

Tasks and proficiency tests: piloting instruments of a study on strategic planning

ABSTRACT

This paper aims at presenting the results of a pilot study which investigated whether some instruments that will be used in a PhD research study are valid and reliable. Two narrative tasks that may serve to collect speech samples in two different conditions were tested in order to guarantee that they are similar in terms of difficulty level, and two proficiency tests – one focused on the listening skill and the other in the speaking skill – were piloted in order to know whether participants who perform both tests have similar scores. The results show that the oral tasks are similar, but one of them may yield learners to produce more complex language. Regarding the proficiency tests, half the participants had different scores in both tests. Moreover, the results also assisted in reflecting on the use of some instruments and in informing some selecting decisions.

KEYWORDS: Speech tasks. Proficiency tests. Piloting phase.

André Luís Specht

andreluisspecht@gmail.com

Universidade Estadual do Centro Oeste,
Itati, Paraná, Brasil.

Raquel Carolina Souza Ferraz

D'Ely

raqueldely@gmail.com

Universidade Federal de Santa Catarina,
Florianópolis, Brasil.

One major concern when conducting quantitative research is the design and/or the selection of instruments that tackle the variables that are being investigated and also that control other variables that may interfere in the research results (DÖRNYEI, 2007). The success of a study may depend on that, especially because data collection, the moment when the designed and/or selected instruments will be put to use, tends to be a process that involves a lot of effort, time, and a great number of people; therefore, if some instrument or technique fail, the entire endeavor may be lost, making it necessary to reassess everything and conduct a new data collection.

As an alternative to avoid such problem, researchers generally conduct a pilot study, that is a mini version of the complete study in which the procedures and instruments are implemented in order to verify them, thus opening the possibility to alter or improve them if necessary (BAILER et al, 2011, p. 130). Moreover, the piloting phase may also be an opportunity for researchers to inform their choice for the use of instruments that are pivotal for the study.

Considering the necessity of this testing movement, this article aims at presenting the results of a piloting phase of a study whose main objective is to investigate the impact of two types of instruction - integrated and isolated¹ - on learners' oral planned performance. Regardless the type of instruction, the idea is to optimize the moment learners have to plan their oral tasks. Within a task-based approach, this planning opportunity time provided to learners prior to task performance is a pre-task condition known as strategic planning (ELLIS, 2005, SKEHAN, 2018). Studies have shown a positive impact of strategic planning on oral performance, especially concerning fluency (the capacity to produce speech in real time); however, some researchers (D'ELY et al, 2019, ELLIS 2009, MEHRNET, 1998) have highlighted, based on mixed results studies have reported, that learners may not be using the pre-task condition effectively. In other words, they may not know how to plan or they may not be familiar with planning. Therefore, instructing them on how to plan may be a way of enhancing the planning condition.

In order to achieve the research objective, the following procedures will be taken: two groups of learners will receive instructional sessions on how to optimize the moment they have for planning their oral tasks. For the integrated group, the instructional sessions will occur during regular lessons, more specifically during speech activities proposed by the textbook they will be using; while for the isolated group, the instructional sessions will occur in one unique separated lesson. Aiming at measuring the impact of the instructions, learners will perform two narrative tasks (prior to and after the instructional sessions) with the opportunity to plan them. In this way, it will be possible to verify whether instruction on strategic planning may cause an impact on the planned performance.

The study execution counts with a great number of instruments that goes from a researcher's journal to interview's guides; however, in this piloting phase², not all the instruments were tested. In fact, only the narrative tasks were piloted in order to verify whether both tasks have the same level of difficulty, so that it will be able to control for task effect in the actual study, and two proficiency tests were applied to the students: a listening and a speaking one, so that we would be able to examine whether students obtain a similar level of proficiency in both

tests; in case they do, it would be less time consuming to use a listening test since it can be administered with all the students in the classroom simultaneously.

THE NARRATIVE TASKS

In general, studies on strategic planning have an experimental nature (ELLIS, 2005), and they basically consist in comparing unplanned and planned speech performance. There are two ways of conducting this speech comparison: between-subject and within-subject. In the former, the comparison is between two groups of participants. One group is the experimental group (i.e. the group that had time to plan), and the other group is the control group (i.e. the group that did not have time to plan). In the latter, the same group performs twice: in the first time with no time to plan (the control condition) and the second time with time to plan (the experimental condition).

Both comparisons have their own purposes, advantages and disadvantages. However, one important factor that must be considered is the task(s) that is (are) used to collect speech data samples for the study, especially for the studies that have a within-subject comparison nature, which require two different tasks. If the benefit of planning time is being investigated, using the same task for both conditions creates a variable that may interfere in the final results. One would not know whether the impact on the speech performance would be due to the opportunity to plan or to task repetition³. The ideal solution would be to use two different tasks, which at the same time should be similar in content and in the type of processing/procedures required from the participants, otherwise, it would not be possible to know whether the strategic planning condition affected the speech performance or whether students performed better in one task rather than the other because it was easier in terms of content, vocabulary they would have to use to narrate the story, so on and so forth.

In order to control for task difficulty, Specht and D'Ely (2017) used three tasks whose structures and topics were similar. Two of them were already used in previous studies on strategic planning, and one task was adapted from a video by them. All of them were picture-cued narrative tasks, containing 8 images with no writing, whose stories were about relationship. In Specht and D'Ely's case, they used three tasks because they had three different conditions: (1) no planning, (2) planning before instruction, and (3) planning after instruction and they had a within-subject comparison design; therefore, besides investigating the impact of strategic planning, they also attempted to understand the impact of instruction on how to plan. The researchers also questioned the participants in the post-task questionnaires they administered after the task performances, and the participants, in general, provided the same evaluation for all the tasks – half of the participants considered all the three tasks easy; while the other half considered them moderate. Nevertheless, Specht and D'Ely did not go further and analyze whether the participants produced similar outcomes in each task, which would mean that the level of difficulty of the tasks was similar. In fact, no study, as far as our knowledge goes, have focused their attention on investigating how similar the level of difficulty two tasks have.

Motivated by the lack of research on the topic, we decided to concentrate our attention on a further analysis in order to guarantee that the two of the tasks

used by Specht and D'Ely have similar levels of difficulty by comparing the performance of participants in both tasks and examining whether they are similar or not in terms of fluency (the capacity to produce speech in real time), complexity (the use of more elaborate language and structures) and accuracy (the ability to avoid mistakes and errors in performance).

THE PROFICIENCY TESTS

The participants' level of proficiency is an important variable to be controlled in studies within SLA, especially considering that it may bias the studies' final results. In this sense, research on strategic planning, which generally aims at investigating the impact of providing learners with the opportunity to plan their speech tasks prior to the actual performance, use proficiency tests to ensure the participants have an even proficiency level, otherwise it would not be possible to know whether the impact in the speech performance was due to the strategic planning condition or the discrepancy in the proficiency levels of the participants (D'ELY, 2006, p. 76).

Almost every research in strategic planning controls for proficiency level; nevertheless, there is not a standard or unique proficiency test used by researchers. As studies on strategic planning are mostly interested in speech production, the oral skill proficiency must be the focus. Considering that, D'Ely and Weissheimer (2004) designed a guide for raters to score speech samples, which was adapted from the First Certificate in English speaking test assessment scale (Cambridge Examination), the Iwashita, McNamara and Elder's scale (2001, apud D'ELY, 2006), and the Royal Society of Arts (RSA) test (in HUGHES, 1989 apud D'ELY, 2006). The candidates have to perform a narrative task orally, and then the product of this task is analyzed by raters. The final results of the raters' analysis is then compared and analyzed for interrater reliability. This proficiency test may be the most indicate if we are in a context where students have not been tested formally for their proficiency level; however, it is time consuming, considering the time spent collecting speech data and the time raters need to evaluate the samples.

Recently, Skehan (2014) published a book with seven studies on planning and speech performance, and one of them (WANG, 2014) used a version of the TOEFL listening subtest which was extracted from Hinkel (2004 apud WANG, 2014). Wang claims that the listening test is a great indicator of general English proficiency, and also that listening involves relatively similar processes to speaking. Administering a listening test in opposition to a speaking test is less time consuming, since all the participants can perform it at the same time and in the same room – an important advantage considering that many research settings might not have a laboratory at the disposal of all researchers in which all the students can record their speech samples simultaneously. Nevertheless, the question is whether a listening test would be in fact measuring oral proficiency. Taking into consideration this issue, students will perform both listening and speaking tests, and the results of both tests will be analyzed and compared in order to examine whether their measurements are similar.

PARTICIPANTS

As the target students that will be invited to participate in the actual study will be level 5 groups from Extracurricular/UFSC, this was the level chosen to pilot the instruments and to make the classes observations. Considering the Common European Framework of Reference for Languages, these students are supposed to have B1 level. We contacted two teachers that were teaching level 5 groups during the second semester of 2015, and they promptly allowed us to observe one of their classes and to invite students to perform the narrative tasks and to do the listening test. They also provided us with (part of) one lesson for the data collection. On the data collection day, only 10 students were presented in each group, and all of them accepted to take part in the study, signing a consent letter and filling in a profile background questionnaire.

In the profile questionnaire, students had to inform their ages, schooling level, and the skills they found the easiest and the most difficult considering English language. Students' age varied from 16 to 51, being 7 students below 20 years old, 7 between 20 and 30, and 4 above 30 years old. Regarding their schooling level, 1 student was in high school, 10 in college, 1 had an undergraduation degree, 2 were in a graduation program, and 6 had a graduation degree. The productive skills (speaking and writing) were reported as being the most difficult skill and the receptive skills (listening and speaking) the easiest ones for almost all the participants. Only two participants checked listening as the hardest skill for them. Interestingly, these participants had some peculiarities. One checked the other receptive skill (reading) as being the easiest one, and the other one was the youngest participants and the only one still in high school, and he checked speaking as being the easiest skill for him.

It is also worth mentioning that, as the time provided by one of the teachers was larger – she offered me a whole lesson - the students from her group were able to perform the narrative tasks and to do the listening test, while the other group only performed the narrative tasks, since the listening test would take extra 30 minutes to be done (see Table 1 for illustration).

Table 1 - Data collection

Group 1 – 10 students	5 students performed Task A and 5 students performed Task B All the students performed the listening test
Group 2 – 10 students	5 students performed Task A and 5 students performed Task B

Source: Own authorship

THE NARRATIVE TASKS

Students from both groups performed the narrative tasks: half of the group performed Task A and the other half performed Task B. Both tasks were used in Specht and D'Ely (2017) and their main topic was relationship. Task A tells a story of a man who shows at the house of his beloved with many presents, but he is always rejected by her; however in the end, she becomes jealous because he found another girl. Task B tells a story of a couple in a restaurant eating their

meal quietly; however the man is imagining cruel things to do to his wife, and he ends up throwing a pea at her and she becomes mad. The performance of the tasks occurred as follows: students had 50 seconds to assimilate the story, and after that, they had to retell the story without having access to it. The students' speech performance was recorded using SONY digital recorders and then transcribed for analysis purpose.

As in D'Ely (2011), the students' narratives were analyzed in terms of fluency (speech rate unpruned), complexity (number of subordinate clauses), and accuracy (number of errors per 100 words). The analysis results from tasks A and B were statistically compared using Mann-Whitney U (FIELD, 2009) in order to examine whether or not the performances of both tasks were similar, so that it could be claimed that students performed both tasks similarly.

THE PROFICIENCY TESTS

The listening test

The listening test was a TOEFL sample test⁴, which contained 34 questions. The maximum score was 36, considering that two questions valued 2 points. The obtained score, then, was contrasted with an official TOEFL table in order to know the listening level of proficiency. This test only measured three levels: low intermediate, intermediate, and high intermediate.

The speaking test

For the speaking test, the students' speech performances in the narrative tasks were used. The audio narratives were sent to two raters who evaluated them with a score based in a guide from D'Ely (2006). The scores went from 0 to 5 with 0,5 intervals, being 0 considered a basic level of proficiency, 3 intermediate, and 5 advanced. In order to know whether both raters were consistent in their evaluation, a Cronbach's Alpha test was run to check interrater-reliability. As their evaluation was reliable, as it can be seen in Table 2 by a high correlation number, the mean of both scores was used as a final score.

Table 2 - Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,900	,963	2

Source: Own authorship

The comparison

As the scores of both tests were different in nature and also taking into account that the listening test only evaluated an intermediate range level of

proficiency, while the speaking test also evaluated levels lower and higher than intermediate, a score adjustment was necessary, so that the results could be compared statistically. Considering the levels of proficiency dealt by both tests, it was possible to claim that 5 levels of proficiency are dealt with: basic – low intermediate – intermediate – high intermediate – advanced. Therefore, the following table was used:

Table 3 - Score adjustment for statistical comparison

Old scores – Listening test	Old scores – Speaking test	New Scores
	0 – 1,4 = basic	1
Low intermediate	1,5 – 2.4 = low intermediate	2
Intermediete	2,5 – 3,4 = intermediate	3
High intermediate	3,5 – 4,4 = high intermediate	4
	4,5 – 5 = advanced	5

Source: Own authorship

Just for sake of illustration, if a participant was classified with a low intermediate level of proficiency in the listening test and scored 4,1 in the speaking test, now the listening test score is 2 and the speaking test score is 4. After converting the scores based on the table, a Spearman’s correlation coefficient test (FIELD, 2009) was run in order to examine whether both tests correlated and measured the same level of proficiency.

RESULTS AND DISCUSSION

The narrative tasks

In order to check whether Tasks A and B yield similar performances, which might indicate that both of them are similar in terms of cognitive demands from task takers, the participants’ performance was transcribed and analyzed in terms of fluency, complexity and accuracy, as it was explained in the Method section. Table 4 presents the mean numbers of the groups’ performance in Task A (group 1) and Task B (group 2), and it can be observed that the performances in terms of fluency and accuracy are more similar than they are in terms of complexity, since Group 2 produced more complex sentences.

Table 4 - Descriptive Statistics

	Groups	N	Mean
Fluency	1	10	11,80
	2	10	9,20
Complexity	1	10	7,65
	2	10	13,35
Accuracy	1	10	9,15
	2	10	11,85

Source: Own authorship

Table 5 presents the results of the Mann-Whitney Test, and it confirms that in fact the groups’ performance was similar regarding fluency and accuracy since

the differences between the results were not significant ($p=0.362$ and $p=0.307$, respectively), and it was different in terms of complexity, considering that the difference between the results was significant ($p=0.03$).

Table 5 - Mann-Whitney Test

	Fluency	Complexity	Accuracy
Mann-Whitney U	37,000	21,500	36,500
Wilcoxon W	92,000	76,500	91,500
Z	-,983	-2,171	-1,022
Asymp. Sig. (2-tailed)	,362	,030	,307

Source: Own authorship

These results may indicate that the performances in Tasks A and B have some similarities, meaning that both groups had similar rates of speech production and produced a similar amount of errors per clauses in both tasks. The difference, though, falls into the number of subordinations produced by group 2, which is larger than the one produced by group 1, which means that their speech is more complex.

One explanation could be related to the nature of the plot. In Task B, participants had to narrate events that were imagined by one of the characters, which do not happen in Task A. This would facilitate the use of more subordination as it could be seen in the following instances: P1TB⁵ – He is thinking about a lot of things to do to hurt her / P2TB⁶ – He is imagining they are fighting. Another explanation could be that, differently from what is advocated in this article, the participants' level of proficiency was not controlled; therefore, through the narrative transcriptions, it was possible to see that some participants from group B had a higher level of proficiency, which may also explain the reason their speech had more subordinations.

Such result reinforces the importance of paying close attention to the choice of tasks as instruments for data collection, especially in strategic planning studies. Even though we were dealing with tasks that were similar in several perspectives such as task structure and task topic, as carefully controlled by Specht and D'Ely (2017), other differences may interfere in the task outcome. Such differences may compromise the results coming from task comparisons. In the case of Specht and D'Ely (2017), specifically, the difference found in this study, were not a problem, because the researchers did not investigate complexity. However, studies that adopt more than one task and intend to investigate several speech dimensions should be attentive to that.

The proficiency tests

As it was aforementioned, the aim of comparing the scores of the two proficiency tests (listening and speaking) is to verify whether both tests measure similar levels of proficiency. From Table 6, it is possible to observe that the correlation coefficient number is not very high ($r=,566$) or significant ($p=,088$), which indicate a not very strong correlation between the two tests. Therefore, it is not possible to assure that both tests measure the same level of proficiency, that is, that they are tapping the same linguistic competence.

Table 6 - Correlations

Spearman's rho	Listening	Correlation Coefficient	Listening	Speaking
		Sig. (2-tailed)	1,000	,566
	N	10	10	
	Speaking	Correlation Coefficient	,566	1,000
		Sig. (2-tailed)	,088	
	N	10	10	

Source: Own authorship

The moderate correlation between the two tests may be further understood if one looks at the individual scores of the participants in Table 7. Half of the participants' scores are the same in both tests; while the other half scored higher or lower in one of the tests. The difference in the performance of the two tests may have some explanations.

Table 7 - Participants' scores

Participants	Listening test scores	Speaking test scores
P1	3	3
P2	4	4
P3	2	4
P4	2	3
P5	2	2
P6	4	3
P7	2	2
P8	2	2
P9	2	3
P10	4	4

Source: Own authorship

One explanation would be the nature of the tests: one is testing a receptive skill and the other a productive one. It may have an impact in the way students face each test. It was possible to observe that some students were more uncomfortable and nervous performing the speaking test, in which they had to narrate a story under a certain amount of pressure. They only had 50 seconds to assimilate the story, and in addition, they did not have access to the story while retelling it. The same could not be observed in the listening test. Even though students had some pressure regarding the amount of time they had to answer the questions based on the audio they had listened, if they did not have time to answer all the answers, they could skip them. Moreover, it is worth recalling that participants reported that the receptive skills were the easiest ones, which corroborate the explanation.

Another explanation could be related to the speaking test execution. As we only had four voice recorders, we had to separate students in groups, and the students that would not record the story were asked to wait outside of the classroom. We arranged four desks, each one on each of the corners of the classroom, and they all had to retell the stories simultaneously. They were also advised to speak in a lower voice, so that they would not disturb one another. Notwithstanding all the effort, some students reported that they had trouble

concentrating due to the other students' speaking, which can also influence their speaking test performance.

All in all, for this study's purpose and in order not to take risks, it can be concluded that the listening test is not a good indicator of speaking proficiency, and as we are dealing with oral proficiency, it would be more advisable to make use of a speaking test to guarantee the target participants have the same level of proficiency, thus, controlling this variable. It is important to highlight that further investigation should be conducted on the topic, since this study had limitations such as number of participants and problems in data collection. Wang (2014) pointed out that a listening test is a good indicator of proficiency in general, being some speaking processes similar to the speaking ones; however, this study was not able to contribute on that.

CONCLUSION

To conclude, it is possible to reinforce that the piloting phase - designing, selecting and piloting research's instruments (or at least the more central ones) - is an important step in a research study. This piloting phase assisted us in: (a) selecting a pair of tasks, which will be used to collect our participants' performance in two different moments: prior to and after the instructional treatment; (b) informing our decision in the use of a proficiency test, which will be used to control our participants' level of proficiency; and (c) being acquainted with the context we will collect our data.

Concerning the participants' performance in the tasks, it was possible to observe that the tasks were similar in terms of fluency and accuracy, but not in terms of complexity. Based on that, it would be advisable to use a different task to compose the pair of tasks in a research study. Preferentially, one in which students would not have to describe what the character(s) is(are) thinking. Another option would be to use the task, but take this issue in consideration when analyzing the results of the study.

Considering the proficiency test, it was not possible to assure that the listening test was a good indicator of speaking proficiency. The results showed that both tests did not correlate strongly, with half of the participants presenting different levels of proficiency in the tests. In our opinion, when the research study deals with oral performance, the safest option is to use the speaking test. The use of the listening test was cogitated because it could be administered with the entire group once, considering that most institutions do not have a language laboratory where various students can record speech samples at the same time. Another solution would be the use of the students' performance in the first narrative task (the one before instruction) also as data for raters' assessment, thus eliminating an extra step. As students will be recording narratives during class time, the less they are taken from the regular class flux the best. It would only be necessary to inform the raters that they would be evaluating planned performances.

All in all, this study do not have the intention to make any generalization on what concerns differences between tasks and proficiency tests, since there are a range of options out there. It intends to shed some light on the choice we will make in a larger-scaled study and to provide a reflection on the importance

of making informed decisions on studies that take a quantitative stance towards data, reinforcing the value of piloting our instruments.

Tarefas e testes de proficiência: instrumentos piloto de um estudo sobre planejamento estratégico

RESUMO

O objetivo desse artigo é apresentar resultados advindos de um estudo piloto que buscou entender se alguns instrumentos que serão utilizados em uma pesquisa de doutoramento são válidos e confiáveis. Duas tarefas narrativas que servirão para coletar amostras de fala em dois momentos diferentes foram testadas com o intuito de garantir que elas tenham níveis de dificuldade similares e dois testes de proficiência - um focado na habilidade auditiva e outro na habilidade oral - foram pilotados para saber se os participantes que fizeram os dois testes apresentam avaliações similares em ambos. Os resultados mostraram que as tarefas orais são parecidas, porém uma delas pode levar o aluno a produzir frases mais complexas. No que diz respeito aos testes de proficiência, metade dos participantes teve avaliações bem diferentes nos testes auditivos e orais. No mais, os resultados também ajudaram na reflexão do uso de alguns instrumentos e na escolha informada de outros.

PALAVRAS-CHAVE: Tarefas orais. Testes de proficiência. Estudo piloto.

NOTAS

¹ There are two major types of strategy instruction; it can be integrated in the language teaching materials or it can be a separated program (ELLIS, 1998, p. 558).

² As this study is not a mini version of the complete study as defined by Bailer et al (2011), but the testing of some instruments that will be used in the study, the term 'pilot study' will be substituted by 'piloting phase' in this article.

³ Task repetition is a task condition in which students perform the same task twice. Repeating the same task may impact learners' speech performance, being the first performance considered a rehearsal of the task (BYGATE, 2001 and D'ELY; MOTA; BYGATE, 2019 are examples of studies that deal with this variable).

⁴ The test sample chosen was not the same used by Wang (2014), because we could not have access to the book; therefore, another TOEFL listening test sample was used.

⁵ P1TB stands for Participant 1 who performed Task B.

⁶ P2TB stands for Participant 2 who performed Task B.

REFERENCES

BAILER, Cyntia; TOMITCH, Leda; D'ELY, Raquel. Planejamento como processo dinâmico: a importância do estudo piloto para uma pesquisa experimental em linguística aplicada. **Revista Intercâmbio**, v. XXIV, p. 129-146, 2011.

BYGATE, Martin. Effects of task repetition on the structure and control of oral language. In: BYGATE, Martin; SKEHAN, Peter; SWAIN, Merrill (Ed.). **Researching pedagogic tasks: Second language learning and testing**. London: Longman, 2001.

D'ELY, Raquel. **A Focus on learners' metacognitive processes: the impact of strategic planning, repetition, strategic planning plus repetition, and strategic planning for repetition on L2 oral performance**. Tese (Doutorado), Universidade Federal de Santa Catarina, Florianópolis, 2006.

D'ELY, Raquel. The impact of familiarity with strategic planning and teacher-led planning on learners' L2 oral performance of focused and unfocused tasks. In: TUMOLO, Celso; FIGUEIREDO, Débora; MORITZ, Maria Ester; D'ELY, Raquel. (Org.). **Contextualized Practices in EFL Teaching and Assessment**. Florianópolis: PGI-UFSC, 2011.

D'ELY, Raquel; MOTA, Mailce; BYGATE, Martin. Strategic planning and repetition as metacognitive processes in task performance: Implications for EFL learners' speech production. In: WEN, Zhisheng; AHMADIAN, Mohammad Javad (Ed.). **Researching L2 Task Performance and Pedagogy: In honour of Peter Skehan**. John Benjamins, 2019. p. 199-228.

D'Ely, R. & Weissheimer, J. **Scale of L2 oral proficiency**. Unpublished research paper. Universidade Federal de Santa Catarina, Florianópolis, 2004.

DÖRNYEI, Zoltan. **Research Methods in Applied Linguistics**. Oxford University Press, 2007.

ELLIS, Rod. **The study of second language acquisition**. Oxford University Press, 1998.

ELLIS, Rod. Planning and task-based performance: Theory and research. In: ELLIS, Rod (Ed.). **Planning and task performance in a second language**. Filadélfia: John Benjamins, 2005. p. 3-34.

ELLIS, Rod. **The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production**. *Applied Linguistics*, v. 30, p.475-509, 2009.

FIELD, Andy. **Discovering statistics using SPSS**. London: Sage Publications, 2009.

LOCKE, Lawrence; SILVERMAN, Stephen; SPIRDUSO, Waneen. **Reading and understanding research**. Thousand Oaks, CA: Sage Publications, 1998.

MEHNERT, Uta. The effects of different lengths of time for planning on second language performance. **Studies in Second Language Acquisition**, v. 20, p.52-83, 1998.

SKEHAN, Peter. **Processing Perspectives on Task Performance**. London: John Benjamins, 2014.

SKEHAN, Peter. **Second Language Task-Based Performance: Theory, Research, Assessment**. New York: Routledge, 2018.

SPECHT, André; D'ELY, Raquel. Planning oral narrative tasks: optimizing strategic planning condition through strategy instruction. **Acta Scientiarum. Language and Culture**, v. 39, p. 203-212, 2017.

WANG, Zhan. On-line time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In: SKEHAN, Peter (Ed.). **Processing Perspectives on Task Performance**. London: John Benjamins, 2014, p. 27-62.

Recebido: 13 fev. 2019

Aprovado: 03 set. 2019

DOI: 10.3895/rl.v21n34.9579

Como citar: SPECHT, André Luís; D'ELY, Raquel Carolina Souza Ferraz. Tasks and proficiency tests: piloting instruments of a study on strategic planning. *R. Letras*, Curitiba, v. 21, n. 34 p. 170-184, set. 2019. Disponível em: <<https://periodicos.utfpr.edu.br/rl>>. Acesso em: XXX.

Direito autorial: Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.

