

Time Pressure and Within-task Variation in EFL Oral Performance

Mohamed Ridha Ben Maad

(ridhamaad@hotmail.com)

University of Manouba, Tunisia

Abstract

This article reports a study of variability in the oral performance of 17 English as a foreign language (EFL) students in a Tunisian university. Time pressure, long ignored by research community as a key element in the process of language learning, was operationalized as the prime factor in order to give evidence of variable output during task engagement. Analysis of the subsequent performances on the same task showed that time pressure pushed learners towards varied results, that is, higher rates of complexity, decrements in accuracy and mixed patterns in fluency. The results were found to be verifiable in light of the established task-based literature and the processing-based theory whose main tenet is attention as a limited resource.

1 Introduction

It is commonly held that the concept variability is a theme strictly researchable in sociolinguistics given the well-documented findings and the plethora of research lines advanced by sociolinguists. However, this concept has been explored beyond the confines of sociolinguistic research in that reverberations in language classroom research appeared in the works of Elaine Tarone (1983, 1985, 1988) and, most prolifically, in task-based models (e.g. Skehan, 1998). Certainly, instances of variation are schematically presented in the differences between task types and task conditions. Despite the fact that this new wave of findings has instilled a psycholinguistic dimension into the study of variability, it is obvious that task-based variation has been strictly limited to cross task-elicitation methodology.

In order to expand the concept beyond the bounds of this research format, the present study seeks to explore variation within tasks per se. Herewith, this type of variation is identified though the effect of time pressure on EFL learners' real time performance of speaking tasks. The study champions a processing-based framework in the interpretation of its findings. As such, this theoretical perspective gives us an incisive account of patterns of change in subjects' performances due to their declining cognitive capacity vis-à-vis the mounting pressure of time. Interestingly enough, evidence of within-task variation may suggest a call for revisiting or looking with caution at the surplus of findings in task-based research which typically envisage variability as a fact only when comparing task types or conditions.

2 Research background

Variability has become an established research construct thanks to the landmark works in the sociolinguistics of the 1970s (e.g. Labov, 1972). Also known as variationist linguistics, it emerged to highlight untouched areas left by the then influential Universal Grammar (UG) based approach

whose research objective has been sternly directed at the sentence and its internal structure. Variationists, like Labov, advanced their theory on the assumption that language use can qualify as a window into the mental representation of language. Thus, viewing language as an observable social and psychological fact, this research line enables researchers to spot patterns of language learning and change within variable data collected in longitudinal and cross-sectional studies. In contrast with a UG blueprint, 'probabilistic knowledge' (Ellis, 1985) captures these patterns by means of units of analysis that can observe the development of grammar from zero to definitive attainment. These formal tools, referred to as variable rules, proved to be empirically viable due to the development of statistical programs such that of VARBRUL (Preston, 1989).

In the 1980s, variability theory gained mainstream status in Second Language Acquisition (SLA) research basically through the seminal works of Tarone (1983; 1985; 1988) and Ellis (1985). Based on Labov's stylistic continuum, Tarone maintained that learners vary in their language use, or shift styles, according to their degree of attention to form, that is, when formal attention peaks learners reach the 'careful style,' and when attention is at its lowest, they eventuate in the 'vernacular style.' It is worth mentioning here that the influence of variationist linguistics on SLA theory has lost its advanced level of abstraction in 1990s among researchers. Yet, the construct of variability has ever since been championed by various scholars of different agendas. For example, Van Patten's (1990) landmark investigation into the form-meaning dichotomy attests to the theoretical weight of variability theory in explaining the pressure underlying second language (L2) learning process. Equally inspired was the emerging task-based approach which attempts to draw up a principled view of task effects out of variable performances elicited across tasks.

Perhaps the involvement of sociolinguistics in the study of L2 variability was one among other factors that cemented its position among L2 researchers. Although Tarone went beyond Labov's 'naturalistic' analysis of interlanguage, adopting tasks as precursors of variation, her reliance on master constructs such as 'style' and 'formality' as a manifestation of standardness and correctness (Beebe, 1988) failed to meet research quality standards set by SLA theorists. In L2 assessment, for example, other aspects of performance (e.g. degree of elaboration, speech rate, prosody, etc.) went unnoticed (Mitchell & Myles, 2004; Wolfram, 1990). Instead, statistical programs such that of VARBRUL (Preston, 1989) were only verifiable through fragmented morphosyntactic representations of interlanguage (see Tarone 1988, pp. 124–131, for further details). Besides, the variationist perspective offers no input on the dynamics of L2 acquisition (Gregg, 1990). Mitchell and Myles (2004) would ascribe this oversight to the absence of 'psycholinguistic plausibility' beyond the simplistic 'coin-tossing' metaphor and its reference to learners as "continuously calculating probabilities and choosing between alternative forms" (p. 178).

2.1 *Second language variability revitalized*

Despite the failure of variationists to build a full-fledged model which accommodates the theoretical and methodological directions of SLA research, the construct of variability has reverberated across many emerging models, especially those related to interlanguage output. For example, one may easily establish the connection between Skehan's (1998) cognitive approach to task and Tarone's paradigm. Like Skehan, Tarone directed variability research to investigating learner production where mainstream SLA focus was mostly oriented towards the 'input' side of language representation, which is the locus of interest for most of the studies based on Skehan's model (e.g. Foster & Skehan, 1996; Ortega, 2005). By way of comparison, it is believed that the same research format was used: different task types and conditions exert differential 'effects' on learners' output. Like the early attempts of variationists, task-based researchers endeavor to encapsulate this substance of variability into generalizations that implicitly appeal to their pedagogy-affiliated agenda.

The main contribution in Skehan (1998) is conceivably psycholinguistic in character. In addition to focus on tasks as precursors to variability, Skehan shared with variationists their focus on the core construct of 'attention'. Perhaps, the interesting novelty which Skehan brought at this point was his treatment of attention from a psycholinguistic perspective, a central problem in variability research. Far from the typical variationist view of attention as a metric for of stylistic vari-

ety, Skehan (1998), inherently building on a good deal of literature from cognitive psychology (e.g. Shiffrin & Schneider, 1977) and SLA (e.g. Van Patten, 1990), referred to attention as a cognitive resource of limited capacity. Herewith, Skehan's model is theoretically capable of explaining the patterns of language processing based on the idea of competitiveness. Equally important is Skehan's focus on what Ellis (2003) dubs as 'discourse analytical method' which broadens research focus on various interlanguage aspects beyond the morphosyntactic level (e.g. temporal features of fluency such as speech rate and reformulations).

2.2 *The problem with task-based variability*

Despite the wide array of studies investigating the effect of task on L2 variation (e.g. Bygate, 1999; Skehan & Foster, 1997, 1999; Tarone, 1988), almost all variability-related findings were documented based on cross-task elicitation. As such, the quality of describing and operationalizing variability has been constrained to comparing differences between task types. For example, cross-task variation was heavily documented through the plethora of planning research (see Ellis, 2005 for details). However, this comparative design cannot verify the complete picture of variability seeing that there is high likelihood of variation within tasks. It is, in this sense, hardly surprising to tap substantial evidence of variability in some learners' performances during task engagement. This line of reasoning has not gone beyond conjecture in the absence of empirical undertaking despite scattered reference to time pressure as a key factor of task variation (Lynch & MacLean, 2000; Skehan, 1996; 1998). Building on this research oversight, the present study was set to probe this trajectory of within-task variation by tapping instances of variability due to speaker participants' growing sensitivity to time pressure.

The findings of the present study were examined within the processing model described in Skehan (1998). In this vein, it was hypothesized that:

1. Learners' oral performance is sensitive to the incremental effect of time pressure during task engagement.
2. Time pressure predicts competition for attentional focus among the areas of fluency, complexity, and accuracy.

3 Context and methodology

Nineteen first-year students, from the Faculty of Letters, Arts and Humanities of Tunis, were to deliver presentations on general topics (e.g. euthanasia, terrorism, pollution, etc.) chosen by their teacher during regular class-scheduled sessions. All the participant speakers had the same native language background (i.e., Tunisian Arabic) and a fairly homogeneous English proficiency level according to exam data collected from their speaking, writing, and reading teachers. Their mean age was 20.8 years and the mean length of EFL education received was 4.9 years. There was no drastic classroom change to the normal classroom routines and physical setup; and the researcher's intervention was minimal during the episodes of data collection (i.e. except in the preparation of tape-recording materials). In this experiment, participants were required to perform oral presentations on topics already assigned by their teacher. Planning time was unlimited seeing that the assignments were given to them in the beginning of the course. They were allowed to use visual prompts and numerical information during their speech performances.

The present study was designed around one single task, topic-based oral presentation, as the sole data elicitation tool. Time pressure was the independent variable which was presumed to peak in influence towards the end of the presentations. In order to ensure better chances for data validity, possible external factors, such as planning type and time, teacher or peer intervention etc. were neutralized¹. An operationalization of time pressure was conceived on the basis of the time length of the speech performances. To test the reliability of the elicitation task in showing evidence of time pressure, the researcher examined performance samples of two non-participants and their feedback on difficulty level and distribution during the speech deliveries. Accordingly, the cutoff

time to be considered in the subsequent phase of data analysis was set at 6 minutes². Two out of the 19 performances were eliminated since the length of production was less than required whereas the majority of recorded performances were cut to match the set time interval.

The dependent variables of the present study were operationalized in keeping with Skehan's (1996, 1998) tripartite model of output. Accordingly, three main variables were identified. First, fluency was represented by three measures, namely Reformulations per minute, Pauses per minute (1 pause > 0.5 second), and Speech rate per minute (syllable number divided by 60). Where the first two measures stand for 'repair fluency' (Skehan, 1996), the third measure refers to velocity. Second, complexity evaluates the organizational quality of oral performance. It engages "greater syntactic patterning" (Foster & Skehan, 1996, p. 304), which reflects risk-taking behavior that encourages the use of 'cutting edge' language. The measures of Subclauses per minute and Words per T-unit were employed here to tap the amount of subordination and lexical density, respectively. Third, accuracy accounts for L2 learners' positive attitudes towards form and the rule system of the target language, thus entailing risk-avoiding behavior (Skehan, 1996). This variable was assessed by measures of error-free clauses per minute (main and/or subclauses without errors), grammatical errors per minute and lexical errors per minute.

4 Method

All the data samples considered for analysis were tape-recorded and transcribed according to standard orthography. A total of 17 transcripts were coded according to a coding scheme designed by the researcher. The six-minute span of every transcript was divided into two phases: an 'onset phase' which covers the first three minutes and a 'cutoff phase' which represents the last three minutes. In order to test the reliability of the coded data, two graduate students of linguistics assisted in reviewing some fragments from the transcripts (around 10 %). Where reliability scores were acceptable for fluency and complexity, i.e. 90.8 % and 95.1 % respectively, accuracy scores showed only 71.3 % of inter-rater agreement. As a result, all coded accuracy data were collectively reviewed and problematic instances were discussed for final convergence of tally. The raw data were subjected to basic descriptive statistics by computing (i) the mean scores of individual / group performances to examine the effect size of time pressure on the direction of outcomes and (ii) the percentage scores to test evidence of trade-off effects among performance measures. In order to test Hypothesis One which predicts increasing negative effect of time pressure on speaking performance, results were discussed according to the main three outcome variables.

Table 1 displays average frequency scores on three measures of fluency: reformulations, pauses and speech rate. As for the measure of reformulations, related scores follow a steady increase in frequency towards the end of performance. The highest of all is the time span between minutes 4 and 5 where the lowest is between minutes 1 and 2. Contrary to reformulations, both pauses and speech rate showed consistent improvement, namely against the predicted direction. The number of pauses per minute, an index of undesirable fluency (Skehan, 1996), decreased gradually especially at the very beginning (minutes 1 and 2) and the very end (minutes 5 and 6). Equally improving was the speed in the informants' performances. However, the improvement of speech rate was subject to fluctuations especially at the time interval of minutes 2 and 3. Improvement, instead, peaked at the level of minutes 4 and 5. In sum, Hypothesis One did not receive full support in regards to fluency due to the unpredictable improvement of speech velocity and decrease in pausing.

Variable/ Minute	Onset Phase				Cutoff Phase			
	1	2	3	%	4	5	6	%
Reformulations	1.23	1.70	2.70	30.91	3.29	4.47	4.82	69.08
Pauses	9.58	8.41	7.58	68.25	6.18	5.58	4.53	31.74
Speech rate	122.40	123.00	120.75	43.32	121.80	148.20	152.40	56.67

Table 1: Mean and percentage scores for fluency measures

Hypothesis One partially predicted that time pressure has detrimental effect on complexity. The findings tabulated below (Table 2) describe results of subordination and word number per T-unit which respectively refer to the degree of elaboration and lexical density. Despite the chronological increase of subordination along the six-minute span, such improvement is characterized by irregularity. More precisely, the average of subordination did not go beyond one sub-clause per T-unit (e.g. 0.53 means score for minute 3) during the onset phase. Conversely, at the cutoff phase use of subordination reached peak level of frequency especially at the minute intervals of 3/4 and 5/6. In the same vein, Word number per T-unit demonstrated steady improvement along the performance span, except at minute 6 where increase reached its highest. Overall, to such improvement at the level of the above complexity measures, complexity results did not corroborate the Hypothesis One which predicted decrements in performance due to time pressure.

Variable/Minute	Onset Phase				Cutoff Phase			
	1	2	3	%	4	5	6	%
Subordination	0.26	0.36	0.53	28.82	0.78	0.92	1.14	71.17
Words/T-unit	12.45	13.60	14.68	43.02	15.94	18.44	19.56	56.97

Table 2: Mean and percentage scores for complexity measures

With a view to testing the hypothesized decrement of performance at the level of accuracy, Table 3, below, presents findings corresponding to error-free clauses, grammatical errors and lexical errors. The results related to error-free clauses per minute showed steady decline, where the lowest score occurred at the time interval of minutes 1 and 2. This finding appears to counter the main pattern of highest and/or lowest scores which mostly occurred at the cutoff phase. The decreasing frequency of error-free clauses was paralleled by a consistent increase of error types. With reference to grammatical errors, the rise in frequency was steady. Yet, the rhythm of change slowed down at the cutoff phase chiefly when differences between mean scores grew narrower (esp. minutes 5 and 6). Such was the pattern of change with lexical errors. Though Table 3 shows that the mean frequency scores of lexical errors was lower than those of grammatical errors at the onset phase, the frequency average of lexical errors in the opening minute (i.e. 0.58) increased more than nine times (i.e. from 0.58 to 5.29). In light of these findings, it can be concluded that time pressured participants into multiple instances of inaccuracy. Hence, Hypothesis One was supported.

Variable/Minute	Onset Phase				Cutoff Phase			
	1	2	3	%	4	5	6	%
Error-free Clause	0.66	0.52	0.44	63.52	0.37	0.32	0.24	36.47
Gram. Errors	2.29	3.29	4.47	39.12	5.00	5.29	5.35	60.87
Lexical Errors	0.58	1.06	1.94	21.93	3.40	4.05	5.29	78.06

Table 3: Mean and percentage scores for accuracy measures

In light of the mixed findings related to the three performance areas, Hypothesis One was neither fully confirmed nor fully nullified due to the different variability trajectories between these areas. Perhaps, the biggest chunk of these results (i.e. complexity and much of fluency) tipped the balance towards the second part of this conclusion.

In order to test Hypothesis Two which predicted that time pressure exerts a trade-off effect on various areas of speaking performance, both percentage and mean scores from the three tables above were discussed. To start with, the percentage figures located in the last column of the three tables represent the proportion of frequency distribution of every measure both at the onset phase and the cutoff phase. The tally of percentage margins among either the highest scores or the lowest scores of the three variables showed insignificant and sometimes non-existent margins (fluency: 12.41 %; complexity: 14.20 %; and accuracy: 17.19 %). The slim margins indicate the proportionate distribution of time pressure effect on the three dependent variables of speaking performance. In addition to the margins of difference between the lowest/highest scores at each of the dependent variables, there are minimal differences between the group proportions of the three measure sets. The lowest and the highest of scores are respectively as follows: fluency: 35.32 % vs. 64.66 %; complexity: 35.92 % vs. 64.07 %; and accuracy: 32.50 % vs. 67.48 %. Building on the two ways of percentage tally, it can be concluded that this balanced effect distribution logs no independent effect of time pressure on a particular outcome variable.

In order to spot the lowest and highest points of time pressure effect on speaking performance, mean scores from the three tables above were re-examined. It was found that almost all the significant data were located at the cutoff phase. The measure of grammatical errors per minute was the exception with no clear highest/lowest instance. It was in minute 5 that most of the highest instances of variation occurred (i.e. five minute intervals related to minute 5 out of the 8 measures had the highest/lowest scores). In sum, based on (i) the findings about the divergent direction of variation between complexity and accuracy and (ii) the fact of determining the cutoff phase as the locus of the highest/lowest instances of time pressure effect, one can understand that a negative effect on one performance prompted positive effect on another. As a result, Hypothesis Two was confirmed.

5 Discussion

The findings of the present study substantiated the hypothesized connection between time pressure and within-task variation. For a fact, some of the investigated measures were found to be more variable than others (cf. Lexical errors and Words/T-unit). Time pressure seemed to yield progressively improved performance in so far as complexity is considered. The converse was the case with formal accuracy. Task performers were unable to evenly allocate their attention to the three areas of performance due the ever-increasing demands ensuing from task engagement and time pressure. In this vein, if one compares scores of every measure along the two time groups, s/he can clearly establish that focus was gradually switching from accuracy to complexity and, partly, to fluency. In some way, this direction reflects a change of strategy use patterns, from a conservative rule-based to a rather relaxed and risk-taking language processing.

Building on these findings, the present study supports the revision of the construct of variability in SLA research. The study draws on a processing-based approach (Skehan, 1998) to conceive of a finely-grained definition of variability beyond the early applications championed by Tarone (1983, 1985, 1988). Counter to the early definition of variationists which assumes that form-focused attention is a dwindling resource (i.e. speakers stretch out their attention to eventuate in a nonstandard form of output), the findings of the present study endorse a psycholinguistic view of variability as an evidence of attentional allocation to a given aspect of performance at the expense of other aspects. Such allocation process is guided by the changeable effect size of task factors. Similarly, this data-driven definition of variability finds support in Robinson's (1995) empirically-based conclusion that attention can be resource-depleting as well as resource-directing. Put it differently, Like Robinson's observation that cognitive task complexity had beneficial effects for fluency and complexity, findings in the present study attested to the same effect size for complexity and, fairly so, for fluency.

The study also corroborated that the relationship between the three performance areas is bound by trade-offs stemming from limited attentional capacity. That is, the improvement at the level of complexity was detrimental to the accuracy aspect of performance. Yet, fluency scores did not seem to be consistent with such attentional distribution. Instead, fluency displayed mixed results, with an increasing number of reformulations (i.e. undesirable fluency) on the one hand, and a dwindling number of pauses together with the improvement of speech rate (i.e. effective fluency) on the other. Perhaps such apparent irregularity can be ascribed to the heterogeneity of fluency components. Understood from a processing-based perspective, it is the increasing speech velocity that pushed the performers to trade their reliance on pauses for more instances of hesitation and reformulation in order to sustain persistent pressure. Again, such a processing pattern can be explained by the change of behavior, from a form-focused conservatism marked by high accuracy in the beginning of the task towards a risk-taking engagement in the form of substantial elaboration by the end of task performance.

6 Conclusion

Researching the relationship between time pressure and within-task variation calls for the need to revisit the wealth of across-task-elicited findings gathered in the past two decades. It may be propitious here to question whether a given task type can decidedly lead to a given result or perhaps if the very relationship can be 'distorted' by moderating effects like time factor or individual differences. Had the focus on this study been on task type, one could have been content with the partial result that the informants were simply form-oriented rather than content-oriented (i.e. at the onset phase). This clipped interpretation of findings, especially stemming from time-consuming tasks, is to give way to a sharper outlook that compels technical adjustment which can capture the mediational effects of factors like time pressure and individual differences.

Notes

¹ There was no experimental manipulation of planning as all participants were assigned the topics one week ahead of their presentations and one type of planning was used which was 'unguided planning' (Sangarun, 2001; Ortega 2005). As for peer intervention, I instructed all the listener participants not to intervene during the presentations to interrupt the time pressure buildup, the primary focus of this study.

² The decision of the cutoff time was motivated by my long informal observation of similar performances prior to this experiment. I noticed that the typical delivery time was around 10 minutes and that performance exacerbation and signs of difficulty emerged considerably from minute four.

References

- Beebe, L.M. (1988). *Issues in second language acquisition*. Boston: Heinle & Heinle Publishers.
 Bygate, M. (1999). Task as context for the framing, reframing and unframing of language. *System*, 27, 33–48.

- Ellis, R. (1985). A variable competence model of second language acquisition. *IRAL*, 23, 47–59.
- Ellis, R. (1990). A response to Gregg. *Applied Linguistics*, 11, 118–131.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
- Ellis, R. (2005). *Planning and task performance in a second language*. Philadelphia: John Benjamins.
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299–323.
- Gregg, K. (1990). The variable competence model of second language acquisition, and why it isn't. *Applied Linguistics*, 11, 364–383.
- Labov, W. (1972). *Language in the inner city*. Oxford: Basil Blackwell.
- Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4(3), 221–250.
- Mitchell, R & Myles, F (2004). *Second Language Learning Theories* (2nd ed.). London: Arnold.
- Ortega, L. (2005). What do learners plan? Learner driven attention to form during pre-task planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 77–109). Philadelphia: Benjamins.
- Preston, D.R. (1989). *Sociolinguistics and Second Language Acquisition*. Oxford: Basil Blackwell.
- Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, 45, 99–140.
- Sangarun, J. (2001). *The effects of pre-task planning on foreign language performance*. Unpublished doctoral thesis, OISE-University of Toronto, Canada.
- Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127–190.
- Skehan, P. (1996). A framework for the implementation of task based instruction. *Applied Linguistics*, 17, 38–62.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, England: Oxford University Press.
- Skehan, P., & Foster, P. (1997). The influence of planning and post-task activities on accuracy and complexity in task-based learning. *Language Teaching Research*, 1, 185–211.
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retelling. *Language Learning*, 49, 93–120.
- Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics*, 4, 142–163.
- Tarone, E. (1985). Variability in interlanguage use: A study of style-shifting in morphology and syntax. *Language Learning*, 35, 373–404.
- Tarone, E. (1988). *Variation in interlanguage*. London: Edward Arnold.
- Tarone, E. (1990). On variation in interlanguage: A response to Gregg. *Applied Linguistics*, 11, 392–400.
- Van Patten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition*, 12, 287–301.
- Wolfram, W. (1990). Incorporating dialect study into the language arts class. *ERIC digest*. Washington, DC: ERIC Clearinghouse on Languages and Linguistics, ED318231.