ADEQUACY IN A STRATEGIC PLANNING STUDY: SCRUTINIZING THE USE OF AN ADDITIONAL SPEECH DIMENSION

A ADEQUAÇÃO EM UM ESTUDO DE PLANEJAMENTO ESTRATÉGICO: INVESTIGANDO O USO DE UMA DIMENSÃO DE FALA ADICIONAL

André Luís Specht[1]

Unicentro

Abstract: This article presents the results of a study that aimed at scrutinizing adequacy adopted as a speech dimension and measure in Specht and D'Ely (2020)'s study. Adequacy is a dimension and a measure little explored in studies that analyze speech performance from a cognitive perspective. In the case of Specht and D'Ely, they decided to include it to tackle more discourse-oriented features of speech performance. The composition of the measure was the sum of raters' scores to five criteria: structure, appeal, clarity, lexical choice and fluency. However, it was not analyzed whether all criteria had an impact on participants' speech performances individually. For that, statistical analyses were run (a) to examine whether adequacy may be considered a separate speech dimension different from other speech dimensions used in the study, complexity, accuracy and fluency (CAF) and (b) to understand whether all criteria had an active role in the measure and if not which one(s) did. The results have shown that adequacy may be considered a specific speech dimension, and that every criteria, except for fluency, had an active role in the measure, being textual organization (structure) a more salient criterion.
Keywords: adequacy; strategic planning; speech performance; CAF.

Resumo: Neste artigo, apresentamos os resultados de uma pesquisa que teve como objetivo analisar a dimensão de fala adequação adotada no estudo de Specht e D'Ely (2020). No estudo citado, adequação foi conceptualizada para medir características da produção oral com um foco mais discursivo. A medida consistia na soma e média das notas dadas por avaliadores em cinco critérios: estrutura, apelo, clareza, escolha lexical e fluência (CAF). Entretanto, não foi analisado se todos esses critérios tiveram um efeito individual na produção oral dos participantes. Para isso, análises estatísticas foram rodadas para (a) examinar se a adequação poderia ser considerada uma dimensão de fala, diferente das outras dimensões utilizadas no estudo: complexidade, acurácia e fluência e (b) entender se todos os critérios tiveram um papel ativo na medida e se não qual(is) teve/tiveram. Os resultados mostraram que a adequação pode ser considerada uma dimensão de fala específica e que todos os critérios, exceto pela fluência, tiveram um papel ativo na medida, sendo a organização textual (estrutura) o critério mais saliente.
Palavras-chave: adequação; planejamento estratégico, produção oral, CAF.

---

[1] Possui Licenciatura em Letras - Inglês e Literatura Correspondente pela UNICENTRO (2007); Mestrado em Letras/Inglês e Literatura Correspondente; linha de pesquisa: Ensino e Aprendizagem, pelo Programa de Pós Graduação em Inglês da Universidade Federal de Santa Catarina, UFSC (2014); Doutorado em Estudos da Linguagem; linha de pesquisa: Ensino e Aprendizagem, pelo mesmo programa (2017). Atualmente é professor adjunto da Universidade Estadual do Centro Oeste, UNICENTRO, lotado no Departamento de Letras, DELET, no campus de Irati. Também faz parte dos grupos de pesquisas AELE: Aprendizagem e Ensino de Língua Estrangeira (UNICENTRO), LabLingue: Laboratório de Psicologia do Bilinguismo (UNICENTRO), Estudos Literários: teoria, crítica e ensino (UNICENTRO) e AQUILES: Aquisição de Inglês como Língua Estrangeira: Questões teóricas, pedagógicas e de metodologia de pesquisa (UFSC). E-mail: alspecht@unicentro.br

Introduction

Studies on strategic planning have conceptualized speech performance as a multifaceted phenomenon, which, for the most part, has been analyzed in terms of complexity, accuracy, and fluency (also known as CAF), as proposed by Skehan (1998). In order to assess each dimension, different measures are operationalized and adopted. For instance, degree of subordination is a measure used to assess complexity; the percentage of error-free clauses is a measure used to assess accuracy; and the number of pauses is a measure used to assess fluency. Such dimensions and measures reinforce comparisons between studies and have proven to be successful in tackling speech performance (SKEHAN, 2014).

Notwithstanding the adoption of CAF measures to assess speech performance, a great deal of discussion on its reliability has been held in the area (HOUSEN; KUIKEN, 2009, LARSEN-FREEMAN, 2009, NORRIS; ORTEGA, 2009, SKEHAN, 2009). Pallotti (2009), for instance, questioned whether these measures effectively represent a proficient performance. According to the author, a performance may be accurate, fluent, and complex, but, at same time, may not be adequate communicatively. The author, then, proposed the use of adequacy as an extra dimension, which would account for whether the performance is adequate regarding the communicative goals of a task.

With the intention of adopting and proceduralizing adequacy along with CAF, Specht and D'Ely (2020), which analyzed the impact that two types of strategy instruction on how to plan has on Brazilian learners' oral planned performance, have shown the construct to be an effective dimension. Adequacy, in their study, was measured by the average scores provided by three raters who listened to the participants' narratives and rated them following some criteria. The participants after strategy instruction in general produced speech performance that was more adequate communicatively in terms of textual organization, story appeal and clarity, lexical choices, and speaking flow. In total, there were five subjective criteria that were tackled by the raters, which, in the end, consisted as parts of a single speech performance measure.

Even though the measure was considered reliable and managed to deepen Specht

and D'Ely's speech performance analysis, bringing an important feature of task speech performance, each criterion assessed by the raters has specific nature. Textual organization has to do with how the participants organize and deliver the information in the text, while story appeal has to do with how appealing the story was to the interlocutor, for instance. With that in mind, this article aims at scrutinizing the adoption of adequacy as an additional dimension in their study.

For organizational purposes, this article has 4 sections apart from this introduction. In the first section, we present a brief discussion on the use of CAF and adequacy, and revisit Specht and D'Ely's study. In the second section, we explain the method adopted to analyze the adequacy measure in more details. In the third section, we present the results and an analysis. And, finally, in the fourth section, we close the article with our final remarks.

1. Review of Literature
1.1. Is CAF enough?

Even though measures that assess complexity, accuracy and fluency in task-based studies have proven to be useful measures of second language performance, according to Skehan (2009), it is undeniable that "(…) the sole use of CAF indices to assess task-based performance is in contrast to the practices of the teaching and testing fields, where the extent to which classroom learners or test-takers have the abilities to function successfully in real-life settings has been given considerable weight" (RÉVÉSZ; EKIERT; TORGERSEN, 2016, p. 830).

This incongruity, however, does not seem to be the greatest issue, which may be partly explained by the fact that tasks used in research and in the classroom are different due to a more controlled nature required by research (FOSTER, 2009). The most problematic issue relies on the fact that the exclusive use of CAF measures may not be sufficient to estimate a successful performance (DE JONG et al., 2012). Pallotti (2009) explains that a performance may be fluent, complex and accurate, but it may not be adequate communicatively. He demonstrates this incoherence with the following example:

> If in an information gap task a learner were to utter unhesitatingly colorless green ideas sleep furiously on the justification where phonemes like to plead vessels for

> diminishing our temperature, her production would score extremely high on CAF, in spite of being completely irrelevant, and probably counterproductive, for task success. In contrast, an utterance such as No put green thing near bottle. Put under table is neither complex nor accurate, and may not be fluent either, but can turn out to be perfectly functional for achieving the speaker's (and the task's) intended communicative goal. (PALLOTTI, 2009, p. 596)

Based on that, Pallotti (2009) proposed the inclusion of a new dimension: adequacy, which may be employed along CAF measures as an extra speech dimension or as a way to interpret CAF measures themselves. Regarding the former, Pallotti suggests the use of qualitative rating forms, in which raters would evaluate the performance through predefined descriptor scales. As an interpreter of CAF measures, adequacy could be used to examine whether the measures adopted in fact reflect their real purposes within a communicative plan.

Even timidly, some studies on adequacy have been conducted, in the attempt to investigate its relation to other speech dimensions; that is, to what extent adequacy may interact with CAF measures. Iwashita et al. (2008), for instance, found that speech rate had a strong impact on speaking proficiency – the measures they used for adequacy -, while grammatical accuracy and unfilled pauses had a moderate effect. Kuiken et al (2010), unlike Iwashita et al. (2008), assessed adequacy by a six-point scale, which measured the writer's ability to fulfill the communicative goal of the task and the impact of the resultant text on the reader. The authors found correlation between adequacy and syntactic complexity, but no correlation was found with accuracy.

In an attempt to extend Kuiken et al.'s study to oral production, Révész, Ekiert and Torgersen (2016) examined the relationship between adequacy and CAF measures. In addition, they also investigated this relationship considering the level of proficiency and task types. For that, they had 100 participants (80 ESL learners divided equally in four proficiency groups: low intermediate, intermediate, low advanced and advanced, and 20 native speakers). Each participant performed 5 tasks, resulting in a total of 500 performances. The 5 tasks involved: (1) a complaint, (2) a refusal, (3) a narrative, (4) advice, and (5) a summary. Adequacy was assessed by 20 raters (10 PhD students and 10 native speakers). Each of them evaluated 50 performances based on a rating scale, which contained information related to message delivery and specific points related to the nature of each task.

Regarding CAF measures, popular ones were adopted, such as index of subordination, number of words, errors per 100 words, frequency of self-repair, to cite but a few. The results showed that breakdown fluency measures were the strongest predictor of adequacy; while the other measures had a small effect. The results also showed that a lower incidence of false starts was related with higher adequacy for the advanced participants. In addition, no significant interaction was found among the different types of tasks.

1.2. Revisiting Specht & D'Ely (2020)

Specht and D'Ely (2020) are the only study on strategic planning that employed adequacy as an independent variable, at the best of our knowledge. Their main objective was to investigate the impact of two types of strategy instruction on learners' oral planned performance. They had two groups of participants that received different types of treatment and produced two oral narrative tasks each with the opportunity to plan them before and after treatment. One group, named integrated group, composed by 11 participants, received strategy instruction within their regular English classes. The strategy instruction sessions were administered within a period of 3 weeks prior to the lesson oral activities. The other group, named isolated group, composed by 12 participants, received strategy instruction separated from their regular English lessons, they received an entire lesson on strategy instruction with strategy presentation and practice. In addition, they also had a control group, composed of 11 participants, that produced two oral narrative tasks with the opportunity to plan without any type of treatment whatsoever. In spite of the group, all participants performed the tasks on the same days. The participants were part of regular intermediate level English learning groups from the Extracurricular language program at UFSC. For the study, all the students took a proficiency test in order to examine their individual proficiency, and only the data from those who presented an intermediate score were used.

The participants' oral performance in the tasks was analyzed in terms of complexity, accuracy, fluency and adequacy, and the results of the analysis were compared within and between groups in order to verify the impact of strategy instruction. Two measures were adopted to grasp complexity: (i) degree of subordination and (ii) number of words per AS-Unit; three for accuracy: (i) percentage of error-free AS-Units,

(ii) number of errors per 100 words and (iii) average of errors per AS-Units; five measures for fluency: (i and ii) speech rate pruned and unpruned, (iii and iv) number of middle and boundary clause unfilled pauses, and (v) number of self-repairs; and one for adequacy, raters' assessment on speech performance based on a scoring table. For the statistical comparison, Mixed ANOVA tests were run in order to verify whether the differences in performance were significant or not.

Results have shown that the groups that received treatment improved their performance in terms of accuracy and adequacy. The isolated group presented significant differences in both dimensions, while the integrated group improved only in terms of adequacy, which indicates an impact of strategy instruction on oral planned performance. Such differences were explained by the authors based on the following premises: (a) guiding students' attention on how to plan may assist them in producing more accurate language (FOSTER; SKEHAN, 1999); (b) isolated type of strategy instruction highlights the strategies and their use, causing major impact on the group that received it. Moreover, the authors also pointed out that

> It is worth highlighting, though, that the impact of the integrated instruction was only possible to be identified by the adoption of adequacy as an extra measure to assess speech performance. If only complexity, accuracy, and fluency measures had been used, no difference between the integrated and the control groups would have been found. This would have been considered a lack of efficiency for the integrated instruction, and by all means it was not the case (SPECHT; D'ELY, 2020, p. 55).

All in all, the study managed to show the importance of pedagogical guidance through strategy instruction for oral planning purposes, since the groups the received treatment improved their planned oral performance and the control group did not. It is also important to emphasize that adequacy played an important role in their study as an extra speech performance dimension.

2. Method

This article, which aims at scrutinizing the adoption of adequacy as an extra dimension in Specht and D'Ely (2020), intends to answer the following research question: How does adequacy function as a dimension to assess speech performance? In order to

do so, this study intends to take a quantitative and qualitative approach towards data that were not examined by Specht and D'Ely.

2.1. How adequacy was operationalized

In order to reach a more discourse-oriented stance of speech production, adequacy was adopted as a fourth dimension in Specht and D'Ely (2020). Pallotti (2009, p. 596) explains that "(…) adequacy represents the degree to which a learners' performance is more or less successful in achieving the task's goals efficiently". The author claims that high scores on complexity, accuracy, and fluency may not prove that learners have produced adequate outcomes. A narrative, for instance, may have few errors, few pauses and several subordinate clauses and at the same time may be confused and poorly organized.

Pallotti (2009) did not present any concrete measure or framework to assess adequacy, which is understandable, considering that adequacy itself is contextual, that it, it depends on the task's characteristics and goals; however, he suggested that it "(…) can be evaluated by means of qualitative rating, using predefined descriptors scales (…) (p. 597). As the tasks used in Specht and D'Ely (2020) were narrative tasks, and it was expected that students told a story based on the set of pictures they were presented, features of an 'adequate' narrative were considered, and a table was designed. The table contained five statements: (1) The story is well organized - It has beginning, middle and end; (2) The story is interesting – It catches my attention; (3) The lexical choices used by the narrator are understandable and compatible to the story; (4) the story is clear – It is easy to understand; and (5) The rhythm and speed the narrator tells the story is good. Each statement was followed by a scale of scores that went from 1 to 5, being 1 very poor, 2 poor, 3 regular, 4 good, and 5 very good. Each story, in the end, had one final score that was the sum of the scores given to each statement; therefore, the minimum score could be 5 and the maximum score could be 25.

The table was piloted in order to examine whether it was clear and did not hinder comprehensibility while being used. Two raters evaluated some stories using the table, and they complained about not having a basis to evaluate the story. The raters suggested that a training rating session would solve the problem. Based on that, three narrative

stories with different adequacy levels were selected from the pilot study (SPECHT; D'ELY, 2019), to be presented to the raters and discussed along with the researcher as a training session.

The 68 stories (two stories produced by 34 participants) were listened and evaluated by three Brazilian raters – two PhD students and one master student with extensive teaching experience. A Cronbach's Alpha test was run in order to check whether the raters' final scores correlated; that is, whether they followed a similar evaluation pattern, and, as it can be seen in Table 1, a relatively high correlation number, close to 1 was obtained (0.786). This shows that raters evaluated the tests similarly, which is quite intriguing, especially considering that some statements in the evaluating table could be quite subjective.

Table 1. Inter-Rater Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,786	,783	3

Source: Specht (2017)

In Specht and D'Ely (2020), the means of the final scores given by the three raters were used for statistical purposes and represented the adequacy measure. In this study, we analyzed the means of the individual scores given by the raters for each criterion to understand which one(s) had more impact and more characterize the adequacy measure. In other words, we intended to examine whether an adequate speech performance in Specht and D'Ely all the qualities assessed in the evaluating score and if not which one(s) highlighted. For the quantitative analysis, Mixed ANOVA was run for each individual score in order to examine which criteria was most affected. The criteria were labeled as following: (1) structure, (2) appeal, (3) vocabulary, (4) clarity, and (5) fluency. Prior to that analysis, a Principal Component Factor Analysis was run in order to confirm that adequacy was in fact a separate dimension, different from CAF. For the qualitative analysis, the narrative transcriptions were used to compare and illustrate the results.

3. Results

3.1. Principal Component Factor Analysis

As already explained, Specht and D'Ely (2020) considered four speech performance dimensions: complexity, accuracy, fluency, and adequacy. For each dimension, two or more measures, except for adequacy, were adopted in order to tackle different aspects of the respectives dimensions. However, in order to verify whether the measures are indeed representative of speech performance and load on different components, a Principal Component Analysis was carried out. For the analysis, the performance of the first task produced by the three groups under the same condition - planning without instructional treatment - was used.

For conducting the analysis, two assumptions should be met: suitability and sphericity of data. Regarding suitability, Kaiser (1974) suggests that in order to conduct a Principal Component Analysis, the result of the KMO test should be at least 0.6. This assumption was not met, since the result of the test was 0.56 as can be seen in Table 2. However, the result of Bartlett's Test was significant, meeting the assumptions for sphericity. Even though the condition for the use of this statistical test was not entirely advisable, we opted to employ it, because it is our understanding that such a test may provide some type of evidence that adequacy can be seen as a fourth dimension.

Table 2. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,561
Bartlett's Test of Sphericity	Approx. Chi-Square	392,002
	DF	55
	Sig.	,000

Source: Specht (2017)

Considering that all the eleven measures were used to assess speech performance, some level of correlation between at least one measure pair is expected; otherwise the measure that does not present any correlation would be measuring a different

phenomenon, which would not be speech performance. Field (2009) suggests that the correlation should be higher than 0.3. At the same time, there should not be a strong correlation between any measure pairs, which would mean that the two measures are assessing alike aspects. As it can be seen in Table 3, all the measures minimally correlate with other(s) measure(s), since they present correlation numbers higher than 0.3. Therefore, it means that all the variables are measuring the same phenomenon. On the other hand, some pairwise correlations, A1/A2, A1/A3, A2/A3 and F1/F2, presented a strong correlation (higher than 0.8). This may indicate that they are measuring similar aspects of speech performance as it was explained. However, in such cases, this strong correlation could be expected since they are assessing related aspects within the same dimension, for instance, both F1 and F2 measure speech rate. The difference between them was that the former considers self-repairs and repetitions, while the latter did not. If this is the case, the statistical results of these variables shall be similar.

Table 3. Correlation Matrix

	C1	C2	A1	A2	A3	F1	F2	F3	F4	F5	Ad1
C1	1,000	,639	-,217	-,005	,255	,068	,022	-,270	-,119	,387	-,108
C2	,639	1,000	-,352	,079	,503	,055	-,010	-,432	-,028	,641	-,122
A1	-,217	-,352	1,000	-,842	-,841	,291	,361	,452	-,422	-,471	,388
A2	-,005	,079	-,842	1,000	,860	-,235	-,295	-,330	,329	,361	-,337
A3	,255	,503	-,841	,860	1,000	-,231	-,309	-,476	,251	,647	-,352
F1	,068	,055	,291	-,235	-,231	1,000	,982	-,161	-,601	-,192	,312
F2	,022	-,010	,361	-,295	-,309	,982	1,000	-,078	-,661	-,349	,327
F3	-,270	-,432	,452	-,330	-,476	-,161	-,078	1,000	,171	-,423	-,046
F4	-,119	-,028	-,422	,329	,251	-,601	-,661	,171	1,000	,372	-,439
F5	,387	,641	-,471	,361	,647	-,192	-,349	-,423	,372	1,000	-,240
Ad1	-,108	-,122	,388	-,337	-,352	,312	,327	-,046	-,439	-,240	1,000

Source: Specht (2017)

Table 4 shows the results of factor analysis, and it is possible to see that the variables can be divided into four components. Accuracy loads highly on the first component, followed by fluency, complexity, and adequacy. Even though this may indicate that the measures adopted were in fact assessing different dimensions of speech performance, the order in which each component loaded was not in line with the one presented by Guará-Tavares (2008). In her study, fluency loaded as a first component, followed by complexity and accuracy. Nevertheless, such results are expected, considering that Guará-Tavares used in the analysis the outcomes of a task performed under no planning condition, while, in Specht and D'Ely's study, both tasks administered were preceded by strategic planning.

Table 4. Rotated Component Matrix

	Components			
	1	2	3	4
A2A	,932			
A3A	,874			
A1A	-,868			
F2A		,970		
F1A		,953		
C2A			,898	
C1A			,847	
Ad1A				,862

Source: Specht (2017)

In sum, the analyses indicate that all the measures were representative of the same phenomenon, that is, speech performance, and that some measures (all accuracy measures

and two fluency measures) were quite similar to each other, indicating that they are probably assessing equivalent aspects of their respective dimensions. Moreover, the speech performance was divided into four dimensions: accuracy, fluency, complexity, and adequacy, considering that the (or some) variables responsible for that dimension loaded in different components. The results, therefore, are an indicative that adequacy may be seen as an independent dimension, assessing specific aspects of speech performance other than the ones assessed by complexity, accuracy, and fluency.

3.2. Mixed ANOVA

Mixed ANOVA is a statistical test employed in this study to examine whether the differences between- and within-group comparisons are significant. Traditionally, most studies on strategic planning only employ measures related to complexity, accuracy, and fluency; however, these measures may leave some qualitative and discourse-oriented aspects of speech performance unattended such as task completion and task type, for instance. In this sense, the adoption of adequacy was opted in order to bring a qualitative view to the analysis of speech performance.

As already presented, the results of Factor analysis showed that adequacy can be characterized as a separated dimension, which indicated that it measured different aspects of speech performance, besides the ones measured by complexity, accuracy and fluency. The issue, though, lies on understanding which specific aspect(s) of adequacy was/were more salient, since five features were analyzed. Table 5 shows the results of the interactions among tasks and groups and it is possible to see that it was found (almost) significant differences in vocabulary (F=8.687 - p. 0.006), structure (F=6.575 – p. 0.004) and clarity (F=5.973 – p. 0.006), and appeal (F=3.159 – p. 0.056). These results indicate that there are significant differences between task and group pairwise comparisons.

Table 5. Tests of Within-Subject Contrasts

Source	Measure	F value	Sig.	Source	Measure	F value	Sig.
Task	Structure	.864	.360	Task*Groups	Structure	6.575	**.004**
	Appeal	1.383	.249		Appeal	3.249	.052

Vocabulary	8.687	**.006**		Vocabulary	3.159	.056
Clarity	1.173	.287		Clarity	5.973	**.006**
Fluency	1.433	.240		Fluency	2.078	.142

Source: Specht (2017)

Interestingly, fluency was the only adequacy feature that did not present any significant differences. Even though fluency as part of adequacy measure was assessed by raters' perception of fluent speech, it is the only adequacy feature that assesses aspects that are similar to the ones assessed by the fluency measures of the other dimensions. This relation is not quite clear when it comes to complexity and accuracy dimensions, in which no adequacy features can be directly related to their measures. If this similarity is taken into consideration, it is possible to see that this lack of impact in adequacy fluency may be in line with the lack of impact in fluency measures. Running a partial correlation analysis between the fluency measures and the fluency scores, a moderate and significant correlation, shown in Table 6, can be found between the following pairs FluencyA/F1A (0.554), FluencyA/F2A (0.592), FluencyB/F1B (0.639), and FluencyB/F2B (0.636).

Table 6. Results of moderate correlation between fluency measures

	F1A	F2A	F1B	F2B
FluencyA	0.554	0.592	x	x
FluencyB	x	x	0.639	0.636

x = low correlation

Source: Specht (2017)

These moderate correlations may reinforce the connection between adequacy fluency and fluency measures, more specifically the ones concerned with speech rate. This may indicate, even if speculatively, that raters' fluency perception is coherent with how speech rate measures are assessed. This result highlights the validity of speech rate measures that have been employed in strategic planning studies. On the other hand, the fact that four adequacy features did not have a more direct connection with measures of another dimension may support the claim of adequacy being an independent dimension.

Returning to the comparative analysis of the adequacy aspects, Table 7 shows the significant differences between group pairwise comparisons. As it is possible to observe, the integrated and isolated groups outperformed the control group in terms of structure, appeal, and clarity. This means that the experimental groups produced well structured (with a clear beginning, middle, and end), appealing, and clear stories. The results of the task pairwise comparison in Table 8 show that not only the control group underperformed the experimental groups, but they also decreased in terms of structure and clarity.

Table 7. Groups comparison – Task B

Measure	Groups	Groups	Sig.
Structure	Int	Iso	.942
		Con	.156
	Iso	Int	.942
		Con	.013
	Con	Int	.156
		Iso	.013
Appeal	Int	Iso	1.000
		Con	.054
	Iso	Int	1.000
		Con	.018
	Con	Int	.054
		Iso	.018
Clarity	Int	Iso	1.000
		Con	.011
	Iso	Int	1.000
		Con	.003
	Con	Int	.011
		Iso	.003

Source: Specht (2017)

Table 8. Pairwise Comparison – Comparison of Tasks A and B in each group

Measure	Groups	Sig.	Measure	Groups	Sig.
Structure	Int	.534	Appeal	Int	.282
	Iso	**.005**		Iso	**.034**

	Con	**.041**		Con	.191	
Vocabulary	Int	**.027**	Clarity	Int	.313	
	Iso	**.004**		Iso	**.009**	
	Con	.723		Con	**.049**	

Source: Specht (2017)

These prominent differences between the Control and the Isolated groups, regarding their stories' structure, appeal and clarity, can be clearly seen in the stories' transcriptions below. Contrasting the first and second stories produced by a participant of the Control group (Con1) and one of the Isolated group (Iso4), at first glance, it is possible to observe that Con1 presented a briefer story in Task B, while Iso4 produced a slightly longer story in Task B. By reading the stories, it becomes more evident that while in the first story produced by Con1, she presented a story with beginning, middle, and end, the same did not occur in Task B. This structural difference was not present in the story produced by Iso4. This lack of information in Con1's story might have led her story to become less clear and as a consequence less interesting for the raters. This same pattern can be seen in other participants' stories. Another indication that structure was the leading aspect followed by clarity and appeal was the raters' verbalizations while assessing the tasks. They mentioned quite often, in their assessing sheets, that some stories were "more complete" or "more well-structured" than others.

Con1 – Task A - |This is a history about one man :: that loves Mary| the man loved Mary| and he offerted some gifts for her| first of all a ring, after a jewel and a dress| but Mary didn't like the gifts| and the man was dissapointed| some times after this moment the man found another person| a simpatic person :: that loves his| and they got married| and they bought a special car| and in some situations they meet Mary| Mary was alone| and the couple were happy| and Mary stay a little bored :: I think| but this is the end of the history|

Con1 – Task B - |A cat called Jerry :: falled in love to Rose| but Rose had a boyfriend| and Jerry saw the couple together| very very happy| and after that Jerry was very depressive| I think :: Rose loved your boyfriend|

Iso4 – Task A - |the history is about a guy :: that wanted to go out with a girl| but the girl didn't care about he| so he started to bring presents to her| but she still didn't care about| in the end of the history he appears with his car with a girl| and cuz he gave up the first girl| and that's it|

Iso4 – Task B - |There are two cats; Tom and a black cat| They are trying to get the attention of a girl| That girl is a female cat| so everything that Tom does :: the black cat does it better| Tom appears with a car| but the black cat appears with a better and more expensive car| and after all the girl stays with the black cat| and Tom get sad| and he are crying and drinking milk :: cuz the the girl doesn't want to stay with him|

These results reinforce the fact that adequacy may be employed as an additional dimension, which deals, at least in this study, with discourse-oriented aspects, such as textual structure along with more qualitative aspects of speech performance.

In short, adequacy as it was conceptualized and proceduralized in this study appears to be a reliable measure to represent an extra speech dimension as well as assess discoursal and qualitative aspects of speech performance. The measure seems to have initially grasped structural aspects of the oral stories, which may seem related to clarity and appealing aspects, considering that a well-structured oral text may be clearer and more appealing to the listener.

Conclusion

The objective of this study was to scrutinize the adoption of adequacy as an extra dimension in Specht and D'Ely (2020), since, in their study, they only explored the final score of the measure, that is, the sum of the five criteria score and the average final sum of the three raters' scores. Their results have shown that adopting adequacy was an important choice. Participants from the integrated group only showed improvement at the level of adequacy. Without it, they would not have evidence that instruction on how to plan improved their participants' use of planning time and, consequently, their speech performance. However, which criteria in fact were responsible for adequacy to have an impact was not clear.

In this line, our results and analyses were able to identify how adequacy functioned in Specht and D'Ely's study. As seen in the previous section, three criteria were responsible for the measure improvement as a whole, they are: structure, clarity and appealing. No difference was found for vocabulary or fluency. Thus, saying that participants improved their speech performance in terms of adequacy can also mean their performance presented a well structured text, with beginning, middle and end, which led to a clearer and more appealing text. These speech features grasp a more discourse-oriented of speech performance, which were not measured by CAF measures, as we could see in the Principal Component Analysis.

These results and this study provide more support and evidence for adequacy to be used in research that investigates speech performance. It is an extra dimension and

measure that can be adopted by researchers that are interested in a more qualitative analysis of speaking. Furthermore, adequacy can also be used by teachers as a reliable assessment tool for their students' speaking activities or tasks. If used by teachers in the classroom, in order to reduce time consuming, less raters would not be a problem, even the teacher can be the rater. However, it is important to remember that every scale assessing sheets should be created contemplating the task's or the activity's requirements. Specht and D'Ely used a narrative task, there are other types of tasks with different characteristics and outcomes' possibilities. Narrating a story is different from leaving a voicemail message.

All in all, this study, besides enlightening the results brought by Specht and D'Ely, also contributes to the field of Task based Learning and Teaching, leaving insights for educators. For future research, we encourage researchers to investigate how adequacy could be conceptualized for different types of tasks, and how and if they correlate, measuring similar and different qualitative aspects of speech performance. Another suggestion would be to consider the language classroom as context of investigation, and analyze how educators deal with adequacy in their practices.

References

DE JONG, N.J.; STEINEL, M.P.; FLORIJN, A.; SCHOONEN, R. and HULSTIJN, J.H. The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. *In*: HOUSEN, A.; KUIKEN, F. AND VEDDER, I. (Eds), *Dimensions of L2 Performance and Proficiency: Investigating Complexity, Accuracy and Fluency in SLA*. Amsterdam: John Benjamins, 2012.

FIELD, A. *Discovering statistics using SPSS*. 3rd edition. London: Sage Publications, 2009.

FOSTER, P. Task-based learning research: expecting too much or too little? *International Journal of Applied Linguistics.* Vol. 19, N. 3, p. 247-263, 2009.

FOSTER, P. AND SKEHAN, P. The influence of source of planning and focus of planning on task-based performance. Language Teaching Research. Vol. 3, p. 215-247, 1999.

GUARÁ -TAVARES, M. G. Pre-task Planning, Working Memory Capacity and L2 Speech Performance. Unpublished Doctoral Thesis, Universidade Federal de Santa Catarina, Brazil, 2008.

HOUSEN, A. AND KUIKEN, F. Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*. Vol. 30, N. 4, p. 461-473, 2009.

IWASHITA, N.; BROWN, A.; MCNAMARA, T. AND O'HAGAN, S. Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*. Vol. 29, N. 1, p. 24–49, 2008.

KAISER, H.F. An index of factorial simplicity. *Psychometrika*. Vol. 39, p. 31-36, 1974.

KUIKEN, F.; VEDDER, I. AND GILABERT, R. Communicative adequacy and linguistic complexity in L2 writing. In: Bartning, I. Martin, M. and Vedder, I. (Eds), *Communicative Proficiency and Linguistic Development*: Intersections between SLA and Language Testing Research. Eurosla Monographs 1. Roma: Eurosla, 2010.

LARSEN-FREEMAN, D. Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, Special Edition, p. 579-589, 2009.

NORRIS, J. M. AND L. ORTEGA. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, Vol. 30, p. 555-578, 2009;

PALLOTTI, G. CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, Vol. 30, N. 4, p. 590-601, 2009.

RÉVÉSZ A.; EKIERT M. and TORGERSEN E. N. The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*. Vol. 37, p. 828–848, 2016.

SKEHAN, P. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press, 1998.

SKEHAN, P. Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*. Vol. 30, N. 4, p. 510–532, 2009.

SKEHAN, P. The context for researching a processing perspective on task performance. *In*: SKEHAN, P. (Ed.), *Processing perspectives on task performance*. Amsterdam: John Benjamins, 2014.

SPECHT, A. L. Is strategic planning enough? Investigating the impact of two types of strategy instruction on students' oral planned performance. Unpublished Doctoral Thesis, Universidade Federal de Santa Catarina, Brazil, 2017.

SPECHT, A. L. AND D'ELY, R. C. S. F. Tasks and Proficiency Tests: Piloting instruments of a study on strategic planning. *Revista de Letras*. Vol. 21, p. 170-184, 2019.

SPECHT, A. L. and D'ELY, R. C. S. F. Enhancing Strategic Planning through Strategy Instruction: The Effect of Two Types of Strategy Instruction on Learners' Oral Planned Performance. *Ilha do Desterro.* Vol. 73, N. 1, p. 43-62, 2020.