The Influence of Task Type on Perceived Fluency

Mirjana M. Kovac[1*]

[1] Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Split, Croatia

[*] Mirjana M. Kovac, E-mail: mirjana.kovac@fesb.hr

Abstract

In foreign language teaching it is necessary to employ a variety of tasks with a different degree of narrative structure to influence learners' performance in anticipated ways, thus enhancing overall fluency as a significant aspect of successful communication. This study examines how different task types influence the perceived fluency of EFL learners. The obtained results confirm that the perception of speakers' fluency varied across tasks. The worst perception of fluency was achieved in the task based on the formulation of unrelated utterances, that is, the most unstructured task led to the least fluent performance. A frequent introduction of new topics demands a great amount of attention. Therefore, in foreign language teaching emphasis should be laid on tasks including frequent introductions of new topics, due to the observed decrease in perceived fluency.

Keywords

task structure, perceived fluency, speech rate

1. Introduction

There are many definitions of fluency because it encompasses different aspects of language. Lennon (2000) assumed that fluency and accent are the main features of the speakers' L2 ability when evaluated by ordinary interlocutors on the street. However, it does not necessarily refer to the actual L2 speaker's level of proficiency. In his view fluency can be described as the listener's impression that the psycholinguistic processes of speech planning and speech production are functioning smoothly and efficiently, in other words, these two processes can be performed nearly simultaneously by the speaker. Lennon (2000, p. 26) explained that fluency might be defined as the "rapid, smooth, accurate, lucid and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing". However, it may not exactly mirror the speaker's knowledge of the language. In Lennon's view, an individual is fluent as long as his/her speech holds the listener's attention, and it is of utmost importance that we achieve a better understanding of how to improve fluency without sacrificing the form. The former view of fluency as an overall speaking proficiency aspect can be described as fluency in the narrow sense, as opposed to linguistic complexity and accuracy of the linguistic forms (Housen & Kuiken, 2009).

Derwing (2007) indicated that oral fluency can be defined as an automatic procedural skill on the part

of the speaker, and a perceptual phenomenon on the part of the listener. Moreover, fluency has been researched taking into account different perspectives, such as the characteristics of fluent versus disfluent speech, the influence of planning, self-monitoring and the task type. Also, a lot of research has been conducted on the effects of a study-abroad experience on fluency. The results suggest that overall fluency improves with greater opportunities to speak outside of the classroom. Disfluencies such as repetitions, false starts, the excessive use of filled pauses, and a slow speaking rate can negatively affect the listener's understanding and attention. According to Derwing (2007), it is important that L2 speakers develop the oral fluency skills at an early stage of L2 acquisition.

Segalowitz (2010) introduced three levels of fluency: i) cognitive, which refers to the ease of mental preparation, ii) utterance fluency or the smoothness of articulation, and iii) perceived fluency or the listener's judgement about the speaker's fluency. Pinget (2011) claims that perceived fluency can be related to Lennon's (1990) view of fluency, describing it as the only tangible fluency or the fluency in the ear of the listener.

As far as speaking tests are concerned, fluency is a score awarded by human raters. Taking into account the correlations between such subjective measures of fluency and the objective ones, the speech rate as the objective measure has been proven to be one of the best predictor of subjective fluency. Kormos and Denes (2004) implied that speech rate is a good predictor of subjective fluency that can be identified as the number of syllables per time unit. Cuccharini, Strik and Boves (2002) indicated that expert fluency ratings can be more accurately predicted in read speech than in spontaneous speech. This is performed by calculating the variables such as the speech rate, the articulation rate, the phonation/time ratio, the number of pauses, the duration of pauses, and the mean length of runs. Cuccharini, Strik and Boves (2002) laid emphasis on the speech rate as the best perceived fluency predictor. However, only in cases of cognitively more demanding tasks the mean length of runs has turned out to be the best predictor of perceived fluency in spontaneous speech. Ullakonoja (2009) focused on the speech rate development of 12 Finnish university students of Russian during their study-abroad experience. It was found that the speech rate is a better indicator of fluency than the articulation rate in non-native read-aloud speech.

Oral fluency can be investigated from another perspective, that is, from the impact of the task design on accuracy, fluency and language complexity (e.g., Foster & Skehan, 1996; Skehan & Foster, 1997, 1999; Robinson, 2001, 2005). The performance of different task types leads to interlanguage changes, that is, the learners attend to and retain specific information about the language in use. It is important to identify the specific features of each task type and how that particular task and its performance influence language processing. By carefully guiding the learner, the teacher can deliberately focus the learner's attention on particular aspects of language. The deliberate use of different task types should eventually influence learners' performance in predictable ways, thus expanding different opportunities and enhancing overall fluency.

The influence of task characteristics and task condition variables on the different aspects of fluency,

accuracy and complexity have been examined by Foster and Skehan (1996), and Skehan and Foster (1997). The researches included three different task types: personal information exchange, decision-making and a narrative. Two of these tasks used in these studies resulted in a more fluent performance under planned conditions which accordingly led to increases in accuracy. Finally, conclusions were drawn that besides task-type and planning, a clear inherent structure of these two tasks helped decrease the processing burden of the task. With regard to these conclusions, Skehan and Foster (1999) emphasized that the structured task led to a more fluent and a more accurate performance, as opposed to complexity which remained unaffected. On the other hand, if the information to be transmitted was unfamiliar, the overall performance was less fluent and less accurate but more complex. This can be explained by the attentional resources, namely, familiar information requires less attention for processing, and as a consequence, the L2 performance is more fluent and accurate.

Tavakoli and Skehan (2005) argued that a task can be regarded as structured if it had a clear time line, a script, a story including a beginning, a middle part and an end, as well as an appeal what is familiar and organized in the speaker's mind. These authors found that the structured tasks provided more accurate and more fluent language than the unstructured tasks. Mehran and Rahi (2010) also showed that tightly structured tasks led to a more accurate performance. Furthermore, they claimed that a task with a tightly structured storyline was significantly connected with greater fluency compared to a task lacking such a structure. A clear time sequence from the beginning to the end decreases the processing requirements and provides greater attention to accurate and fluent performance.

In Robinson's (2001) view, narratives are more complex than picture descriptions. In a picture description task speakers have visual support and the memory is less loaded than in a narrative task. A number of previous research studies which focused on speech errors and repairs (Levelt, 1983; Verhoeven, 1989) mainly included the description of static objects and constellations, in particular spatial relationships (e.g., up-down, left-right). Thus, special attention was paid to the linearization process of macro-planning, that is, to the organization of natural order. The content of three-dimensional spatial information must be linearly organized since there is no natural order to proceed, as it is in the case of retelling a chronological order of events. Whenever speakers want to express any communicative intention, regardless of its complexity, they must solve the linearization problem. Levelt (1989, p. 138) explained it in the following way: "Deciding what to say first, what to say next, and so on ...". In order to determine an order, the speakers need to memorize what has already been said and what needs to be said. When describing spatial relationships, Levelt (1989) concluded that the participants follow three major principles which apply to spatial and non-spatial description areas, since they reflect the most general characteristics of perception and memory. *The Principle of connectivity* is a general ordering principle in perception and memory, and the dominant way of describing hierarchically organized structures. However, it is not always possible to introduce new items of information without repeating the old items. The *Stack principle* is known in psychology of problem solving and is a dominant way of keeping track of hierarchically organized structures. The speaker's bookkeeping for return address is like putting them

on a stack and always returning to the top item on the stack after reaching the end of a connected string (Levelt, 1989, p. 143). Finally, *the Minimal load principle* is applied when there are more alternatives, and the principle says "Do the simplest thing first" (Levelt, 1989, p. 144), implying that the speakers describe less demanding constellations first. However, it is difficult to apply the principle of natural order to the description of dynamic constellations, and the reasons can be found in the dynamic aspects of the event, as opposed to the static objects and constellations. If the information to be conveyed is complex involving several consecutive speech acts, the speakers must decide how to organize that information. This is the speakers' linearization problem (Levelt, 1989), and the principle of natural order requires that the speech acts must have a chronological order of appearance. Studies based on tasks involving the description of spatial constellations (Levelt, 1982) aimed to determine how the static spatial structure is used for the construction of a linear order. Tasks given to participants in earlier research mainly included picture description, spatial description, interview tasks, storytelling, and information gap activities.

The data collection procedures and the task selection employed in this study are described in the next chapter. It is followed by the study results obtained from the use of appropriate statistical tests. Finally, the findings of this study are summarized in the discussion.

2. Data Collection Procedures and Task Selection

The sample group consisted of 101 participants, first-year students of technical studies in Croatia. The learners received 8-9 years of formal English instruction and it was presumed that the participants' instructional background was very similar, since all of them claimed to have received a mixed form-focused and communicative syllabus as a method of instruction in their primary and secondary school. Therefore, the speakers should have reached the B level of English proficiency according to the Common European Framework of Reference for Languages. The participants were randomly selected to participate in the study. The data were collected by means of five different tasks which were performed in English. All data were collected in an ordinary office with no special facilities at the Faculty. It was preferred to an experimentation room because it would contribute to the informal atmosphere in which the tasks were performed. The subjects were seated opposite the researcher at the same table. The computer and the microphone were placed between the subjects and the researcher. Each student was individually audio-recorded and afterwards the speech samples were transcribed and categorized by two coders to avoid inconsistency and data bias. When there was a discrepancy in the data analysis between the two coders, the third coder was consulted in order to come to a final agreement. Time parameters were measured by means of a speech analysis programme.

Before carrying out the selected speech tasks, the researcher gave explicit instructions for each task. In the story retelling task (CARTOON) the students watched the cartoon Johnny Bravo in the English language. This cartoon was chosen under the assumption that it was relatively unknown to the subjects and that its content represented a significant cognitive effort in terms of discourse organization. After watching the cartoon which lasted for six minutes, each participant described the chronological order of

244

events in their own words in the English language. The participant had one minute to prepare and was not time limited.

In the second task the students had to describe a room picture (ROOM1) that had six pieces of furniture in such a way that someone who could not see the picture could produce a global setting on the basis of the description. The third task (ROOM2) was almost identical to the second one, except for the furniture that was differently arranged.

In the fourth task (UTTER) the subjects had to form utterances based on different semantically unrelated drawings. The syntactic frame was not defined by the researcher, therefore, the only requirement was that the drawing and the corresponding colour appeared in the utterance (e.g., *The rose is red*; or *The blue shirt is very modern*).

The fifth task (STORY) was an invented story narration. The subject had to make up a story based on five unrelated drawings and none of them were allowed to be omitted. As in the case of previous tasks, each participant had one minute to prepare.

In conclusion, the selected tasks consisted of descriptions of static objects and constellations with visual support (ROOM1, ROOM2, UTTER), the invented story narration with visual support (STORY), as well as story telling (CARTOON).

3. Study Results

The participants of the study produced a total sum of 56 862 words, Table 1, that is, almost ten hours of recorded speech were collected.

Table 1. Word Count and Speech Duration

	ALL TASKS	CARTOON	ROOM1	ROOM2	UTTER	STORY
r	**56,862**	24,536	6656	6118	13,068	6484
$t_m[min]$	**598.058**	233.803	61.865	47.472	186.775	68.143

Note. r—word count, t_m—speech duration in minutes.

The analysis of the influence of task type on speech rate, being one of the best predictors of perceived fluency, is performed on the same subjects in each task (101 undergraduate students). Consequently, there are five dependent samples.

Table 2 shows descriptive statistics for the speech rate x in L2 defined as the number of words per minute (*words/min*). The main reason for such a detailed data presentation is a requirement for a good sample knowledge, in order to decide which samples to compare with each other, if the statistical test results display that the populations from which the samples arise, are not identical.

The Shapiro-Wilk test for normality (Marques de Sá, 2007) is selected, Table 3, highlighted in a number of papers as the most efficient normality test (Thode, 2002; Zhang & Yuehua, 2005; Keskin,

2006; Henderson, 2006; Coin, 2008). According to this test, it may be noticed that the distributions of some populations (ROOM1, STORY) significantly deviate from the normal distribution, Table 3. Therefore, the Friedman test is selected (Field, 2005; Marques de Sá, 2007; Demšar, 2006), as probably the most popular (Van de Wiel, 2004) nonparametric test for repeated measures in the case of three or more dependent samples, Table 4. Nonparametric tests are known as distribution-free tests, which require neither normally distributed data, nor homogeneity of variance (Montgomery & Runger, 2003). However, the drawback of the nonparametric tests is that they are less efficient than the parametric ones, but only if the assumptions for using parametric tests are fulfilled, which is not the case with regard to the results of the Shapiro-Wilk test, Table 3. If the assumptions for using parametric tests are not met, the nonparametric tests often provide considerable improvement (Montgomery & Runger, 2003). The p-value of the Friedman test is lower than 0.0001, Table 4, displaying that significant differences in the rate of speech among individual tasks have been obtained. Thus, all populations are not identical considering that the p-value is lower than or equal to the threshold value α (p < 0.0001 ≤ α = 0.05).

Table 2. Descriptive Statistics for the Speech Rate in L2

Speech rate	*CARTOON*	*ROOM1*	*ROOM2*	*UTTER*	*STORY*
n	101	101	101	101	101
x_{min}	59.062	55.365	72.941	31.510	39.574
D_1	73.062	72.162	95.961	51.087	69.201
Q_1	88.794	90.554	109.687	58.527	84.858
Me	105.106	107.865	132.719	71.555	98.438
Q_3	125.077	135.155	153.778	86.526	121.120
D_9	137.997	162.361	173.502	96.704	144.793
x_{max}	197.336	231.148	241.148	123.328	222.615
d	138.274	175.783	168.207	91.818	183.041
d_Q	52.015	44.601	44.091	27.999	36.262
$\bar{X}$	106.658	114.717	134.037	73.002	103.019
s	25.082	34.056	30.669	19.085	29.727
s^2	629.107	1159.811	940.588	364.237	883.695
$S_{\bar{X}}$	2.496	3.389	3.052	1.899	2.958
V	23.52%	29.69%	22.88%	26.14%	28.86%
α_3	0.473	0.622	0.566	0.388	0.845
α_4	0.601	0.427	0.561	-0.090	1.831

Note. n—sample size, x_{min}—sample minimum, D_1—the first decile, Q_1—lower quartile, Me—median, Q_3—upper quartile, D_9—the ninth decile, x_{max}—sample maximum, d—sample range, d_Q—interquartile range, $\bar{X}$—sample mean, s—sample standard deviation, s^2—sample variance, $S_{\bar{X}}$—standard error of

the mean, V—coefficient of variation, α_3—skewness, α_4—kurtosis.

Table 3. Shapiro-Wilk Test for the Speech Rate in L2

Shapiro-Wilk	CARTOON	ROOM1	ROOM2	UTTER	STORY
W	0.9793	0.9696	0.9773	0.9810	0.9596
p	0.1134	0.0197	0.0788	0.1530	0.0035

Note. W—Shapiro-Wilk test statistic, p—the p value.

Table 4. Friedman Test for the Speech Rate in L2

Friedman	n	df	F_r	p
	101	4	249.306	< 0.0001

Note. n—sample size, df—degrees of freedom, F_r—Friedman's test statistic, p—the p value.

The results of the Friedman test reveal that the null hypothesis of identical populations should be rejected, yet, it does not answer the question which tasks, when compared, display significant differences in speaking rates. Dunn's multiple comparison test (Daniel, 1990; Pett, 1997), an effective test with a careful assessment of statistically significant differences between compared pairs (Pett, 1997), answers this question. Comparisons can be performed only for selected pairs of tasks or for all possible pair combinations. In the case of implementing multiple comparisons with a threshold value α = 0.05 for each pair, it is necessary to take into account that the greater number of comparisons considerably increases the overall probability of error type I, that is, the probability of rejecting the null hypothesis when it is true in at least one case of comparison (e.g., it is 0.226 for 5 compared pairs, whereas in the case of 10 comparisons, which is the maximum number in the case of 5 samples, it exceeds 0.4). The solution may be to lower the threshold of significance for each comparison to a value which ensures that the overall probability of error type I does not exceed 0.05. On the other hand, by reducing the threshold value for each comparison, the risk of error type II increases (null-hypothesis is not rejected, even though the alternative hypothesis is correct), which leads to the conclusion that it is useful to compare only the chosen pairs. Thus, the risk of error type II is reduced. The decision which pairs to compare can be made based on good knowledge of descriptive statistics related to each sample, Table 2. In conclusion, it is important to be restrictive in choosing comparison pairs (Marques de Sá, 2007).

Smaller discrepancies of sample values can be noticed between CARTOON and ROOM1 as well as between CARTOON and STORY, Table 2. Therefore, Dunn's multiple comparison test is performed for all other task pairs, Table 5. As can be seen, significant differences between the speech rates, and thus the perceived fluencies, are obtained for all analyzed pairs.

Table 5. Dunn's Multiple Comparison Test for the Speech Rate in L2

Dunn	Difference in rank sum	Significant difference
CARTOON-ROOM2	-149	Yes
CARTOON-UTTER	197	Yes
ROOM1-ROOM2	-117	Yes
ROOM1-UTTER	229	Yes
ROOM1-STORY	71	Yes
ROOM2-UTTER	346	Yes
ROOM2-STORY	188	Yes
UTTER-STORY	-158	Yes

4. Discussion

The obtained results on perceived fluency by comparing different tasks confirm that the perception of L2 speakers' fluency varied across tasks.

The slowest speech rate has been noticed in the formulation of utterances based on semantically unrelated drawings (UTTER), Table 2. Also, a significant difference is obtained for the speech rate between UTTER and all other tasks, Table 5. In UTTER, the participants required more planning time to formulate unrelated utterances, and since some of the words were low-frequency words, even more planning time was needed to search for the particular lexical unit in the mental lexicon. However, it must be emphasized that other researchers pointed out the existence of a strong and consistent bond between the topic of the utterance and the fluency.

In Levelt's (1989) terms, the model of speech production includes four main processes: a) the conceptualization or the planning of the content; b) the formulation, which includes the grammatical, lexical and phonological coding of messages; c) the articulation, which is the production of words; and d) the monitoring which involves the verification of the accuracy or appropriateness of the produced utterance. The preverbal message is forwarded to the formulator and the conceptualizer can start working on the next part of the message, regardless of whether the previous part of the message is still being processed. As a consequence, the articulation of the utterance can begin much earlier than the speaker finished with the planning of the whole message. In other words, the conceptualizer does not hold the process of message planning waiting for the formulator and articulator to catch up with it. Native speakers achieve parallel processing of planning ideas and articulated speech because most of the mechanisms in the production of L1, particularly in the coding phase, are fully automated. Therefore, the incremental, parallel and automated processing nature can explain the high speech rate. On the other hand, speaking a foreign language involves many processes, some of which are not automatized, and which are competing for the limited capacity resources of the cognitive system. Also, foreign language learners do not have automatized linguistic knowledge and fast retrieval of

prefabricated language formulae. Therefore, the speech rate can significantly decrease when the conceptualizer, formulator and articulator compete for the limited attentional resources. The speakers alternate between phases of low fluency, that is, between the phase when macro-planning occurs, and high fluency, when the conceptual structure or the macro-plan is translated into a linguistic structure.

Skehan's (1998) limited-attention model argues that there is a large burden on the learner's attention and choices must be made which aspect of performance to prioritize. A frequent introduction of different topics, which is the case in UTTER, demands a great amount of attention in the conceptualization phase.

By way of analogy, when the attentional load gets too great, the learners need to increase their pausing time since the formulator needs to catch up with the conceptualizer. When topics frequently change, the phases of low and high fluency alternate, causing the different processes to compete for the capacity of shared resources. Since speech production includes message planning and execution, the speech slows down during the planning phase, whereas fluency increases during the phase of speech execution.

Roberts and Kirsner (as cited in Kormos, 2006) explained the cyclic nature of speech production. In their view, in spontaneous speech fluent and disfluent phases interchange, and those cycles are regular and periodical. They also concluded that the speech rate decreases before the introduction of a new topic, which is the case in UTTER, since every sentence represents a new topic, whereas the production speeds up after the new topic has already been introduced. Their results indicate that speech planning is a controlled process requiring attention and therefore slow. Conversely, speech processing in L1 is largely automatic and fast, and the difference between automatic and controlled processing explains the exchange of fluent and nonfluent cycles. The results of this study confirm the findings of Roberts and Kirsner (as cited in Kormos, 2006).

The fastest speech rates can be observed in the description of a room picture, that is, in the description of static objects and constellations (ROOM1 and ROOM2). In the room description tasks special attention is paid to the linearization process of macro-planning, that is, to the organization of natural order (Levelt, 1989). Such tightly structured tasks require less attentional capacity and therefore attention can be directed towards other aspects of performance, such as fluency. According to Levelt, the processing burden for the conceptualizer decreases due to the structured nature of the task, consequently, more attention is left for the formulator and articulator.

Also, the significant difference for the speech rate is achieved between ROOM2 and all other tasks, Table 5. In ROOM2, which is a repeated task with a predetermined structure and a visual support, the memory is less loaded than in other tasks. A positive influence of the task repetition can be observed in terms of speedy and smooth delivery. Also, the already familiar information requires less attention to process and thus more fluent L2 performance can be achieved. Bygate (2001) also claims that task repetition could influence learners' oral performance by relocating their focus of attention.

In STORY the speakers had to invent a story based on five unrelated drawings. There was no tight storyline structure including the beginning, the middle and the end of the story, which would certainly

decrease the processing load. Therefore, the cognitive complexity and the needed planning time considerably contributed to a performance that was significantly less fluent compared to the description of a room picture (ROOM1).

It is important to identify the specific features of distinctive tasks. In foreign language teaching it is necessary to employ a variety of different tasks (familiar vs. unfamiliar, structured vs. unstructured, repeated vs. unrepeated tasks, etc.) that set in motion different skills in order to enhance overall fluency as a vital and integral aspect of successful communication. The findings of this study suggest that emphasis should be laid on tasks including frequent introductions of new topics, due to the observed decrease in perceived fluency.

References

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks second language learning and testing* (pp. 23-48). Harlow: Longman.

Coin, D. (2007). A goodness-of-fit test for normality based on polynomial regression. *Computational Statistics and Dana Analysis*, *52*, 2185-2198.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America*, *111*, 2862-2873.

Daniel, W. W. (1990). *Applied Nonparametric Statistics*. PWS-Kent Publishing Company.

Demsar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research*, *7*, 1-30.

Derwing, T. M., Murray, J. M., & Thomson, R. I. (2007). A longitudinal study of ESL learners' fluency and comprehansibility development. *Applied Linguistics*, *29*, 359-380.

Field, A. (2005). *Discovering Statistics using SPSS*. SAGE Publications: UK.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performances. *Studies in Second Language Acquisition*, *18*, 299-323.

Henderson, A. R. (2006). Testing experimental data for univariate normality. *Clinica Chimica Acta*, *366*, 112-129.

Housen, A., & Kuiken F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, *30*, 461-473.

Keskin, S. (2006). Comparison of Several Univariate Normality Tests Regarding Type I Error Rate and Power of the Test in Simulation based Small Samples. *Journal of Applied Science Research*, *2*(5), 296-300.

Kormos, J. (2006). *Speech Production and Second Language Acquisition*. Lawrence Erlbaum Associates, New York London.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of

second language learners. *System*, *32*, 145-164.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25-42). Ann Arbor: University of Michigan Press.

Levelt, W. J. M. (1982). Linearization in describing spatial networks. In S. Peters, & E. Saarinen (Eds.), *Processes, beliefs, and questions*. Dordrecht: Reidel.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Marques de Sá J. P. (2007). *Applied Statistics Using SPSS, STATISTICA, MATLAB and R*. Springer Verlag: Berlin Heidelberg.

Mehrang, F., & Rahimpour, M. (2010). The impact of task structure and planning conditions on oral performance of EFL learners. *Procedia Social and Behavioral Sciences*, *2*, 3678-3686.

Montgomery, D. C., & Runger, G. C. (2003). *Applied Statistics and Probability for Engineers*. John Wiley and Sons: USA.

Pett, M. A. (1997). *Nonparametric Statistics for Health Care Research* (2nd ed.). Sage Publications.

Pinget, A.-F. (2011). *Native speakers' perceptions of fluency and accent in L2 speech*. MA Thesis, Utrecht University.

Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). Cambridge: Cambridge University Press.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching (IRAL)*, *43*(1), 1-32.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

Segalowitz, N., & Freed, B. F. (2004). Context, contact and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, *Special Issue*, *26*(2), 173-199.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influence on foreign language performance. *Language Teaching Research*, *1*, 185-211.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*(1), 93-120.

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-277). Amsterdam: Benjamins.

Thode, H. C. (2002). *Testing for normality*. Marcel Dekker: New York.

Ullakonoja, R. (2009). Speech rate as an indicator of fluency in the Russian of Finnish learners. In M. O'Dell, & T. Nieminen (Eds.), *The Phonetics Symposium 2008. Tampere Studies in Language, Translation and Culture, Series B* (pp. 97-109).

251

van de Wiel, M. A. (2004). Exact null distributions of quadratic distribution-free statistics for two-way classification. *Journal of Statistical Planning and Inference*, *120*, 29-40.

Zhang, J., & Yuehua, W. (2005). Likelihood-ratio tests for normality. *Computational Statistics and Data Analysis*, *49*, 709-721.