



## L2 Learners' and Professionals' Perceptions of Cognitive Task Complexity: Towards Developing an Index of Task Difficulty

Anas Awwad

*Isra University, Jordan*

This paper attempts to broaden our understanding on how cognitive task complexity (TC) contributes to L2 learners' and language professionals' perceptions of task difficulty (TD). This study was conducted to allow a deeper investigation of the extent to which manipulating TC can influence perceptions of TD, what TC-factors contribute to perceptions of TD, and whether there is a perceptual match between L2 learners and language professionals. The study employed a mixed-method design in which 68 students and 26 language professionals evaluated the TD of two video-based narrative tasks. The students, who were L2 learners of English at a Jordanian secondary school, performed two tasks with increased TC manipulated by intentional reasoning (IR) demands. Data on learners' perceptions of TD were collected qualitatively and quantitatively using a retrospective questionnaire. The language professionals, who were researchers with a TESOL background, read the tasks instructions and watched the two clips before deciding which task would be considered more difficult and why. Results showed a match between learners and language professionals on their perceptions of TD. Both groups rated the more complex task (+IR) as more difficult. Two categories of higher cognitive demands imposed by IR requirement, i.e. task-induced and task-inherent, were found to be the major contributors to the perceptions of TD. The findings have significant implications for the development of an index of TD, which is considered a priority for language teaching, testing and syllabus design.

**Keywords:** task complexity, task difficulty, second language, speech performance, task-based language teaching

### Introduction

Despite the substantial efforts in investigating cognitive task complexity (TC) within Task-Based Language Teaching (TBLT), less attention has been devoted to examining how TC contributes to key stakeholders' perceptions of task difficulty (TD). While TC can be defined as "the cognitive load of a second language (L2) communication task" (Sasayama, 2016, p. 231) imposed on learners, TD is seen as the extent to which a task is perceived difficult based on learners' individual factors (Robinson, 2015). Research on perceptions of TD (e.g., Ishikawa, 2011; Révész & Gurzynski-Weiss, 2016; Robinson, 2001; Robinson, 2007; Sasayama, 2016; Tavakoli, 2009a; Tavakoli, 2009b; Tavakoli & Skehan, 2005) has focused on exploring the impact of different task characteristics and conditions on learners' and/or teachers' perceptions of TD. This line of research has been motivated by two TC models: Skehan's (1998, 2015) Limited Attentional Capacity Model and Robinson's (2007, 2015) Cognition Hypothesis. The two models offer different predictions and interpretations regarding the impact of TC on language performance and learners' perceptions of TD by exploring the interaction between the cognitive requirements of tasks and learners' attentional and memory resources, as well as learners' individual

differences. While this type of research has confirmed that perception of TD is evidently associated with increased TC, exploring more systematically the factors that contribute to this perception is still an under-researched area requiring more collaborative efforts. Hence, researchers are encouraged to investigate what TC-oriented factors affect or contribute to TD perceptions of key stakeholders' (e.g., L2 learners, teachers, researchers and syllabus designers) through mixed-methods investigations. This type of TD research collects retrospective data from L2 learners or teachers quantitatively and qualitatively, to identify whether specific tasks are easy or difficult to perform and why. The major research findings emerging from this type of data can offer useful insight not only into second language teaching, learning, and testing, but also for syllabus design.

A more in-depth examination of TD is assumed to widen the current understanding of TC-TD interaction, and subsequently assist "language educators in designing and employing more effective language teaching materials" (Tavakoli, 2009a, p. 2). The findings of this kind of research are therefore expected to have implications for language teaching, testing and syllabus design by contributing to the efforts to establish a much needed index of TD (Breen, 1987; Candlin, 1987; Nunan, 1989; Skehan, 1998). The absence of systematic operationalisations of TC factors has led to inconsistent findings which, consequently, have constrained the opportunity for agreement on an index of TD. Furthermore, treating the inspection of TD only as a secondary aim in previous research has not furthered this aim. The current study will therefore attempt to address this issue by offering a more systematic approach to operationalise TC and by focusing primarily on exploring the factors that contribute to perceptions of TD using a mixed-method study design. Such a systematic and multi-dimensional investigation of TD can assist stakeholders in recognising the unpredictable nature of task or test performances (Kumaravadivelu, 1993). Examining the factors contributing to TD can also help to minimise the perceptual mismatch between task stakeholders, and this can lead to a more effective pedagogic intervention (Kumaravadivelu, 2006). Moreover, data gathered regarding perceptions of TD can aid researchers to establish the validity of TC manipulations in their studies (Révész, Michel, & Gilabert, 2016). Consequently, this study is motivated by the urgent need to shed more light on second language stakeholders' standpoints on TD, to examine whether the current models of TC-TD mirror their perceptions, and to provide more findings that can support the establishment of an index of TD.

## Literature Review

### Task Complexity and Task Difficulty

Central to the TBLT research agenda are the issues of TC-TD distinction and interaction. Within this research area, the two constructs are often paired or used interchangeably, whereas in fact each refers to a different construct (Bulté & Housen, 2012). Cognitive psychologists define the two constructs from a resource-oriented standpoint, and posit that the degree of TD of a task is associated with the amount of cognitive load that is imposed on task performers by the task design or characteristics (Liu & Li, 2012). Increasing information load and task demands are assumed to tax task performers' cognitive, attentional and memory resources, thus escalate the degree of TC and influence perceptions of TD (Campbell, 1988). This suggests a cause-effect relationship between the two constructs with TC as the cause and TD as the effect. From an interactionist stance, TC and TD are distinguished due to the intercommunication between the factors contributing to each construct. Hence, TC and TD can be perceived as "a product of the interaction between task and task performer characteristics" (Liu & Lu, 2012, p. 555).

According to Sasayama (2016), TD is largely associated with "the actual effort engaged in by the individual as opposed to the supposed demand imposed by a task" (p. 232). This suggests a possible interaction between the learner's mental effort needed during their performance and the manipulation of task design. Thus, TC can be denoted as the objective complexity due to inherent complex features of a task, whereas TD is considered as the subjective complexity due to how task stakeholders weigh a task in

terms of difficulty based on the individual characteristics of task performers. Robinson (2001) distinguishes between TC and TD by defining the former as any attentional or information processing demands that are enforced on task-performers by different degrees of inherent task requirements. TD is however, associated with task performers' perceptions of the level of difficulty of performing a language task based on any individual variations in their ability skills and/or affective domains (Robinson, 2011). This views TC as task-dependent, demonstrating within-learner variation in language performance, whereas TD is learner-dependent, and exhibits between-learner variation in successfully completing the same task (Robinson, 2015). Making a distinction between TC and TD can have pedagogical implications by exploring what learners can bring to task performance, thus enabling prediction of the language they can produce. This can enhance the effectiveness of teaching interventions to support learners with low abilities and also "can be used to match learners to sequences of simple to complex tasks" (Robinson, 2015, p. 107).

According to Skehan (2014), TD is, however, seen "as inherent in the task rather than learner dependent" (p. 6). Henceforth, Skehan (2015) argues against making any clear-cut distinction between TC and TD. Instead, he calls to ground any effect of TC on perception of TD to the effects of task demands on each stage of speech production, i.e. Conceptualisation, Formulation and Articulation. Within this psycholinguistic perspective, Skehan (1998, p. 99) suggests the following scheme that can be employed to classify TC and identify TD: a) code complexity (language required); b) cognitive complexity (thinking required), which includes cognitive familiarity and cognitive processing; and c) communicative stress (task conditions). In line with this scheme, Skehan (2015) defines TC as the amount of attention a task requires from learners to achieve a successful outcome with the presumption that increasing TC consumes limited available attentional resources. This assumption is central to Skehan's (1998) Limited Attentional Capacity Hypothesis which further predicts that TC can result in a trade-off between different aspects of language performance, i.e. complexity and accuracy, due to learners' limited nature of attention. Linking TC to its impact on learners' limited attention is still challenged by Robinson's (2001, 2007) Cognition Hypothesis which suggests that learners have multiple pools of attention which can be extended and directed to attend to different aspects of language performance alongside tasks' increasing complexity. This debate has so far occupied the heart of TC-TD research and motivated its agenda.

The Cognition Hypothesis is a taxonomy of task demands which proposes a distinction between TC and TD. Robinson (2007) identifies three categories of task demands to imitate real-life task performance: 1) task complexity (cognitive factors); 2) task conditions (interactive factors); and 3) task difficulty (learner factors). These criteria classify, operationalise and sequence tasks based on their complexity. Hence, TC can be manipulated via a set of resource-directing variables (e.g., reasoning) versus a set of resource-dispersing variables (e.g., planning). Manipulating resource-directing variables generates different cognitive demands and directs learners' attention to the aspects of language required to meet the increased demands. This results in promoting form, i.e. accuracy and complexity, at the cost of fluency. Alternatively, increasing TC along resource-dispersing variables (e.g., no planning), creates performative and procedural demands, thus disperses learners' attention to non-linguistics areas of speech production (Robinson & Gilabert, 2007). TD is however, affected chiefly by learner ability and affective factors (e.g., working memory, aptitude, motivation) which cannot be manipulated in a way comparable to TC variables (Robinson, 2015). Instead, learner factors can be diagnosed prior to task performance via background questionnaires and/or interviews, then their interaction with TC can be identified after their performance through retrospective questionnaires and/or interviews. The Cognition Hypothesis further predicts that tasks which impose higher cognitive demands (e.g., +reasoning) are likely to be perceived as more difficult to perform than tasks requiring fewer cognitive demands (e.g., -reasoning) (Robinson, 2011). TC is operationalised in the current study through intentional reasoning (IR) demands to respond to: a) the scarcity of studies investigating IR as a TC factor; and b) the absence of systematic frameworks operationalising IR (Awwad, Tavakoli, & Wright, 2017). IR in this study involves "reading other

people's thoughts and understanding their desires, beliefs and motives to draw true conclusions about what they intend to do, why and what consequences follow" (Awwad et al., pp. 44).

## Perceptions of Task Difficulty

Following the inclusion of task as the basic unit in L2 syllabi, task-related issues have occupied a central position in language testing, teaching and learning. Such growing interest in the sufficient utilisation of task has created pressing needs to evaluate whether the designed tasks serve their functions, whether they are presented in a right sequence, whether they are at a suitable level of complexity and whether they are perceived equally in terms of difficulty by stakeholders. These issues are still open research questions as the findings of previous studies have not adequately satisfied these inquiries. Previous studies have either focused on exploring how sequences of complex tasks are perceived with respect to TD by learners (e.g., Ishikawa, 2011; Robinson, 2001, 2007; Sasayama 2016; Tavakoli & Skehan, 2005), or by teachers (e.g., Révész & Gurzynski-Weiss, 2016), or both learners and teachers (e.g., Tavakoli 2009a). However, only a few studies have primarily explored what TC-oriented factors contribute to learners' and teachers' perceptions of TD and to what extent a perceptual match exists between task stakeholders.

Robinson (2001) explored the perceptions of TD of 44 Japanese learners using a retrospective questionnaire that weighed degree of difficulty and other affective factors. The participants used a 0-9 Likert scale to rate how difficult they found performing two map tasks manipulated by the amount of information and prior knowledge. Robinson also investigated the interaction between TC sequences (simple-complex versus complex-simple) and learners' perceptions of TD. The results revealed that the participants perceived the more complex task as more difficult, thus confirming the Cognition Hypothesis predictions. However, task sequencing appeared to have no effect on perceived TD. Sasayama (2016) also used a nine-point Likert scale questionnaire to investigate how 53 Japanese learners would rate the TD of four narrative tasks with increased number of characters. The findings confirmed that Task 1, with the fewest number of characters, was perceived as the least difficult, whereas narrating Task 4, which included the highest number of characters, was perceived as the most difficult. The difference was only significant with a large effect size between the perceptions of the TD of Task 1 versus Task 4, whereas no significant differences were found between the other tasks. These findings emphasise the need to consider salient variations between tasks in terms of TC demands to guarantee discrete differences among these tasks regarding language performance and perceptions of TD. Tavakoli and Skehan (2005), who operationalised TC through different degrees of narrative structure, measured perceptions of TD of 80 EFL learners using a retrospective questionnaire. The participants constantly rated the unstructured tasks as more difficult than the structured ones. This confirmed the researchers' predictions that the absence of storylines in narrative tasks influenced how tasks could be weighed in terms of TD.

Manipulating TC through the amount of intentional reasoning (IR) demands of oral narrative tasks, Robinson (2007) asked his participants to rate the TD of performing the tasks using a retrospective questionnaire. The L2 learners of English (N = 42) rated the tasks that required more reasoning as more difficult and stressful to perform. The findings support those obtained from Robinson (2001) and confirm the predictions of the Cognition Hypothesis, that TC cognitive demands are the main contributors to learners' perceptions of TD. Ishikawa (2011), who also operationalised TC through IR at three levels, employed the same questionnaire as Robinson (2001) to tackle L2 learners' perceptions of TD. The participants (N = 24), who were of different levels of language proficiency (low to upper-intermediate), performed three monologic tasks which required them to report hypothetical changes in human relationships. Perceiving the tasks that required more IR as more difficult, the findings fully support those obtained by Robinson (2007), confirming that IR as a TC variable has a steady impact on learners' perceptions of TD. However, in tackling TD only as a secondary aim and not exploring the factors contributing to perceptions of TD, the studies reported here missed the opportunity to capture a more holistic depiction of this complex construct.

Tavakoli (2009a) was one of the earliest attempts to qualitatively investigate both L2 learners' and teachers' perceptions of TD. She employed retrospective interviews to provide a thorough exploration of the perceptions of TD of L2 learners (N = 10) and ESOL teachers (N = 10). Following Tavakoli and Skehan (2005), this study adopted four retelling picture tasks that varied in their degree of story structure. Data analysis revealed that both groups rated the least structured tasks as more difficult and identified the same factor affecting their perceptions, but they differed on how frequently those factors were mentioned. Both groups mentioned six factors that influenced their perceptions of TD, i.e., *cognitive demands*, *linguistic demands*, *clarity of prompts*, *amount of information* and *task structure*. This study can be criticised for the small sample size (N = 10) which could reduce the generalisability of the results. However, the findings of Tavakoli and Skehan (2005) contradict those obtained in previous studies which found perceptual mismatches between learners and teachers (e.g., Barkhuizen, 1998; Hawkey, 2006; Kumaravadivelu, 1991, 2006). It is worth mentioning that these studies did not focus on exploring the interaction between TC and TD. Instead, they were designed with a wider scope to compare learners' and teachers' interpretations of classroom activities in terms of their nature, aims and the demands they imposed on learners. The perceptual mismatch in these studies was found to be associated with a wide range of factors (e.g., *cognitive requirements*, *communicative competence*, *linguistic competence*, and *pedagogic strategies*).

Révész and Gurzynski-Weiss (2016) employed a bottom-up approach to investigate the sources contributing to TD from L2 teachers' perspectives. Employing a novel method which combined think-aloud strategies with eye-tracking techniques, this study collected data from 16 teachers on how they could evaluate linguistic demands of a set of tasks, i.e., decision making and gap-filling, and how they could gauge TD to meet learners' different levels of language proficiency. The results endorsed *linguistic and cognitive demands* as prime criteria for evaluating, increasing or decreasing TD. Other factors included *modality*, *task outcome* and *interactional demands*. The eye-tracking data were to some extent in harmony with the teachers' think-aloud annotations. Longer and more eye-fixations were captured on tasks instructions rather than on the prompts when the teachers were thinking about assessing the suitability of proficiency level to match each task as compared to increase or decrease TD. However, this study can be also criticised for the relatively small sample size and for overlooking learners' perceptions to triangulate those obtained from teachers' perspectives.

Interestingly, some of the previous studies reviewed here are in harmony with respect to endorsing cognitive and linguistic demands as key factors contributing to TD. The majority of TD-factors detected in these studies are incorporated in either Robinson's (2007) or Skehan's (1998) taxonomies of TC, which indicates that both frameworks are, to some extent, reliable tools to identify and operationalise TC and TD. However, it is worth noting that linguistic demands as a key factor associated with perceptions of TD in previous studies does not conform to the factors in Robinson's framework of TC, unlike Skehan's framework. This again calls more attention to the ongoing debate between these two predominant TC frameworks on what factors are more salient in determining TC and TD. Furthermore, a review of the literature shows that only a few studies investigated both learners' and teachers' perceptions of TD quantitatively and qualitatively. Henceforth, more collaborative research efforts are needed to adopt multi-perspective approaches to identify the aspects contributing to perceived TD if a well-established index of TD is to be proposed.

Though previous studies revealed consistent results regarding rating TD or the factors that contributed to perceptions of TD, the reliability of the findings can still be questioned due to the subjectivity of rating or the unsystematic operationalisations of TC variables. The abstract nature of the variables contributing to TC and TD, and the procedural difficulties involved in defining, manipulating and exploring the two constructs, can fairly explicate the scarcity of studies with respect to TD. It is therefore crucial to adopt more valid frameworks to operationalise TC and mixed-method study designs that incorporate collecting quantitative and qualitative data from different stakeholders to triangulate the findings regarding TD.

Responding to the aforementioned issues, this study is designed to help pave the way to establish the much-needed index of TD by considering the interaction between learner factor diagnosed through

quantitative and qualitative perceptual data and task factor signified by a careful manipulation of TC through intentional reasoning (IR). Learners' perceptions of TD are examined alongside language professionals' judgement in order to address learners' rating subjectivity, and thus increase the reliability of the findings. Hence, the originality of this study lies in four aspects: 1) it is the only study to explore both L2 learners' and language professionals' perceptions of TD on tasks manipulated by IR as a TC variable; 2) it is the only study to adopt a mixed method approach to explore the interaction between IR as a TC factor and TD factors; 3) it proposes a two-level framework to operationalise IR, i.e., task instructions and task content; and 4) it exploits the findings as a validity evidence of the operationalisation and manipulation of TC. Following the aforementioned aims, the current study addresses the following research questions:

1. How difficult do L2 learners and professionals perceive performing two tasks with two levels of TC (-IR vs +IR)?
2. What factors contribute to L2 learners' and professionals' perceptions of TD?
3. Is there a perceptual match between L2 learners and professionals?

## Method

This study employed a mixed-methods approach and a within-participant research design which incorporated analyses of retrospective data on the perceptions of TD of L2 learners and ESOL professionals. Data were collected quantitatively and qualitatively by administering a retrospective questionnaire to rate (learners) or evaluate (language professionals) two video-based narrative tasks (-IR vs +IR) with respect to their degree of TD. Throughout this procedure, the two tasks were presented in a counterbalanced order.

## Participants

The participants were enrolled from two different contexts. The L2 learners were 68 students at a secondary school in Jordan. All were male and spoke Arabic as a first language. They were 16 years old and had been learning English for 10 years at school. They reported they had never lived in an English-speaking country before. The learners belonged to four levels of language proficiency that corresponded to CEFR A2-C1 based on their school internal language assessment and the administering of an Oxford Placement Test (Allan, 2004). However, most learners were in the B1-B2 range (N = 58). The twenty-six language professionals who participated in this study were researchers in Applied Linguistics from a university in the UK with different levels of experience in teaching ESOL (2-17 years). The language professionals were 20 females and 6 males aged between 25 and 64 from 9 first language backgrounds (Arabic, Chinese, English, Dutch, Farsi, Greek, Kazakh, Turkish, Urdu). All the participants volunteered to take part in this study.

## Tasks and Materials

The present study employed two video-based narrative tasks to elicit L2 speech performance. The two video clips were selected from the Pat & Mat television series (Beneš & Jiránek, 1976) featuring two handymen who frequently run into problems and complex situations. The characters, Pat and Mat, normally try out different creative and unpredictable approaches to overcome a sequence of obstacles. The selected clips offer a rich stimulus for a watch-and-tell condition as viewers can narrate the events while watching the story unfold. The silent nature of the episodes makes them appropriate tasks to operationalise IR in which the learners can reason the characters' intentions, read their thoughts, predict their actions, and justify their decisions. De Jong and Vercellotti's (2016) framework of prompts selection

was implemented wherein the structure, storyline complexity and number of elements were carefully considered to ensure the comparability of the content of the two clips. Hence, the two selected clips incorporated schematic chronological structure (Tavakoli, 2009a), comparable storyline complexity (Tavakoli & Foster, 2008), and equal number of characters (De Jong & Vercellotti, 2016). However, they varied in terms of the amount of IR demands each prompt required. IR was employed in this study as a variable to manipulate TC.

Attempting to investigate intentional reasoning (IR) more systematically and differently from previous studies (e.g., Ishikawa, 2006; 2008; Sattarpour & Farrokhi, 2017), a two-level operationalisation was proposed at instruction-level and content-level (see Awwad et al., 2017 and Awwad & Tavakoli, 2019, for further details). Task instructions were manipulated to encourage the participants to only tell and describe the story of the -IR clip and to further read the characters' minds, reason their intentions, and predict and justify their actions under the +IR condition. To manipulate IR at content-level, a careful selection process was considered to ensure the content of the two selected clips involved a different degree of IR requirements. The content of the -IR video clip which required less reasoning showed Pat and Mat enjoying a sunny day when they decided to cook lunch outdoors, but suddenly it started raining. They kept trying to overcome this problem until they finally decided to cook inside their cottage, where they encountered a new set of challenges. The events of this video were clear and did not include unpredictable ideas or imaginative actions. Therefore, the clip did not require the participants to reason much about intentions while narrating the story (see Appendix). The content of the +IR video clip which required more reasoning was about Pat and Mat trying to fly with their car like an airplane. They came up with weird but creative ideas to make their car fly. The content of the +IR condition was more imaginative and unpredictable, and therefore required reading the characters' thoughts and predicting what they intended to do to make their car fly, considering why they would do this, and what consequences would follow (see Appendix).

To collect quantitative and qualitative data from the L2 learners on their perceptions of TD, a retrospective questionnaire was administered immediately after performing the two narrative tasks (-IR/+IR). Following Robinson, (2001) and Tavakoli (2009a), the TD questionnaire included two scale-questions to rate the difficulty level of each task by ranking them as 1 (very easy), 2 (easy), 3 (difficult), and 4 (very difficult). Additionally, two open-ended questions were included in order for the learners to justify their judgement. The language professionals' TD questionnaire required them to evaluate the TD of the same tasks by rating them as 1 (easy) or 2 (difficult), and then justify their judgement through two open-ended questions. The retrospective questionnaire was employed as a data collection tool to stimulate the participants to immediately recall and verbalise their feelings and thoughts about the level of TD after performing the narrative tasks. The instant recall of perceptions is assumed to be a reliable and valid technique to collect retrospective data as the relevant information is retrieved from the participants' memory after a short time interval to minimise the effect of memory decline (Dörnyei, 2007).

## **Procedure**

Data were collected from the participants individually in a quiet room during a regular school day. All the learners performed the two video-based tasks (-IR/+IR) in a counterbalanced order to reduce any effect of practice and order. The instructions were presented to the participants in their L2 and L1 before each video was shown. The researcher set a voice recorder and played the video clips (120 seconds each) while the participants narrated the events of each video clip as they were watching them. After performing the two tasks, each participant completed a questionnaire on his perceptions of TD, in which he described how difficult each task (-IR vs +IR) was and justified his judgement through two open-ended questions. Regarding the language professionals, data were collected during one session at the same time. The researcher presented the instructions of each task in English and then played the two video clips. After watching the two clips, the language professionals were requested to rate which task (-IR vs +IR) would be more difficult for L2 learners to perform and what factors could contribute to the

perceptions of TD. They completed a TD evaluation form by choosing 1 (easy) or 2 (difficult) and answered two follow-up questions.

## Data Analysis

The output of the L2 learners' and professionals' questionnaires were considered as primary data to answer the research questions raised earlier. For the quantitative data, the descriptive analysis was followed by running paired-sample t-tests to locate any statistically significant differences between the participants' perceptions of the TD of the -IR versus the +IR task. Cohen's *d* effect size was calculated when significant differences were observed (Cohen, 2013). In order to interpret the findings in terms of their effect sizes more accurately, Plonsky and Oswald's (2014) thresholds for within-group comparisons were adopted, i.e., small (0.6), medium (1.00) and large (1.40).

For the qualitative analysis, the data gathered from the learners' and professionals' responses were separately exposed to a systematic thematic analysis (Creswell, 2015). Furthermore, the process of the qualitative analysis of the responses emerging from the questionnaires adhered to the procedures suggested by Dörnyei (2007). All the learners and professionals' answers to the two open-ended retrospective questions in which they justified their rating of TD were transcribed separately. Some participants mentioned more than one reason for their TD rating. Therefore, each different reason was included in a global enumerated list for each group of participants. A colour coding scheme was employed to trace similar themes in each category. The list of preliminary themes of each group was then filtered by grouping similar or interrelated themes together to identify the most commonly occurring factors that affected the perceptions of TD. Common patterns in the responses of each group about their perceptions of TD were identified, examined and grouped together to form a number of themes. Next, another experienced researcher retraced the data to check whether the identified themes matched the participants' responses. Then, the grouped sets of themes of each type of respondent, i.e. learners and language professionals, were compared to detect whether there was a degree of TD perceptual match or mismatch between the two types of respondent. Finally, the identified factors were further matched against the two-level framework employed by this study to operationalise IR to validate the way TC was manipulated.

## Results

### Quantitative Analysis

In order to answer the first research question regarding how difficult the L2 learners and professionals perceived two tasks varying in TC through two levels of IR, paired samples t-test comparisons were conducted, and the effect sizes were calculated. The learners rated each task in terms of difficulty on a four-point scale (1 = very easy, 2 = easy, 3 = difficult, 4 = very difficult). As shown in Figure 1 below, the descriptive analysis of the learners' rating of TD indicated that the +IR task was perceived as more difficult ( $M = 2.70$ ,  $SD = .55$ ) than the -IR task ( $M = 1.90$ ,  $SD = .64$ ). The output of the paired-samples t-test detected a statistically significant difference between the learners' perceptions of the TD of the two tasks, with a fairly large effect size in favour of the +IR task ( $t = 8.005$ ,  $p = .000$ ,  $d = 1.34$ ).

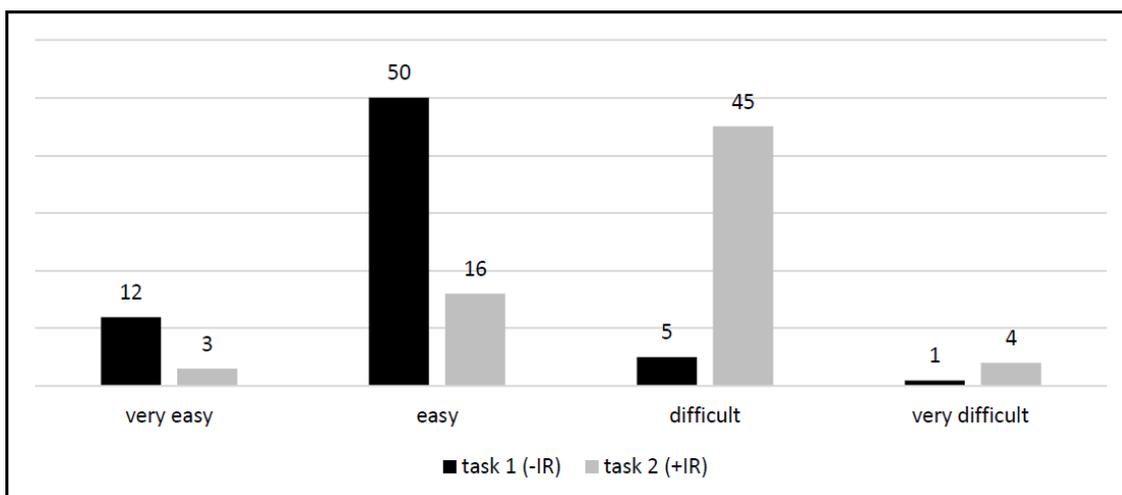


Figure 1. Learners' perceptions of task difficulty.

Turning now to the results of the language professionals' perceptions of the TD, they evaluated the same two tasks as 1 = easy or 2 = difficult. The descriptive analysis showed that the majority of language professionals rated the +IR task as more difficult ( $M = 1.85$ ,  $SD = .37$ ) than the -IR task ( $M = 1.15$ ,  $SD = .37$ ) with 22 out of 26 language professionals rating the +IR task as more difficult. A t-test also revealed a statistically significant difference in favour of the +IR task with a medium effect size ( $t = 4.79$ ,  $p = .000$ ,  $d = 1.189$ ). The significant differences shown by the t-tests confirmed that both L2 learners and language professionals perceived the more complex task that required higher IR demands as more difficult than the less complex task (-IR).

### Qualitative Analysis

To answer the second research question concerning the factors that contributed to the learners' and language professionals' perceptions of TD, the qualitative data gathered from the questionnaires were exposed to a thematic analysis. The analysis revealed a number of common factors that affected both learners' and professionals' judgement with respect to TD. Four key themes emerged from the learners' responses, whereas only three themes were unfolded from the language professionals' comments. Concerning learners, the most frequently cited themes were two cognitive-related factors that stemmed from their perceptions of the higher cognitive requirements imposed in the +IR condition. As shown in Table 1 below, the learners' comments about the cognitive demands were divided into two categories, i.e. task-induced cognitive demands as triggered by task instructions (42% of the comments) and task-inherent demands as imposed by task content (41%). The non-cognitive-related factors affecting learners' perceptions of TD were linguistic demands (11%) and time pressure (6%). The table below shows the frequency and percentages of each theme supported by examples from the learners' responses.

The majority of the responses that were collected from the 68 learners (83%) attributed their perceptions of TD to the existence or absence of the higher cognitive demands as imposed by either each task content or instructions. Sixty-nine responses out of a total of 164 comments (42%) which were mentioned by the learners attributed the perceptions of TD to factors associated with task-induced cognitive demands. That is the requirements of the +IR condition to read the characters' thoughts, explain their intentions, predict their behaviours, and justify their actions and reactions as stated explicitly in the +IR task instructions. Conversely, many learners found the -IR task easier because they were only asked by the -IR task instructions to tell and describe what was happening. The second most frequently mentioned theme was the task-inherent cognitive demands. Sixty-eight responses out of a total of 164 comments (41%) gathered from the learners related their perceptions of the difficulty of the two tasks to the degree of topic familiarity, logic, clarity, and predictability. For example, a number of participants

rated performing the +IR task as more difficult because the story included unfamiliar, unreal, unpredictable or imaginative events. Alternatively, those who rated the -IR task as easier, mentioned that the chains of events of the -IR story were easier to predict and narrate because they were familiar with the topic or they found the story plot simpler, clearer or more coherent as it contained more logical and real-life events than the +IR story.

TABLE 1  
*Thematic Analysis of the Learners' Perceptions of TD*

Themes	Examples from the data	No. of times	%
Task-induced cognitive demands: caused by task instructions (e.g., <i>read minds, reason, predict and justify actions</i> )	- Difficult because you have to think and predict what is happening and explain their intentions. (+IR) - I found it difficult because I have to read their minds and what they are thinking about. (+IR) - Easy because describing and telling what's happening is easier than foretelling. (-IR)	69	42%
Task-inherent cognitive demands: caused by task content (e.g. <i>clarity, logic, familiarity, predictability</i> )	- It was difficult because it's not a real-life story and there's a lot imagination. (+IR) - It's difficult because that's not common in normal life. So I needed some thinking about it. (+IR) -The task was easy because all actions were easier to understand and predict. (-IR)	68	41%
Linguistic demands	- I find it difficult because I didn't know all the words. (+IR) - Very easy because the words that I used are easy. (-IR)	18	11%
Time pressure	- It is not very easy because sometimes I don't have the time to say the correct explanation of the scene. (+IR) - I needed more time to analyse the events. (+IR)	9	6%

*Note.* Total number of participants 68. Total number of responses: 164

The linguistic requirement was the third category that influenced the learners' perceptions of TD receiving 11% of the responses (18 comments). Some participants mentioned that they needed to use more complex lexical items or language to express intentionality or narrate the story of the +IR task. Others mentioned that they could not find the suitable words. Very few comments stated that narrating the -IR task was easier because the required lexical items were easier, simpler, or more frequently used. The final emerging theme was associated with the effects of increasing the TC on time pressure. Only nine comments (6%) mentioned that time was not enough to meet the cognitive and linguistic demands of the +IR task and that they needed more time to successfully complete the task. Even though the two tasks were equal in terms of speaking in real time conditions and that time pressure is a feature of video-based narratives, the negative effect of time pressure was more salient in the +IR condition.

Moving to the findings of the qualitative analysis of the language professionals' comments, the thematic investigation revealed three central themes on why L2 learners would find the +IR task more difficult than the -IR counterpart. The key themes emerging from the language professionals' responses are shown in Table 2 below supported by examples from their comments. In line with the L2 learners, the language professionals attributed their perceptions of TD largely to the increased cognitive demands of the +IR condition that were inherent in the content of the +IR story. This theme appeared in 15 out of 31 comments (48%). The majority of language professionals labelled the content of the +IR story as more difficult to narrate because it included unrealistic and complicated events with unpredictable consequences, unlike the -IR story. The other dominating theme was related to factors that were initiated by the instructions of each task. These factors were traced in 14 of the language professionals' responses from a total of 31 (45%). They confirmed that instructing the learners to read others' minds, reason about intentions, justify actions and interpret reactions would increase the cognitive requirements of the +IR task, resulting in higher perceptions of TD than in the -IR condition. Only two comments ascribed the perceptions of TD to the linguistic demands of the -IR task, which required a higher number of different lexis than the +IR task. However, none of the language professionals suggested time pressure as a possible factor affecting learners' perceptions of TD.

TABLE 2  
*Thematic Analysis of the Language Professionals' Perceptions of TD*

Themes	Examples from the data	No. of times	%
Task-induced cognitive demands: caused by task instructions (e.g., <i>read minds, reason, predict and justify actions</i> )	- More effort is required in the +IR task to interpret and explain the characters' reactions. - Video 2 will be more difficult because you need to ask more 'why-questions' while describing the actions. (+IR)	14	45%
Task-inherent cognitive demands: caused by task content (e.g. <i>clarity, logic, familiarity, predictability</i> )	- Clip 1 is easier because it is more realistic and is linked to our daily life. (-IR) - Their reactions in the second clip (+IR) are not predictable and not what one would normally expect.	15	48%
Linguistic demands	- It involves the use of a lot of vocabulary and requires more communication (+IR).	2	6%

Note. Total number of participants 26. Total number of responses: 31

## L2 Learners' and Professionals' Perceptual Match

The third research question investigated whether there was a match between L2 learners and professionals on their perceptions of TD. The findings obtained from the quantitative data clearly confirmed a perceptual match between the two stakeholders, who rated the more complex task (+IR) as more difficult to perform than the less complex one (-IR). Furthermore, the thematic analysis of the qualitative data revealed a match regarding their justification of the factors that affected the judgement of TD. Both groups identified the same two categories of cognitive demands, i.e. task-induced and task-inherent as focal contributors to the perceptions of TD. A high percentage of the learners' comments (83%) and language professionals' comments (94%) attributed the aforementioned two categories of cognitive demands as having a direct impact on learners' perceived TD. Moreover, the linguistic demands of the task were regarded by both groups as another possible factor to influence weighing TD. However, linguistic demand was mentioned in the responses infrequently by both learners (11%) and language professionals (6%). As regards time pressure, only a very low percentage of the learners' comments (6%) indicated it as another probable criterion of perceiving TD, whereas none of the language professionals regarded time pressure as influencing their perceptions of TD when tasks were increased in TC.

## Discussion

This study is the first to investigate perceptions of task difficulty of both L2 learners and language professionals on tasks manipulated by intentional reasoning as a TC variable. In light of the limitations of previous research on perceptions of TD, the current study took the construct of TD as a prime focus and employed a mixed-methods approach to tackle more thoroughly the multi-faceted nature of TD from the perspectives of both learners and language professionals. This study also explored the extent to which both groups' perceptions of TD matched, as well as their reasons for these perceptions. The current study strove further to operationalise TC more systematically to ensure that the designed tasks were distinctive with respect to TC requirements. Moreover, the study attempted to exploit the findings to validate the way TC was operationalised through IR at two levels, i.e., task content and task instruction.

To summarise the findings presented earlier, it was found that both learners and language professionals perceived the more complex task (+IR) as more difficult than the less complex task (-IR). These findings support the predictions of both Robinson's Cognition Hypothesis (2001) and Skehan's Limited Attentional Capacity (1998) that tasks with increased TC will be rated as more difficult by task performers. The results of this study are also in line with most of previous studies which found that increasing TC had a direct impact on perceptions of TD (e.g., Robinson, 2001; Tavakoli & Skehan, 2005).

The findings are further in harmony with those of Sasayama (2016), who found a significant difference with a large effect size between learners' perceptions of TD of Task 1 (designed-to-be-simplest) versus Task 4 (designed-to-be-most-complex). However, the current study found a significant difference with a large effect size by employing only two levels of TC (-IR vs +IR), instead of four levels as in Sasayama (2016). This can be an indication of the successful manipulation of TC in this study by designing the two tasks with a prominent benchmark of cognitive demands. Hence, this study advocates considering such a benchmark of TC requirements when designing a continuum of complex tasks to ensure attaining significant results regarding language performance and perceptions of TD across all levels of TC. The findings of this study lend full support to the two studies which manipulated TC similarly through IR demands, i.e. Ishikawa (2011) and Robinson (2007). It is worth mentioning that these two studies did not explore which IR-related factors affected learners' perceptions of TD. This research gap was addressed by the current study, which investigated these factors not only from L2 learners' perspectives but also from language professionals' viewpoints.

By investigating the perception of TD qualitatively, the results revealed two TC-related factors as prime contributors to the perception of TD of both learners and language professionals, namely task-induced cognitive demands originated by task instruction and task-inherent cognitive demands imposed by task content. Moreover, this study designated the linguistic demands of task as a less predominant factor influencing TD. These results lend more support to the Cognition Hypothesis which does not designate the language required to perform a task as a key factor affecting learners' perceptions of TD. The qualitative findings of this study offer partial support to Tavakoli (2009a), who also found that the cognitive demands of a task were the main factor contributing to TD, as acknowledged by L2 learners and teachers. Whilst this study attributed linguistic demands as a secondary TD factor, Tavakoli (2009a) endorsed lexical requirements alongside other factors, such as task structure, information quantity and clarity of prompts as the main features of TD. The discrepancy regarding the factors that were identified in both studies can be related to the dissimilar ways each study operationalised TC. Likewise, the findings of this study are in line with Révész and Gurzynski-Weiss (2016), who identified similar sources contributing to TD from L2 teachers' perspectives, that is manipulation of task instructions, cognitive demands (e.g., reasoning and familiarity) and linguistic demands. However, language required was the most frequently cited factor by the teachers in Révész and Gurzynski-Weiss study, whereas it was the least frequently mentioned theme by the teachers in the current study. Combining quantitative and qualitative data as a verification of data on TD in this study, Révész and Gurzynski-Weiss (2016) mapped teachers' eye-tracking data against their think-aloud annotations to attain more reliable findings. This cross-checking approach in data gathering and analysis in both studies can be regarded as one of their strengths.

Turning now to discuss the findings regarding the perceptual match of TD between learners and language professionals, the results suggested a match between both groups with respect to rating TD and identifying the factors contributing to it. Both learners and language professionals recommended induced and inherent cognitive demands as key factors in determining the TD of tasks with increased TC. However, the language professionals mentioned these factors more frequently than the learners. While both groups attributed linguistic demands as a secondary TD-factor, only learners mentioned time pressure as a TD-factor. This awareness of time pressure in the case of L2 learners can be justified by the fact that they performed the two tasks in real-time before evaluating their level of TD retrospectively. This was not the case regarding the language professionals who only watched the two video clips and then rated them in terms of the predicted TD level for L2 learners. This may explain why few learners attributed their perceptions of TD to the challenges of performing narrative tasks under time constraints, which was more salient in the more complex task (+IR).

The findings of this study were further employed as validation of the way TC was operationalised through two levels of IR demands and the way each task was selected and incorporated as a stimulus of speech production. The current study thus responded to the calls for researchers to validate their TC manipulation (Révész et al., 2016) and prompts selection (De Jong & Vercellotti, 2016). Attempting to

offer a more systematic operationalisation of TC variables, the present study proposed a two-level operationalisation of IR at the levels of task instructions and task content (for further detail, see Awwad et al., 2017, and Awwad & Tavakoli, 2019). Task instructions were designed carefully to encourage the participants to primarily tell and describe the events in the -IR task, whereas the instructions of the +IR task were manipulated to induce the participants to further read intentions, reason, predict, explain and justify actions. Task content was operated through a careful selection of the video clips to ensure their comparability except for the amount of IR involved in each clip. Both learners and language professionals rated the +IR task as more difficult than the -IR task. They further identified task-induced cognitive demands as oriented by task instructions, and task-inherent cognitive demands as stemming from the content of each clip as the key factors contributing to the perception of TD. Therefore, the findings provide robust evidence of the validity of IR operationalisation at two levels and also the careful designation and implementation of the two video-based tasks in this study.

Examining stakeholders' judgement of TD through combining quantitative and qualitative data, as in the present study, is argued to be crucial for broadening our knowledge on how task design and/or conditions interact with learner factors to explain how simple-to-complex sequences of tasks can be approached, performed and perceived differently by task performers or test takers. Therefore, the findings are assumed to have implication for language learning, teaching, testing and syllabus design by helping educators establish an index of task difficulty. The results of the present study suggest that L2 language professionals and syllabus designers are guided to consider the induced and inherent cognitive demands when designing, choosing or implementing L2 tasks. Moreover, the results obtained from this study have important implications for L2 testing. Language testing professionals need to consider carefully cognitive demands are required by different tasks, as well as the linguistic and time pressure demands. Tasks that target learners across a continuum of language proficiency levels need to adopt a prominent benchmark of cognitive demands which can challenge learners differently and thus distinguish them across different proficiency levels.

Finally, the limitations of this study need to be acknowledged to serve as potential considerations for future studies. The study employed a 2x2 research design which paired the -IR task with the -IR instruction and the +IR task with the +IR instruction. Employing a 4x4 study design counterbalancing the two tasks, i.e., less complex (-IR) and more complex (+IR) with the instruction manipulation would have extended our understanding of whether the content and the instruction manipulation interact to influence perceptions of task difficulty. Mapping the learners' perceptions of TD with their actual speech performance would have also offered deeper insights into the interaction between actual learners' language performance and subjective observation of TD. Future research is encouraged to further explore the interaction between learners' perceptions of TD and their language proficiency which was beyond the scope of this study. In addition, the current study collected data on TD by only administering pen-and-paper questionnaires. However, including retrospective interviews with L2 learners and teachers would provide richer data that tap into different facets of the construct under investigation. This may widen our understanding of the factors that can shape the desired index of TD.

## **Conclusion**

The findings of the present study provide empirical evidence regarding the direct effect of IR as a TC factor on the perceptions of TD. Identifying the TC-related factors contributing to TD and the perceptual match between learners and language professionals leads to the suggestion that task stakeholders, including researchers, teachers and task designers, need to consider both induced and inherent cognitive demands of task as central to any process that involves selecting, designing, sequencing, and employing tasks within L2 classrooms. Hence, the factors which have been identified in this study, i.e. task-inherent and task-induced cognitive demands, should be incorporated in any index of TD. It can be further

concluded that TC-oriented factors affecting TD may not be purely task-dependent or learner-dependent as they can be a combination of both task-internal and task-external factors.

Researchers need to scrutinize more thoroughly and systematically the ways TC variables are being conceptualised and operationalised before any conclusions can be made regarding the interaction between TC and TD. Providing independent evidence to support the validity of researchers' interventions is desirable to increase the generalisability of any findings. A benchmark of TC demands needs to be employed to distinguish tasks that are designed and sequenced within a simple-to-complex continuum to ensure that each task is different enough to achieve significant variances regarding language performance and perceptions of TD. The interaction between TC and TD remains an open area of enquiry which requires more collaborative efforts that consider involving all task stakeholders as sources of primary data to help attain a more holistic picture of the two constructs. Data on perceptions of TD should be collected by using a wide range of instruments, including questionnaires and interviews, to enhance the reliability of the findings. Future studies are encouraged to treat TD-related issues as a prime focus to help this line of research move forward by proposing a set of factors or criteria that can form the foundation of the much-needed index of TD. More comparable studies are inevitably required before such an index can be established.

### Acknowledgements

I would like to express my sincere gratitude to Dr. Parvaneh Tavakoli for her guidance and valuable comments on the first drafts of this manuscript. I would like also to extend my gratitude to the anonymous reviewers who have helped me improve this work.

### The Author

Anas Awwad is an Assistant Professor at the Department of English Language & Literature at Isra University, Jordan. His current research interests include Second Language Acquisition, Task-Based Language Teaching, Language Testing and TESOL. His recent publications investigated the effects of task complexity on speech performance and the interaction between task complexity and learner individual differences.

Department of English Language & Literature  
Faculty of Arts  
Isra University, Jordan  
PO Box 33/11622, Amman  
Tel: 00962-6-4711710  
Email: anas.awwad@iu.edu.jo; anasawwad2@gmail.com

### References

- Allan, D. (2004). *Oxford Placement Test. University of Cambridge Local Examination Syndicate*. Oxford: Oxford University Press.
- Awwad, A., Tavakoli, P., & Wright, C. (2017). "I think that's what he's doing": Effects of intentional reasoning on second language (L2) speech performance. *System*, 67, 158-169.
- Awwad, A., & Tavakoli, P. (2019). Task complexity, language proficiency and working memory: Interaction effects on second language speech performance. *International Review of Applied Linguistics in Language Teaching*. "Online early view". doi:10.1515/iral-2018-0378

- Barkhuizen, G. (1998). Discovering learners' perceptions of ESL classroom teaching/learning activities in a South African context. *TESOL Quarterly*, 32, 85-108.
- Beneš, L., & Jiránek, V. (1976). Pat & Mat, Czech stop-motion animated series. Retrieved from <http://en.patmat.cz/home-pat-and-mat/>
- Breen, M. (1987). Learner contributions to task design. In C. Candlin & D. Murphy (Eds.), *Language learning tasks* (pp. 23-46). London: Prentice Hall.
- Brindley, G. (1987). Factors affecting task difficulty. In D. Nunan (Ed.), *Guidelines for the development of curriculum resources* (pp. 45-56). Adelaide: National Curriculum Resource Centre.
- Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32, pp. 21-46). Amsterdam: John Benjamins.
- Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of management review*, 13(1), 40-52.
- Candlin, C. (1987). Towards task-based language learning. In C. Candlin & D. Murphy (Eds.), *Language learning tasks* (pp. 5-22). London: Prentice Hall.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Creswell, J. W. (2015). *A concise introduction to mixed methods research*. London: Sage Publications.
- De Jong, N., & Vercellotti, M. (2016). Similar prompts may not be similar in the performance they elicit: Examining fluency, complexity, accuracy, and lexis in narratives from five picture prompts. *Language Teaching Research*, 20(3), 387-404.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- Hawkey, R. (2006). Teacher and learner perceptions of language learning activity. *ELT journal*, 60, 242-252.
- Ishikawa, T. (2006). The effect of task complexity and language proficiency on task-based language performance. *The Journal of Asia TEFL*, 3(4), 193-225.
- Ishikawa, T. (2008). The effect of task demands of intentional reasoning on L2 speech performance. *The Journal of Asia TEFL*, 5(1), 29-63.
- Ishikawa, T. (2011). Examining the influence of intentional reasoning demands on learner perceptions of task difficulty and L2 monologic speech. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 307-330). Amsterdam: John Benjamins.
- Kumaravadivelu, B. (1991). Language-learning tasks: Teacher intention and learner interpretation. *ELT journal*, 45, 98-107.
- Kumaravadivelu, B. (2006). TESOL methods: Changing tracks, challenging trends. *TESOL Quarterly*, 40, 59-81.
- Liu, P., & Li, Z. (2012). Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics*, 42, 553-568.
- Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge: Cambridge University Press.
- Plonsky, L., & Oswald, F. 2014. How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912.
- 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.
- Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgements: A validation study. *Studies in second language acquisition*, 38(4), 703-737.
- Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27-57.

- Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *IRAL-International Review of Applied Linguistics in Language Teaching*, 45(3), 193-213.
- Robinson, P. (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (Vol. 2). Amsterdam: John Benjamins Publishing.
- Robinson, P. (2015). The Cognition Hypothesis, second language task demands, and the SSARC model of pedagogic task sequencing. In M. Bygate (Ed.), *Domains and directions in the development of TBLT* (Vol. 8, pp. 87-121). Amsterdam: John Benjamins.
- Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *IRAL-International Review of Applied Linguistics in Language Teaching*, 45, 161-176.
- Sasayama, S. (2016). Is a 'complex' task really complex? Validating the assumption of cognitive task complexity. *The Modern Language Journal*, 100(1), 231-254.
- Sattarpour, S., & Farrokhi, F. (2017). Exploring the interplay of planning time, reasoning demands, and language learning aptitude in Iranian EFL learners' written production. *The Journal of Asia TEFL*, 14(4), 736-754.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, P. (2014). The context for researching a processing perspective on task performance. In P. Skehan (Ed.), *Processing perspectives on task performance* (Vol. 5, pp. 1-26). Amsterdam: John Benjamins.
- Skehan, P. (2015). Limited attention capacity and cognition. In M. Bygate (Ed.), *Domains and directions in the development of TBLT* (Vol. 8, pp. 123-155). Amsterdam: John Benjamins Publishing.
- Tavakoli, P. (2009a). Investigating task difficulty: Learners' and teachers' perceptions. *International Journal of Applied Linguistics*, 19(1), 1-25.
- Tavakoli, P. (2009b). Researching task difficulty: Towards understanding L2 proficiency. In A. Benati (Ed.), *Issues in second language proficiency* (pp. 216-232). London: Continuum.
- Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58, 439-473.
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (Vol. 11, pp. 239-273). Amsterdam: John Benjamins.

## Appendix

1. A link to the video clip that has been used in the less complex task (- *intentional reasoning*) can be found below.

<https://www.youtube.com/watch?v=hh4yQzsXmLE>

2. A link to the video clip that has been used in the more complex task (+ *intentional reasoning*) can be found below.

<https://www.youtube.com/watch?v=GPgnJFDr4Lw>