What makes a task difficult? Pre-service teacher analysis of task difficulty

Author 1: Parvaneh Tavakoli, University of Reading, UK

Author 2: Farahnaz Faez, Western University, Canada

To be published in **TASK |** Journal on Task-Based Language Teaching and Learning

Title: What makes a task difficult? Pre-service teacher analysis of task difficulty

Abstract

Two conceptual models of Task Complexity, Cognition Hypothesis (Robinson, 2001, 2007) and Limited Attentional Capacity (Skehan, 1998; 2003; 2018) have been proposed and widely debated in the task-based language teaching (TBLT) literature. However, little empirical evidence exists to suggest either of the models is based on teacher input or being used by teachers for classroom use. Drawing on pre-service teacher analysis of task difficulty, the study aimed to develop an in-depth understanding of task features they consider when evaluating task difficulty. Participants, 127 pre-service teachers at the end of their one-year MA TESOL program in Ontario, Canada, evaluated two sets of sample tasks, ranking them according to their degree of difficulty and identifying the features that contributed to this difficulty. 727 pieces of raw data, extracted from the task difficulty analysis, were categorized. Five main categories of task difficulty were identified, namely 1) linguistic demand, 2) cognitive operational demand, 3) design features, 4) informational demand, and 5) communicative demand. Learner related factors, external to task design, were also suggested as issues related to difficulty. We propose a set of task difficulty features that can be used in replication and validation studies to help with the development of a teacher evidence-based model of task difficulty.

Key words: task difficulty, task features, pre-service teacher

Introduction

Using ‘tasks’ in second language (L2) teaching has become a prominent approach in L2 instructional settings around the world. More teachers and schools are using tasks in either a task-supported approach (i.e., where tasks are used as a smaller unit of curriculum to support

teaching objectives) or a task-based teaching approach (i.e., where the entire curriculum is based on tasks) (Ellis, 2019; Tavakoli & Jones, 2018). Tasks are similarly popular in other approaches to language teaching including Content and Language Integrated Language Learning, CLIL, (Dalton-Puffer, 2011; Tavakoli & Jones, 2018). The rationale for using tasks in L2 teaching is strongly supported by the psycholinguistic and pedagogic evidence that suggests tasks effectively promote second language learning in instructional settings (East, 2021; Ellis, 2019). From a psycholinguistic perspective, it is believed that tasks promote L2 learning by providing opportunities for meaningful communication that enhance naturalistic acquisitional processes (Révész & Gurzynski-Weiss, 2016; Skehan, 1998). From a pedagogic perspective, tasks are reported to promote a learner-centred approach to pedagogy and enhance learner autonomy through learning-by-doing (East, 2021; Ellis, 2003, 2019; Erlam, 2016; Shintani, 2016; Van den Branden, 2006). Using tasks in L2 teaching inevitably requires teachers to have adequate skills and knowledge to analyze, evaluate, and sequence tasks (Erlam & Tolosa, 2022). To this aim, teachers are commonly engaged in the process of analyzing and evaluating tasks in relation to both learner needs (e.g., difficulty level, interest, engagement) and teaching objectives (e.g., development of a certain skill or acquisition of a language function).

Determining task difficulty and developing a task sequencing plan are an integral part of the task selection process. In many pedagogic contexts, task selection and task sequencing decisions are guided by a needs analysis. Needs analysis, in this sense, can provide rich information about task frequency, difficulty, and training requirements, which would, in turn, "aid the highly complex decision regarding how tasks may be selected" (Gilabert & Malicka, 2021, p. 226). Needs analysis, however, is only possible if the characteristics of the target learner group(s) and their needs and aims are known in advance of the syllabus design. This information, although much needed and valued, is often not available to most teachers and teacher educators.

Selecting appropriate tasks that meet learning objectives is crucial to the success of TBLT instruction, yet there seems to be little information for teachers on what criteria to consider and how to determine task difficulty and suitability. While more experienced teachers draw on their previous experiences and/or rely on their own intuition (Ellis, Skehan, Shintani & Lambert, 2019; Erlam & Tolosa, 2022), less experienced teachers may find analyzing tasks and evaluating their difficulty demanding. Therefore, teachers would benefit from a scheme to help with task

analysis and evaluation for teaching purposes. Skehan (2018, p. 28) argues that developing such a scheme is essential as it would "enable teachers to choose the best for their learners" helping them "avoid making bad choices" that may adversely affect the development of learner interlanguage.

Since TBLT early days, there have been repeated calls for research in task difficulty to help teachers, material writers and syllabus designers to evaluate, grade and sequence tasks (Baralt, Gilabert, & Robinson, 2014; Baralt, Harmath-de Lemmos & Werfelli, 2014; Ellis et al., 2019; Prabhu, 1987; Révész & Gurzynski-Weiss, 2016; Skehan, 1998, 2003, 2018; Tavakoli, 2009a). Three decades after the initial calls, task difficulty and its relationship to task sequencing seem to be an under-researched area of TBLT. Ellis et al. (2019, p. 14) argue that issues related to task grading and sequencing remain to be "a major challenge in TBLT". Skehan (2018) claims that developing a scheme for defining and analyzing task difficulty will be a major step towards developing a systematic approach to task sequencing. Supporting the rationale for developing a task difficulty scheme, Baralt, Gilabert and Robinson (2014, p.1) reiterate that to date "there is still no widely agreed-upon set of criteria that can be used to grade and sequence tasks." The study reported in this paper is an attempt to help fill this gap by enhancing a more in-depth understanding of task difficulty and developing a set of task difficulty features based on evidence collected from less-experienced teachers' analysis and evaluation of task difficulty.

Research on task difficulty

The concept of task difficulty, or the demands and challenges performing a task imposes on L2 learners, has been central to TBLT research from its early emergence. Among the several proposed models of task difficulty (e.g., Candlin, 1987; Ellis, 2003; Prabhu, 1987), the two widely cited models are Skehan's (1998, 2003, 2018) limited attentional capacity (LAC) and Robinson's (2001, 2007, 2011) triadic componential framework (TCF). Skehan's LAC model proposes that task difficulty can be explained in terms of three dimensions of *code complexity* (the linguistic demands associated with performing a task), *cognitive complexity* (the cognitive processes stimulated by a task) and *communicative stress* (the pressure prompted by the conditions under which a task is performed). In LAC, code complexity considers a range of

factors such as grammatical difficulty, vocabulary load and text density; cognitive complexity refers to the demands associated with cognitive familiarity (e.g., whether the task, its topic and discourse are familiar) and cognitive processing (e.g., whether the information is well organized and requires much transformation and manipulation); communicative pressure, highlights the role of performance conditions that give rise to pressure linked with task performance (e.g., stakes, modality, and number of and relationship between interlocutors). LAC postulates that learners will find a task more difficult to perform if it involves high linguistic demands, low cognitive familiarity, high cognitive processing, and high communicative pressure. More recently, Skehan (2018) revisits his model and adds some key information about the principles that should be considered. Linked to our focus on task difficulty in this paper, Skehan refers to Levelt's (1989) model of speech production and argues that "task difficulty needs to be analyzed distinctly for the Conceptualizer and the Formulator" (Skehan, 2018, p. 27). Levelt (1989) proposes that the speech production process involves three stages of Conceptualizer (where pre-verbal message is generated), Formulator (where ideas are transformed to linguistic forms), and Articulator (where linguistic forms change to overt speech).

Robinson's (2001, 2011) TCF model proposes that task difficulty should be explained along three dimensions of task complexity (i.e., cognitive demands induced in a task), task conditions (i.e., interactive demands associated with task performance) and task difficulty (i.e., individual learner dependent factors). In Robinson's model, task complexity refers to the inherent cognitive demands of a task and is divided into *resource-directing* and *resource-dispersing* elements with the former concentrating on conceptual demands (e.g., if there are causal reasoning demands), and the latter focusing on procedural demands (e.g., whether the task demonstrates a clear structure). Task conditions include factors about the nature and amount of interaction between the interlocutors including the kind of participation required or the amount of contribution expected. And finally, task difficulty encapsulates learner-dependent variables such as motivation and anxiety that potentially affect task difficulty.

In a revised version of the model, known as SSARC (*stabalize*, *simplify*, *automatize*, *restructure* and *complexify)*, Robinson (2010) aims to develop the TCF model in a pedagogic direction with task sequencing in its heart. Robinson (2010, 2015) proposes that in order to use tasks in a syllabus, the five steps (i.e., *stabalize*, *simplify*, *automatize*, *restructure* and *complexify*) should be considered (see Robinson, 2015, for further details). The SSARC model, in principle contends

that learners should first perform tasks that are simple on all aspects of the TCF model parameters before performing tasks with an increasing level of cognitive complexity. Robinson (2015) argues that the SSARC model provides a solid foundation for task sequencing, but he proposes two principles to be considered and observed. First, only cognitive complexity (both resource-directing and resource-dispersing) should be considered for task sequencing purposes. He argues that issues related to task difficulty and task condition cannot be considered relevant to task sequencing. Learner factors, for example, are usually not known at the syllabus design stage, and therefore, they may not play an important role in the decision-making process of task selection and sequencing. Second, resource-dispersing factors should increase before resource-directing factors as this would allow for the move from knowing to automatizing and complexifying.

As can be seen above, the two models, although both rooted in a cognitive approach to L2 acquisition, vary in several regards. First, LAC considers task difficulty only in relation to characteristics inherent within the task and its performance conditions, but TCF extends the concept to include factors such as learner individual variables. A second difference is in how the components of each model are organized. LAC considers three constructs of cognitive demands, linguistic demands and performance condition demands as distinct but inter-related constructs. However, TCF considers cognitive demands as the major component of the model assuming some of the performance conditions to be categorized under the cognitive category (e.g., planning time) and others under task conditions (e.g., one-way or two-way exchange of information), with a potential for overlapping effects. The most important difference, relating to the focus of the current paper, is that while LAC postulates that task sequencing should rely on all three task difficulty factors (code complexity, cognitive complexity and communicative pressure), TCF suggests cognitive complexity alone should be taken as the basis for task grading and sequencing.

It is necessary to note that the two models use terminology differently. In Skehan's model task difficulty is used to refer to the overall demands of a task, but Robinson uses the term differently distinguishing between task complexity (task internal) and task difficulty (task external) factors to define the demands associated with learner characteristics. In this article, we follow Skehan's use of the term *task difficulty* to refer to the overall demands of a task associated with task performance.

Teacher analysis of task difficulty

While substantial research has been invested in investigating the impact of task design on task difficulty in relation to learner performance and L2 acquisition (Foster & Skehan, 1996; Gilabert, 2007; Michel, 2011; Révész, 2009; Robinson, 2001, to name a few), little research has focused on teachers' analysis and evaluation of task difficulty for the purpose of teaching or material design. One of the earlier studies investigating teachers' perceptions of task difficulty is Tavakoli (2009a). Working with both English language teachers and learners, Tavakoli (2009b) investigated their perceptions of task difficulty through task performance and retrospective interviews to identify factors affecting perceptions of task difficulty. Performing four different oral narrative tasks, the learners considered cognitive and linguistic demands as two determining factors affecting task difficulty. While both teachers and learners were concerned about issues related to cognitive complexity (e.g., organization and clarity of information), teachers also described learner factors (e.g., learner attention) and teacher factors (e.g., teacher instructions) as variables affecting task difficulty. While this study provided a useful insight in teacher and learner perceptions of task difficulty, its findings were limited as its design did not allow for an evaluation of a range of different tasks or with a larger sample of teachers.

A second study to report here is Révész and Gurzynski-Weiss (2016) who used think-aloud protocols and eye-tracking techniques to examine teachers' perspective on task difficulty. The participants, 16 English language teachers, were asked to identify sources of difficulty in decision making and information gap tasks and to discuss the linguistic ability needed to perform the tasks. The teachers were also asked to discuss how they would manipulate the task to make them suitable for their learners. The findings suggested that the teachers were primarily concerned about linguistic demands when analyzing the tasks. When discussing how to make the task more suitable for their learners, they relied on conceptual demands as the factor increasing task difficulty. The study's contribution was unique in that it used a methodological novel perspective to examining teacher perspectives to task difficulty. The authors acknowledged that focusing on only two task types (i.e., decision-making and information gap task) and working with a small sample size were key limitations of the study.

In sum, there is limited research on what task features teachers consider when choosing tasks or evaluating and analyzing task difficulty for the purpose of their teaching. As can be seen above, the few studies examining this topic have focused on a limited number of task types (narratives in Tavakoli, 2009a, and decision-making and information gap in Révész and Gurzynski-Weiss, 2016), or data collected from a small group of teachers, two important research design elements that might have had a restricting impact on their findings. In this study, we aim to recruit a larger group of (pre-service) teachers and more varied task types which would enable us to elicit rich data on teacher analysis of task difficulty; the study is interested in the reasoning the teachers provide and criteria they consider when analyzing task difficulty. This will further complement previous research as it will employ a new approach to examining task difficulty paving the way for task difficulty ranking and sequencing. Including pre-service teachers, an important teaching practitioner population, in this study is a valuable contribution to the field as the findings will offer important pedagogic implications for teacher training programmes and potentially help promote change in "current teaching traditions" (Maijala et el., 2023).

Aims of the study

The primary aim of the current study was to investigate pre-service TESOL teachers' understanding of task design and how they analyze and evaluate task difficulty in language teaching. The study specifically sought to develop an insight into how less-experienced TESOL teachers analyze tasks and what features they suggest affect task difficulty. The study's secondary aim was to compare these features with the existing models of task difficulty to determine the extent to which these models capture the teachers' criteria for task difficulty. The proposed set of features can be used by other researchers, teachers, and teacher training programmes to provide a basis for further evaluation and validation. The question guiding the current study is 'What features do pre-service English language teachers consider when analyzing and evaluating task difficulty?'

Participants

Participants in this study were 127 preservice teachers completing a course-based MA TESOL program in Ontario, Canada. Participants were all international students, with 95% originally from China. All participants had a proficiency level of B2/C1 at the start of the MA program. As the data were collected towards the end of their one-year program in Ontario, it is assumed that their proficiency either improved or remained at the same level. Biographical data indicated that the majority had no prior formal teaching experience. Limited occasional teaching experience was reported by a small group and ranged between 1-12 months. The MA TESOL program includes eight courses, and each course consists of 36 hours of face-to-face instruction completed in 12-weeks. The program did not include a course designated to task-based language teaching (TBLT), but participants were introduced to TBLT, the rationale for using it in the classroom, key principles of TBLT (e.g., Ellis, 2003; Faez & Tavakoli, 2019), task features, and task types in an introductory applied linguistics course. Students were also introduced to the Common European Framework of Reference (CEFR) in which its can-do statements and action-oriented approach lend themselves to a TBLT approach (Little, 2006). The MA TESOL program is intended to be 'hands on' and engaging in discussing reflective questions related to a topic is as a general practice in all courses. Students engaged in discussions related to advantages and limitations of using tasks in various types of English language classrooms. They also engaged in activities in which they had to identify and/or design appropriate tasks for prospective classrooms. However, it should be noted that introduction to TBLT in the MA program is still considered rudimentary and that none of the existing models of task difficulty was introduced or discussed in their curriculum. Neither did the training involve asking students to analyze task difficulty before the data collection started. Data was collected when participants were in their final semester of the program and had only a few weeks left to complete the degree requirements and graduate from the program. Given most of our participants came from one linguistic and cultural background with limited teaching experience, we are aware that the data collected are not representative of the target population of pre-service teachers globally, and as such, the results should not be considered generalizable.

Data Collection

Data was collected via a three-part questionnaire. The first part included questions to help apprehend teachers' understanding of a language teaching task. The second part, which is the

focus of this paper, included two similar questions to gauge their analysis of task difficulty and features that contribute to task difficulty from their perspective. In each question, participants were provided with a context and asked to rank three tasks appropriate for that context according to their level of difficulty from 1-3. In question 1, the three tasks included, a) An oral narrative based on a picture story, b) A spot-the-difference task, and c) Describe your perfect holiday. In question 2, the tasks included a) A discussion task of the advantages and disadvantages of travelling alone and travelling in a group, b) A balloon task in which participants had to choose which three people to keep on the balloon in case of a crash, c) A role-play job interview task[1]. Subsequently, they were asked which task was most difficult, why they thought a particular task was difficult, and factors that contributed to the difficulty of their selected task (see Appendix 1 for the two questions). Emphasis was made that there was no right or wrong answer. The third part of the questionnaire sought demographic information.

The choice of tasks was informed by some important decisions. Primarily, we were interested in speaking tasks to allow us to map the findings to Levelt's model, as discussed above. Second, to collect rich data, it was necessary to provide the participants with a range of speaking tasks of different types in terms of mode (e.g., monologic, dialogic), input (e.g., visual, textual), cognitive processes (e.g., comparing, analyzing, reasoning), and task outcomes (e.g., telling a story, finding a list of differences). Given the exploratory nature of the study, it was important to use a wide range of tasks that allow the participants to discuss a variety of different factors contributing to task difficulty. Third, to ensure the participants' familiarity, tasks were chosen from a pool of speaking tasks referred to in previous lectures on the course when introducing and discussing pedagogic tasks (e.g., Lambert & Oliver, 2020; Willis & Willis, 2007). These tasks, however, had not been analyzed in previous lectures. Finally, it was decided that the tasks were deliberately described minimally and presented in a 'task skeletons' format in order to leave a scope for a degree of teachers' personal interpretation of the tasks. A degree of openness to interpretation was believed to encourage teachers to think of and identify a wider range of factors affecting task difficulty. These tasks could have been replaced by any other speaking tasks if

[1] The task names used in this article act as labels to identify each task. They were not aimed at distinguishing tasks based on their type or topic.

they met the criteria above. Providing the participants with the actual tasks, although a more ecologically valid alternative, was not possible in the research design (e.g., limited class time).

Data Analysis

Data from both questions 1 and 2, totalling 254 sets of answers, were coded. On average participants provided three reasons within a range of 2-5; the length of each participant's analysis varied from a short sentence to explain the reason to writing a short paragraph justifying their choice of the difficult task. To begin the coding process, key information that provided a reason for task difficulty was entered into an excel sheet. All reasons provided by participants were included regardless of how many reasons each participant provided. We started by coding a subset of five participants to create a coding legend. We discussed our individual coding and legend through Zoom meetings and assigned data from five more participants to each of us to code independently for the next meeting. The first cohort we coded was discussed in four meetings, five participants for each of the first three meetings, and 10 participants in the fourth meeting. During each meeting, the team would discuss and compare each item and suggest any changes to be made to the coding legend and continued until there was a reached consensus. Each of the subsequent cohorts were discussed in a separate meeting. There were nine data-analysis discussion meetings in total.

Results

Before analyzing the data for features of task difficulty, we examined teachers' choices of difficult tasks (see Questions 1 and 2 in Appendix 1, and Table 1 below). When comparing three tasks of A) Oral narrative based on a picture story, B) Spot-the-difference task, and C) Describe your perfect holiday in Question 1, participants perceived task C (Describe your perfect holiday, 58.27%) to be most difficult, followed by Task A (Oral narrative, 28.35%) and Task B (Spot the difference,13.39%). For the second question, comparing three tasks of A) Discussion task, B) Balloon task, and C) Role-play job interview task, participants interpreted task C (Role play job interview task, 47.24%) as most difficult, followed by Task B (Discussion task, 41.73%), and task A (Balloon task, 11.02%).

Table 1

Participants' Ranking of Most Difficult Task

Cohorts	**Question 1**			**Question 2**		
	A Oral Narrative	B Spot-the-difference	C Your perfect holiday	A Discussion	B Balloon task	C Role-play job interview
C1 = 14	1	3	10	0	3	11
C2 = 25	8	2	15	1	15	9
C3 = 20	6	2	12	3	4	13
C4 = 23	7	5	11	3	11	9
C5 = 23	9	3	11	3	14	6
C6 = 22	5	2	15	4	6	12
Total = 127	36	17	74	14	53	60
Percentage	28.35	13.39	58.27	11.02	41.73	47.24

To understand what features pre-service English language teachers consider when analyzing and evaluating task difficulty, we worked with 727 pieces of information identified by the participants. Of these pieces, 15 were not clear enough for coding.

The broad categories identified through our coding and their frequency are reported in Table 2. The most frequent feature (28.79%) contributing to task difficulty seems to be the *linguistic demand* of a task and the least recurrent feature (8.71%) is task *communicative demand*. It is worth noting that our coding was driven by the reasons mentioned by the participants and is therefore totally based on their justifications as to why they perceived a particular task more difficult. We did not start with any preconceived categories or codes, and we constantly changed, expanded, and regrouped our categories and codes as we worked through the data. A small group of the features proposed could be linked to more than one category. In such cases, we used the context of the teacher's evaluation to make a decision about which category seemed the most suitable.

Table 2

Categorization of Task Difficulty

Codes	**Frequency**	**Percentage**
Linguistic Demand	205	28.79%
Cognitive Operational Demand	160	22.47%
Learner Background	105	14.74%
Design Features	101	14.18%
Informational Demand	79	11.10%
Communicative Demand	62	8.71%
Total	712	100%

Each of the six categories were derived from more specific codes that were grouped into broad categories outlined in Table 2. The specific codes that formed each of the broad categories are included in Tables 3-8 below. Each of the Tables include a description of the specific code along with their frequency and sample excerpts from the data. The sub-components of each category are presented in the order of their frequency.

Linguistic Demand. Participants overwhelmingly mentioned various aspects of linguistic demand that contributed to their perception of task difficulty. Their concerns about linguistic demand spread a wide spectrum of features from the general concerns about output demands and familiarity with genres to specific and more detailed features such as lexical and grammatical items. The four main sub-categories were *output demands*, *vocabulary*/phrases, *grammar,* and *discourse/genre*. Reasons such as, 'complete sentences are required to complete this task', 'output (speaking and writing) is more difficult than input (reading and listening)', and 'students might not know the proper words, expressions, phrases, grammar or structures for this task', are some examples that led us to the four codes under linguistic demand seen in Table 3.

Table 3

Linguistic Demand

Codes	Frequency	Description	Excerpts from the Data
1	119	Output demand/specific language skills	A lot of language skills will be tested in this task such as making a comprehensive dialogue. They need the language to describe the holiday.
2	53	Vocabulary/phrases	Describing a picture story requires using unfamiliar vocabulary.
3	24	Grammar	Using the right tense is necessary.
4	9	Discourse/genre	They don't know the genre of job interview discourse.

Cognitive Operational Demand. The second most frequently cited category of features in relation to task difficulty relates to concerns about the cognitive operational demands of a task. In this category, eight groups of features emerged, highlighting a range of cognitive operations that potentially increase the demands associated with processing task information. Issues reflecting the need to have organized and critical thinking skills, a good memory, attentional capacity, and ability to generate ideas, solve problems and make decisions were among the more frequent features identified as sources of task difficulty. As can be seen in Table 4, the most frequent sub-categories were *reasoning*, *critical thinking*, and *idea generation/creativity*, respectively, while *problem solving*, and *decision making* were mentioned least frequently. Overall, the participants perceived task as more difficult and demanding if they needed more cognitive operations (e.g., more attention was needed). Table 4 outlines this category, providing excerpts from the data related to each feature.

Table 4

Cognitive Operational Demand

Codes	Frequency	Description	Excerpts from the Data
1	43	Reasoning	They have to debate, state, and reason why they have to stay on the balloon.
2	31	Critical Thinking	The critical thinking requirements contribute to the difficulty of task X.

3	21	Idea generation/ creativity	It requires students to create something out of the picture story – use their imagination.
3	18	Organization of thoughts	They have to organize the story on their own without hints.
5	17	Reliance on memory/attention	Task X needs a lot of attention.
6	11	Compare/contrast	They have to compare two pictures.
7	10	Problem solving	Students need problem solving skills.
8	9	Decision making	Learners need to decide who to keep on the balloon.

Learner Background. Concerns about learner background, was the next most frequently mentioned feature associated with task difficulty (Table 5). In this category, the participants were mainly concerned about the individual characteristics which learners bring to task performance and how such individual learner variables affect task difficulty. Participants mostly noted issues related to *background and experience of the learner*, for example, a learner with no previous job experience will find the 'job interview' task difficult. The second group of features focused on students' *level of language proficiency*. Here we distinguish 'level of proficiency' from *linguistic demand* as the former is perceived by the participants as an individual learner variable, whereas the latter is considered a task-internal feature. There were several references to *age* as a factor affecting task difficulty, for example, school-age teenagers may find the job interview difficult. Only one participant noted personality issues such as self-confidence as a feature affecting task difficulty. As can be seen in the next section, we have not included *Learner Background* features in the criteria for task difficulty as they represent features that are external to task and its design. Table 5, below, shows the four sub-categories of *Learner Background* features with examples from the participants' analysis of task difficulty.

Table 5

Learner Background

Codes	Frequency	Description	Excerpts from the Data
1	55	Background and/or experience	For learners who have no job experience, the content and situation of job interview is new.
2	40	Level of language proficiency	Learners' limited language proficiency makes it difficult.
3	9	Age	Learners' age contributes to task difficulty.
4	1	Confidence	Some students will be shy.

Design Features. This category (Table 6) relates to general features of task design and task implementation that affect task difficulty. In this category, participants discussed task features teachers commonly consider when designing tasks to facilitate performance and optimize learning opportunities. The features included *familiarity with task type* (whether the participants are familiar with the task type), *support provided by the teacher* (whether the task requires support from the teacher), *clarity and sufficiency of task instruction*s (whether task instructions are adequate and clear), *task appropriateness and interest* (whether tasks are interesting and appropriate for the particular group of learners), and *taskness* (whether tasks are authentic and relevant to learners' lives outside class). Table 6 below provides a list of the sub-categories for *Design features* along with the frequency of each sub-category and excerpts from participants' analysis.

Table 6

Design Features

Codes	Frequency	Description	Excerpts from the Data
1	35	Familiarity with task type	Students will need to know about the professions to have an interview.
2	22	Teacher support needed	All the elements by themselves without any help.
3	13	Clarity and sufficiency of task instructions	Description is limited to a picture.
4	12	Task appropriateness	Whether they love to share their own life.

5	10	Taskness	Task is not related to students' real life.
6	9	Task interest	Learners may not be interested in the topic.

Informational Demand. This category (Table 7) emerged from comments participants made about the informational demands associated with task performance and task completion. In this category, the participants were mainly concerned about the extent to which the amount, quality and structure of information provided in a task and what learners are expected to do with it affects task difficulty. It included five sub-categories *of amount of information/content/language* (whether the task contained adequate amount of information or content), *close-endedness versus open-endedness* (whether the task asked for one decision/outcome to emerge or whether different outcomes were possible), *clarity of task content/information* (whether the information provided was clear), *organization of information/ task structure* (whether the information was presented in a structured manner), and *familiarity with content/information* presented in a task. *Informational demand* is different from the previous category, *Design features*, as the former specifically focuses on the content and information provided in a task, whereas the latter highlights the design and implementation features that can change before and during task implementation for pedagogic reasons. Table 7 below, provides more information on this category.

Table 7

Task informational demand

Codes	**Frequency**	**Description**	**Excerpts from the Data**
1	36	Amount of information/content provided in the task	No specific content is provided to complete the task.
2	30	Close-endedness versus open-endedness	The topic is "free" and people would provide different answers.
3	5	Clarity and sufficiency of task content	Task X is difficult to follow.
4	4	Organization of information/ task structure	It is hard for elementary learners to organize the content.

5	4	Familiarity with task content/information	It requires learners to be familiar with the content.

Communicative Demand. The final category (Table 8) highlights the participants' concerns about challenges associated with the conditions under which a task is performed (e.g., time pressure and anxiety) or the communicative requirements of task performance (e.g., collaboration and interaction requirement). In this category, possibility of collaborating with someone to complete tasks was considered a facilitating factor and time pressure a hindering element. The data contained several references to issues related to difficulty being associated with tasks that should be completed individually or under time restriction. Having to deal with anxiety from such communicative demands was also considered a source of difficulty. Regarding task anxiety, participants considered the communicative context, not the learner, as the source of anxiety, which makes this category different from individual learner variables indicated in *learner background* category above.

Table 8

Task Communicative Demand

Codes	**Frequency**	**Description**	**Excerpts from the Data**
1	25	Collaboration requirement	X is an interactive task which requires collaboration of two people.
2	25	Time pressure/ no time to prepare	X provides no preparation time and students cannot prepare in advance.
3	12	Task anxiety	Having the nerves to speak in front of a big class.

Discussion

Issues related to task difficulty have remained at the heart of debates in TBLT for decades, with several researchers asking teachers and course designers to rely on 'their experience and

intuitions" (Ellis et al., 2019, p. 14) when evaluating task difficulty. While this might be, at least partially, feasible for more experienced teachers, such as those in Erlam and Tolosa (2022), many less experienced teachers would find evaluating task difficulty overwhelming and challenging. The task difficulty criteria/features developed in the current study are aimed at helping teachers, both novice and experienced, to have a practical and structured approach to understanding different features that potentially affect task difficulty.

The results indicated the variety and depth of teachers' consideration of variables affecting task difficulty. The findings demonstrated that participating teachers considered a broad range of features affecting task difficulty that can be categorized into six groups (Table 2 above). The most frequently mentioned category was Linguistic Demand (28.79%), implying the participants were predominantly concerned about linguistic demands a task imposes on learners. Tasks in which students are required to produce long stretches of language or use unknown linguistic items and structures (e.g., describe your perfect holiday) were deemed particularly difficult. Interestingly, the participants referred to a range of both macro-linguistic (e.g., output demands and type of genre) and micro-linguistic features (e.g., vocabulary and grammar), highlighting their awareness of the significant role linguistic issues play in task performance. This finding is in line with previous research (Révész & Gurzynski-Weiss, 2016; Tavakoli, 2009a) as teachers have shown concerns about the impact of linguistic demands on task difficulty. The range of linguistic issues mentioned by the participants in this study, however, is unique. We attribute this diversity to the large data set, the range of tasks, and participants' linguistic training and background (MA TESOL). The participants' concern about linguistic demands can also be explained in the light of their linguistic background (e.g., the language distance between Chinese and English). Gilabert and Castellví (2019) have argued that linguistic complexity (e.g., morphological complexity) of the target language should be considered as a factor affecting task difficulty when designing and sequencing tasks.

The second most frequent category of features (22.47%), Cognitive Operational Demand, highlighted the participants' attention to the challenges cognitive processes impose on learners and included a broad range of cognitive processes from reasoning to critical thinking and creativity. The attention participants paid to cognitive demands was not surprising as it is in line with previous research. Révész and Gurzynski-Weiss (2016) reported conceptual demands as the

second most frequent factor considered by teachers when evaluating task difficulty. However, Révész and Gurzynski-Weiss's (2016) conceptual demand category was broader and included cognitive processing factors (e.g., reasoning and number of elements), learner-related factors (e.g., learner background knowledge) and task familiarity.

The third most frequent category (14.61%) was *Learner Background*. In this category, the participants suggested learner individual variables such as professional background, amount of experience, proficiency, and age play a role in determining task difficulty. Such factors predominantly represent individual learner variables that are external to task design and therefore deemed as not manipulable in task design and development. Such factors, although crucial in task selection for pedagogic purposes, are not commonly included in a task design framework as accommodating an infinite number of differences between individual learners or specific groups of learners may not be feasible in a task difficulty framework or model intended for everyday teacher use. The particular attention our participants paid to *Learner Background* features could have been triggered by the openness-to-interpretation characteristic of our sample tasks that encouraged them to interpret the task in a flexible manner and ask these questions (e.g., *how experienced is the learner*). The data analysis suggested that in examples categorized under *Learner background*, the teachers were in principle asking for more information about the kind of learners that were about to perform the tasks, implying the same task may not be of the same difficulty level to learners of varying characteristics (e.g., age, proficiency, confidence). Such learner-related features have also been proposed in previous research (Révész & Gurzynski-Weiss, 2016; Tavakoli, 2009a), although they are not considered as an inherent characteristic of task difficulty.

The next category reported in the dataset (14.33%) was *Design Features*. This category includes features related to task design (e.g., clarity of instructions) that are independent of task cognitive or informational demands. Features such as clarity of instructions, and task interest have been previously reported to affect task difficulty (Révész & Gurzynski-Weiss, 2016; Tavakoli, 2009a, 2009b). A new feature labeled as 'taskness' also emerged in the analysis, implying the participants questioned whether a task meets the criteria to qualify as a task (e.g., relates to the students' life outside classroom) and consider it as an aspect of task difficulty. It is possible to interpret teachers' comments about 'taskness" and its relationship to difficulty in terms of

motivation to engage with the task and complete it. Their attention to 'taskness' may, in effect, suggest students are more motivated to perform tasks that reflect real life and focus on meaning than tasks which do not meet task criteria. It is also possible to interpret this finding in relation to the teachers training on their MA course.

Informational Demand (11.10%) represented features related to the quality, quantity and structure of information provided in the task. In this category, the participants focused on how information is presented (e.g., amount, clarity, and structure of information) and what manipulation it involves (e.g., open-ended or close-ended). This category was distinct from others as it predominantly questioned the impact the quality/quantity of information and what was expected to do with it would have on task difficulty and learner performance. The category of open-ended versus close-ended in the current study is similar to Ellis et al.'s (2021) *open versus closed tasks* in which a number of outcomes are possible in the former, while a single outcome is expected in the latter task type. The difference between these classifications is that Ellis et al.'s category primarily focuses on task outcomes.

The least frequently mentioned category (8.71%) was *Communicative Demand*, highlighting teachers' concerns about the interactional requirements and time-related factors affecting task difficulty. The participants generally considered collaborative tasks and tasks without a time pressure as less demanding. Interactional requirements have also been reported by Révész and Gurzynski-Weiss (2016) as a source of difficulty, but we are surprised to see that in our data this category received the lowest frequency, implying the participants did not pay as frequent attention to it as they did in other categories.

Figure 1 below provides a visual representation of the different aspects of task difficulty and some examples of the sub-categories proposed by our participants. It is worth noting that Learner Background is not included here as the features proposed seemed to be external to task design and task difficulty.

Figure 1

Teacher criteria for analyzing task difficulty

Features teachers consider		Examples
Linguistic demand	→	Vocab, grammar, discourse demands
Cognitive operational feature	→	Amount of reasoning, critical thinking, problem solving needed
Design features	→	Task instructions, interest, familiarity
Informational demand	→	Amount and structure of information provided
Communicative demand	→	Number of interlocutors, time pressure, task anxiety

Baralt and colleagues (2014) asserted that there is not a widely agreed-upon set of criteria to be used for grading and sequencing tasks. This is problematic when considering the amount of research interest and effort put into the existing frameworks of task difficulty, Skehan's LAC and Robinson's TCF discussed earlier. In what follows we compare the features proposed by our teachers with these models to determine the extent to which they support teachers' perceptions of task difficulty.

As discussed earlier, Skehan's (2018) LAC model includes three categories of *code complexity*, *cognitive complexity*, and *communicative stress*. Code complexity seems to be directly related with *linguistic demand* category derived from the data analysis in the current study. Skehan's model emphasizes grammatical difficulty, vocabulary load, and text density as key features affecting task difficulty, while our data offer a broader range of linguistic factors including macro and micro-linguistic elements. The fact that teachers put substantial emphasis on linguistic demands suggests teachers are predominantly concerned about the impact of such demands on task difficulty. C*ommunicative demand* category, in our study, also corresponds to Skehan's communicative stress, that is, the pressure prompted by the conditions under which a task is performed. In our results, factors such as time pressure, collaborative requirements, and task anxiety represent the communicative demands. In Skehan's model, however, communicative stress also includes features such as task modality and task degree of control, features that we have included in task design features. Our data suggest Design Features is a distinct category of

features primarily concerned with elements that can change without a major impact on task cognitive operations or informational demand.

In Skehan's LAC model, *cognitive complexity* includes cognitive processing (e.g., whether the information is well organized) and cognitive familiarity (e.g., whether the task, its information and content is familiar). Similar to LAC, our results include two categories of features related to demands associated with cognitive complexity: *Cognitive operational* demands, and *Informational* demands. While these two categories seem related to Skehan's cognitive complexity, they are different in the way they have been defined. Our *Cognitive operational demands* underline the impact of cognitive processes (e.g., reasoning, critical thinking, problem solving) that are perceived as difficult and therefore predicted to affect task difficulty. In contrast, Skehan's model, although interested in cognitive complexity, does not emphasize the role of such cognitive processes, and instead focuses on the way information is organized. Our *Informational* demand category, however, seems closely related to Skehan's cognitive processing as both focus on information presentation and organization and the kind of manipulation required to complete a task. As discussed earlier, we have attributed the participating teachers' attention to *Learner Background* features to the fact that our tasks did not include specific learner information, and as such the teachers used the opportunity to seek information about the learners, arguing task difficulty can only be fully evaluated if the learners' characteristics are known. While we have considered this category as external to task design and difficulty, future research will need to carefully examine teachers' rationale in how learner-related features are considered to affect task difficulty. Future research will further need to examine to what extent learner-related factors can effectively be included in task and syllabus design.

Skehan (2018) argues that linking task difficulty to Levelt's (1989) model in terms of factors affecting task difficult in terms of Conceptualization (e.g., the demands related to accessing and manipulating the information) in contrast to those affecting task difficult in terms of Formulation (e.g., having a large repertoire of linguistic items and automatic access to this) would facilitate the development of a task difficulty model. Given our findings, it is plausible to claim that while linguistic demand is related to processes involved in the Formulator, the other categories (e.g., Cognitive operational and Task informational demands) are linked to the work of the

Conceptualizer, although perhaps to a varying degree, as they affect the way the pre-verbal message is developed and shaped. For example, it is easy to understand that the degree of cognitive operational demand, informational demand, and details of design features affect the way the pre-verbal message is formulated. Similarly, it can be claimed that communicative demand (e.g., considering the interlocutor's needs) can affect the operations of the Conceptualizer in different ways (e.g., having more time to develop the pre-verbal message in a dialogue).

When comparing our findings with Robinson's TCF, a few similarities and differences are worth discussing. TCF includes three categories of *task complexity*, *task conditions* and *task difficulty*. *Task condition*, in TCF, refers to interactional needs associated with task performance, which is similar to *Communicative demand* in the current study highlighting interactional needs of the task as well as other sources of communicative pressure (e.g., time pressure). In terms of Task complexity category, Robinson's TCF explains two sets of factors: *Resource-directing* (conceptual demands) and *resource-dispersing* (procedural demands) factors. It seems that the *cognitive operational* demand relates to TCF's resource-directing factors in that both discuss the cognitive processes activated during task performance and the impact they have on task difficulty. *Task informational* demand (e.g., structure of information), also seems related to Robinson's Resource-dispersing features that underline the importance of procedural features during task performance.

In terms of the differences between our findings and Robinson's model, three points are worth discussing. One of the differences is that while Robinson considers learner factors as a source of task difficulty (his *task difficulty* category), we have considered such factors (*Learner Background* in our data) as a task external category, arguing such features, although significant, cannot be accounted for in a general task difficulty framework. Individual learner differences (e.g., age, proficiency, confidence, anxiety) are only known for specific individuals or groups of learners, while a task difficulty framework should provide a solid foundation and an overarching scheme for task difficulty analysis adaptable for use with specific groups of learners in diverse contexts. Second, and perhaps most important, is the participants' predominant concern about linguistic demands, the largest category of task difficulty features in our study, which seems to be missing in TCF. Given the fact that all previous studies in this area (Révész & Gurzynski-

Weiss, 2016; Tavakoli, 2009a, 2009b) as well as Skehan's model highlight the importance of linguistic demands in task difficulty, it is surprising to see this category is still missing in TCF. The final difference is TCF's lack of attention to issues related to *task design* category (e.g., clarity of task instructions). This category focuses on task features that can be manipulated to control for task difficulty without compromising the cognitive and informational demands of the task. Although TCF considers a range of cognitive features associated with task design, it does not consider the design features independent of cognitive factors.

Overall, there are several similarities between our findings and Skehan's LAC model and to a lesser extent to Robinson's TCF model. The set of criteria presented in this paper provides a more nuanced understanding of the broad categories in Skehan's framework and offers a new category of *Design Features*, highlighting important task features that should be considered when evaluating task difficulty. The findings of the current study, however, needs validation from future studies including replication studies and studies with other populations.

Pedagogical Implications

The impetus for this study was to understand criteria *teachers* consider when analyzing and evaluating task difficulty. The issue of how teachers select appropriate tasks and how they grade and sequence tasks to enhance language acquisition is central to the successful implementation of TBLT. The findings of this study have significant implications for preservice and in-service teachers as well as teacher education programs and other professional development programs.

Several TBLT scholars have highlighted that advance planning and sequencing are crucial to the success of task-based pedagogy (Chan, 2012; Erlam, 2016). The features discussed in our framework can be used during such advance planning to identify aspects of task that may need modification or learner preparation. The breakdown of linguistic demand in our data set, for example, can help draw teachers' attention to a range of features to be considered and planned for in advance. Teachers can evaluate a task for their classroom considering the required grammar and vocabulary (micro elements) versus the discourse/genre and the output requirements (macro elements) of the task. It is necessary to point out that linguistic demands vary not only in different tasks but in different languages, given their morphological and

grammatical structures. To illustrate, it is possible to speculate that linguistic demands play a more crucial role in determining task difficulty in a morphologically complex language such as Persian. Linguistic demands also contribute to cognitive operational demands, another aspect of task difficulty, to be carefully considered by the teachers and modified, if needed, to ensure the task demands are reasonable to the specific groups of learners.

These findings also have implications for language teacher education programs. Language teacher education should develop teachers' understanding of how languages are learned and taught and raise teachers' awareness of how their practices impact students' language learning (Scarino, 2014). TBLT researchers have argued that keeping teachers "involved in a sustained effort over a longer period of time" (Van den Brandon & Van Gorp, 2021, p. 21) is of prime importance in teacher education programs if a continued approach to professional development is intended. This framework can be introduced in teacher education programs and professional development workshops as a way of engaging teachers in the process of task evaluation, selection and design. The two widely cited models of task difficulty, Skehan's LAC model and Robinson's TCF, are conceptual models and do not provide a pedagogically practical model to help teachers determine task difficulty and suitability for classroom use. The broad categories of task features identified in the current study through teacher evaluations might be easier for teachers to remember and engage with compared to the existing conceptual models. The substantial teacher cognition literature (e.g., Borg, 2006) indicates how teachers have unique ways of accepting new professional knowledge and integrating it into their practice. With regards to implementing task-based pedagogy, Erlam and Tolosa (2022) emphasize that in presenting such new professional knowledge, it is important that teachers recognize what is similar or the same and what is different. According to Van den Branden (2009) teachers gradually develop the required skills in using tasks efficiently in their classrooms. Moving beyond task difficulty, a related concern in TBLT approach, is how to sequence tasks for classroom use (Baralt, Gilabert and Robinson, 2014). We argue that the framework of task difficulty developed in this study can pave the way for a systemic approach to task grading and sequencing for teachers and designing curriculum.

Conclusions

The present study was aimed at investigating task difficulty from a novice teachers' perspective and developing a set of task difficulty features that can be used by teachers and practitioners when evaluating and analyzing task difficulty. Given the growing prominence of TBLT, examining task difficulty is a particularly important aspect of L2 teachers' concerns when lesson planning, selecting, and preparing materials and/or designing curriculum. The study was motivated by calls for research on teachers' views on task difficulty to help develop a framework to be used for selecting, grading, and sequencing tasks (Ellis, et al., 2019). The study has three important findings. First, the results suggest that teachers, even those with limited teaching experience, have analytic skills in evaluating task difficulty that resemble criteria frequently mentioned in conceptual models of task difficulty in the literature. The results of the study suggest the participating teachers drew on their linguistic and TESOL training to provide a careful analysis of the various sources of task difficulty. Second, the features and categories emerging from teacher analysis correspond, to a great extent, to research-oriented models of task difficulty in this area. Our data analysis suggests that a large majority of the factors identified as sources of task difficulty have been highlighted in Skehan's (1998) LAC model and to a lesser extent in Robinson's (2001, 2011) TCF models; therefore, it is reassuring to find out what research-oriented models offer in this regard are, to a great extent, in line with teachers' analysis of task difficulty. Teacher analysis of task difficulty, however, went beyond these models to highlight components that seemed important to teachers but were missing in the existing models, thereby offering a single more comprehensive model. Our findings suggest the lack of attention to such features, particularly in the case of linguistic demands in TCF, will make the model impractical and difficult for teachers to work with. Third, the results have offered a new perspective to teachers and researchers to adopt a similar structured and analyzable approach to unwinding and scrutinizing task difficulty, enabling them to dissect different aspects of task to evaluate its difficulty. We acknowledge that our study should be considered a first step in developing a work-in-progress framework that could offer teachers and researchers with an opportunity to analyse and evaluate tasks for pedagogic purposes in their contexts. The next step in this process is to replicate the study with different groups of teachers (e.g., in-service, more experienced, diverse linguistic backgrounds), different tasks (e.g., writing) and different types of data (e.g., observations).

Our study was limited for a number of reasons. First, the data were collected from only teachers with limited teaching experience and predominantly from the same language and educational background. Lack of diversity, in experience and background, among the participants might have affected the findings. Data from more experienced and a diverse group of teachers may help shed more light on the complex construct of task difficulty. Also, more information about their teaching orientation could have helped us evaluate their task difficulty analysis more critically. Second, our data relied on teachers' analysis of only six speaking tasks in a 'skeletons' format; asking teachers to analyze a broader range of tasks and to work with the actual tasks could have provided a richer opportunity for understanding teachers' criteria for task difficulty in more depth and breadth. For this reason, we suggest the findings are interpreted carefully. The data would have been richer if the written analysis was accompanied by follow up interviews, observations, and performance-based data. Finally, future research should consider developing an index of task difficulty to be used for task grading and sequencing.

References

Baralt, M., Gilabert, R., & Robinson, P. (2014). *Task Sequencing and Instructed Second Language Learning*. Bloomsbury Academic. http://dx.doi.org/10.5040/9781472593665

Baralt, Harmath-de Lemmos & Werfelli, (2014). Teachers' application of Cognition Hypothesis when lesson planning: A case study. In Baralt, Gilabert & Robinson (Eds.) *Task Sequencing and Instructed Second Language Learning.* Bloomsbury Academic. http://dx.doi.org/10.5040/9781472593665.

Candlin, C. N. (1987). Towards task-based language learning. In C. N. Candlin & D. F. Murphy (Eds.), *Language learning tasks* (pp. 5–22). Prentice-Hall.

Chan, H.Y. (2012). Values and the advance care planning framework in New Zealand. *BMJ Supportive & Palliative Care, 2*(2), 203–203. https://doi.org/10.1136/bmjspcare-2012-000250.120

Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics, 31*, 182-204. https://doi.org/10.1017/S0267190511000092

East, M. (2021). *Foundational principles of task-based language teaching*. Routledge. https://doi.org/10.4324/9781003039709

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

Ellis, R. (2019). Towards a modular language curriculum for using tasks. *Language Teaching Research,* 23(4), 454-475. https://doi.org/10.1177/1362168818765315

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2019). *Task-based language teaching: Theory and practice*. Cambridge University Press.

Erlam, R. (2016). "I'm still not sure what a task is": Teachers designing language tasks. *Language Teaching Research.* 20(3), 279-299. https://doi.org/10.1177/1362168814566087

Erlam, R., & Tolosa, C. (2022). *Pedagogical realities of implementing task-based language teaching*. John Benjamins.

Faez, & Tavakoli, P. (2019). *Task-based language teaching*. TESOL Press.

Foster, P., & Skehan, P. (1996). The Influence of Planning and Task Type on Second Language Performance. *Studies in Second Language Acquisition*, *18*(3), 299–323. https://doi.org/10.1017/S0272263100015047

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching, IRAL*, *45*(3), 215–240. https://doi.org/10.1515/iral.2007.010

Gilabert, R., & Castellví, J. (2019). Task and syllabus design for morphologically complex languages. In J. Schwieter & A. Benati (Eds.), *The Cambridge Handbook of Language Learning* (pp. 527-549). Cambridge University Press.

Gilabert, R., & Malicka, A. (2021) From needs analysis to task selection, design, and sequencing. In: M.J. Ahmadian & M.H. Long (eds.). *The Cambridge Handbook of Task-Based Language Teaching* (pp. 226-249). Cambridge University Press.

Lambert, C., & Oliver, R. (2020). Using tasks in language teaching: Practice in diverse contexts. Multilingual Matters.

Levelt, W. (1989). *Speaking: From Intention to Articulation.* MIT Press.

Little, D. (2006). The Common European Framework of Reference for Languages: Content, purpose, origin, reception and impact. *Language Teaching*, *39*(3), 167–190. https://doi.org/10.1017/S0261444806003557

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley Blackwell.

Michel, M. C. (2011). Effects of task complexity and interaction on L2 performance. In P. Robinson (Ed.) Second language task complexity: Researching the Cognition Hypothesis of language learning and performance (pp. 141-173). John Benjamins.

Maijala, M., Heikkola, L. M., Kuusalu, S.-R., Laine, P., Mutta, M., & Mäntylä, K. (2023). Pre-service language teachers' perceptions of sustainability and its implementation in language teaching. *Language Teaching Research*. https://doi.org/10.1177/13621688231170682

Prabhu, N.S. (1987). Second language pedagogy. Oxford University Press.

Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second language acquisition*, *31*(3), 437-470. https://doi.org/10.1017/S0272263109090366

Révész, A. & Gurzynski-Weiss, L. (2016). Teachers' Perspectives on Second Language Task Difficulty: Insights from Think-Alouds and Eye Tracking. *Annual Review of Applied Linguistics*, *36*, 182–204. https://doi.org/10.1017/S0267190515000124

Robinson, P. (2001). Task complexity, task difficulty, and task production: exploring interactions in a componential framework. *Applied Linguistics*, *22*(1), 27–57. https://doi.org/10.1093/applin/22.1.27

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching, IRAL*, *45*(3), 193–213. https://doi.org/10.1515/iral.2007.009

Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Pütz & L. Sicola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 243–268). John Benjamins

Robinson, P. (2011). *Second language task complexity : Researching the cognition hypothesis of language learning and performance*. John Benjamins.

Robinson, P. (2015). The Cognition Hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. *Domains and directions in the development of TBLT*, *8*, 87-121.

Scarino, A. (2014). Learning as reciprocal, interpretive meaning-making: A view from collaborative research into the professional learning of teachers of languages. *The Modern Language Journal*, 98(1), 386–401. https://doi.org/10.1111/j.1540-4781.2014.12068.x

Shintani, N. (2016). *Input-based tasks in foreign language instruction for young learners*. John Benjamins. https://doi.org/10.1075/tblt.9

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

Skehan, P. (2003). Task-based instruction. *Language Teaching*, *36*(1), 1–14. https://doi.org/10.1017/S026144480200188X

Skehan, P. (2018). *Second language task-based performance: theory, research, assessment* (First edition.). Routledge, an imprint of Taylor and Francis.

Tavakoli, P. (2009a). Investigating task difficulty: learners' and teachers' perceptions. *International Journal of Applied Linguistics, 19(1),* 1-25. https://doi.org/10.1111/j.1473-4192.2009.00216.x

Tavakoli, P. (2009b). Assessing L2 task performance: Understanding effects of task design. *System (Linköping)*, *37*(3), 482–495. https://doi.org/10.1016/j.system.2009.02.013

Tavakoli, P. & Jones, R. (2018) *An overview of approaches to second language acquisition and instructional practices.* Report. Welsh Government (Research & Statistics), Cardiff, Wales. ISBN 9781789034882

Van den Branden, K. (2006). Training teachers: Task-based as well? In K. Van den Branden (Ed.), *Task-based language education: From theory to practice* (pp. 217–248). Cambridge University Press. https://doi.org/10.1017/CBO9780511667282.011

Van den Branden, K. (2009). Mediating between predetermined order and chaos: The role of the teacher in task-based language education. *International Journal of Applied Linguistics*, 19(3), 264–285. https://doi.org/10.1111/j.1473-4192.2009.00241.x

Van den Branden, K., & Van Gorp, K. (2021). Implementing task-based language education in primary education: Lessons learnt from the Flemish experience. *Language Teaching for Young learners 3*(1), 3-27. https://doi.org/10.1075/ltyl.20013.bra

Appendix 1:

Question 1: You are teaching a group of elementary teenage learners learning English as a foreign language in Frankfurt, Germany. Consider the following three tasks and rank them according to their difficulty from 1-3. **Please note there is no right or wrong answer.**

The three tasks are:

a) An oral narrative based on a picture story (look at the picture story you have been given and narrate the story that is happening).
b) A spot-the-difference task (work with a partner to identify and list the differences in the two pictures).
c) Describe your perfect holiday (orally).

Which task is the most difficult for your learners?

Why do you think this task is the most difficult? What factors do you think have contributed to the difficulty of this task?

Question 2: You are teaching a group of intermediate learners in their early 20s performing the tasks as part of their university exam in a small town in China. Consider the following three tasks and rank them according to their difficulty from 1-3. **Please note there is no right or wrong answer.**

The three tasks:

a) Discuss (in pairs) the advantages and disadvantages of travelling alone and travelling in a group.

b) Balloon task: A doctor, a mechanic, a teacher, a lawyer, and a sailor are riding in a balloon when it falls into trouble. To save the balloon from crashing down, you can only keep three people in the balloon. Work with a partner to choose which three people to keep on the balloon?
c) Take part in a role-play job interview with your teacher. Answer her questions and explain why you are the best candidate to take part in the job.

Which task is the most difficult for your learners?

Why do you think this task is the most difficult? What factors do you think have contributed to the difficulty of this task?