validity of the current manipulation of conceptualising demands must first be discussed. Notably, we observed the higher frequency of filled pauses in the argumentative task, compared to the PicN task. This finding is in line with previous studies reporting that speakers produced more filled pauses with enhanced conceptualisation demands, set by providing many alternative choices (Christenfeld 1994) and by discourse transitions (Fraundorf and Watson 2014; Greene and Cappella 1986; Roberts and Kirsner 2000). Accordingly, the current argumentative task may have partially succeeded in elevating conceptualisation demands. Moreover, the responses in the post-speaking questionnaire may also support the high conceptualising demands in the argumentative task, as evidenced by the slightly higher number of participants who reported conceptualising demands in the argumentative task compared to in the PicN task ($n = 12$ vs 8; see Table 5). The following excerpts³ indicated that in the argumentative task, they were engaged with content generation and discourse organisation – planning their opinions, examples and supporting information, while maintaining the coherence of information:

³ All excerpts in this article were translated from Japanese into English by the first author. The original Japanese responses and their English translations are available in the supplementary material.

- *I was not able to communicate my reasons very clearly when expressing my opinion. It was very difficult to speak clearly and coherently about my opinions.* (Participant ID 3004)
- *I had difficulty coming up with specific examples.* (Participant ID 3002)

Taken together, it seems plausible to argue that the current argumentative task encouraged students to engage with content elaboration. Alternatively, the observed fluency advantage in the argumentative task might be explained by the interplay between the open-ended nature of the task and pre-task planning time. First, despite the demands on content generation, open-ended tasks may allow speakers to select only the information that they can express with their own resources, which can preemptively reduce breakdowns in speaking performance (Préfontaine and Kormos 2015). Hence, even if students conceptualised a highly complex or elaborated message, they could modify or simplify the message so as to express it effortlessly with their own linguistic repertoires. Students could rely on retrieving readily available linguistic resources from memory (cf. Barrett et al. 2004) rather than advanced lexical items and constructions corresponding to an elaborated message, which are not likely to be fully automatised and requires conscious effort and time to retrieve (i.e., controlled processing). Unlike in the argumentative task, some speakers mentioned that even though the visual prompt of the PicN task helped them to readily specify what to express, they had difficulty with retrieving the corresponding vocabulary items. This indicates that due to the lack of accessible lemmas corresponding to their intended message, they might have needed to consciously select alternative lemmas. According to the dual-process theory (Evans and Stanovich 2013), such conscious effort might have increased the demands on their limited attentional resources, subsequently lowering fluency. These characteristics of the PicN task are mentioned in the following excerpts from the post-speaking task questionnaire:

- *The original story was there, so it was easy to get to the point and decide to speak.* (Participant ID 2039)
- *It was difficult to explain some pictures. I didn't know the English translation of Katei Saien (home gardening).* (Participant ID 3005)
- *What I found challenging was to describe Noujo ga kakudai siteiku (the expansion of a farm).* (Participant ID 3003)

Second, three minutes were provided for pre-task planning in all the speaking tasks in the current study. Given that longer pre-task planning improves fluency (Tavakoli and Skehan 2005), due to the three-minute planning time in the current argumentative task, the participants could have sufficiently prepared the speech content they could communicate with their own linguistic repertoires. In other words, the effects

of conceptualising demands in the argumentative task may have been counteracted by sufficient pre-task planning time. Meanwhile, the predefined content of speech in the PicN task might have pushed them to describe some information that was difficult but essential for task accomplishment, thus lowering the fluency of their performance.

A close examination of each fluency measure can paint a more nuanced picture of how enhanced conceptualising demands are reflected in fluency. End-clause pauses have been found to reflect conceptualisation-related processes (De Jong 2016; Tavakoli 2011). It can thus be hypothesised that end-clause pauses will be more frequent in open tasks where conceptualising demands are elevated, compared to closed tasks. However, the current study observed the opposite pattern. The frequency of end-clause pauses was higher in the PicN task (i.e., more closed task) than in the argumentative task (i.e., more open task), indicating that the participants were pushed to modify the preverbal message more frequently. In the PicN task, they could not avoid expressing certain information to accomplish the PicN task, even if the corresponding linguistic items were not available in their linguistic resources. When they fail to retrieve corresponding lemmas, they need to modify the planned message so that the intended message can be adequately communicated by their own resources. Since modification of the preverbal message consumes attentional resources (Kormos 2006; Levelt 1989), the amount of attentional resources for conceptualisation might have been reduced (cf. Felker et al. 2019), which may explain the higher frequency of breakdowns at clausal boundaries in the PicN task.

6.2 Speech processing demands on formulation

To examine the effects of two types of formulation demands – different activation levels of lexical and phonological representations – on speech production, the current study compared students' fluency between PicN and RtoS tasks and between RtoS and RwltoS tasks, respectively. The results demonstrated that the enhanced activation of lexical representations led to a slower articulation rate, more mid-clause pauses, fewer but longer end-clause pauses, more filled pauses, fewer repetitions, and more false starts. Meanwhile, the enhanced activation of phonological representations resulted only in longer mid-clause pauses.

Considering the nature of activation spreading in speech production (Kormos 2006; Levelt 1989), the enhanced activation of lexical representations is supposed to facilitate the retrieval of linguistic items during speech production and thus improve fluency. However, participants' speaking performance in the RtoS task was characterised by a slower articulation rate and more mid-clause pauses, compared to the PicN task (i.e., lower fluency). Contrary to the expected facilitative effects, the results

here suggest that speech production in the RtoS task may have been impeded by the enhanced activation of task-relevant linguistic items prior to speech. One possible explanation for this finding is a potential competition between the activation routes from speakers' internal conceptualisation and the external source text of the RtoS task. Although the source text should externally activate task-relevant linguistic items in speaker's mental lexicon, the activated linguistic items may not necessarily have matched the internal activation from the speaker's conceptual plan of the message. Similarly, another possible reason for the lack of such facilitative effects in the current study may lie in the size of speech planning unit (Levelt 1989). The majority of studies reporting priming effects commonly elicits speakers' production in a controlled manner, such as single words and short sentences (see McDonough and Trofimovich 2008). In priming research, possibly due to its experimental nature, participants typically do not plan their production over a longer stretch of discourse (see Kim et al. 2019). Therefore, a relatively long discourse required in the current RtoS task might have inhibited the expected facilitation of retrieving relevant linguistic items.

Although the content of speech is largely predefined in both the PicN task and the RtoS task, the frequency of filled pauses was higher in the RtoS task. This finding suggests that despite the similar nature of predefined speech content, the RtoS task might have imposed slightly higher conceptualising demands than the PicN task (see Fraundorf and Watson 2014; Greene and Cappella 1986; Roberts and Kirsner 2000). Consistently, more participants reported conceptualising demands in the RtoS task, compared to the PicN task ($n = 16$ vs 8; see Table 5). The following participants' responses also suggest that the slightly higher conceptualising demands in the RtoS task may have derived from the necessity to recall and select the content of source texts:

- *I thought I understood the text largely, but I had difficulty with remembering what it said when I was speaking.* (Participant ID 2028)
- *I wanted to choose the information that would make it easier to understand, but I couldn't.* (Participant ID 3007)

The slightly higher conceptualising demands in the RtoS task might also be evidenced by the higher number of false starts and the longer duration of end-clause pauses. In the light of speech production mechanisms, false starts may indicate the need for allocating more attentional resources for conceptual processing, including discourse management (Williams and Korko 2019). Similarly, end-clause pauses are closely related to conceptualisation processes such as content planning (De Jong 2016; Götz 2013; Lambert et al. 2017). It thus seems plausible that in the RtoS task, our participants might have experienced a shortage of attentional resources and needed more time for content planning due to their active engagement with recalling and selecting

information from the source text. However, the frequency of end-clause pauses was lower in the RtoS task than in the PicN task. Albeit speculative, one possible explanation is that the frequency and duration of end-clause pauses may reflect different processes of conceptualisation, that is, macroplanning and microplanning (Kormos 2006; Levelt 1989). The aforementioned demands on recalling and selecting information in the RtoS task may be associated with macroplanning difficulty. It can thus be postulated that the longer duration of end-clauses was reflective of the difficulty in macroplanning processes. However, to comprehend the source text, participants were supposed to have parsed the text and have specified the informational aspects of the text (i.e., the creation of a situation model; see van Dijk and Kintsch 1983). In other words, reading comprehension in the RtoS task could have created a memory trace of such informational aspects of the text in students' short-term memory. As a result, such memory traces of information that are specified in microplanning (see Levelt 1989) might have helped them reduce breakdowns at clausal boundaries.

By comparing fluency measures between RtoS and RwLtoS tasks, there was an examination of which fluency measures are associated with the enhanced activation of phonological representations operationalised by a bimodal source text (i.e., reading-while-listening in text comprehension). The results showed a significant effect of bimodal input only in mid-clause pause duration. Students produced longer mid-clause pauses in the RwLtoS task than in the RtoS task, indicating that the enhanced activation of phonological representations resulted in lower fluency performance. As with the contrast between the PicN task and the RtoS task, this finding can be explained by the potential competition between internal and external activation. In addition, the results indicated that further enhanced competition from bimodal input may have extended the latency of retrieving linguistic items. Theoretically, mid-clause pauses are reflective of disruptions in the retrieval of L2-specific linguistic knowledge (De Jong 2016; Götz 2013; Lambert et al. 2017). The longer mid-clause pauses in the RwLtoS task might thus be interpreted as the evidence of a delay in linguistic retrieval, possibly due to the further external activation of in-text linguistic items by the bimodal source text.

7 Conclusions and limitations

In combination with the dual-process theory (Evans and Stanovich 2013), the current study aimed to extend the framework of speech processing demands (Skehan 2009, 2014; Skehan et al. 2012) and speech production mechanisms (Kormos 2006; Levelt 1989, 1999) by comparing L2 learners' fluency across four speaking tasks, which differed in the demands on conceptualisation and formulation. For the sake of the ecological validity of findings, we intentionally adopted spontaneous speaking tasks

widely used in pedagogical and assessment contexts. The current findings can provide insights into the construct validity of fluency measures in L2 research. Due to the serial nature of speech production, conceptualisation processes may be reflected in the whole range of utterance fluency measures, while speakers' engagement with content planning can be particularly observed in the frequency of filled pauses (Fraundorf and Watson 2014; Greene and Cappella 1986; Roberts and Kirsner 2000). Meanwhile, formulation processes (i.e., L2-specific linguistic encoding) may be primarily associated with the frequency and length of mid-clause pauses (De Jong 2016; Lambert et al. 2017; Suzuki and Kormos 2020).

The current study has several methodological limitations. First, due to the focus on task characteristics, the 3-min pre-task planning was allowed consistently across tasks. However, the findings indicated a potential interplay between task characteristics and pre-task planning time. Considering the multifaceted nature of planning in speech production (for a review, see Skehan 2014; Skehan et al. 2012), the current study should be replicated with varying conditions for pre-task and on-line planning. Second, the demands on content generation and discourse organisation may not fully control for the complexity of preverbal messages and the subsequent linguistic demands of messages. Considering the serial nature of speech production (Kormos 2006; Levelt 1989), it might be virtually impossible to separate the demands on conceptualisation and formulation. Therefore, future studies should seek to further clarify the effects of conceptualising demands on fluency with valid measurement of the complexity of preverbal messages (cf. Bulté and Housen 2012; Skehan 2009; Vasylets et al. 2017). Finally, due to the current interest in the ecological validity of findings, we adopted four common speaking task types. However, those tasks could differ in various aspects beyond speech processing demands (e.g., genre, topic difficulty), and thus future research may also be recommended to explore methodologies to control for those variables (for challenges in this issue, see Skehan 2009).

Despite these limitations, the current study represents the first systematic attempt to disentangle the effects of conceptualisation and formulation demands on L2 fluency using ecologically valid speaking tasks. In particular, the introduction of the argumentative task as a means of operationalising conceptualisation demands offers a viable approach. The study also highlights a methodological challenge for task-based performance research. The manipulation of task demands by researchers may not always exert the intended effects, as learners can flexibly and purposefully allocate their attentional resources across different speech production processes. This variability underscores the importance of individual differences in how learners approach tasks both in speaking performance assessment and task-based pedagogies (Bryfonski et al. 2024; see also Révész et al. 2024).

Acknowledgments: This study is based on part of the first author's PhD thesis at Lancaster University, supervised by the second author.

Research ethics: The study reported in our manuscript is part of the first author's PhD thesis (Suzuki, 2021), which was supervised by the second author, and the ethical approval for the PhD project was granted by the Faculty of Arts and Social Sciences and Lancaster Management School's Research Ethics Committee of Lancaster University in January 2019.

Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors state no conflict of interest.

Research funding: This study was supported by Grant-in-Aid for Early-Career Scientists from Japan Society for the Promotion of Science (22K13181) to the first author.

Data availability: Upon acceptance, anonymised dataset will be available on IRIS repository.

References

- Baddeley, Alan. 2003. Working memory and language: An overview. *Journal of Communication Disorders* 36(3). 189–208.
- Barr, Dale J. 2013. Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology* 4(June). 3–4.
- Barrett, Lisa Feldman, Michele M. Tugade & Randall W. Engle. 2004. Individual Differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin* 130(4). 553–573.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). <https://doi.org/10.18637/jss.v067.i01>.
- Bock, J. Kathryn. 1982. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review* 89(1). 1–47.
- Boersma, Paul & David Weenink. 2012. Praat: Doing phonetics by computer [Computer software].
- Bosker, Hans Rutger, Anne-France Pinget, Hugo Quené, Ted Sanders & Nivja H. De Jong. 2013. What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing* 30(2). 159–175.
- Bryfonski, Lara, Yunjung Ku & Alison Mackey. 2024. Research methods for IDs and TBLT: A substantive and methodological review. *Studies in Second Language Acquisition*. 617–643. <https://doi.org/10.1017/S0272263124000135>.
- Bulté, Bram & Alex Housen. 2012. Defining and operationalising L2 complexity. In Alex Housen, Folkert Kuiken & Ineke Vedder (eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, 23–46. Amsterdam: John Benjamins.
- Chandler, Paul & John Sweller. 1991. Cognitive load theory and the format of instruction. *Cognition and Instruction* 8(4). 293–332.
- Christenfeld, Nicholas. 1994. Options and UMS. *Journal of Language and Social Psychology* 13(2). 192–199.
- Council of Europe. 2001. *The common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

- Coupé, Christophe. 2018. Modeling linguistic variables with regression models: Addressing non-gaussian distributions, non-independent observations, and non-linear predictors with random effects and generalized additive models for location, scale, and shape. *Frontiers in Psychology* 9(April). 1–21.
- de Bot, Kees. 1992. A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics* 13. 1–24.
- De Jong, Nivja H. 2016. Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching* 54(2). 113–132.
- De Jong, Nivja H. & Hans Rutger Bosker. 2013. Choosing a threshold for silent pauses to measure second language fluency. *DiSS 2013. Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech* (January 2013), 17–20.
- Educational Testing Service. 2013. TextEvaluator®. Available at: <https://textevaluator.ets.org/>.
- Evans, Jonathan St. B. T. & Keith E. Stanovich. 2013. Dual-process theories of higher cognition. *Perspectives on Psychological Science* 8(3). 223–241.
- Felker, Emily, Heidi Klockmann & Nivja H. De Jong. 2019. How conceptualizing influences fluency in first and second language speech production. *Applied Psycholinguistics* 40(1). 111–136.
- Ferrand, Ludovic & Jonathan Grainger. 1993. The time course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the Psychonomic Society* 31. 119–122.
- Foster, Pauline & Parvaneh Tavakoli. 2009. Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning* 59(4). 866–896.
- Foster, Pauline, Alan Tonkyn & Gillian Wigglesworth. 2000. Measuring spoken language: A unit for all reasons. *Applied Linguistics* 21(3). 354–375.
- Fraundorf, Scott H. & Duane G. Watson. 2014. Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience* 29(9). 1083–1096.
- Götz, Sandra. 2013. *Fluency in native and nonnative English speech*. Amsterdam: John Benjamins.
- Greene, John & Joseph N. Cappella. 1986. Cognition and talk: The relationship of semantic units of temporal patterns of fluency in spontaneous speech. *Language and Speech* 29(2). 141–157.
- Housen, Alex, Folkert Kuiken & Ineke Vedder. 2012. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins Publishing Company.
- Indefrey, Peter & Willem J. M. Levelt. 2004. The spatial and temporal signatures of word production components. *Cognition* 92(1–2). 101–144.
- JACET. 2003. *JACET List of 8000 Basic Words*. Tokyo: JACET English Vocabulary SIG.
- Jackson, Daniel O. & Sakol Suethanapornkul. 2013. The cognition hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning* 63(2). 330–367.
- Kim, You Jin, Yeon Joo Jung & Stephen Skalicky. 2019. Linguistic alignment, learner characteristics, and the production of stranded prepositions in relative clauses: Comparing FTF and SCMC contexts. *Studies in Second Language Acquisition* 41(5). 937–969.
- Kormos, Judit. 2006. *Speech production and second language acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Kormos, Judit, Shungo Suzuki & Masaki Eguchi. 2022. The role of input modality and vocabulary knowledge in alignment in reading-to-speaking tasks. *System* 108. 102854.
- Košák-Babuder, Milena, Judit Kormos, Michael Ratajczak & Karmen Pižorn. 2019. The effect of read-aloud assistance on the text comprehension of dyslexic and non-dyslexic English language learners. *Language Testing* 36(1). 51–75.
- Lambert, Craig & Judit Kormos. 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.

- Lambert, Craig, Judit Kormos & Danny Minn. 2017. Task repetition and second language speech processing. *Studies in Second Language Acquisition* 39(1). 167–196.
- Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, Mass: MIT Press.
- Levelt, W. J. M. 1999. Language production: A blueprint of the speaker. In C. Brown & P. Hagoort (eds.), *Neurocognition of language*, 83–122. Oxford: Oxford University Press.
- Liu, Yeu Ting & Andrew Graeme Todd. 2014. Dual-modality input in repeated reading for foreign language learners with different learning styles. *Foreign Language Annals* 47(4). 684–706.
- Luoma, Sari. 2004. *Assessing speaking*. Cambridge: Cambridge University Press.
- McDonough, Kim & Pavel Trofimovich. 2008. *Using priming methods in second language research*. New York, NY: Routledge.
- McNamara, Danielle S., Arthur C. Graesser, Philip M. McCarthy & Zhiqiang Cai. 2014. *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.
- Millington, Neil. 2019. Dreamreader.net. <http://dreamreader.net/> (accessed 10 May).
- Mirdamadi, Farhad & Nivja H. De Jong. 2015. The effect of syntactic complexity on fluency: Comparing actives and passives in L1 and L2 speech. *Second Language Research* 31(1). 105–116.
- Pallotti, Gabriele. 2020. Measuring complexity, accuracy, and fluency (CAF). In Paula Winke & Tineke Brunfaut (eds.), *The routledge handbook of second language acquisition and language testing*, 201–210. New York: Routledge.
- Peng, Peng, Marcia Barnes, CuiCui Wang, Wei Wang, Shan Li, H. Lee Swanson, William Dardick & Sha Tao. 2018. A meta-analysis on the relation between reading and working memory. *Psychological Bulletin* 144(1). 48–76.
- Préfontaine, Yvonne & Judit Kormos. 2015. The relationship between task difficulty and second language fluency in French: A mixed methods approach. *The Modern Language Journal* 99(1). 96–112.
- R development Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Available at: <http://www.r-project.org/>.
- Révész, Andrea, Hyeonjeong Jeong, Shungo Suzuki, Haining Cui, Shunsui Matsuura, Kazuya Saito & Motoaki Sugiura. 2024. Task-generated processes in second language speech production: Exploring the neural correlates of task complexity during silent pauses. *Studies in Second Language Acquisition* 46(4). 1179–1205.
- Roberts, Benjamin & Kim Kirsner. 2000. Temporal cycles in speech production. *Language and Cognitive Processes* 15(2). 129–157.
- Robinson, Peter. 2011. *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. Amsterdam: John Benjamins.
- Schmidt, D. 2020. sylcount: Syllable counting and readability measurements. R package version 0.2-2. Available at: <https://CRAN.R-project.org/package=sylcount>.
- Segalowitz, Norman. 2010. *Cognitive bases of second language fluency*. London & New York: Routledge.
- Segalowitz, Norman. 2016. Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching* 54(2). 79–95.
- Skehan, Peter. 1996. A framework for the implementation of task-based instruction. *Applied Linguistics* 17(1). 38–62.
- Skehan, Peter. 2009. Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics* 30(4). 510–532.
- Skehan, Peter. 2014. Limited attentional capacity, second language performance, and task-based pedagogy. In Peter Skehan (ed.), *Processing perspectives on task performance*, 211–260. Amsterdam: John Benjamins.
- Skehan, Peter, Xiaoyue Bei, Qian Li & Zhan Wang. 2012. The task is not enough: Processing approaches to task-based performance. *Language Teaching Research* 16(2). 170–187.

- Sprenger, Simone A., Willem J. M. Levelt & Gerard Kempen. 2006. Lexical access during the production of idiomatic phrases. *Journal of Memory and Language* 54(2). 161–184.
- Suzuki, Shungo. 2021. The multidimensionality of second language oral fluency: The interface between cognitive, utterance, and perceived fluency (Doctoral thesis). Lancaster University, UK.
- Suzuki, Shungo & Judit Kormos. 2020. Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition* 42(1). 143–167.
- Suzuki, Shungo & Judit Kormos. 2023. The multidimensionality of second language oral fluency: Interfacing cognitive fluency and utterance fluency. *Studies in Second Language Acquisition* 45(1). 38–64.
- Suzuki, Shungo, Judit Kormos & Takumi Uchihara. 2021. The relationship between utterance and perceived fluency: A meta-analysis of correlational studies. *The Modern Language Journal* 105(2). 435–463.
- Suzuki, Shungo & Andrea Révész. 2023. Measuring speaking and writing fluency: A methodological synthesis focusing on automaticity. In Yuichi Suzuki (ed.), *Practice and automatization in second language research: Theory, methods, and pedagogical implications*, 247–266. New York: Routledge.
- Suzuki, Shungo, Toshinori Yasuda, Keiko Hanzawa & Judit Kormos. 2022. How does creativity affect second language speech production? The moderating role of speaking task type. *Tesol Quarterly* 56(4). 1320–1344.
- Tavakoli, Parvaneh. 2011. Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal* 65(1). 71–79.
- Tavakoli, Parvaneh & Pauline Foster. 2008. Task design and second language performance: The effect of narrative type on learner output. *Language Learning* 61(2). 37–72.
- Tavakoli, Parvaneh & Ann-Marie Hunter. 2018. Is fluency being ‘neglected’ in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research* 22(3). 330–349.
- Tavakoli, Parvaneh, Fumiyo Nakatsuhara & Ann-Marie Hunter. 2020. Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal* 104(1). 169–191.
- Tavakoli, Parvaneh & Peter Skehan. 2005. Strategic planning, task structure, and performance testing. In Rod Ellis (ed.), *Planning and task performance in a second language*, 239–273. Amsterdam: John Benjamins.
- van Dijk, Teun A. & Walter Kintsch. 1983. *Strategies of discourse comprehension*. New York, NY: Academic Press.
- Vasylets, Olena, Roger Gilabert & Rosa M. Manchón. 2017. The effects of mode and task complexity on second language production. *Language Learning* 67(2). 394–430.
- Williams, Simon A. & Malgorzata Korko. 2019. Pause behavior within reformulations and the proficiency level of second language learners of English. *Applied Psycholinguistics* 40(3). 723–742.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/iral-2024-0185>).