



# *Predicting L2 fluency from L1 fluency behaviour: the case of L1 Turkish and L2 English speakers*

Article

Accepted Version

Duran-Karaoz, Z. and Tavakoli, P. (2020) Predicting L2 fluency from L1 fluency behaviour: the case of L1 Turkish and L2 English speakers. *Studies in Second Language Acquisition*. ISSN 1470-1545 doi: <https://doi.org/10.1017/S0272263119000755> Available at <http://centaur.reading.ac.uk/87564/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1017/S0272263119000755>

Publisher: Cambridge University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in

the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# **Predicting L2 fluency from L1 fluency behaviour: The case of L1 Turkish and L2 English speakers**

## **Abstract**

The article reports on the findings of a study investigating the relationship between first language (L1) and second language (L2) fluency behaviour. Drawing on data collected from Turkish learners of English, the study also addresses the question of whether proficiency level mediates the relationship, if any. The data were coded for a range of breakdown, repair, speed and composite measures. Language proficiency was measured by means of two tests: Oxford Placement Test and an Elicited Imitation Task. The results show that some breakdown and repair measures were positively correlated in L1 and L2, but no correlations were observed for articulation rate and speech rate. The relationships were not mediated by proficiency level. Regression analyses show that a number of models predicted L2 fluency. L1 fluency contributed significantly to models predicting pausing behaviour; EIT scores predicted L2 speech rate; and L1 fluency and OPT scores predicted L2 reformulation and end-clause pauses. The important implications of the findings for fluency research and second language pedagogy are discussed.

Keywords: fluency; second language oral performance, second language proficiency

## 1. Introduction

Research in second language (L2) oral fluency is inspired by at least two convincing rationales. First, recognizing the importance of fluency as a key construct of L2 ability, researchers have attempted to examine fluency in relation to other dimensions of linguistic output such as syntactic complexity and accuracy. Second, researchers are interested in fluency as a means towards developing a more profound understanding of L2 processing and production. Regardless of rationale and motivation, researchers in this area are most likely to agree that a) fluency has a complex nature that is difficult to conceptualize and measure, b) our current knowledge of what psychological and social factors contribute to fluency is limited (see Segalowitz, 2016, for a detailed discussion), and c) more research is needed to extend our understanding of fluency.

In its short history, oral fluency research has yielded some important findings that help us develop not only a more in-depth understanding of fluency and its components, but also a more reliable measurement of it. These findings also point at emerging evidence to suggest that L2 fluency behaviour is, at least to some extent, influenced by L1 behaviour and individual differences (de Jong, Groenhout, Schoonen, & Hulstijn, 2015; Derwing, Munro, Thomson, & Rossiter, 2009; Huensch & Tracy-Ventura, 2016). This body of research implies that the amount of pausing and hesitation in L2 speech might, to some extent, mirror L1 fluency behaviour, and that L2 fluency can, at least to some degree, be predicted from L1 fluency behaviour. These findings, however, remain rather limited as few languages, language typologies and individual differences have been examined so far.

One of the shortcomings in the current understanding of fluency is its relationship with proficiency level (Tavakoli, Nakatsuhara & Hunter, 2017). While it

is commonly agreed that fluency increases as proficiency improves, little research has examined whether the relationship is linear (*ibid.*). More central to the focus of this paper is the question of whether L2 proficiency level plays a role in the relationship between L1 and L2 fluency. To the best of our knowledge, there has been little research investigating whether the relationship between L1 and L2 fluency persists consistently across assessed levels of proficiency, or to what extent L2 speakers' fluency behaviour deviates from their L1 fluency behaviour when they become more proficient L2 users. The current study aims to help fill this gap by examining the relationship between L1 Turkish and L2 English fluency behaviour, and by probing whether proficiency level mediates this relationship.

## **2. L2 Fluency Research**

Fluency in its broad sense is 'a cover term for oral proficiency' (Lennon, 1990, p. 398), and is sometimes interchangeably used for overall proficiency. In the narrow sense of the term, however, fluency is regarded as a component of oral communicative ability, both temporal and non-temporal aspects, that influence how smoothly and effortlessly speech is delivered (Derwing et al., 2009). The present study is interested in fluency in the latter sense of the term. While we do not aim to add another definition to the literature, we find Koponen and Riggensbach's (2000) definition useful as it considers fluency as "flow, continuity, automaticity, or smoothness of speech" (Koponen & Riggensbach, 2000, p. 6).

Research in L2 fluency has shed light on the complex and multifaceted nature of fluency. Segalowitz (2010), in his triadic framework, proposes that fluency consists of three aspects of cognitive, utterance, and perceived fluency. *Cognitive fluency* is the "ability to efficiently mobilize and integrate the underlying cognitive processes

responsible for producing utterances” (Segalowitz, 2010, p. 48); *utterance fluency* refers to the temporal characteristics of speech that can be measured objectively (e.g., speed, pauses, or hesitations); and *perceived fluency* concerns the listener impression of the cognitive fluency of a speaker. While these three aspects are equally important and are inevitably inter-related, *utterance fluency* has been central to research in second language acquisition (SLA) (e.g., de Jong et al., 2016; Huensch & Tracy-Ventura, 2017; Kormos & Denes, 2004; Thai & Boer, 2016; Wright, 2013 to name a few) as its concrete and measurable nature allows objective and systematic examination. A new classification of different aspects of fluency is proposed by Skehan, Foster and Shum (2016) who argue that utterance fluency should be divided to speed, clause-level and discourse-level. They propose that clause processing, which focuses on lexical choices occurring within a clause, is linked to mid-clause pausing and influenced by the work of the Formulator and Articulator in Levelt’s model (Levelt, 1989, see below for a brief discussion). In contrast, discourse-level processing, occurring above the clause level and dealing with issues and problems related to conjoining linguistic units, should be considered a discourse level fluency. The authors maintain that discourse-level disfluencies are linked to the Conceptualizer’s demands, while the clause-level disfluencies are evidence of micro-planning and connected to the Formulator. Given that Skehan’s et al.’s proposal still needs empirical evidence from L2 research to be validated, we will adopt Segalowitz’s (2010) classification of fluency in this study.

For measurement purposes, *utterance fluency* is generally divided into three categories: breakdown, speed and repair (Skehan, 2003; Tavakoli & Skehan, 2005). *Breakdown fluency* concerns the flow of speech and examines pauses in terms of *amount*, *location* (mid versus end-clause position) and *character* (filled versus silent).

*Speed fluency* refers to the speed with which speech is produced (e.g., number of syllables per second), and *repair fluency* indicates repair strategies speakers employ to monitor and modify their utterances (e.g., false starts, repetitions or corrections). A useful distinction made in this area is between pure and composite measures of fluency (de Jong et al., 2015; Hunter, 2017; Tavakoli et al., 2017). While composite measures, i.e. those combining two or more aspects of fluency (e.g., speed *and* breakdown fluency in Mean Length of Run), are known to be linked with human judgement of fluency (e.g., Boer, Eyckmans, Kappel, Stengers & Demecheleer, 2006; Kormos & Denes, 2004; Prefontaine & Kormos, 2016), pure measures, i.e. those examining only one aspect of fluency (Skehan, 2014), are believed to reveal underlying processes of speech formulation and production (e.g., Huensch & Tracy-Ventura, 2017) and can help provide a more nuanced picture of the speech production processes.

L2 fluency research has also gained currency in view of the need for developing an L2 speech production model. Many researchers (de Bot, 1992; Kahng, 2014; Kormos, 2006; Skehan, 2009) have taken Levelt's (1989, 1999) model of L1 speech production as a starting point. According to this model, during the process of speech production a speaker goes through three stages of *conceptualization*, *formulation* and *articulation*. During *conceptualization*, the speaker generates a pre-verbal message that then moves to the *formulation* stage, where an original linguistic message is formulated through grammatical and phonological encoding of lemmas accessed in the lexicon. This then moves to the *articulation* stage, where the linguistic plan is converted into overt speech. de Bot (1992) and Kormos (2006) claim that although similar processes may be in place between the two language processing systems, there is a range of L2-related phenomena that should be carefully considered

in the development of an L2 speech processing model, including differences in L1 and L2 speech development and cross-linguistic influences. Kormos (2006) argues that for L2 processing, since L2 speech may not yet have become automatic, an additional L2-specific knowledge store is needed to act as a source of declarative memory to process ‘knowledge of syntactic and phonological rules’ (p.178). In effect, the degree of automaticity in speech production highlights one of the key differences between L1 and L2 processing (Kormos, 2006; Tavakoli, 2019).

In order to develop a better conceptualization of L2 speech processing, therefore, it is necessary to examine the extent of the relationship between L1 and L2 fluency and the factors that may affect this relationship, e.g., individual differences such as language proficiency. Derwing et al. (2009) argue that insight can be gained through an investigation of automatic processing in both languages or through a careful investigation of fluency in L1 and L2 speech production. The present study is primarily interested in the latter, i.e., examining fluency in L1 Turkish and L2 English speakers. A brief overview of relevant research in this area is presented below.

### *2.1. Research in L1 and L2 Fluency*

Studies focusing on the relationship between L1 and L2 fluency are relatively scarce (e.g., de Jong et al., 2015; Derwing et al., 2009; Huensch & Tracy-Ventura, 2016; Riazantseva, 2001; Towell & Dewaele, 2005; Towell, Howkins & Bazergui, 1996). One of the earliest studies on this topic is Riazantseva (2001). Comparing data elicited from advanced and intermediate Russian learners of L2 English, Riazantseva examined the *pausing phenomena* in L1 Russian and L2 English. Measuring pauses of longer than 100ms, she found that the intermediate group exhibited more L1-like behaviour than the advanced group, and interpreted the differences in the light of L1



transfer, concluding that L2 speakers may overcome these effects as their L2 proficiency develops. The study design, however, did not include low-proficiency learners, nor did it examine the relationship between pausing behaviour and other aspects of fluency.

In a longitudinal study of L2 immigrant learners in Canada, Derwing et al. (2009) investigated the extent to which temporal features of L1 speech can be related to L2 fluency characteristics. They collected data over 2 years (at 2 months, 10 months and 1 year of residency in Canada) with 16 L1 Slavic and 16 L1 Mandarin speakers of L2 English, performing the same task in their L1 and L2. The data were analysed for a number of temporal measures (i.e. *number of pauses per second*, *speech rate*, *pruned syllables per second*) and rated by trained judges. The fluency ratings demonstrated correlations between L1 and L2 measures, highlighting a stronger correlation for the Slavic L1 group than for the Mandarin L1 group. The results also indicated that L1 and L2 measures correlated significantly for both Slavic and Mandarin languages during the initial stages of language exposure, but at later stages only for the Slavic group. The authors concluded that the group differences might be attributed to different amounts of exposure to English or to a ‘closer relationship between Slavic languages and English than between Mandarin and English’ (p.534).

Perhaps the strongest evidence about the relationship between L1 and L2 fluency behaviour comes from two recent studies: de Jong et al. (2015) and Huensch and Tracy-Ventura (2016). Working with learners of Dutch as an L2, de Jong et al. (2015) used two types of fluency measures, i.e. uncorrected measures and measures corrected for L1 behaviour, to explore the relationship between L1 and L2 fluency. They postulated that since the original L2 measures could be testing both personal speaking styles and learners’ L2-specific skills, correcting these measures by

adjusting for L1 fluency behaviour would lead to a more accurate measurement of fluency. They examined two typologically different L1s, i.e. English and Turkish, and elicited language samples through eight different tasks from intermediate to advanced proficiency learners. Their results showed that all L2 fluency measures were, to a certain extent, related to L1 fluency, with the variance ranging from 21% to 57%. The authors also reported that although all measures (i.e. both corrected and uncorrected) predicted proficiency significantly, the corrected measure of *syllable duration* predicted L2 proficiency more strongly than its original uncorrected equivalence, concluding that adjusting this measure for L1 fluency would lead to more accurate results.

In another study, Huensch and Tracy-Ventura (2016) examined the contribution of L1 fluency behaviour, cross-linguistic differences and L2 proficiency to L2 fluency behaviour over time in a Study Abroad context. The participants, L1 English speakers who were studying Spanish and French for their university degree, performed an oral narrative task three times: twice in their L2s (once before travelling abroad and once after residing abroad for 5 months), and a last time in their L1 after they had returned home from Study Abroad. The results suggested that “L1 fluency behaviour, cross-linguistic differences, and proficiency differentially contributed to explaining L2 fluency behaviour prior to and during immersion” (Huensch and Tracy-Ventura, 2016, p. 2). They reported that proficiency level predicted L2 fluency measures of mean syllable duration and mean silent pause duration before residing abroad; however, after 5 months of study-abroad no influence of proficiency level was observed for any of the L2 fluency measures employed. These results are particularly noteworthy as they show a clear relationship between L1 and L2 fluency and moreover broaden our perspective on the change in fluency behaviour as proficiency

increases. However, this study neither examined different proficiency levels systematically nor did it establish whether proficiency level mediates the relationship between L1 and L2 behaviour. Like Huensch and Tracy-Ventura (2016), most studies examining the relationship between L1 and L2 fluency were performed with learners in a Study Abroad context or participants using and learning L2 in the target language community for communication purposes (de Jong et al., 2015; Derwing et al., 2009; Leonard & Shea, 2017; Di Silvio, Diao & Donovan, 2016). Given the research evidence that fluency is one of the key aspects of the L2 use to benefit from Study Abroad or life in the L2 community (Mora & Valls-Ferrer, 2012; Tavakoli, 2018; Wright, 2013; 2018), it seems necessary to conduct studies that investigate the relationship between L1 and L2 fluency among speakers who have not had substantial exposure to the L2 or authentic opportunities for using it for genuine communication purposes.

A final study to report here is Peltonen (2018) who examined L1 and L2 fluency behaviours among Finnish children learning English at school. Working with 42 participants from two school levels, she examined a range of fluency measures and reported positive correlations between the majority of temporal L1 and L2 measures. In addition, the study suggested that many L2 temporal measures can be predicted from L1 behaviour. While the participants in this study belonged to two levels, Peltonen (2018) acknowledged that the two groups may not be distinctive in their proficiency level, and therefore, the results may not show whether proficiency level mediates relationship between L1 and L2 fluency behaviour

Taken together, the results of these studies suggest there is a link between L1 and L2 fluency behaviour, and imply that relationships between the two would change over time as proficiency increases. However, these results are neither conclusive nor

generalizable to different L1s (de Jong et al., 2015; Derwing et al., 2009). Further research is needed to demonstrate whether and in what ways the L1-L2 fluency relationship changes when learners' proficiency improves. Before discussing the study, it is necessary to highlight the role of proficiency in understanding the relationship between L1 and L2 fluency.

### *2.2. Second language proficiency and fluency*

Language proficiency is an individual learner difference commonly examined in L2 studies due to its direct relationship with L2 acquisition and development. Language proficiency, or as Hulstijn (2015) defines it, “knowledge of language and the ability to access, retrieve and use that knowledge in listening, speaking, reading and writing” (pg.21), is of particular interest to the current study as it can help develop a better understanding of the relationship between L1 and L2 fluency. In L1, speech production relies largely on incremental, parallel and automatic processing (Levelt, 1989). The parallel and automatic nature of L1 processing helps make speech fluent with little undue hesitation or disruption. In contrast, L2 speech production, particularly at lower levels of proficiency, does not operate on automatic processing. Working with incomplete linguistic knowledge, the L2 speech production process faces several challenges including access and retrieval of linguistic units and monitoring the language during and after production (Kormos, 2006). This controlled and serial, rather than parallel, processing is often marked by signs of disfluency, e.g., slower speech rate and frequent pauses. When proficiency improves, whether due to exposure, practice or instruction, linguistic knowledge expands and parallel processing becomes possible, the production process becomes more automatic and speech becomes more fluent. Therefore, it can be hypothesized that at higher levels

of proficiency an increase in automaticity makes parallel processing possible, freeing up attentional resources available to the speaker, and L2 speech processing becomes more similar to L1 speech processing. This implies that as language proficiency improves, one might expect a stronger relationship between L1 and L2 fluency behaviour. In contrast, for less proficient speakers, this relationship may be overshadowed by the L2 speech production challenges (e.g., serial processing, and access and retrieval demands), and therefore, two different processing systems can be observed. While research in this area already **recognizes** that L2 fluency is at least partly related to L1 fluency (de Jong, et al., 2015; Peltonen, 2018), it is necessary to understand whether this relationship persists as proficiency develops. This is what the current paper aims to investigate.

### **3. Research Aims**

The current study attempts to expand our knowledge of the relationship between L1 and L2 fluency in three regards. First, it aims to examine the relationship between L1 Turkish and L2 English in order to contribute to the current understanding of fluency research in this area. Previous studies have shown that typological similarities and differences between two languages seem to affect the relationship between L1 and L2 fluency behaviour (de Jong et al., 2015; Huensch & Tracy-Ventura, 2016). Structural differences between Turkish and English can also be expected to affect the results. In Turkish, a highly agglutinative language, words can have several affixes (mostly suffixes) to reflect different meanings or grammatical functions. For example, an English sentence or phrase, such as ‘they will come’, can be expressed in a single word in Turkish, with ‘come’ as the morpheme (i.e. ‘gelecekler’). Given that speed is usually calculated in terms of number of syllables per minute, it is not expected to be

affected by the structure of words and utterances. However, in an agglutinative language, it is common to produce a word of several syllables to represent a whole sentence, and this is likely to result in speaker's producing more syllables in Turkish than in English in the same amount of time, which might have an influence on speed. Given that the current study is not focusing on cross-linguistic differences between the two languages, this is one of the several questions about L1-L2 fluency relationship that future research will need to address.

Second, the study aims to examine whether L2 proficiency level mediates the relationship between L1 and L2 fluency. Compared to most previous studies that have looked at only one or two levels of proficiency (de Jong et al, 2015; Derwing et al., 2009; Huensch & Tracy-Ventura, 2016), the current study is one of the few studies in which participants come from three different proficiency levels (A2, B1 and B2 of the CEFR). In addition, from a methodological point of view, we are extending the existing research framework by examining proficiency level from a broader perspective. Research in SLA suggests there are two underlying constructs in language proficiency, i.e. declarative and procedural knowledge. However, most of the research in fluency studies examined proficiency in terms of either of the two kinds. This is a shortcoming of research in this area since recent research (e.g. [Tavakoli et al., 2017](#)) showed that certain aspects of fluency might be linked to different kinds of linguistic knowledge. Our rationale for using measures of language proficiency that test both kinds of knowledge was based on the assumption that a more complete profile of the learner proficiency would provide us with a more valid interpretation of the role of proficiency in the relationship between L1 and L2 fluency.

Following research in this area (Gaillard & Tremblay, 2016), we employ the Elicited Imitation Task (EIT) and the grammar section of the Oxford Placement Test

(OPT). We assume that the first is more likely to test the participants' procedural knowledge while the latter it is more likely to test their declarative knowledge.

The following research questions guide our study:

(1) Is there a relationship between L1 Turkish and L2 English fluency behaviour? If so,

(1a) To what extent does language proficiency mediate the relationship between L1 and L2 fluency behaviour?

(2) To what extent can L2 fluency measures be predicted from L1 fluency behaviour?

#### **4. Method**

The study had a factorial design with language proficiency as the independent variable, and different measures of fluency as the dependent variables of the study. Language proficiency was between-participant variables, representing three levels (A2, B1 and B2 according to the CEFR).

##### *4.1. Participants*

The data was initially collected from 44 participants who volunteered to take part in the study at the time of the data collection; however, data from two of the participants were removed at a later point since they did not meet all task completion requirements. The data reported therefore here comes from 42 native speakers of Turkish undergraduate students (25 females and 17 males) aged between 19 and 25. They were taking English courses at a state university in Turkey either as part of their foundation classes or their degree programme (i.e. as a compulsory course of General English).

They had varying levels of English at the start of their programmes and had been on these courses for five months at the time of the study.

Previous research (Mora & Valls-Ferrer, 2012; Tavakoli, Campbell & McCormack, 2016) has shown that life/study abroad experiences and professional use of L2 would have a positive impact on fluency helping it develop fast. Therefore, in order to explore our research questions and to ensure homogeneity of the sample, we needed a group of participants whose L2 fluency was not affected by such experiences. To control for this, a short demographic questionnaire, available on IRIS (<https://www.iris-database.org/iris/app/home/index>), was used before the data collection, and data from anyone with such a profile were excluded.

#### *4.2. Instruments*

*Oral Narrative Tasks.* For the purposes of this study, two oral narrative tasks were used (see Appendix 1). Following de Jong and Vercelloti's (2016) recommendations for careful selection of "features that constitute task complexity" (p. 387), a number of criteria were considered when selecting the narratives. These included number of characters and props (de Jong & Vercelloti, 2016), amount of contextual support (Revesz, 2009), similar storyline complexity (Tavakoli & Foster, 2011), and similar amount of intentional reasoning (Awwad, Tavakoli & Wright, 2017). Although care was taken to choose comparable tasks, we are aware that the two narratives were different and as such a task design impact might be detected. We ran t-tests to check for any possible effects of task design.

*Proficiency Tests.* Proficiency is measured rather variably in L2 studies. This measurement ranges from studies using C-tests and vocabulary size tests to those



employing internationally standardized and validated four-skills (reading, writing, speaking and listening) tests. In the current study, our measurement of proficiency was based on the definition of proficiency provided earlier (Hulstijn, 2015: 21). Following Hulstijn's (2015) definition, we examined language proficiency along the two dimensions of "knowledge of language" or *declarative knowledge* and "the ability to use the language" or *procedural knowledge* (for further discussion see Section 3 below). To achieve this aim, we used the grammar section of the *Oxford Placement Test* (Alan, 2004), and the *Elicited Imitation Task* (Ortega, Iwashita, Norris & Rabie, 2002; Tracy-Ventura, McManus, Norris & Ortega, 2014) to measure the participants declarative and procedural knowledge respectively<sup>1</sup>. EITs are increasingly used in L2 studies to measure *procedural knowledge* (Yan, Maeda, Lv & Ginther, 2015) and as evidence of the participants' interlanguage (Ellis, 2005; Kim, Tracy-Ventura & Yung, 2016; Yan et al., 2015). In an EIT, participants are asked to repeat a set of sentences of varying length and complexity usually after listening to it once. Also, since it is dependent on fast language processing and producing speech in real time, the EIT is arguably more suited to measuring procedural oral language ability, i.e., both implicit and speeded-up explicit knowledge types (Suzuki & DeKeyser, 2019)<sup>2</sup>.

The EIT used in the current study consisted of a total of 30 sentences in English designed to test L2 proficiency with a combination of grammatical features, syntax and vocabulary. These sentences varied in length (between seven to nineteen

---

<sup>1</sup> While a grammar test is primarily a means of activating declarative knowledge, we are aware that it can, at least to some extent, tap into procedural knowledge as well. For research purposes in our field, however, a grammar test is usually used as a measure of declarative knowledge (references).

<sup>2</sup> Given that we do not distinguish between implicit knowledge and speeded-up explicit knowledge in this study, the use of EIT seems well justified for measuring procedural knowledge of both kinds.

syllables), ordered from the fewest syllables to the most, and were adopted from Ortega et al. (2002). The task also involved a practice session using additional 5 Turkish sentences at the beginning of the test to make sure the procedures were well-understood and followed. The results of the proficiency tests are discussed below.

#### *4.3.Procedures*

The data were collected on two separate days. On the first day, all participants sat the OPT in a classroom, and the EIT individually in a different room. Individual meetings were then arranged for the participants in the same week to perform the oral narratives in L1 and L2. In the implementation of the EITs, the procedures in Gaillard and Tremblay (2016) were followed. Following the practice session in Turkish, the participants were presented with the English sentences one by one while they were recorded. They were asked to listen to each of the sentences, which was followed by a beep sound and then were required to repeat it. There were two seconds between the end of each sentence and the beginning of the beep sound, and the participants had only one attempt to repeat the sentences. The rationale for this arrangement came from previous studies (Gaillard & Tremblay, 2016; Ortega et al., 2002) that suggested this procedure in order to ensure the test takers do not merely mimic the stimuli, but rather they would process them.

On a separate day, the participants performed the tasks. They were divided randomly into two equal groups, with one group performing Task A and the other performing Task B<sup>3</sup>. They were given the picture prompts and 30 seconds of planning time, followed by 90 seconds to retell the story. No additional information about the

---

<sup>3</sup> Using two different tasks was aimed at increasing test security.

tasks or vocabulary to be used was provided. Their speech was digitally recorded as they performed the tasks. Each participant had to narrate the story twice, once in Turkish and once in English. To control for any practice effect, a counter-balanced design was used for the language in which they performed the task, i.e., half of the participants narrated the story first in L1 and then in L2, while the other half performed the tasks in the reverse order of first L2 and second L1. The details of the counterbalanced design are provided in Table 1. All the information about the purposes of the study and the instructions were provided in Turkish.

Insert Table 1 here

In order to determine the power of the study's sample size, we used GPower 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) to run a post-hoc power analysis. Running the analysis for a linear regression fixed model, we calculated the power of each individually significant regression model with an alpha level of 0.05, an effect size of 0.15, and a sample size of 42. The results showed a power of .98 for *number of mid-clause filled pauses*, .99 for *length of end-clause silent pauses*, .97 for *number of reformulations*, and .99 for *length of end-clause silent pauses*. The only significant model achieving a power below the .80 threshold level was for *number of mid-clause silent pauses* ( $1-\beta = .67$ ). The results of the power analysis suggest a good level of confidence could be maintained in the findings.

#### 4.4.Data Analysis

The OPT comprised 100 questions in total, and for each correct item, 1 point was awarded. The scoring of the EIT sentences was done based on a holistic scale (i.e.

scores ranging from 0 to 4), which was adapted from Ortega et al. (2002) and has since been employed in several studies (e.g. Kim et al., 2016; Gaillard & Tremblay, 2016; Tracy-Ventura et al., 2014; Wu & Ortega, 2013). Based on this scale, four points were given for exact repetition, three points for accurate repetitions of the sentences keeping the content meaning but including small structural changes, two for repetitions which included changes in grammar that could affect the meaning of the sentences, one for repetition of the half of the sentence, and zero when less than half of the sentence or no repetition was provided. After providing a second rater a training session on the scoring system, a total of 20% of the data was scored by this researcher. The correlation coefficient between the two sets of scores was .98. The maximum score one could obtain on the EIT was 120. For comparability reasons, the OPT scores were converted into 120 as well. There was a positive moderate correlation between the OPT and the EIT scores ( $r = .51, p < .000$ ), suggesting that a high score on the OPT was associated with a high score of the EIT. This correlation demonstrates that about 26% of the variance in participants' OPT scores can be accounted for by their EIT scores.

The mean of the EIT and the OPT scores was calculated, and this combined score was used to group the participants into proficiency levels. The groupings were based on the scoring system of the Oxford Online Placement Test i.e., each 20-scores corresponds to one level of proficiency (A1: 1-20; A2: 21-40, etc.). Table 2 shows the proficiency categorisation of **this placement test** and the participants in this study. While the maximum score one could obtain was on this test 120, our participants' combined scores ranged between 21 and 80, suggesting they belonged to the three levels of A2, B1 and B2 of the CEFR (Council of Europe, 2009), with 15, 15 and 12 participants in each group, respectively (see Oxford English Testing website for

further information about how scores are interpreted in terms of the CEFR levels). Although we did not use the proficiency level categorisation in our data analysis, we believe placing the participants into the corresponding CEFR levels of proficiency is helpful when discussing the results.

**Insert Table 2 here**

In the narrative tasks, the participants produced speech samples of varying length (Appendix 3); however, 60 seconds was used as the cut-off point for data analysis. To have comparable measures of analysis, analyses of shorter samples were corrected for time to a 60-seconds level. Following Foster, Tonkyn and Wigglesworth (2000), the data were transcribed and coded for AS-units and clause boundaries. Instances of repair fluency (i.e. repetitions, replacements, false starts and reformulations) were marked with a set of conventional symbols. Our choice of repair measures follows previous research (Hunter, 2017; Kormos, 1998; Skehan, 2003, 2009, 2015) that suggests repetitions, replacements, false starts and reformulations are examples of repair strategies that L2 learners use. Whether these repairs occur because of the conceptualization demands (e.g. amount of information) or formulation demands (e.g. lexical retrieval or syntactic encoding), they suggest that speakers are engaged in a repair process. Also, unpruned data were used in the analysis; that is to say, the syllables in repetitions, replacements, reformulations and false starts were included in the analysis.

Following Thai and Boers (2016), we used the Syllable Count programme ([www.syllablecount.com](http://www.syllablecount.com)) for English transcripts to calculate the syllable counts. We used a similar programme, Hece Hesaplama ([www.hesapla.online.com](http://www.hesapla.online.com)), to calculate

the syllable counts in the Turkish transcripts. To ensure the reliability of these programmes, 10% of these calculations for both languages were double checked by manual counts, and an inter-rater of 0.98 (Cohen's Kappa) was achieved. In order to ensure the inter-reliability for the data coding, 20% of the coded data were checked by a second bilingual (Turkish-English). Again, Cohen's Kappa statistics was performed to determine the consistency on the placement of clause boundaries, pauses and repair fluency instances and types (i.e. repetitions, replacements, reformulations and false starts), combining all of the second-coded data. For each of these, a reliability of above 0.95 was achieved.

The audio files were analysed for silent pauses and the length of all pauses using PRAAT software, which allowed for accurate measurement and multiple revisions. The data were segmented using the Annotate to TextGrid (silences) command (Boersma & Weenink, 2010). A threshold of 250 ms was chosen for all the pauses (de Jong et al., 2015).

To capture an accurate and complete picture of fluency, it was necessary to use measures that were shown to represent each aspect of utterance fluency reliably. Following recent research in this area, we chose the following seven measures. For breakdown fluency, frequency of pauses, both filled and silent, at mid-clause and end-clause positions were chosen (Bosker et al., 2013; Huensch & Tracy-Ventura, 2017). Fluency research (de Jong et al., 2015; Hunter, 2017; Tavakoli et al., 2017) emphasizes that in order to understand speakers' breakdown behaviour, it is necessary to examine pause character (filled or unfilled) and location (mid-clause and end-clause). Speech rate was selected to represent composite (speed and pausing aspects combined) fluency as it has been reported to be one of the most sensitive aspects of

utterance fluency to demonstrate change in fluency behaviour (Huensch & Tracy-Ventura, 2017; Mora & Valls-Ferrer, 2012; Tavakoli, 2018; Tavakoli et al., 2017). Articulation rate was also included in our measures as it is the only pure measure of speed that excludes pauses to provide a measure of how fast someone speaks regardless of their pausing behaviour. Total number of repairs, i.e., total number of repetitions, replacements, false starts and reformulations, has also been suggested as a valid measure to demonstrate L2 speakers' repair behaviour and monitoring processes (Hunter, 2017; Kahng, 2014; Skehan, 2009). Table 3 below demonstrates the utterance fluency measures used in this study.

Insert Table 3 here.

## 5. Analysis and Results

Before answering the research questions, the descriptive statistics provided in Table 4 for means and standard deviations are discussed for the different fluency measures across the two languages. As shown in Table 4, the figures in Turkish as L1 and English as L2 were very similar for *number of end-clause filled pauses*. For a number of other fluency measures, the figures seemed different in the two languages. These included measures of breakdown and repair fluency, and speed and composite measures.

In order to see whether these differences are statistically significant, paired samples t-tests were run (Table 4). Bonferroni-corrected  $p$  values were also considered to minimize Type I errors. To estimate the magnitude of the differences, effect sizes (Cohen, 1998) were calculated, and the results were interpreted using the field-specific benchmarks suggested by Plonsky and Oswald (2014), i.e., for within

group comparisons and correlations, small ( $d = .60$ ;  $r = .25$ ), medium ( $d = 1.00$ ;  $r = .40$ ) and large ( $d = 1.40$ ;  $r = .60$ ). The results showed a statistically significant increase, with medium to large effect sizes, from Turkish to English in the number of mid-clause filled pauses ( $t = -5.79$ ,  $p < .001$ ,  $d = .80$ ) and number of mid-clause silent pauses ( $t = -10.93$ ,  $p < .001$ ,  $d = 1.92$ ) while a statistically significant decrease, with large effect sizes, from Turkish to English was observed in number of end-clause silent pauses ( $t = 6.35$ ,  $p < .001$ ,  $d = 1.42$ ), articulation rate ( $t = 16.66$ ,  $p < .001$ ,  $d = 3.50$ ) and speech rate ( $t = 17.23$ ,  $p < .001$ ,  $d = 3.29$ ). The lower articulation and speech rates and more frequent mid-clause pauses (both silent and filled) imply that overall the speakers were more fluent in their L1. The effect sizes further indicate that the differences in fluency measures between Turkish and English are important.

We also ran a number of t-tests to examine any possible effects of the two different narratives on the speakers' performance (Table S1, see the supplementary information online). The results showed no statistically significant differences between fluency of performances in the two tasks.

**Insert Table 4 here.**

Research question 1 examined whether there was a relationship between L1 fluency and L2 fluency behaviour. In order to address this question, Pearson product-moment correlations were run between L1 and L2 fluency measures for the four aspects of fluency (i.e. breakdown, repair, speed and composite) separately. Preliminary analyses were performed to ensure that there were no violations of the assumptions of normality, linearity and homoscedasticity. As can be seen in **Table 5**, the results demonstrate moderate to strong positive correlations for some breakdown



measures. Regarding breakdown fluency, there was weak (under  $r = .4$ ) to moderate (under  $r = .7$ ) (see Plonsky & Oswald, 2014) positive correlations between the two languages for *number of mid-clause filled pauses* ( $r = .60, p = .001$ ), *number of end-clause filled pauses* ( $r = .30, p = .048$ ), *number of mid-clause silent pauses* ( $r = .34, p = .024$ ) and for number of repair ( $r = .45, p = .003$ ). No significant correlations were found for speed and composite measures (Figure S1, see the supplementary information online).

Insert Table 5 here

In order to examine the extent to which the relationship between L1 and L2 fluency was moderated by L2 proficiency level (represented by the mean score of EIT and OPT scores), partial correlations were carried out separately for this variable as the controlling factor. Once again, it was ensured that the assumptions of normality, linearity and homoscedasticity were not violated. Weak to moderate correlations were maintained for all the significant results. An inspection of zero order correlations (Table 6) suggested that controlling for language proficiency had little impact on the strength of relationship between the measures; number of mid-clause filled pauses ( $r = .60, p < .001$ ), number of end-clause filled pauses ( $r = .32, p < .040$ ), number of mid-clause silent pauses ( $r = .36, p < .020$ ) and number of repair ( $r = .46, p < .001$ ) achieved a statistically significant level, with zero order correlations of  $r = .60, p < .001$ ;  $r = .30, p < .048$ ;  $r = .34, p < .024$ ; and  $r = .45, p < .003$ , respectively.

Insert Table 6 here.

Research question 2 examined the extent to which measures of L2 fluency can be predicted from measures of L1 fluency and language proficiency. To address this question, multiple regressions were run for all fluency measures (variables entered simultaneously) with English fluency measure as the dependent variable, and Turkish fluency measure and language proficiency measured by the OPT and the EIT scores as independent variables. Inspections of the SPSS output, e.g., normal distribution of residuals, independence of observations, multicollinearity among the independent variables, and linearity between the dependent and independent variables suggested that all assumptions of regression analysis were met. The results from the multiple regressions are presented in **Table 7**.

**Insert table 7 here.**

As indicated in **Table 7**, the results imply that for a number of measures, models predicting L2 fluency from L1 fluency and language proficiency reached a statistically significant level. As for breakdown fluency, the significant models were *number of mid-clause filled pauses*,  $F(3,38) = 9.10, p < .001$ , and *number of mid-clause silent pauses*  $F(3,38) = 4.11, p < .01$ , explaining 41% and 24% of the variance in the participants' performance respectively (adjusted  $R^2 = .37$  and  $.18$ , respectively). While the models for *number of end-clause silent pauses*,  $F(3,38) = .75, p < .525$ , and *articulation rate*,  $F(3,38) = 1.17, p < .371$ , did not reach a statistically significant level, the model for *number of end-clause filled pauses*  $F(3,38) = 2.63, p < .06$  missed reaching a significant level, although 17% of the variance in the number of end-clause filled pauses was explained by this model (adjusted  $R^2 = .10$ ). For number of mid-clause filled pauses, Turkish fluency measure significantly contributed to the model

( $p < .001$ ) while language proficiency did not. For number of mid-clause silent pauses, all measures (i.e. L1 measure, OPT and EIT scores) significantly contributed to the model ( $p < .012$ ,  $p < .03$  and  $p < .049$ , respectively).

Other models reaching a significant level was number of repair,  $F(3,38) = 5.88$ ,  $p < .002$  and speech rate,  $F(3,38) = 11.70$ ,  $p < .001$ . For the model of number of repair, while Turkish number of repair ( $p < .007$ ) and EIT scores ( $p < .02$ ) made a significant contribution to the model, OPT scores did not. The variance explained by this measure was 31 % (adjusted  $R^2 = .26$ ). Interestingly, for the model of speech rate, the contribution made by OPT scores was non-significant, while the EIT scores and L1 fluency measure made a significant contribution to the model ( $p < .001$  and  $p < .036$ , respectively). The amount of variance explained by speech rate was 48% (adjusted  $R^2 = .43$ ), suggesting that L2 speech rate can be predicted from language proficiency assessed by the EIT scores and Turkish speech rate. To interpret the strength of the adjusted  $R^2$  values, we followed Plonsky and Ghanbar's (2018) recent proposal in which values of up to .20 are considered small and those above .50 are regarded as large.

## **6. Discussion**

The prime aim of the current study was to investigate whether L1 and L2 fluency behaviours were related, and whether this relationship was mediated by proficiency level. The study also aimed to examine to what extent L1 fluency behaviour and language proficiency predicted L2 fluency in a group of Turkish learners of English. Our descriptive statistics highlighted a number of similarities and differences between L1 and L2 fluency behaviour. It was interesting to see that the participants produced similar amount of end-clause filled pauses in the two languages, while they showed

large differences in their pausing behaviour for number of mid-clause pauses (both filled and silent) and end-clause silent pauses. This is similar to previous research in this area (Skehan, 2014; Tavakoli, 2011) in that L2 speakers make more frequent mid-clause and fewer end-clause pauses. We believe providing descriptive statistics about fluency measures in different languages would offer a baseline for cross-linguistic analysis of fluency in future studies. As anticipated, the results of the t-tests showed that the participants were overall more fluent in their L1. The higher figures for repair measure in English and a much higher speed in Turkish are examples of their higher fluency levels in L1.

The results also indicated that overall L2 speed measure was lower than those reported in previous studies (e.g. Mora & Valls-Ferrer, 2012; Tavakoli et al., 2016). The lower L2 speed measure in our study may be explained by the fact that several of our participants were from an A2 proficiency level, and none of them had either studied or lived abroad in an English-speaking community, whereas in both Tavakoli et al. (2016) and Mora and Valls-Ferrer (2012), the participants were on Study-Abroad courses and had higher proficiency levels.

The results of the correlation analyses suggested that some of the breakdown measures and total number of repairs were positively correlated in L1 and L2, although the strength of associations varied. The results showed significant positive correlations for three breakdown measures, i.e., number of mid-clause filled pauses ( $r = .60$ ), number of end-clause filled pauses ( $r = .31$ ), and number of mid-clause silent pauses ( $r = .35$ ), suggesting that the L2 speakers' frequency of pausing was to some extent a function of their L1 pausing behaviour. These correlations replicate the findings of Peltonen (2018) in which moderate to high correlations of ( $r = .68$ ) and ( $r = .41$ ) were reported for end-clause silent pauses and mid-clause filled pauses.

Interestingly, the highest correlation found in our results is for mid-clause filled pauses ( $r = .60$ ), indicating a high degree of association between L1 and L2 pausing behaviour. Previous research has suggested that mid-clause pausing is a key characteristic of L2 speech production, especially when the production process is not yet automatic (Kormos, 2006; Segalowitz, 2010; Tavakoli, 2011; Tavakoli & Skehan, 2005). The high correlation obtained for mid-clause filled pauses, however, suggests that L2 pausing behaviour in the current study is also to some extent a function of one's personal style carried over from L1 behaviour. Overall, these findings corroborate the findings of previous research (de Jong et al., 2015; Peltonen, 2018) regarding the correlations between breakdown measures in L1 and L2.

The results also indicated a positive correlation for total number of repairs ( $r = .45$ ) in the two languages, confirming the hypothesis that L2 repair behaviour might be related to L1 repair behaviour. This finding is in contrast with that of Peltonen (2018) where weak correlations between L1 and L2 repetitions were found. The different results might have been caused by the way repair fluency is operationalized in the two studies. While Peltonen (2018) measured repair in terms of repetitions, in our study we examined a range of repair measures that the SLA literature suggests as reliable representatives of repair processes. Examining a wider range of repair measures, we argue, has given us a rich opportunity to investigate the L1 and L2 repair behaviours and to explore the possible relationship between them.

With regards to L2 speakers' use of repairs across different proficiency levels, some emerging research evidence (Tavakoli et al., 2017) suggests that use of repairs is not related to proficiency level, implying that a kind of personal style might be at work. This finding is particularly important as repair is often perceived as an indicator of dysfluency in second language pedagogy, e.g., language benchmarks such as

CEFR, and language testing scales such as IELTS. L2 teachers regularly associate repair with monitoring processes involved in L2 production, and often link it to the development of the learner's interlanguage system. The results presented here suggest that while repair measures represent the speaker's process of repairing an utterance, e.g., correcting an L2 error or reformulating an L2 structure and concept, it can to some extent demonstrate the speaker's personal speaking style. This is an important point to be included in L2 teacher training programmes as previous research (Tavakoli & Hunter, 2018) has highlighted the crucial role that teachers' understanding of the relationship between L1 and L2 fluency plays in their classroom practice and their ability to help their learners.

As for speech rate and articulation rate, however, the results did not show any significant correlations between L1 and L2 fluency behaviour. This is in contrast to de Jong et al. (2015), Derwing et al. (2009) and Huensch and Tracy-Ventura (2016) who did report correlations for speed measures. We interpret these contradictory results in the light of the differences between our participants and those in the other studies. Our study included 15 participants at A2 (CEFR) level of proficiency. As discussed earlier, learners at lower proficiency levels often speak at a slower speed with more mid-clause pauses, hesitations and interruptions. This means their speech rate and articulation rate are typically slower than those in their L1; the different L2 speaking patterns in the A2 level participants may have made it difficult to obtain a significant correlation between the two speed patterns. It is also important to note that our participants had never lived/studied abroad or used English for professional purposes in their everyday life. Arguably, they had fewer opportunities for developing fluent speaking skills compared to those in de Jong et al. (2015), Derwing et al. (2009) and Huensch and Tracy-Ventura (2016) who had lived/studied abroad.

When controlled for proficiency level, the significant results of the partial correlations suggested that the relationship between L1 and L2 fluency was maintained across different proficiency levels and, indicating that the relationship between fluency behaviour in the two languages was not mediated by proficiency. This suggests that the relationship between L1 and L2 fluency behaviour persists regardless of how proficient the speakers are, highlighting the fact that the L2 speaker's fluency behaviour is, at least to some extent, related to their L1 fluency behaviour.

Our second aim was to investigate whether L2 fluency behaviour can be predicted from L1 fluency behaviour and language proficiency. The results of multiple regression analyses indicated that four models predicted L2 fluency, with L1 fluency making a significant contribution to all of them, OPT to one and EIT to three. When interpreting the findings, we have followed Plonsky and Ghanbar's (2018) proposal in which  $R^2$  values in the realm of .20 (or below) and .50 (and above) are considered as small and large, respectively. The models predicting L2 fluency were mid-clause filled and mid-clause silent pauses, total number of repairs and speech rate. Two of the significant models where L1 fluency predicted L2 fluency behaviour were number of mid-clause filled pauses and number of mid-clause silent pauses ( $R^2$  values of 41% and 24% respectively which are regarded medium and small) with L1 making a significant contribution to both ( $p < .001$  and  $p < .01$ ). These findings are in line with Huensch and Tracy-Ventura's (2016) results where frequency of L1 pauses predicted L2 fluency behaviour. However, a key difference between the two studies is that instead of measuring pauses at clause boundary, which is recommended in the literature, Huensch and Tracy-Ventura (2016) examined pauses at AS-unit boundary. The literature on fluency studies (Skehan, 2009, 2014; Peltonen, 2018) considers this

a limitation which has been addressed in the current study as we distinguished between mid and end clause pauses. The more detailed analysis of pause location provides a more useful insight into the differences between L1 and L2 pausing behaviour, and seems to be crucial for an in-depth understanding of the differences between L1 and L2 processing.

The highest significant level achieved was for speech rate ( $R^2 = 48\%$ ) in which the EIT scores made a significant contribution to the model ( $p < .001$ ), whereas L1 measure made a significant but modest contribution ( $p < .04$ ). This finding is interesting as it suggests that the L2 speech rate can be predicted by a measure of procedural knowledge. It is necessary to note that articulation rate, which is a pure measure of speed, failed to show a significant model, suggesting speed of performance in L1 cannot predict L2 speed. This finding highlights the importance of including both pure and composite measures of speed when analysing fluency. For total number of repairs, both L1 behaviour and the EIT scores made significant contributions to the model ( $p < .007$  and  $p < .02$ , respectively). The adjusted  $R^2$  figure of 26% in this model implies that L2 repair can also be predicted from L1 repair and L2 procedural knowledge. It is worth noting that the  $R^2$  figures obtained for our significant models ranged between small level of .18 to a medium level of .48 (Plonsky & Ghanbar, 2018) indicating the models were able to explain up to 48% of the variance in learners' performance.

The results also revealed that in a number of models, language proficiency predicted L2 fluency behaviour. Our interest in examining the role of language proficiency was motivated by previous research (de Jong, 2016; de Jong et al., 2015; Huensch & Tracy-Ventura, 2016) that questioned whether language proficiency mediated the relationship between L1 and L2 fluency behaviour. We were also



inspired by research (e.g., de Jong, Steinel, Florijn, Schoonen & Hulstijn, 2012, and Revesz, Ekiert and Torgergen, 2016) that claimed some L2 fluency measures were reliable predictors of proficiency. For speech rate, the model reached significance with the EIT scores making a significant contribution to the model ( $R^2$  figure of 48% and  $p < .001$ ).

The results indicated that the EIT scores were a good predictor of a number of L2 fluency measures, i.e., speech rate, number of repairs and mid-clause silent pauses. This finding highlights the role of procedural knowledge in producing fast and uninterrupted speech. The OPT scores, on the other hand, predicted mid-clause silent pauses suggesting that L2 declarative knowledge may encourage silent pauses in mid-clause position. That is to say, mid-clause silent pauses are opportunities the speakers employ to use their declarative knowledge of the language during the production process. Research in this area (Kahng, 2014; Skehan, 2014; Skehan et al., 2016) reports that mid-clause pausing is linked to the Formulation (Levelt, 1989) stage of speech processing. Taking the two points together, it is plausible to argue that L2 learners who rely on their declarative knowledge are expected to have more mid-clause pausing. Given the relatively small sample size of the study, however, we suggest the results are considered with caution.

## **7. Final Remarks**

Our study has been the first to examine the role of proficiency level in L1 and L2 fluency behaviour systematically across three proficiency levels. In addition, it is one of the first studies to examine proficiency level through more than one means. For these reasons, our findings will make a valuable contribution to helping develop a more reliable understanding of the relationship between L1 and L2 fluency behaviour.

Overall, our results are in line with recent research in oral fluency (Huensch & Tracy-Ventura, 2016; Peltonen, 2018), and reveal that L1 and L2 fluency behaviours are related for breakdown and repair measures, and that the relationship persists across different proficiency levels. Unlike Peltonen (2018), our results suggest that L2 pure speed measure cannot be predicted from L1 speed. We interpret this difference in relation to how the participants' proficiency was assessed as the current study adopted a more systematic approach to measuring proficiency.

The results also reveal that the pausing and breakdown aspects of L2 fluency can, to some extent, be predicted from L1 fluency behaviour. The significant models for speech rate, number of mid-clause filled and silent pauses and total number of repairs suggest that a considerable amount of the variance in L2 fluency behaviour can be explained by L1 fluency. It may be inferred from the results that pure speed measure, i.e., articulation rate, is independent in the two languages. The results also indicate that L2 procedural knowledge predicts speech rate, whereas the participants' declarative knowledge contributes to the models predicting mid-clause silent pausing. The results of the study have significant implications for the field of language testing, particularly for developing speaking rating scales. Rating scales for the assessment of speaking in many international language tests, e.g., the British Council's IELTS and APTIS, frequently refer to features such as pauses and reformulations as indicators of disfluency. The results of the current study suggest that frequent use of such features may represent, at least to some extent, the speaker's L1 behaviour rather than their L2 disfluency.

An important methodological contribution of this study concerns the assessment of proficiency level as well as the impact that the choice of test might have on the results. We used two measures of proficiency to tap into different underlying

constructs of L2 speaking proficiency, i.e. L2 declarative and procedural knowledge. This allowed us to see which aspects of fluency are linked to declarative and which to procedural types of linguistic knowledge. The findings clearly suggest that a more complete profile of the learner proficiency can help shed light on the role of proficiency in fluency behaviour.

## References

- Awwad, A., Tavakoli, P. & Wright, C. (2017). "I think that's what he's doing": Effects of intentional reasoning on second language (L2) speech performance. *System*, 67(1): 158-169
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, 10(3), 245-261.
- Boersma, P., & Weenink, D. (2010). *Praat: Doing phonetics by computer [Version 5.1. 35, Computer program]*. Retrieved June 10, 2010.
- de Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, 13 (1), 1–24.
- de Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, 54(2), 113–132.
- de Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behaviour. *Applied Psycholinguistics*, 36(02), 223–243.
- de Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2012). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In A. Housen, F. Kuiken & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (p.121–142).
- de Jong, N., & Vercellotti, M. L. (2016). Similar prompts may not be similar in the performance they elicit: Examining fluency, complexity, accuracy, and lexis in

- narratives from five picture prompts. *Language Teaching Research*, 20(3), 387–404.
- Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31, 533–557.
- Di Silvio, F., Diao, W., & Donovan, A. (2016). The Development of L2 Fluency During Study Abroad: A Cross-Language Study. *The Modern Language Journal*, 100(3), 610–624.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in second language acquisition*, 27(02), 141–172.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behaviour Research Methods*, 41(4), 1149-1160.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354–375.
- Gaillard, S., & Tremblay, A. (2016). Linguistic Proficiency Assessment in Second Language Acquisition Research: The Elicited Imitation Task. *Language Learning*, 66 (2), 419–447
- Huensch, A., & Tracy-Ventura, N. (2016). Understanding second language fluency behaviour: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, 1–31.
- Huensch, A., & Tracy–Ventura, N. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *The Modern Language Journal*, 101(2), 275-293.

- Hulstijn, J. (2015). *Language Proficiency in Native and Non-native Speakers: Theory and research*. Amsterdam: Benjamins Publishing.
- Hunter, A. M. (2017). *Fluency development in the ESL classroom: The impact of immediate task repetition and procedural repetition on learners' oral fluency* (Doctoral Thesis). St. Mary's University, Twickenham, London. Retrieved from <https://research.stmarys.ac.uk/1868/1/Final%20version%2023.10.2017.pdf>
- Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, 64(4), 809–854.
- Kim, Y., Tracy–Ventura, N., & Jung, Y. (2016). A Measure of Proficiency or Short-Term Memory? Validation of an Elicited Imitation Test for SLA Research. *The Modern Language Journal*, 100(3), 655–673.
- Koponen, M., & Riegenbach, H. (2000). Overview: Varying perspectives on fluency. In *Perspectives on fluency* (pp. 5–24). University of Michigan.
- Kormos, J. (1998). *Self-repairs in the speech of Hungarian learners of English*. PhD Dissertation, Eotvos University, Budapest.
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145-164.
- Lennon, P. (1990). Investigating fluency in EFL: a quantitative approach. *Language Learning*, 40 (3), 387–417.

- Leonard, K. R., & Shea, C. E. (2017). L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *The Modern Language Journal*, 101(1), 179–193.
- Levelt, W. J. M. (1989). *Speaking from intention to articulation*. Cambridge, Mass: Massachusetts Institute of Technology Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Sciences*, 22, 1–38.
- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 46(4), 610–641.
- Official Website of Oxford English Testing. Accessed 28 March 2018 at:  
[CUhttps://www.oxfordenglishtesting.com/uploadedfiles/buy\\_tests/oopt\\_meaning.pdf](https://www.oxfordenglishtesting.com/uploadedfiles/buy_tests/oopt_meaning.pdf).
- Official Website of Syllable Count. Accessed 28 March 2018 at:  
<http://www.syllablecount.com/>
- Official Website of Hece Hesaplama. Accessed 28 March 2018 at:  
<https://www.hesapla.online/hece-hesaplama/>
- Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002). *An investigation of elicited imitation tasks in cross linguistic SLA research*. Paper presented at the Second Language Research Forum, Toronto, October 3–6.
- Peltonen, P. (2018). Exploring connections between first and second language fluency: A mixed methods approach. *Modern Language Journal*, 102(4), 676–692.
- Plonsky, L., & Oswald, F. L. (2014). How Big Is “Big”? Interpreting Effect Sizes in L2 Research. *Language Learning*, 64(4), 878–912. doi:10.1111/lang.12079

- Plonsky, L. & Ghanbar, H. (2018), Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal*, 102: 713-731.
- Préfontaine, Y., & Kormos, J. (2015). The relationship between task difficulty and second language fluency in French: A mixed methods approach. *The Modern Language Journal*, 99(1), 96-112.
- Préfontaine, Y., & Kormos, J. (2016). A qualitative analysis of perceptions of fluency in second language French. *International Review of Applied Linguistics in Language Teaching*, 54(2), 151-169.
- Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second language acquisition*, 31(3), 437-470.
- Révész, A., Ekiert, M., & Torgersen, E. N. (2016). The effects of complexity, accuracy, and fluency on communicative adequacy in oral task performance. *Applied Linguistics*, 37(6), 828–848.
- Riazantseva, A. (2001). Second language proficiency and pausing a study of Russian speakers of English. *Studies in Second Language Acquisition*, 23(04), 497–526.
- Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. New York: Routledge.
- Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching*, 54(2), 79-95.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36, 1–14.
- Skehan, P. (2009). Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics*, 30 (4), 510–532.



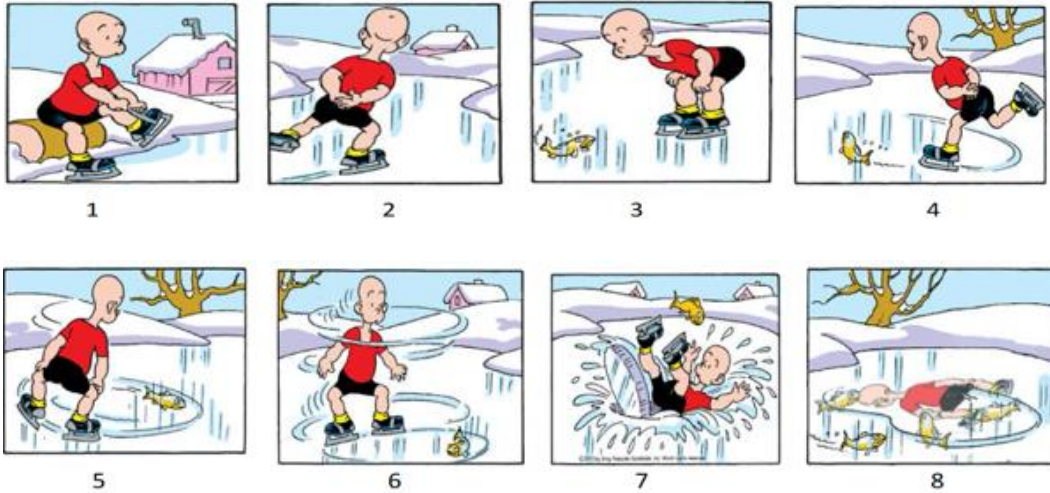
- Skehan, P. (2014). Limited attentional capacity, second language performance, and task-based pedagogy. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 211–260). Amsterdam, the Netherlands: John Benjamins.
- Skehan, P. (2015, November). Limited Attention Capacity and Cognition. In M. Bygate (Ed.), *Domains and Directions in the Development of TBLT: A decade of plenaries from the international conference* (Vol. 8, p. 123). John Benjamins Publishing Company.
- Skehan, P., Foster, P. & Shum, S. (2016). Ladders and Snakes in Second Language Fluency. *International Review of Applied Linguistics in Language Teaching*, 54(2), pp. 97-111.
- Suzuki, Y. & DeKeyser, R. (2019). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, 64(4): 747-790.
- Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers, *ELT Journal*. 65(1): 71-79.
- Tavakoli, P. (2016). Speech fluency in monologic and dialogic task performance. *International Review of Applied Linguistics (IRAL): Special Issue on fluency*. 54(2): 131-150.
- Tavakoli, P. (2018). L2 Development in an intensive Study Abroad EAP context. *System*, 72(1), 62-74.
- Tavakoli, P. (2019). Automaticity, fluency and second language performance. In Z. Wen & M. Ahmadian (Eds.) *Researching L2 task performance and pedagogy*. (pp. 39-52). Amsterdam: John Benjamins.

- Tavakoli, P. & Hunter, A-M. (2018). Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*. 22(3): 330-349.
- Tavakoli, P. and 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: John Benjamins.
- Tavakoli, P. Campbell, C. & McCormack, J. (2016). Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly*. 50(2): 447-471.
- Tavakoli, P. Nakatsuhara, F. & Hunter, A-M. (2017). *Scoring validity of the Aptis Speaking test: Investigating fluency across tasks and levels of proficiency*. ARAGs Research Reports Online. ISSN 2057-5203 London: British Council.
- Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, 50(2), 369-393.
- Trachte, D. (1948). Henry comics retrieved from <https://www.mycomicshop.com>.
- Towell, R., & Dewaele, J. M. (2005). The role of psycholinguistic factors in the development of fluency amongst advanced learners of French. Focus on French as a foreign language: *Multidisciplinary approaches*, 210–239.
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17(1), 84–119.
- Tracy-Ventura, N., McManus, K., Norris, J., & Ortega, L. (2014). Repeat as much as you can': Elicited imitation as a measure of oral proficiency in L2 French. *Measuring L2 proficiency: Perspectives from SLA*, 143–166.

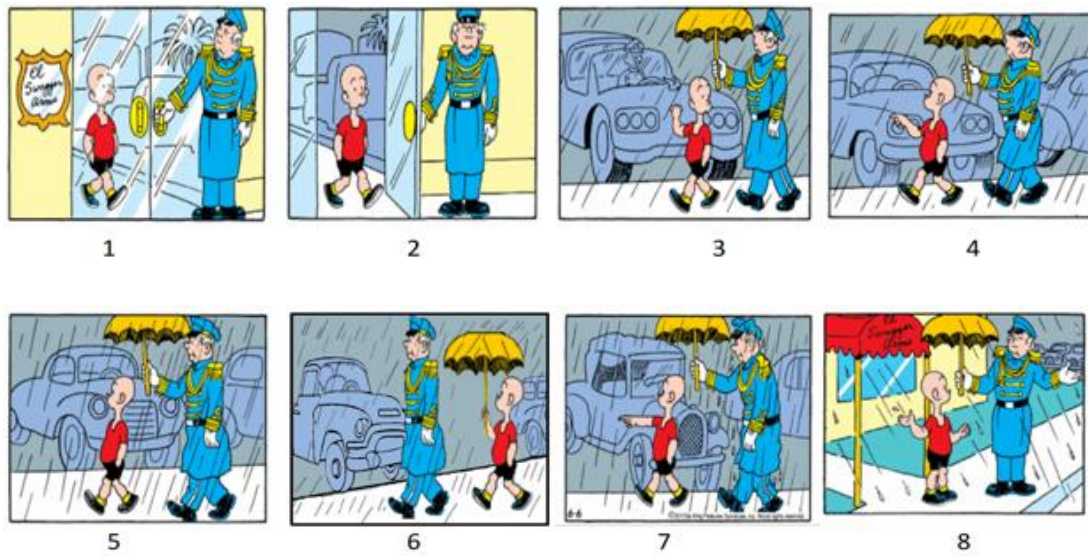
- Wright, C. (2013). An investigation of working memory effects on oral grammatical accuracy and fluency in producing questions in English. *Tesol Quarterly*, 47(2), 352-374.
- Wright, C. (2018). Effects of study abroad on L2 Mandarin Chinese: Effects of Time and Task. In C.Sanz & A. Morales-Front (Eds.), *The Routledge Handbook of Study Abroad Research and Practice* (pp.1-17). New York: Routledge.
- Wu, S. L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals*, 46(4), 680-704.
- Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2015). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, 33(4), 497–528.

**Appendix 1. Oral Narrative Tasks (Henry comics by Don Trachte)**

**Task A. The boy going ice-skating**



**Task B. The boy looking for his car**



**Appendix 2. Descriptive statistics for OPT and EIT scores (N=42)**

	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Std. Deviation</b>	<b>CI (95% for Mean)</b>	
<b>EIT scores</b>	9.0	105.0	36.381	26.5109	Lower Bound	28.12
					Upper Bound	44.64
<b>OPT scores</b>	37.0	88.0	64.938	14.2540	Lower Bound	60.49
					Upper Bound	69.38

**Appendix 3. Descriptive statistics for the length of speech samples\* produced in L1 and L2 (N=42)**

	<b>Mean</b>	<b>Std. Deviation</b>	<b>CI (95% for Mean)</b>	
<b>Turkish</b>	36.47	13.09	Lower Bound	32.39
			Upper Bound	40.55
<b>English</b>	50.45	11.88	Lower Bound	46.74
			Upper Bound	54.15

*\*The lengths of speech samples that were used for data analysis*

Table 1. The groupings of the participants across tasks (N=42).

<b>Tasks</b>	<b>Sub-groups</b>	
<b>Task A.</b> (The boy going ice-skating) n=21	<b>Task A-Group 1</b> (English first-Turkish second) n= 10	<b>Task B-Group 2</b> (Turkish first-English second) n=11
<b>Task B.</b> (The boy looking for his car) n= 21	<b>Task B- Group 1</b> (English first-Turkish second) n=11	<b>Task B-Group 2</b> (Turkish first-English second) n= 10

Table 2. The proficiency groupings of the participants based on Oxford Online Placement Test scoring system (N=42).

<b>CEFR Levels</b>	<b>Mean score of OPT and EIT*</b>	<b>n</b>
C2	81-100	-
C1	61-80	-
B2	41-60	12
B1	21-40	15
A2	1-20	15

*\*This is the combined score of OPT and EIT scores (50% of each), and the maximum score is 120.*



Table 3. Utterance fluency measures and calculations (adopted from Kormos, 2006, p.163)

<b>Breakdown fluency measures (pauses above 0.25 sec)</b>	
Number of <i>mid-clause filled</i> pauses (MCFP) / min.	The total number of filled pauses within the clauses divided by the total speaking time
Number of <i>end-clause filled</i> pauses (ECFP) / min.	The total number of filled pauses between the clauses divided by the total speaking time
Number of <i>mid-clause silent</i> pauses (MCSP) / min.	The total number of silent pauses within the clauses divided by the total speaking time
Number of <i>end-clause silent</i> pauses (ECSP) / min.	The total number of silent pauses between the clauses divided by the total speaking time
<b>Repair fluency measures</b>	
Number of repair	The total number of Repetitions, replacements, reformulations and false starts.
<b>Speed fluency measures</b>	
<i>Articulation rate</i>	Number of syllables divided by total speaking time in seconds (excluding pause time) / Number of syllables per minute
<b>Composite measures</b>	
<i>Speech rate</i>	Number of syllables produced divided by total speaking time in seconds (including pause time) / Number of syllables per minute.

Table 4. Descriptive statistics and t-test results for L1 and L2 fluency measures (N=42)

Fluency measures	Mean	Std. Deviation	t	Sig. (2-tailed)	CI (95% for Mean)	Cohen's d
<b>Breakdown measures</b>						
Tur number of MC FP	2.67	3.61	-5.79	.001	Lower Bound	1.54
					Upper Bound	3.80
Eng number of MC FP	5.75	4.07			Lower Bound	4.47
					Upper Bound	7.02
Tur number of ECFP	3.68	3.71	.72	.471	Lower Bound	2.52
					Upper Bound	4.83
Eng number of ECFP	3.24	2.71			Lower Bound	2.39
					Upper Bound	4.09
Tur number of MCSP	9.24	5.31	-10.93	.001	Lower Bound	7.58
					Upper Bound	10.90
Eng number of MCSP	19.51	5.34			Lower Bound	17.85
					Upper Bound	21.18
Tur number of ECSP	15.34	4.71	6.35	.001	Lower Bound	13.87
					Upper Bound	16.81
Eng number of ECSP	9.80	2.84			Lower Bound	8.91
					Upper Bound	10.69
<b>Repair measures</b>						
Tur number of repair	2.82	2.46	-2.07	.044	Lower Bound	2.05
					Upper Bound	3.59
Eng number of repair	3.77	3.07			Lower Bound	2.81
					Upper Bound	4.73

<b>Speed measures</b>						
Tur articulation rate	378.52	59.51	16.66	.001	Lower Bound	359.97
					Upper Bound	397.07
Eng articulation rate	192.69	45.46			Lower Bound	178.52
					Upper Bound	206.86
<b>Composite measures</b>						
Tur speech rate	224.16	53.48	17.23	.001	Lower Bound	207.49
					Upper Bound	240.82
Eng speech rate	79.42	31.72			Lower Bound	69.59
					Upper Bound	89.31

Note: *SD*, Standard Deviation; *L1*, first language; *L2* second/foreign language; *MCFP*, mid-clause filled pauses; *MCSP*, mid-clause silent pauses; *ECFP*, end-clause filled pauses; *ECSP*, end-clause silent pauses

Table 5. Correlations between L1 and L2 fluency measures for all groups (N=42)

<b>Fluency measures</b>	<b>r</b>	<b>P</b>
<b>Breakdown fluency measures</b>		
Number of MCFP	.606**	.001
Number of ECFP	.306*	.048
Number of MCSP	.348*	.024
Number of ECSP	-.056	.727
<b>Repair fluency measures</b>		
Number of repair	.447**	.003
<b>Speed fluency measures</b>		
Articulation rate	.071	.653
<b>Composite measures</b>		
Speech rate	.266	.088

\*Correlation is significant at the 0.05 level (2-tailed)

\*\*Correlation is significant at the 0.01 level (2-tailed)

Table 6. Partial correlations for significant results, controlling for language proficiency level

<b>Fluency measures</b>	<b>r</b>	<b>P</b>	<b>Zero order (Pearson) correlation</b>
Number of MCFP	.604	.001	.606, $p < .001$
Number of ECFP	.322	.040	.306, $p < .048$
Number of MCSP	.361	.020	.348, $p < .024$
Number of repair	.466	.001	.447, $p < .003$

Table 7. Multiple regressions models predicting L2 fluency from L1 fluency, OPT score and EIT score

Measures	Test Statistics	B	SE	$\beta$	<i>F</i>	<i>P</i>	R Square	Adjusted R <sup>2</sup>	Effect size (Cohen's <i>f</i> <sup>2</sup> )
<b>Breakdown fluency measures</b>									
Number of MCFP	Intercept	.690	2.42		F (3, 38) = 9.10	.001	.41	.37	.69
	L1 measure	.606	.146	.538		.001			
	OPT score	.072	.043	.253		.100			
	EIT score	-.035	.023	-.225		.138			
Number of ECFP	Intercept	5.75	1.92		F (3, 38) = 2.63	.064	.17	.10	.20
	L1 measure	.262	.115	.359		.028			
	OPT score	-.053	.035	-.277		.136			
	EIT score	-.001	.018	-.014		.938			
Number of MCSP	Intercept	9.06	4.33		F (3,38) = 4.11	.013	.24	.18	.31
	L1 measure	.391	.149	.389		.012			

	OPT score	.143	.063	.382		.03			
	EIT score	-.068	.033	-.336		.049			
Number of ECSP	Intercept	9.52	2.33		F(3,38)= .758	.525	.05	-.01	.05
	L1 measure	-.021	.100	-.035		.836			
	OPT score	-.005	.039	-.027		.889			
	EIT score	.026	.020	.245		.202			
<b>Repair fluency measures</b>									
Number of repair	Intercept	-.411	1.99		F(3,38) = 5.88	.002	.31	.26	.44
	L1 measure	.522	.182	.419		.007			
	OPT score	.066	.036	.308		.070			
	EIT score	-.044	.018	-.380		.020			
<b>Speed fluency measures</b>									
Articulation rate	Intercept	136.78	58.50		F(3,38) = 1.07	.371	.07	.00	.07

	L1 measure	.100	.122	.131				.416	
	OPT score	.012	.581	.004				.983	
	EIT score	.472	.318	.275				.147	
<b>Composite measures</b>									
Speech rate	Intercept	-.859	22.07		F (3, 38) =11.70	.001	.48	.43	.92
	L1 measure	.153	.070	.258				.036	
	OPT score	.340	.309	.153				.277	
	EIT score	.655	.164	.547				.001	



