Review and Critique of Current Theories of Task Complexity: The Lack of a 'Social Dimension' in Language Assessment

Victoria Clark

American University in Cairo, Egypt

0211

The Asian Conference on Language Learning 2013

Official Conference Proceedings 2013

Abstract

The growing interest in task-based language assessment is not without concerns. Tasks used in language assessments vary in terms of their complexity and the language they elicit. L2 learners' performance varies from task to task. Thus, one of the major challenges facing those concerned with gauging the influence of task characteristics and performance conditions on candidate performance is how to determine the complexity of tasks (Elder et al. 2002). Identifying characteristics and performance conditions that determine task complexity is necessary to ensure that appropriate tasks are selected and can be sequenced to improve the reliability of the task-based assessment and to ensure that the interpretations and uses that are made based on the test results are valid. However, current studies conducted to investigate the effects of task features on L2 performance have mainly focused on the effects of the cognitive (psycholinguistic) features of tasks and findings have been inconsistent and proven to be relatively insensitive when applied to testing context. This paper will review the current research on task complexity and it will be argued that this research has failed to take into account social elements and their effect on task complexity and L2 performance. The paper concludes with the view that current research into task complexity to inform tasked-based language assessment design can be improved through taking into consideration social as well as cognitive variables.

Introduction

Since the introduction of communicative language teaching (Widdowson 1978) and models of communicative competence (Canale and Swain 1980; Bachman 1990; Bachman and Palmer 1996), approaches to testing have evolved. Instead of tests being based on traditional psychometric methods of measuring isolated pieces of grammar and vocabulary knowledge, known as discrete-point testing, tests are increasingly focusing on measuring the ability to use language in social contexts. L2 learners are tested on their ability to use language accurately and appropriately in communicative contexts (Skehan 1998). This has led to a growing interest in task-based language assessment (TBLA) which Brindley (1994, p. 74) defines as "the process of evaluating, in relation to a set of explicitly stated criteria, the quality of the communicative performances elicited from learners as part of goal-directed, meaning-focused language use requiring the integration of skills and knowledge." In TBLA, a broader conception of communicative competence is embraced and test tasks require test-takers to show competence of topical, social, and/or pragmatic knowledge as well as knowledge of the formal elements of language (Mislevy et al. 2002). It also allows assessment to be aligned with task-based instruction, has positive washback effects on instruction, and reduces the limitations of discrete-point assessments (Long and Norris 2000).

However, as Ellis (2003, p. 288) points out, "if task-based tests are to be used to infer the abilities of test-takers to predict performance and to generalise from context to context, it will be necessary to understand how the choice of task influences the way the testee performs." Indeed, the issue of task types and variation in L2 performance has been growing in interest for the last few decades. Researchers have argued that learners' performance differs from task to task, i.e. there is "task-induced variation" (Ellis 2008).

Task Complexity and the Cognitive Approach

One of the major challenges facing those concerned with gauging the influence of task characteristics and performance conditions on candidate performance is how to determine the complexity of tasks (Elder et al. 2002). However, in task-based language teaching and assessment literature "a principled and empirically supported conceptualization of task complexity has long been considered a primary goal and has, for just as long, proved to be elusive" (Norris et al. 1998. P. 39). Since Long (1985) first contended that the notion of complexity in task-based language pedagogy is an important consideration, many attempts have been made to address this issue.

Current research into task complexity is dominated by research into cognitive (psycholinguistic) categories (e.g. Robinson et al. 1995; Robinson 1995, 2001; Foster and Skehan 1996; Skehan 1996, 1998). This approach focuses on the effects on L2 performance of the cognitive features of tasks and is based on the assumption that cognitive complexity of a specific task influences the learners' task performance. Predictions have been made concerning how cognitive features increase or decrease task complexity and how these differing task demands affect learner accuracy,

complexity and fluency. Two main models have emerged from this approach. One model was developed by Skehan and his associates (Skehan 1998, 2001, 2003; Skehan and Foster 1999, 2001) and presents a framework of factors which they claim affect the complexity of a task. This framework is known as the Limited Attentional Capacity Model (LACM). The other influential model for determining task complexity is Robinson's Triadic Componential Framework (2001a, 2001b, 2003, 2005, 2007). These two models offer competing views on how attention is deployed during task performance and on how the manipulation of cognitive elements in tasks impacts L2 production.

Skehan's Limited Attentional Capacity Model (LACM)

For Skehan (1998), task complexity is the amount of attention that the task demands from the learners; difficult tasks require more attention than easy ones. The basic assumption underlying this view is that attentional resources are limited and that L2 learners must prioritise where they allocate their attention. When task demands are increased, attention allocated to particular aspects of task performance may be reduced as demand in other areas increases. Skehan (1998) claims that when L2 learners reach their attentional limits, processing for meaning will be prioritised over processing for language form. This is due to the fact that the demands of processing task content and the demands of task performance are in competition with one another (Skehan and Foster 2001); a more complex task will require L2 learners to devote more attention to content and thus cause them to over-emphasise fluency at the expense of accuracy and complexity. A less complex task, on the other hand, will allow an increase in attention to linguistic code and thus result in gains for focus on language forms. Skehan's concept of limited attentional resources leads to a trade-off between attention to form and attention to meaning during task performance. Furthermore, Skehan (2009) specifies that when attentional resources are limited, not only is there a trade-off effect between meaning and form but there is also a trade-off effect between accuracy and complexity. Performing a task may lead to gains in two out of three of the performance dimensions, but not typically in all three. In other words, fluency and accuracy or fluency and complexity may be increased (Skehan 2001, 2009). The theoretical basis for this model stems in part from findings by VanPatten (1990, 1994) who argued that form and meaning compete for learners' attention. Also, Schmidt (2001) argued that due to the constraints of the working memory, attention is limited and when attending to one area, the other areas are forced to operate with less attentional resources.

Skehan (2001, p. 194) claims that assessing a task's complexity "is crucial to understanding how it might be performed." He designed a model of task complexity in which the factors that affect task complexity are divided into three dimensions: code complexity, cognitive complexity and communicative stress. Skehan (1998) claimed these factors may influence a learners' allocation of attentional resources during a task as well as their linguistic performance. Table 1 below shows Skehan's model of task complexity.

Table 1: ***Skehan's model of task complexity (based on Skehan 1998)***

Code complexity	Cognitive complexity	Communicative stress
Linguistic complexity and variety Vocabulary load and variety	Cognitive familiarity - Familiarity of topic - Familiarity of discourse genre - Familiarity of task Cognitive processing - Information organisation - Amount of computation - Clarity of information - Sufficiency of information	Time pressure Scale Modality Stakes Opportunity for control

Code complexity deals with the linguistic demand of the task, i.e. the language needed to complete the task. More complex tasks are hypothesised to be those which require more advanced and a wider range of grammatical structures and lexical items from the task performer. Cognitive complexity distinguishes two areas: cognitive familiarity and cognitive processing. Cognitive familiarity concerns the extent to which the learner can draw upon previous experiences of performing such a task or similar ones. Thus, if the task itself or the topic of the task is not familiar to the learner, it is hypothesised to be more complex. Cognitive processing concerns the thinking that is required to perform the task. The more the learner needs to organise the information or the more steps needed to complete the task, the more demanding the task will be and thus more complex. The last set of factors is referred to under the term 'communicative stress'. These factors are concerned with the performance conditions for accomplishing the task (Skehan 1998). Time pressure refers to the amount of pressure exerted on the learner to perform a task quickly, as little or no planning time may make the task more complex. The factor referred to as 'scale' refers to the number of participants and relationships in the task, increases in the number of which will lead to increased complexity. The concept of modality refers simply to whether the task is a speaking or listening task or a reading or writing task. Speaking is assumed to exert more pressure on the learner than writing, whilst a listening task is believed to be more demanding than a reading task. Stakes refer to the importance of performing the task and of performing it well. The higher the stakes, the more demanding and thus complex the task is argued to be. Finally, control is concerned with the extent to which the participants of the task can influence the performance of the task. If a learner is allowed to ask for clarification and comprehension checks, for example, then task complexity is hypothesised to be lower

than if the learner has no influence on the speed of delivery of the input. In addition to these three categories, Skehan recognises that learner characteristics, such as the learner's intelligence, breadth of imagination and personal experience may also interact with the essential complexity of the task to influence its difficulty for a particular learner.

Robinson's Cognition Hypothesis and Triadic Componential Framework (TCF)

Robinson (2001a, 2001b, 2003, 2005, 2007) holds a different view on the effect of cognitive task complexity on linguistic performance. Robinson argues that the limits of attention are not as fixed as Skehan believes. Rather, he sees attention as a resource which is expandable**.** In his Cognition Hypothesis (2001, 2003, 2005), Robinson predicted that increasing the cognitive demands of tasks along certain dimensions would not result in a trade-off relationship between aspects of speaker production, but could actually direct the learners' attentional resources to linguistic forms, and thus improve the complexity and accuracy of learner output. Robinson supported the assumption upon which his claims are based with findings from Givon (1985, 1995) who argued that structural complexity tends to accompany functional complexity. In other words, when faced with a complex task, i.e. a task with increased cognitive demands, L2 speakers are pushed to use more complex language and achieve greater accuracy to ensure that communication is effective. In contrast, with a simple task, learner language does not need to be as linguistically accurate or complex for it to be performed successfully. This view is also motivated by claims that tasks which have high communicative and cognitive demands can lead learners to push production (Swain 1985), and stretch their interlanguage (Long 1989).

To address the claims of the Cognition Hypothesis, Robinson proposed the Triadic Componential Framework, shown in Table 2, in which he distinguished three dimensions which interact to influence task performance and learning: (1) task complexity, which corresponds to Skehan's 'cognitive complexity' category; (2) task conditions, which is comparable to Skehan's 'communicative stress' category; (3) and task difficulty, which is lacking in Skehan's model (Kuiken and Vedder 2007) but acknowledged by Skehan to affect complexity. For Robinson (2003, p. 56), the term 'task complexity' is not used synonymously with 'task difficulty' as with many other researchers, but instead refers specifically to "the intrinsic cognitive demands of the task", and can explain variation in performance by a learner on any two tasks. On the other hand, the category of 'task difficulty' is understood by Robinson as the learners' perceptions of task demands which are affected by both their ability, such as working memory capacity, and other affective variables, such as motivation. This category accounts for variation in performance by two learners performing the same task. Finally, task conditions are concerned with the participation factors (e.g. one- way vs. two-way) as well as participant factors (e.g. interlocutor is familiar vs. unfamiliar).

Table 2: ***Robinson's Triadic Componential Framework (Robinson 2007)***

Task Complexity (cognitive factors)	**Task Conditions (interactive factors)**	**Task Difficulty (learner factors)**
a) resource-directing variables making cognitive/conceptual demands +/- few elements +/- here-and-now -/+causal reasoning -/+ spatial reasoning -/+ intentional reasoning -/+ perspective-taking	**a) participation variables making interactional demands** +/- one-way flow +/- convergent solution +/- open solution +/- few participants +/- few contributions needed +/- negotiation not needed	**a) ability variables and task-relevant resource differentials** h/l working memory h/l reasoning h/l task-switching h/l aptitude h/l field independence h/l mind/intention reading
b)resource-dispersing variables making performance/procedural demands +/- planning +/- single task +/- task structure +/- few steps +/- independency of steps +/- prior knowledge	**b) participant variables making interactant demands** +/- same gender +/-familiar +/- same proficiency +/- shared content knowledge +/- equal status and role +/-shared cultural knowledge	**b) affective variables and task-relevant state-trait differentials** h/l openness to experience h/l control of emotion h/l task motivation h/l processing anxiety h/l willingness to communicate h/l self-efficacy

In his TCF, each factor can be +/- or h/l (high/low). Factors that are '+' or 'h' are assumed to make a task less complex and easier, whilst tasks which are '-' or 'l' are assumed to make the task more complex or difficult.

Robinson argued that dimensions of task complexity can be manipulated to either increase or decrease the cognitive demands that the task imposes on the task performer. Robinson distinguishes in his category of task complexity between resource-directing and resource-dispersing dimensions, and made separate claims for them regarding the way these dimensions affect resource allocation during L2 performance. Increasing task complexity along the resource-directing dimensions does not degrade linguistic output, but instead may improve the accuracy and complexity of the language produced. The reason given for the improvement is that to ensure effective communication, more accurate and specific linguistic features, such as logical connectors, embedding and subordination, are necessary. Robinson (2001, p. 35) sees increasing tasks along these dimensions as "a means of *directing* resources to a wider range of functional and linguistic requirements." Fluency, on the other hand, suffers from increased task complexity since tasks with higher cognitive demands increase the need for conscious language processing, thus affecting procedural dimensions like fluency (Robinson 2005). In contrast to increasing task complexity along the resource-directing dimensions, increasing task complexity along the

resource-dispersing dimensions has the effect of degrading accuracy, complexity and fluency because by making these dimensions more complex, greater demands on the working memory and attention are imposed and resources are, therefore, not directed to specific linguistic features (Kuiken and Vedder 2007).

These predictions for the effects of task complexity are for monologic tasks only. For dialogic, interactive, tasks Robinson (2005) claimed that greater complexity would most likely lead to greater amounts of negotiation for meaning, i.e. more clarification requests and comprehensions checks; as a consequence, the overall length and complexity of the utterance would be reduced on more complex tasks, as this increased amount of interaction will prevent learners from directing attention to syntactic complexity. For fluency and accuracy, Robinson made similar predictions for dialogic tasks as for monologic tasks. Table 3 shows the predictions for monologic and dialogic tasks.

Table 3: ***Effects of task complexity along resource-directing dimensions (based on Robinson 2011)***

Monologic simple + fluency,- complexity, - accuracy	**Monologic complex** - fluency, +complexity,+ accuracy
Dialogic simple + fluency,- complexity, - accuracy	**Dialogic simple** - fluency, -complexity ,+ accuracy

Empirical Studies and Testing Contexts

Empirical research has been carried out in classroom and laboratory settings to investigate the possible influence of cognitive task features on task complexity and on L2 performance. In a number of studies, manipulating task complexity along cognitive variables has led to systematic influences upon performance, though there is some disagreement as to the type of influence on L2 performance. In fact, the studies conducted on cognitive variables and their influence on learners' accuracy, complexity and fluency have yielded conflicting results. Take for example the cognitive variables of planning time which forms the bulk of the research and has offered inconsistent results. Mehnert's (1998) study found that that overall planning time had positive effects on performance, and that when allocated 10 minutes of planning time, learners were more fluent, more accurate and more lexically dense than non-planners. However, there were no significance changes in complexity. Other researchers have also found gains in fluency if learners are given time to prepare the task (Foster and Skehan 1996; Skehan and Foster 1997; Ortega 1999, Yuan and Ellis 2003). The same is true for structural complexity (Foster and Skehan 1996; Ortega 1999; Yuan and Ellis 2003), though no significant effects were found in the majority of studies for lexical complexity (Ortega 1999; Yuan and Ellis 2003) except in

Gilabert's (2005) study which found that lexical complexity (but not structural complexity) increased with planning time, As for accuracy, with planning time Yuan and Ellis (2003) found no differences in accuracy whilst Ortega's (1999) study revealed mixed results.

The findings of studies in which cognitive elements were manipulated in testing contexts have also proven to be somewhat conflicting. In a study conducted by Wigglesworth (1997), L2 speakers performed tasks in a planned and unplanned condition in a testing context. Performance was measured by analytical speech measures and analytical rating scales. Interestingly, no significant differences in the scores assigned using the rating scales were evident, but significant differences were shown in the analytical speech measures for fluency, accuracy and complexity. Wigglesworth concluded that planning time may affect the performance of test-takers positively but external ratings may well be insensitive to this effect. In a further study by Wigglesworth (2001), in which performance was measured only by analytical rating scales, task structure and task familiarity were investigated and no significant effects were shown on L2 performance. In addition, planning time was also explored and appeared to have an adverse effect on performance of both structured and unstructured tasks. The findings were inconsistent with those in non-testing situations and it was once again suggested that the external ratings were at fault. However, Iwashita, Elder and McNamara (2001) used both analytical rating scales and discourse measures when investigating the effect of planning time as well as other task characteristics (perspective, immediacy and adequacy) on performance. Their findings showed that planning time had no impact on oral performance or test scores. Elder and Iwashita's (2005**)** results also found no evidence of any effects and failed to confirm the findings of previous research on cognitive features and their effect on L2 performance. The inconsistent findings have even lead to claims that these findings "present a challenge to those who think that the task-based literature can make contributions to assessment" (Tavakoli and Skehan 2005, p. 244).

Limitations of the Cognitive Approach

Most of the previous research on task complexity has been from a psycholinguistic perspective (Taguchi 2007), and the research has not provided us with a clear picture as to the effect of manipulating cognitive variables on task complexity and L2 production. Several theories have been put forward to explain the inconsistent findings and insensitivity to testing contexts such as methodological flaws of the studies (such as small participation population and inconsistent task types), and the fact that task variables can potentially interact in complex ways and this can affect L2 production. Also, the inconsistent results could be due to factors such as test-takers' own characteristics (O'Sullivan 2002; Lumley and O'Sullivan 2005). Finally, concerning the inconsistent results, Elder et al. (2002, p. 362) have suggested that the results in testing contexts differ so markedly from those of the previous research because the testing context and pedagogic contexts are different "with the former producing a cognitive focus on display rather than on task fulfillment or getting the message across." Therefore, learners may be more concerned with accuracy than fluency during tests.

In sum, psycholinguistic categories have been very useful in investigating task complexity and have been shown to affect performance. However, replications of the psycholinguistic approach in a language testing context have not resulted in any corroboration of any statistically significant evidence found in research conducted in laboratory or instructional contexts. It is clear that further research into identifying dimensions of tasks that affect task difficulty and hence task complexity is needed. In addition, there is a need to look into new categories which may be useful in predicting task complexity and which could complement and extend the research on cognitive elements of tasks.

Social (Pragmatic) Approach to Task Complexity

A new approach was proposed by Fulcher and Reiter (2003), and later investigated by Taguchi (2007) and Clark (2012), in which the social (pragmatic) features of tasks are manipulated instead of the psycholinguistic (cognitive) features as "replications of the psycholinguistic approach have shown the categories of Skehan's model to be insensitive in a language testing context" (Fulcher and Reiter 2003, p. 328). Their studies were attempts to find new categories that may be useful in predicting task complexity for "what is lacking in the existing literature is the inclusion of pragmatic conditions in defining task difficulty" (Taguchi 2007, p. 114). However, psycholinguistic and pragmatic dimensions are not distinct and indeed social factors can influence cognitive processing in L2 performance. In fact, it has been argued that cognitive processing is possibly socially or environmentally driven (O'Sullivan 2000; Dörnyei 2009). On the basis of their findings, Fulcher and Reiter (2003) and Taguchi (2007) suggested that pragmatic features can determine task complexity and affect task performance and test scores.

This approach to determining task complexity is based on the findings from the field of pragmatics, in particular Brown and Levinson's (1987) politeness theory which predicts that speakers' and hearers' power relationships, and social and psychological distance, and the degree of imposition involved in speech acts, constrain communication (Blum-Kulka and House 1989), and relative power (P), social distance (D) and the degree of imposition of a task (I) play an important role in speech act behaviour (Brown and Levinson 1987). The power status and social distance of the addressee influences our linguistic choices as does the degree of imposition of the task. Speaking to someone who is of a higher power status (+P) and of a more distant relationship (+D) requires a higher level of politeness than when speaking to person who is –P and –D. The same is assumed for speech acts which involve a high degree of imposition (+ I) compared to – I acts (i.e. asking to borrow $100 [+I] compared to asking to borrow a pen [-I]). This higher degree of politeness is assumed to make higher demands on the speaker and hence +P, +D and +I situations are believed to be more complex (Taguchi 2007).

Studies by Fulcher and Reiter (2003) and Taguchi (2007) have focused on exploring pragmatic variables to determine and operationalise task complexity and investigate

the effects of manipulating these variables in tasks on L2 performance. Fulcher and Reiter (2003) manipulated the pragmatic features of social power and imposition and investigated the difficulty of the tasks as perceived by native and non-native speakers. 23 Spanish and 32 English-speaking students performed, in their native language, role-play tasks whose pragmatic features had been manipulated to increase or reduce the demands of the task. Afterwards, the participants watched video recordings of their role-play performance of requests and judged, on a 10-point scale, how successful they perceived their performance to have been. In one set of tasks, the social power of the speaker was lower than the hearer (S<H); in another set, it was equal (S=H), and in a final set it was higher (S>H). For imposition, each task was marked as being either 'high' or 'low'. The aim of this study was to investigate to what extent the different tasks described above and the participants' own first language cultural background can account for the differences in the assessment of task achievement when the participants perform the tasks in their native language. The perceived degree of task difficulty was high for tasks in which the social power was (S<H) and the degree of imposition was high. This study suggested that social factors such as power and the degree of imposition could serve as useful factors in predicting task difficulty and hence task complexity. In addition, the findings suggested that this increase or decrease in task complexity could affect output in measurable ways.

In a related study, Taguchi (2007) investigated the effects of tasks which had differing levels of the P, D and I sociolinguistic variables. In one task the power status difference, social distance and degree of imposition were small, i.e. PDI (Low) task, and in the other the power status difference, social distance and degree of imposition were large, i.e. PDI (High) task. The subjects were 59 Japanese L2 learners of two different proficiencies. The tasks were role-plays in which requests and refusals were the chosen speech acts. The oral performance was analysed for overall appropriateness, planning time and speech rate. A rating scale was used to assess appropriateness, while planning time was operationalised as the time taken to prepare for each role-play, and speech rate was assessed by the number of words spoken per minute. The results showed that for the L2 learners, PDI (High) tasks were more difficult to produce as seen from the lower scores awarded for appropriateness, with lower proficiency learners having more difficulty than higher proficiency learners. As for speech rate, the L2 learners spoke at a quicker rate in PDI (Low) tasks than in PDI (High) tasks, with the lower proficiency learners speaking slower than the higher proficiency ones in both cases. L2 learners also planned PDI (Low) tasks more quickly than PDI (High) tasks. The proficiency level, however, had no significant effect on planning time. This study showed that pragmatic variables could be useful criteria in distinguishing between tasks, and that when the pragmatic variables of power, distance and degree of imposition are manipulated, task demands can be either increased or decreased.

In a related study by Clark (2012), in which the test-takers were assessed in the L2 and their output analysed for accuracy, fluency and complexity so as to be more in line with previous studies conducted on task complexity, tasks with increased P, D, I were deemed to be more complex by the test-takers and though mixed results were found for fluency when performing PDI (High) tasks, gains were found in complexity

(lexical and structural). It appears that test-takers may perform better in more demanding tasks in terms of their linguistic complexity (see Robinson 2001).

Importance of Social Factors

The above discussion has highlighted the importance of investigating social factors and their influence on performance. First, very few studies have been carried out which investigate pragmatic features of task and their effect on production and studies conducted on pragmatic features suffer from being small in scale and having methodological limitations. Second, SLA and language pedagogy research findings concerning the effect of manipulating cognitive variables have not been replicated in a testing context. Therefore, there is a need to investigate the claims made by Fulcher and Reiter (2003), Taguchi (2007) and Clark (2012) that sociopragmatic features could be useful criteria in distinguishing between tasks, and that when the sociopragmatic variables of power, distance and degree of imposition are manipulated, task demands, and thus complexity, can be either increased or decreased. Finally, as Chalhoub-Deville (2001) points out, while the task has been the focal point of discussion and empirical investigation in second language literature for a few decades, the same cannot be said of the task used in performance testing and assessment.

Conclusion

In conclusion, the concept of task complexity, which has mainly been approached from a psycholinguistic/cognitive perspective (e.g. Skehan 1998; Robinson 2001, 2003, 2007, 2009), has been shown to be somewhat restrictive as it has failed to take into consideration the social context and how this affects learners' perception of task difficulty and the intrinsic demands of the task itself (i.e. task complexity) and how these affect L2 performance. The frameworks provided by Skehan (e.g. 1998) and by Robinson (e.g. 2007) do not provide a complete insight into task demands, or task complexity. It is necessary to add a social component to existing theory and to approach the construct of task complexity from a more socio-cognitive perspective.

When designing tasks for assessment purposes, it is essential that the task characteristics and performance conditions that determine task complexity are identified so that appropriate tasks are selected and sequenced in a principled manner. Task complexity does not reside in one characteristic of a task that can be manipulated to make it more or less complex, but resides in several components of tasks, which interact in ways which are yet not fully understood. The ability to select and sequence appropriate tasks also has implications for ensuring that parallel forms of testing are indeed equivalent in terms of task complexity. Ensuring test equivalency would provide a more reliable assessment of L2 ability and would improve the validity of the interpretations of test scores (Tavakoli 2009).

References

Bachman, L. F 1990, *Fundamental considerations in language testing*, Oxford University Press, Oxford.

Bachman, LF and Palmer, AS 1996, *Language testing in practice*, Oxford University Press, Oxford.

Blum-Kulka, S and House, J 1989, 'Cross-cultural and situational variation in requestive behavior in five languages', in S Blum-Kulka, J House, & G Kasper (eds), *Cross-cultural pragmatics,* NJ: Ablex, Norwood, pp. 123-154.

Brindley, G 1994, 'The appropriacy of psychometric measurement models for testing second language listening comprehension', *Language Testing*, vol.11, pp. 145-170.

Brown, P and Levinson, S 1987, *Politeness: some universals in language use*, Cambridge University Press, Cambridge.

Canale, M and Swain, M 1980, 'Theoretical bases of communicative approaches to second-language teaching and testing', *Applied Linguistics,* vol. 1, no. 1, pp. 1-47.

Chalhoub-Deville, M 2001, 'Task-based assessment: a link to second language instruction', in M Bygate, P Skehan, and M Swain (eds), *Researching pedagogic tasks: second language learning, teaching and testing,* Longman, Harlow, pp. 210-28.

Clark, V 2012, *The effect of manipulating sociopragmatic variables on task complexity and second language output,* Unpublished Ph.D. dissertation, University of Reading, England.

Dörnyei, Z 2009, 'Individual differences: interplay of learner characteristics and learning environment', *Language Learning*, vol. 59, pp. 230–248.

Elder , C, Iwashita, N and McNamara, T 2002, 'Estimating the difficulty of oral proficiency tasks: what does the test-taker have to offer?', *Language Testing,* vol. *19*, no. 4, pp. 347-368.

Elder, C and Iwashita N 2005, 'Planning for test performance: does it make a difference?', in R Ellis (ed.), *Planning and Task Performance in a Second Language*, John Benjamins, Amsterdam, pp. 219- 237.

Ellis, R 2003, *Task-based Language Learning and Teaching*, Oxford University Press, Oxford.

Ellis, R 2008, *Principles of instructed second language acquisition*, Center for Applied Linguistics, Washington.

Foster, P and Skehan, P 1996, 'The influence of planning and task type on second language performance', *Studies in Second Language Acquisition,* vol.18, pp. 299-323.

Fulcher, G and Reiter, RM 2003, 'Task Difficulty in Speaking Test', *Language Testing,* vol. 20, pp. 321-344.

Gilabert, R 2005, *Task complexity and L2 narrative oral production*, Unpublished Ph.D. dissertation, University of Barcelona, Spain.

Givon, T 1985, 'Function, structure, and language acquisition', in D Slobin (ed*.), The crosslinguistic study of language acquisition:* vol *1*, NJ: Lawrence Erlbaum, Hillsdale, pp. 1008-1025.

Givon, T 1995, *Functionalism and grammar,* Benjamins, Amsterdam.

Iwashita, N, McNamara, T and Elder, C 2001, 'Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design', *Language Learning,* vol. 51, no. 3, pp. 401-436.

Kuiken, F and Vedder, I 2007, 'Cognitive task complexity and linguistic performance in French L2 writing', in M García Mayo (Ed.), *Investigating tasks in formal language learning*, Clevedon, UK, Multilingual Matters, pp. 117–135.

Long, MH 1985, 'A role for instruction in second language acquisition: task-based language training', in K Hyltenstam and M Pienemann (eds), *Modelling and assessing second language acquisition* Clevedon, UK, Multilingual Matters, pp. 77-99.

Long, M. H 1989, 'Task, group, and task-group interaction', *University of Hawaii Working Papers in English as a Second Language*, vol. 8, pp. 1–26.

Long, MH and Norris, JM 2000, 'Task-based teaching and assessment', in M Byram (ed.), *Encyclopedia of language teaching,* Routledge, London, pp. 597-603.

Lumley,T and O'Sullivan, B 200, 'The impact of test taker characteristics on speaking test task performance, *Language Testing*, vol. 22, no. 4, pp. 415-437.

Mehnert, U 1998, 'The effects of different lengths of time for planning on second language performance', *Studies in Second Language Acquisition,* vol. *20*, pp. 52-83.

Mislevy, RJ, Steinberg, LS, Breyer, FJ, Almond, RG and Johnson, L 2002, 'Making sense of data from complex assessment', *Applied Measurement in Education*, vol. 15, pp. 363-378.

Norris, J, Brown, JD, Hudson, T and Yoshioka, J 1998, *Designing second language performance assessments* (Technical Report 18), University of Hawaii, Honolulu.

Ortega, L 1999, 'Planning and focus on form in L2 Oral Performanc', *Studies in Second Language Acquisition,* vol. 21, pp. 109-148.

O'Sullivan, B. 2000, 'Exploring gender and oral proficiency interview performance', *System,* vol. 28, no. 3, pp. 373-386.

O'Sullivan, B 2002, 'Learner acquaintanceship and oral proficiency test pair-task performance', *Language Testing*, vol. 19, no. 3, pp. 277–95.

Robinson, P 1995, 'Task complexity and second language narrative discourse', *Language Learning,* vol. 45, pp. 99-140.

Robinson, P 2001a, 'Task complexity, cognitive resources, and syllabus design: a triadic framework for examining task influences on SLA', in P Robinson (ed.), *Cognition and Second Language Instruction,* Cambridge University Press, Cambridge, pp. 287–318.

Robinson, P 2001b, 'Task complexity, task difficulty, and task production: Exploring interactions in a componential framework', *Applied Linguistics,* vol. 22, pp. 27–57.

Robinson, P 2003, 'Attention and memory during SLA', in *The handbook of second language acquisition*, in Catherine Doughty and Michael Long (Eds), Blackwell, Oxford, pp. 631–678.

Robinson, P 2005, 'Cognitive complexity and task sequencing: Studies in a componential framework for second language task design', *International Review of Applied Linguistics,* vol. 43, pp. 1–32.

Robinson, P 2007a, 'Task complexity, theory of mind, and intentional reasoning: effects on speech production, interaction, uptake of premodified input and perceptions of task difficulty', *International Review of Applied Linguistics in Language Teaching (IRAL),* vol 45, no. 3, pp. 195-215.

Robinson, P 2007b, 'Criteria for grading and sequencing pedagogic task', in MP García Mayo (ed.), *Investigating tasks in formal language learning,* Multilingual Matters, Clevedon, pp.7-27.

Robinson, P 2011, *Second language task complexity: researching the cognition hypothesis of language learning and performance*. John Benjamins, Philadelphia/Amsterdam.

Robinson, P, Ting, S and Urwin, J 1995, 'Investigating second language task complexity', *RELC Journal*, vol. 26, pp. 62-79.

Schmidt, R 2001, 'Attention', in P Robinson (ed.), *Cognition and second language instruction,* Cambridge, Cambridge University Press, pp. 3-32.

Skehan, P 1996, 'A framework for the implementation of task-based instruction', *Applied Linguistics*, vol. 17, no. 1, pp. 38-62.

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

Skehan, P 2001, 'Tasks and language performance assessment', in Bygate, M, Skehan, P & Swain, M (eds), *Researching pedagogic tasks,* Longman, London, pp. 167-185.

Skehan, P 2003, 'Task-based instruction', *Language Teaching,* vol. 36, pp. 1–14.

Skehan, P 2009, 'Modeling second language performance: Integrating complexity, accuracy, fluency, and lexi', *Applied Linguistics*, vol. 30, pp. 510–32.

Skehan P and Foster P 1997, 'The influence of planning and post-task activities on accuracy and complexity in task based learning', *Language Teaching Research,* vol. 1, no. 3, pp. 185-211.

Skehan, P and Foster, P 1999, 'The influence of task structure and processing conditions on narrative retellings', *Language Learning*, vo. 49, no. 1, pp. 93-120.

Skehan, P and Foster, P 2001, 'Cognition and tasks', in P Robinson (ed.), *Cognition and second language acquisition*. Cambridge University Press, Cambridge, pp. 183–205.

Swain, M 1985, 'Communicative competence: some roles of comprehensible input and comprehensible output in its development', in SM Gass & CG Madden (eds), *Input in second language acquisition,* Newbury House, Rowley, MA, pp. 235-253.

Taguchi, N 2007, 'Task difficulty in oral speech act production', *Applied Linguistics,* vol. 28, pp. 113-135.

Tavakoli, P 2009, 'Assessing L2 task performance: Understanding the effects of task design', *System,* vol. 37, no. 3, pp. 482-495.

Tavakoli, P and Skehan P 2005, 'Strategic planning, task structure and performance testing', *Planning and task performance in a second language*, in R Ellis (ed.), John Benjamins, Amsterdam, pp. 239-273.

Van Patten, B 1990, 'Attending to content and form in the input: an experiment in consciousness', *Studies in Second Language Acquisition,* vol. 12, pp. 287-301.

Van Patten, B 1994, 'Evaluating the role of consciousness in SLA: terms, linguistic features, and research methodology', *AILA Review,* vol. 11, pp. 27-36.

Widdowson, HG 1978, *Teaching language as communication*, Oxford University Press, Oxford.

Wigglesworth, G 1997, 'An investigation of planning time and proficiency level on oral test discourse', *Language Testing,* vol. 14, pp. 85-106.

Wigglesworth, G 2001, 'Influences on performance in task-based oral assessments', in M Bygate, P Skehan, and M Swain (eds), *Task based learning,* Longman, Addison Wesley.

Yuan, F and Ellis, R 2003, 'The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production', *Applied Linguistics*, vol. 24, pp. 1-27.

iafor

iafor