

Examining Foreign Persian Test-Takers' Speech Fluency: The Role of Repetition and False Start across Different Persian Proficiency Levels

Masoomeh Estaji *

Corresponding Author, Associate, Department of English Language and Literature, Allameh Tabataba'i University, Tehran, Iran. E-mail: estaji@atu.ac.ir

Reza Morad Sahraee 

Professor, Department of Linguistics, Allameh Tabataba'i University, Tehran, Iran. E-mail: sahraei@atu.ac.ir

Monireh Shahbaz 

Ph.D. Candidate of Teaching Persian to Speakers of Other Languages (TPSOL), Allameh Tabataba'i University, Tehran, Iran. E-mail: shahbaz.mnr@gmail.com

Abstract

The present study investigated repair fluency, one of the three main aspects of fluency in Skehan's model (2003; 2009), across different proficiency levels using the given test-taker data in a speaking test. There are three main aspects of fluency in the model: Speed fluency, breakdown fluency, and repair fluency. To date, there is no research aiming to measure the speech fluency of foreign Persian test takers. Therefore, this study sought to achieve an appropriate rating scale for speech fluency. The study, in particular, has examined some parts of the repair fluency construct in the mentioned model, specifically repetition and false starts, and attempted to determine the extent to which repetitions and false starts varied across different Persian proficiency levels of foreign Persian test-takers. This study transcribed, rated, and labeled repetitions and false starts of 23 foreign Persian language learners taking a speaking test. The spoken data were classified into four proficiency levels according to the Persian Teaching Standard Reference (PTSR), namely pre-intermediate, intermediate, upper-intermediate, and advanced. The duration of recordings was 44 minutes in total (11 minutes for each level). The results were analyzed, using the Kruskal-Wallis test, and revealed that false starts and repetitions did not make a distinction across Persian proficiency levels. The findings imply that raters and rating scale designers do not need to consider repetitions and false starts for identifying or describing different Persian proficiency levels.

Keywords: Speech fluency, Rating scale, Repair fluency, False start, Repetition

Cite this Article: Estaji, M., Sahraee, R. M., & Shahbaz, M. (2024). Examining Foreign Persian Test-Takers' Speech Fluency: The Role of Repetition and False Start across Different Persian Proficiency Levels. *Educational Measurement*, 14(56), 69-102. <https://doi.org/10.22054/jem.2024.73106.3454>



© 2016 by Allameh Tabataba'i University Press

Publisher: Allameh Tabataba'i University Press

DOI: <https://doi.org/10.22054/jem.2024.73106.3454>

1. Introduction

For most people, knowing a language is synonymous with the ability to speak that language. The reason for this may be that speaking is the most visible and evident manifestation of linguistic proficiency. Spoken language has gained a higher position in linguistic sciences in the 21st century to the extent that it has become intertwined with the term “language” in applied linguistics and broader research domains (Hughes, 2011, p. 5). It is worth noting that in terms of speech fluency in Persian as a first language, and in general, research has mostly been done in the fields of cognitive sciences, medicine, education, and rehabilitation. The present research can fill a small part of the existing gap in providing the necessary materials for the development of assessment scales. With the findings of this research (and other data-driven and corpus-based studies that scientifically examine and measure the influential components of speech assessment), one can hope for more fair and objective judgments of the speech performance of Persian language learners. In particular, conducting this research is a step toward developing a reliable assessment scale, and its absence, at least in two cases, namely placement tests and proficiency tests can lead to incorrect and problematic decision-making. Erroneous decisions in these two cases (especially the latter) can cause confusion and distrust among test-takers and other stakeholders, and ultimately even damage the reputation of the test, undermining the credibility of the test and its administrators.

Research Questions

Pursuing the research objectives, the following research questions were raised in this study.

1. To what extent does false start change among Persian language examinees at different proficiency levels?
2. To what extent does the frequency of incomplete and complete repetitions (including repetition of sentences, phrases, and words) change among Persian language examinees at different proficiency levels?

2. Literature Review

In this section, several important definitions of speech fluency in a second language are presented. As can be guessed, different aspects have been adopted so far to describe this speech component. For example, some definitions have been influenced by definitions of

fluency in the first language; some have had a holistic and categorical view, while others have had an analytical view. Evidently, there are also overlaps among these definitions. Fillmore's (1979) definition of this element, although focusing on speech fluency in the first language, is considered one of the groundbreaking perspectives on speech fluency, according to Tavakoli and Wright (2020). In fact, this holistic view of fluency formed the basis for definitions that were later presented in the 1980s regarding speech fluency in a second language (Tavakoli & Wright, 2020). In this view (Fillmore, 1979), one "type" of fluency is the ability to speak for a long time with fewer pauses and fill the time with speech. Someone equipped with this type of fluency does not have to pause frequently to think about what they want to say or how to express it. One of the best examples of this type of fluency can be found among sports commentators.

The second type of fluency is the ability to speak in a continuous, logical, and deeply meaningful way. The most important component of this ability is mastery of the semantics and syntax of the language. Individuals with this ability do not need to fill their speech with a large amount of meaningless content. The third type of fluency is the ability to use the appropriate words in a range of contexts. Someone who is fluent in this regard always uses the most appropriate content and has no linguistic problems in different conversational contexts. In contrast, someone who is not fluent in this regard may feel completely comfortable in familiar and informal contexts, but becomes silent in the presence of strangers or when faced with an unexpected crisis in human interaction.

The fourth type of fluency is the ability to be creative and imaginative in language use, expressing ideas in novel ways, making jokes, playing with the sounds of words regardless of their meaning, changing style, creating and using metaphors, among others. The impression the listener has of such a speaker is that they quickly edit what they are stating, and are able to quickly consider a very diverse range of alternative methods in response to a situation and choose the one that seems the most impressive and masterful. Therefore, a truly proficient language user is someone who possesses all four of the mentioned skills.

Lennon (1990) presented two concepts of fluency: A macro concept and a micro concept. In his macro concept, he considers fluency as "an umbrella term for oral proficiency" (p. 389). In his micro concept, he

regards fluency as “a component, possibly separable, of oral proficiency” (p. 389), alongside accuracy, coherence, appropriateness, pronunciation, lexical resource, and the like.

The most comprehensive classification of the definitions presented regarding fluency and its different dimensions belongs to Tavakoli and Hunter (2018). Based on the quantitative and qualitative data obtained from 84 second language teachers in England, they concluded that there are four approaches to fluency among them: One very broad approach that defines fluency in terms of mastery and general proficiency; another broad approach that equates fluency with the ability to speak a second language; a narrow approach in which fluency is synonymous with ease, smoothness, and continuity, and is distinct from accuracy and complexity; and finally, a very narrow approach with three aspects of speed, breakdown, and repair. This last approach is the basis for conducting this research, and there were two reasons for choosing it: One is that this is the most detailed approach among the four mentioned approaches, and the other is that this approach has an objective and empirical view of fluency and enumerates components that are all scientifically and empirically measurable.

The variables calculated in each type of fluency have a considerable diversity. The present research only focuses on repair fluency, the most common variables of which are self-repair, false start, and incomplete and complete repetition (of words, phrases, and sentences). Self-repair refers to the learner’s correction of their own linguistic errors in pronunciation, grammar, word choice, and the like (Pishghadam et al., 2011, as cited in Sultana, 2009; Pishghadam et al., 2011, as cited in Lam, 2006). In the present research, whenever the Persian language learner noticed their own mistake and immediately corrected it, it was considered one instance of self-repair.

In false start, the individual starts their utterance and then, as if a new idea has occurred to them, suddenly stops it before it is completed and begins a new utterance (Fox Tree, 1995; Institute of Phonetics and Speech Processing, n.d., paragraph 1). The terms “false start” and “restarting” are used to refer to the unfinished information and the new information, respectively (Fox Tree, 1995). In this research, whenever the Persian language learner started an utterance and then suddenly stopped it before it was completed and began a new and different or unrelated utterance, it was considered one instance of false start.

Repetition here does not have a significantly different meaning and definition compared to the general definition, and it is more or less the same as what we all know. This term refers to the immediate and repeated saying of words or phrases (Houston, 2016; Skehan et al., 2016). The only thing to keep in mind is that the Persian language learner sometimes repeats their utterance incompletely and sometimes completely. It should be noted that the present research examines the two variables of false start and repetition among these three variables.

Considering its goal and questions, this research has chosen the most common measurable components in the repair fluency, namely incomplete and complete repetition of words, phrases, and sentences, and false start, based on Tavakoli and Wright (2020), to be measured. This research aimed to implement the oral data of the examiner-examinee in test conditions, label them, and measure one of the three broad aspects of fluency, i.e., repair, and its constituent components at different levels. In this way, it can provide a small part of the ideal scale mentioned earlier for evaluating Persian learners' speech. Accordingly, the ultimate goal of this research and similar studies that each complete a piece of this puzzle is to facilitate the process of examining the speech fluency of foreign Persian learners and make the decisions based on test results as fair as possible.

3. Methodology

In this research, the speech of 23 Persian language learners was examined using the Saadi Foundation's speaking assessment scale. At this stage, the speeches were selected in which the researchers had the highest confidence about the language level; that is, they reached a conclusion about the level of proficiency with much greater confidence. Of the total 23 examined speeches, 8 were placed at the pre-intermediate level, 6 at the intermediate level, 6 at the upper-intermediate level, and 3 at the advanced level. The duration of the speeches at each level was 11 minutes, totaling 44 minutes. Three of these speeches were recorded during an in-person exam, and 20 were recorded during an online exam. As for their gender, 8 of the examinees were male, and 15 were female.

After determining the levels, all speeches were transcribed with all details (including repeated and incomplete words, errors, and fillers) and included in the Saadi Foundation's international corpus along with the metadata. To answer the questions, 7 fluency labels were applied to

the recorded speeches after defining them in the Saadi Foundation's international corpus. These labels were derived from Tavakoli and Wright (2020). One label was for false start, and 6 labels were for different types of repetition.

1. False start
2. Incomplete word repetition
3. Incomplete phrase repetition
4. Incomplete sentence repetition
5. Complete word repetition
6. Complete phrase repetition
7. Complete sentence repetition

The Saadi Foundation's speaking test and speaking assessment scale were the instruments of this research. The Saadi Foundation's speaking test contained questions about global and lived human experiences, such as music, family, human relationships, education, holidays, sports, books and reading, history, vehicles, traffic, food, movies, parties, gifts, television, travel, and the like, and therefore, only measured the general language proficiency of Persian language learners. In this test, the speech performance of the Persian language learners was examined based on the Saadi Foundation's assessment scale, which considers four aspects of speech as follows:

1. Fluency and cohesion
2. Lexical resource
3. Grammatical accuracy and complexity
4. Pronunciation

Each of these four components was assigned a level from 1 to 7 based on the Saadi Foundation's speaking assessment scale, and the average of the four criteria, rounded down, was considered the individual's level in the speaking test.

4. Conclusion

This research aimed to measure a part of the repair fluency construct as one of the three aspects of the speech fluency model (Skehan, 2003; 2009), which considers fluency to have three aspects: Speed, breakdown, and repair. Examining the speech of 23 Persian language learners at four levels - pre-intermediate, intermediate, upper-intermediate, and advanced - from this perspective, showed that none of the seven repair fluency variables considered in this research, namely false start, incomplete word repetition, incomplete phrase repetition,

incomplete sentence repetition, complete word repetition, complete phrase repetition, and complete sentence repetition, could differentiate the levels of Persian language proficiency. Although in two of them (false start and complete word repetition) a significant frequency difference was observed between the first two levels on the one hand and the second two levels on the other. Overall, based on the Kruskal-Wallis test result, the difference in these variables among the individual levels was not significant.

This finding is useful for at least two groups of Persian language education practitioners. One is for those who, due to their activity in this field, occasionally need to evaluate (proficiency level) the speech of Persian language learners. These individuals do not need to pay attention to the extent to which the examinee makes false start or repeats words and sentences in their evaluation. The second is for groups or organizations that intend to develop a speaking assessment scale. These groups or organizations can create a scale regardless of the variables examined in this research and look for other variables that can differentiate the proficiency levels.

This research, like many others, had limitations. The most significant limitation of this research was the lack of a suitable assessment scale for determining or assessing oral proficiency. In particular, this research was a step toward developing a reliable assessment scale. To reduce the effect of this limitation, only the speeches with the least uncertainty about the assessment result were selected. Another limitation of this research was that not even a similar research sample had been conducted on foreign Persian learners, so that researchers could refer to it and gain a clearer understanding of what they should do in the research process. Even the number of similar foreign studies to the present research was few, making it more difficult for researchers. Although the few existing studies had considerable similarities with this research in terms of data collection conditions and the linguistic and national diversity of the examinees, only a part of them was devoted to repetition and false start. Finally, it is suggested that the variables of this research be measured in other research conditions, such as in non-test conditions, with the same test questions (e.g., only asking about traffic from the examinees), controlling variables such as nationality, gender, and age, with a longer speech duration per level, and with a different assessment scale.

ارزیابی روانی گفتار آزمون‌دهندگان فارسی آموز خارجی: نقش تکرار و خطای شروع در سطوح مختلف بسندگی زبان فارسی

معصومه استاجی *

نویسنده مسئول، دانشیار زبانشناسی کاربردی، گروه زبان و ادبیات انگلیسی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: estaji@atu.ac.ir

رضامراد صحرایی

استاد زبانشناسی، گروه زبانشناسی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: sahraei@atu.ac.ir

منیره شهباز

دانشجوی دکتری آموزش زبان فارسی به غیرفارسی‌زبانان، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: shahbaz.mnr@gmail.com

چکیده

پژوهش حاضر با استفاده از داده‌های شفاهی آزمونگر-آزمون‌دهنده، یکی از سه وجه کلان روانی در انگاره Skehan (2009; 2003)، یعنی روانی اصلاح، را در سطوح مختلف بسندگی اندازه‌گیری کرده است. در این انگاره، روانی سه وجه دارد که عبارت‌اند از: روانی سرعت، روانی شکست، و روانی اصلاح. تاکنون هیچ پژوهشی با هدف اندازه‌گیری روانی گفتار آزمون‌دهندگان فارسی آموز خارجی انجام نشده است. لذا انجام چنین پژوهشی مخصوصاً برای رسیدن به مقیاسی مناسب برای ارزیابی روانی گفتار لازم است. این پژوهش بخشی از سازه روانی اصلاح در انگاره پیش‌گفته، یعنی تکرار و خطای شروع را مورد بررسی قرار داد و به دنبال پاسخ‌گویی به این پرسش بود که خطای شروع و تکرار تا چه اندازه در سطوح مختلف فارسی‌آموزی تغییر می‌کند. بدین منظور، گفتار ۲۳ فارسی‌آموز خارجی، که در شرایط آزمونی جمع‌آوری شده بود، پیاده‌سازی، ارزیابی (=تعیین سطح)، و برچسب‌زنی شد. مجموع مدت‌زمان گفتارها، که بر اساس چارچوب مرجع بنیاد سعدی (صحرایی و مرصوص، ۱۳۹۵)، از چهار سطح بسندگی پیش‌میان، میان، فوق‌میان، و پیشرفته بودند، ۴۴ دقیقه (هر سطح ۱۱ دقیقه) بود. نتایج با استفاده از آزمون کراس‌کال-والیس تحلیل شد و مشخص شد خطای شروع و تکرار سطوح بسندگی فارسی را از یکدیگر متمایز نمی‌کنند؛ بنابراین ارزیابان و طراحان مقیاس‌های ارزیابی، هنگام تشخیص یا توصیف سطوح بسندگی، نیازی به لحاظ کردن دو مؤلفه تکرار و خطای شروع ندارند.

کلیدواژه‌ها: آزمودنی‌های فارسی‌آموز، اصلاح، روانی گفتار، سرعت، سطح بسندگی فارسی، شکست، مقیاس ارزیابی

استناد به این مقاله: استاجی، معصومه، صحرایی، رضامراد، و شهباز، منیره. (۱۴۰۳). ارزیابی روانی گفتار آزمون‌دهندگان فارسی آموز خارجی: نقش تکرار و خطای شروع در سطوح مختلف بسندگی زبان فارسی. فصلنامه اندازه‌گیری تربیتی، ۱۴(۵۶)، ۶۹-۱۰۲. <https://doi.org/10.22054/jem.2024.73106.3454>

مقدمه

بلد بودن یک زبان از دید اکثر قریب به اتفاق افراد، مساوی با توانایی صحبت کردن به آن زبان است. دلیل این امر شاید آن باشد که نمود حقیقی و آشکار زبان‌وری سخن گفتن است. در واقع، بسیاری از ما وقتی به افراد چندزبانه غبطه می‌خوریم که توانایی آن‌ها در صحبت کردن به چند زبان را می‌بینیم (نه مثلاً زمانی که توانایی آن‌ها در خواندن به چند زبان بر ما روشن می‌شود) و حتی «کسانی که زبان می‌دانند را گویشور زبان می‌نامیم» (Ur, 2012, p. 117). گونه گفتاری در قرن بیست و یکم جایگاه بالاتری در علوم زبانی کسب کرده تا جایی که در زبان‌شناسی کاربردی و گستره‌های وسیع‌تر پژوهشی با واژه زبان درآمیخته است (Hughes, 2011, p. 5). دلیل دیگری که باعث می‌شود توانایی صحبت کردن یک مهارت عالی و مطلوب به نظر بیاید، ممکن است به چستی آن مربوط باشد؛ مثلاً این که صحبت کردن مهارتی منفک و مجزا نیست و با حوزه‌های گوناگون دیگری همچون واژه، دستور، تلفظ، فرهنگ و گفتمان در ارتباط است (ibid, p. 6) و از طرفی، برخلاف نوشتن (به‌عنوان دیگر مهارت تولیدی زبانی)، در لحظه رخ می‌دهد که این خود پیچیدگی و حساسیت این مهارت را بالا می‌برد. دسته‌ای از پیچیدگی‌های مهارت صحبت کردن به آزمون این مهارت و مخصوصاً اندازه‌گیری و ارزیابی روا و پایای آن مربوط می‌شود. برای این که بتوان گفتار را درست و عادلانه ارزیابی کرد، باید مجهز به ابزاری به نام مقیاس ارزیابی بود. Davies و همکاران (1999, pp. 153-154) مقیاس ارزیابی را این‌گونه تعریف می‌کنند:

مقیاس ارزیابی شامل یک سری سطوح طرح‌ریزی شده است که عملکرد زبان‌آموز بر اساس آن قضاوت می‌شود. مقیاس بسندگی مانند یک آزمون تعریفی عملیاتی از سازه‌ای زبانی همچون بسندگی ارائه می‌دهد. دامنه این مقیاس‌ها نوعاً از تسلط صفر تا یک نقطه پایانی، یعنی گویشور بومی و باصلاحیت، است. سطوح برحسب آنچه آزمون‌دهنده می‌تواند با زبان انجام دهد (یعنی همان تکلیف‌های آزمونی) و به کمک مشخصه‌های زبانی (همچون واژه، دستور، روانی و مانند این‌ها)، تعریف می‌شوند.

برای تدوین مقیاس‌های ارزیابی پنج روش وجود دارد که می‌توان آن‌ها را در دودسته جای داد: یکی روش‌هایی که تجربه و دانش کارگروهی از متخصصان را در تهیه مقیاس به کار می‌گیرند و تسلط در استفاده از مقیاس‌ها حاصل معاشرت، تعامل و هماهنگی ارزیابان خواهد بود (Fulcher, 2012; Fulcher & Davidson, 2007) و دیگری روش‌هایی که

جدیدتر هستند و وجه مشترک آن‌ها داده‌محور بودن است یعنی «نمونه‌هایی از گفتار یا نوشتار را گفتمان کاوی^۱ می‌کنند تا مشخصه‌های کلیدی عملکردها را در سطوح معینی از بسندگی شناسایی کنند» (Fulcher, 2012, p. 384) و «اندازه‌گیری عملکرد به کمک آن‌ها نتایج ثابتی به دست می‌دهد» (ibid, p. 385) بدیهی است که مقیاس‌های ارزیابی با هر روشی که تدوین شده باشند، یکدست کردن فرایند تعیین سطح و ارزیابی را مدنظر دارند و حاوی مؤلفه‌ها و متغیرهایی هستند که به کارگیری درست آن‌ها در نهایت یک «متن»، اعم از نوشتاری یا گفتاری، را تشکیل می‌دهد.

باید خاطر نشان کرد که در این پژوهش، مقیاس ارزیابی واژه‌ای پوششی برای اشاره به هر خروجی جدول گونه‌ای است که در بردارنده توصیفاتی اندازه‌گیری شده از بسندگی زبانی (در اینجا گفتار) در ابعاد مختلف آن از جمله دستور، واژه، روانی، انسجام و پیوستگی و مانند این‌ها باشد و در ارزیابی تولیدات زبانی آزمون‌دهنده مبنای قضاوت قرار بگیرد. توضیح آن‌که در ادبیات پژوهشی، اصطلاحات دیگری همچون شیوه‌نامه^۲ نمره‌دهی^۳، توصیفگرهای سطح^۴، مقیاس بسندگی^۵ و مانند این‌ها نیز در اشاره به توصیف گفته شده استفاده می‌شوند که در اینجا، با هدف ایجاد یکدستی و جلوگیری از ایجاد ابهام و سردرگمی، از به کارگیری آن‌ها پرهیز می‌کنیم و همه آن‌ها را مقیاس ارزیابی می‌نامیم زیرا تفاوت‌های احتمالی میان آن‌ها در اینجا مطمح نظر نیست.

همان‌طور که می‌توان حدس زد، بعضی از مؤلفه‌ها و متغیرهای زبانی که معمولاً در مقیاس‌های ارزیابی لحاظ می‌شوند، میان دو مهارت نوشتن و صحبت کردن مشترک و بعضی دیگر مختص یکی از این دو هستند. مثلاً غنای دستوری و منبع واژگانی در هر دو مهارت ذکر شده قابل بررسی و ارزیابی هستند اما پاراگراف‌بندی مخصوص مهارت نوشتن و تلفظ مخصوص مهارت صحبت کردن است. در مورد ارزیابی گفتار، یکی از مؤلفه‌های تعیین‌کننده، مؤلفه روانی^۵ است. تاکنون در خصوص چگونگی تعیین سطح روانی گفتار فارسی آموزان خارجی پژوهشی انجام نشده است؛ گو این‌که هیچ‌یک از دیگر مؤلفه‌های پیش‌گفته نیز در طول سطوح مختلف فارسی‌آموزی اندازه‌گیری نشده‌اند و بنابراین، نوع و

1. discourse analysis
2. scoring rubric
3. band/level descriptors
4. proficiency scale
5. fluency

میزان تغییرات این مؤلفه‌ها در سطوح بسندگی بر ما روشن نیست. در نتیجه، مواد خام تهیه یک مقیاس برای ارزیابی و تعیین سطح گفتار (و نیز نوشتار) فارسی‌آموزان خارجی در دست نیست.

در حیطه آزمون‌سازی زبان دوم، نام روانی با مقیاس‌های ارزیابی گره خورده است و اکثر قریب به اتفاق پژوهش‌ها حول محور تهیه یک مقیاس ارزیابی می‌گردد. تلقی از روانی به‌عنوان یکی از نشانگرهای توانایی در صحبت کردن سابقه‌ای طولانی دارد و در قدیمی‌ترین تا جدیدترین آزمون‌های صحبت کردن، یعنی از سال ۱۹۳۰ بدین سو، (و در واقع در مقیاس‌های ارزیابی آن‌ها که طبیعتاً در ابتدا ساده بودند و رفته‌رفته دقیق‌تر شدند) دیده می‌شود (Fulcher, 2010). باینکه امروزه شاهد برگزاری آزمون‌های مقیاس بالا و معتبر یا بسیار معتبری همچون آپتیس^۱، پی.تی.ای^۲، تافل^۳ و در رأس آن‌ها آیلتس^۴ در سراسر جهان هستیم، اما این به این معنی نیست که مقیاس‌های ارزیابی آن‌ها عاری از هر اشکالی است. دست کم دو اشکال عمده به این مقیاس‌ها وارد است. یکی این که زبان به کار رفته در آن‌ها مبهم است و مشخص نیست در این مقیاس‌ها منظور از صفت‌ها و قیدهایی چون کمی، تا حدی، عموماً، قابل توجه، طولانی (مثلاً مکث طولانی یا خطاهای قابل توجه) و مانند این‌ها چیست و دیگر این که فرض می‌شود رابطه میان بسندگی و روانی خطی است یعنی لزوماً با افزایش سطح بسندگی، روانی سرعت افزایش و روانی شکست و اصلاح کاهش می‌یابد (Tavakoli & Wright, 2020). این اشکال اخیر مسئله‌ای است بسیار حائز اهمیت که در ادامه این نوشتار نیز مورد توجه خواهد بود.

در مقیاس‌های ارزیابی آزمون‌های مقیاس بالای زبان فارسی نیز این خلأها، بلکه به شکل خیلی جدی‌تر وجود دارد. آزمون‌سازی زبان فارسی به‌عنوان زبان دوم از لحاظ تعداد و تنوع پژوهش‌ها در مراحل بسیار ابتدایی‌تر از آزمون‌سازی زبان انگلیسی قرار دارد و تاکنون هیچ پژوهش پیکره بنیادی پیرامون اندازه‌گیری روانی در طول سطوح مختلف بسندگی زبان فارسی انجام نشده است. شایان ذکر است که از سال ۱۳۷۹ تا سال ۱۴۰۱، جمعاً ۳۱ پژوهش (برگرفته از ۹ پایگاه داده‌های فارسی‌زبان و وبگاه‌های مجلات مرتبط با آموزش فارسی به غیرفارسی‌زبانان، زبان‌شناسی و آموزش زبان فارسی و در قالب پایان‌نامه کارشناسی ارشد و

1. Aptis
2. PTE
3. TOEFL
4. IELTS

دکتری و مقاله پژوهشی و همایشی) پیرامون آزمون‌سازی فارسی به‌عنوان زبان دوم انجام شده است که ۲۰ عدد از آن‌ها به بررسی یا روایی‌سنجی آزمون‌های موجود یا محقق‌ساخته پرداخته‌اند؛ ۵ پژوهش به ضرورت‌ها، راهکارها و روش‌ها در این حیطه نظر افکنده‌اند و مابقی (یعنی ۶ عدد) نیز موضوعات متفرقه دارند اما هیچ‌کدام از آن‌ها به موضوع روانی در صحبت کردن به زبان فارسی نپرداخته‌اند. شاید بتوان گفت مرتبط‌ترین آن‌ها با پژوهش حاضر (تنها در مقام مقایسه)، از آن رودمعجنی و قبول (۱۳۹۹) باشد. آن دو با استفاده از نتایج به‌دست آمده از یکی از آزمون‌های برگزارشده در مرکز بین‌المللی آموزش زبان فارسی به غیرفارسی‌زبانان دانشگاه فردوسی مشهد، میزان اعتبار معیار نمره‌دهی صحبت کردن در این آزمون را اندازه‌گیری کردند و به این نتیجه رسیدند که این معیار دارای پایایی ۹۷ است و ارزیابان نیز می‌توانند این مقیاس را برای آزمون‌دهندگان با توانایی‌های مختلف به کار گیرند چون این مقیاس توانایی تمییز میان فارسی‌آموزان ضعیف، متوسط و قوی از یکدیگر را دارد. هرچند تعدادی از فارسی‌آموزان نیز هستند که هیچ نمره‌ای مناسب سطح بسندگی‌شان نیست. همچنین، هر سه سازه شیوه بیان، کیفیت زبان و بسط موضوع که در این معیار لحاظ شده‌اند، متناسب و دقیق بودند.

Estaji and Kia Shemshaki (2021) نیز برای بررسی نوع، بسامد و دقت استفاده از ابزارهای انسجامی در نگارش فارسی‌آموزان خارجی، نوشته‌های آن‌ها را ابتدا بر اساس مقیاس ارزیابی نوشتار بنیاد سعدی ارزیابی کردند. سپس فهرستی از ابزارهای انسجامی فارسی را از چهار پیکره بیجن‌خان، همشهری، سراجی و بنیاد سعدی استخراج و به سه دسته پرسامد، میان‌سامد و کم‌سامد تقسیم کردند. آن دو با بررسی ۲۰ نوشته از لحاظ نوع و بسامد ابزارهای انسجامی و درستی کاربرد آن‌ها، به این نتیجه رسیدند که استفاده از ابزارهای انسجامی از لحاظ تعداد و نوع، با افزایش سطح بسندگی بالا می‌رود اما ارتباط قابل توجهی میان سطح زبانی و کاربرد ابزارهای انسجامی پرسامد یا درستی استفاده از این ابزارها وجود ندارد. بررسی مقیاس ارزیابی نیز نشان داد پیش‌بینی‌هایی که در سطوح مختلف در خصوص ابزارهای انسجامی صورت می‌گیرد، جامع و کامل نیست.

همچنین جلیلی (۱۳۹۶) با اشاره به طاق‌فرسابودن تصحیح و ارزیابی تولیدات نوشتاری زبان‌آموزان، یک چارچوب ارزیابی جزئی‌نگر شامل سه مقوله دستور، دایره واژگانی و درست‌نویسی طراحی کرده است. او برای این کار از حدود ۱۰۰ متن نگارش‌یافته توسط

فارسی‌آموزان سطح پیشرفته مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام خمینی (ره) استفاده نموده و در چارچوب پیشنهادی خود، خطاهای نوشتاری فارسی‌آموزان را به چند دسته تقسیم کرده است.

گفتنی است در خصوص روانی گفتار در زبان فارسی به‌عنوان زبان اول و به‌صورت عام هم پژوهش‌ها بیشتر در حوزه علوم شناختی، پزشکی، تربیتی و توان‌بخشی انجام شده است. در حوزه زبان‌شناسی نیز مرتبط‌ترین پژوهش‌های یافت‌شده بیشتر ناظر به لایه آواشناسی و پیرامون موضوع کلی‌تر گفتار است که از آن میان می‌توان به زارع و همکاران (۱۳۹۹) و صادقی (۱۳۹۹) اشاره کرد.

از طرف دیگر، در حال حاضر، در هیچ نهاد یا مؤسسه‌ای، برای ارزیابی مهارت صحبت کردن فارسی‌آموزان خارجی از مقیاس‌های ارزیابی دسته دوم (= داده‌محور و پیکره‌بنیاد) استفاده نمی‌شود؛ هرچند در بعضی مؤسسه‌ها همچون بنیاد سعدی و مرکز آموزش زبان فارسی به غیرفارسی‌زبانان دانشگاه علامه طباطبائی از مقیاس‌های ارزیابی دسته نخست (= مبتنی بر دانش و تجربه) استفاده می‌شود. این دست مقیاس‌ها اگرچه از سختی ارزیابی می‌کاهند اما همان‌طور که ذکرش رفت، به‌هیچ‌وجه به اندازه مقیاس‌های داده‌محور دقیق نیستند و همین امر، احتمال اتخاذ تصمیم‌های نادرست را افزایش می‌دهد. پس جای یک مطالعه علمی و پیکره‌بنیاد که روانی را در سطوح بسندگی فارسی بررسی کند، در میان پژوهش‌های آزمون‌سازی زبان فارسی به‌عنوان زبان دوم خالی است.

در واکنش به مسئله معرفی‌شده، این پژوهش بر آن است تا با تمرکز بر داده‌های حاصل از مصاحبه شفاهی آزمونگر-آزمون‌دهنده در شرایط آزمونی، مؤلفه روانی در صحبت کردن را در آزمون‌دهندگان سطوح مختلف فارسی‌آموزی اندازه‌گیری کند و از این رهگذر، گوشه‌ای از مواد خام لازم برای تهیه یک مقیاس داده‌محور برای ارزیابی گفتار فارسی‌آموزان را فراهم نماید. این پژوهش پژوهشی است پیکره‌بنیاد که از سه وجه مهم و کلان روانی، یعنی سرعت^۱، شکست^۲ (مانند مکث^۳) و اصلاح^۴ (مانند خوداصلاحی^۵)، تکرار^۶

-
1. speed
 2. breakdown
 3. pause
 4. repair
 5. self-correction
 6. repetition

و خطای شروع^۱) (Skehan, 2003; Skehan, 2009; Tavakoli & Skehan, 2005; (Tavakoli & Wright, 2020)، بخشی از وجه سوم (=اصلاح)، یعنی تکرار و خطای شروع، را مورد مطالعه قرار می‌دهد.

پژوهش حاضر می‌تواند گوشه کوچکی از خلأ موجود در زمینه تأمین مواد لازم برای تهیه مقیاس‌های ارزیابی را پر کند. با در دست داشتن یافته‌های این پژوهش (و دیگر پژوهش‌های داده‌محور و پیکره‌بنیادی که هر کدام یکی از مؤلفه‌های تأثیرگذار در ارزیابی گفتار را به‌طور علمی بررسی و اندازه‌گیری کنند)، می‌توان به انجام قضاوت‌های عادلانه‌تر گفتار آزمون‌دهندگان فارسی‌آموز امیدوار بود. در واقع، انجام این پژوهش گامی در جهت تهیه یک مقیاس ارزیابی پایاست و بدیهی است که انجام‌نشدن آن به‌طور خاص، دست کم در دو جا، یعنی در خصوص دو نوع آزمون، موجب تصمیم‌گیری‌های نادرست و مسئله‌ساز می‌شود؛ یکی در آزمون‌های تعیین سطح که در ابتدای دوره‌های فارسی‌آموزی برگزار می‌شود و بر اساس نتیجه‌ای که از آن‌ها حاصل می‌شود، فارسی‌آموزان در سطوح گوناگون توزیع می‌شوند تا محتوایی متناسب با دانش و توانایی زبانی خود دریافت کنند و دیگری در آزمون‌های بسندگی (همچون آمفا) که نتیجه حاصل از آن‌ها بسیار سرنوشت‌ساز است. می‌توان نتیجه گرفت که تصمیم‌گیری اشتباه در این دو مورد (و علی‌الخصوص مورد دوم) موجب سردرگمی و بی‌اعتمادی آزمون‌دهندگان و سایر ذی‌نفعان و درنهایت، حتی خراب‌شدن وجه آزمون می‌شود و اعتبار آزمون و متولیان آن زیر سؤال می‌رود.

پیشینه پژوهش

در این بخش، به ارائه چند تعریف مهم از روانی گفتار در زبان دوم می‌پردازیم. همان‌طور که می‌توان حدس زد، تابه‌حال زوایای متفاوتی برای توصیف این مؤلفه گفتار اتخاذ شده است. مثلاً برخی تعاریف، متأثر از تعاریف روانی در زبان اول بوده‌اند؛ بعضی دید کل‌نگر و دسته‌ای نیز دید جزءنگر داشته‌اند. البته هم‌پوشانی‌هایی هم در میان این تعاریف دیده می‌شود. تعریف Fillmore (1979: 93) از این عنصر، اگرچه روانی گفتار در زبان اول را مدنظر دارد، اما به گواه Tavakoli and Wright (2020, P. 4)، یکی از دیدگاه‌های دوران‌ساز نسبت به روانی گفتار شمرده می‌شود. در واقع، این دیدگاه، که نگاهی کل‌گرا به روانی دارد،

1. false start

پایه و اساس تعاریفی شد که بعدتر و در سال‌های ۱۹۸۰ در خصوص روانی گفتار در زبان دوم ارائه شدند (Tavakoli & Wright, 2020, P. 5).

در این دیدگاه (Fillmore, 1979, P. 93)، یک «نوع» (تأکید از نگارنده است) از روانی توانایی حرف‌زدن به مدت طولانی با مکث‌های کم تعداد و پرکردن زمان با سخن است. کسی که به این نوع از روانی مجهز است، مجبور نیست به دفعات زیاد مکث کند تا به آنچه می‌خواهد بگوید یا به چگونگی بیان آن فکر کند. یکی از بهترین مثال‌ها از این نوع روانی را می‌توان در میان گزارشگران ورزشی پیدا کرد.

نوع دوم روانی توانایی حرف‌زدن به صورت پیوسته، منطقی و عمیقاً پرمعنی است. مهم‌ترین جزء سازنده این نوع توانایی تسلط بر معنی‌شناسی و نحو زبان است. افراد دارای این توانایی نیاز ندارند گفته خود را با تعداد زیادی محتوای خالی از معنی پر کنند.

نوع سوم روانی توانایی گفتن حرف‌های مناسب در گستره‌ای از بافت‌هاست. کسی که از این لحاظ روان است، همیشه درست‌ترین حرف را می‌زند و در بافت‌های مکالمه‌ای مختلف، از لحاظ زبانی مشکلی ندارد. در مقابل، کسی که از این لحاظ روان نیست، ممکن است در بافت‌های آشنا و خودمانی کاملاً احساس راحتی کند اما در حضور غریبه‌ها یا در هنگام رویارویی با بحرانی غیرمنتظره در تعامل انسانی ساکت می‌شود.

نوع چهارم روانی توانایی خلاقیت و خیال‌پردازی در کاربرد زبان، بیان ایده‌ها به روش‌های بدیع، ساختن لطیفه، پرداختن به آواهای واژه‌ها فارغ از معنی آن‌ها (یکی از بهترین نمونه‌های آن ساختن زبان‌گیره‌ها^۱ است)، تغییر دادن سبک، خلق استعاره‌ها و استفاده از آن‌ها و غیره است. برداشتی که شنونده چنین گوینده‌ای دارد، این است که او آنچه را که می‌گوید، از پیش سریعاً ویرایش می‌کند و قادر است گستره بسیار متنوعی از روش‌های جایگزین در واکنش به یک موقعیت را به سرعت بررسی کند و بهترین روش را که پرطمطراق‌ترین و استادانه‌ترین آن‌ها به نظر می‌رسد، برگزیند؛ بنابراین، یک زبان‌ور کاملاً با استعداد کسی است که هر چهار مهارت گفته‌شده را داشته باشد.

این تعریف مقدمه اولین تعاریف روانی زبان دوم قرار گرفت. اولین تعاریفی که از روانی گفتار در زبان دوم مطرح شدند، مربوط به اواسط دهه ۱۹۸۰ هستند. این تعاریف دیدی کلی نسبت به روانی داشتند (Tavakoli & Wright, 2020) و بر طبیعی بودن و عادی بودن کلام تأکید داشتند (Brumfit, 1984; De Keyser & Zamel, 1978 as cited in Buitrago,)

2017, p. 140). البته در همان زمان هم دیدگاه‌های بسیار جزئی نسبت به روانی وجود داشت. مثلاً Möhle (1984) تحت تأثیر مطالعات انجام‌شده در زمینه آسیب‌شناسی گفتار و زبان^۱، روانی را متشکل از اجزایی قابل اندازه‌گیری همچون نرخ گفتار^۲، طول و جایگاه مکث‌های خالی^۳، بسامد و توزیع مکث‌های پرشده^۴، تکرار، خوداصلاحی^۵ و مانند آن قلمداد می‌کند (Tavakoli & Wright, 2020; Lennon, 1990)؛ اجزایی که هنوز هم کم‌وبیش در مطالعات روانی اندازه‌گیری می‌شوند.

Lennon (1990) دو نوع مفهوم از روانی ارائه داد: مفهوم کلان و مفهوم خرد. او روانی را در مفهوم کلان خود، «اصطلاحی پوششی برای بسندگی شفاهی»^۶ (p. 389) می‌داند. همچنین روانی را در مفهوم خرد خود، در کنار درستی^۷، ارتباط^۸، تناسب، تلفظ، منبع واژگانی و مانند این‌ها، «جزئی احتمالاً جدایی‌پذیر از بسندگی شفاهی» (ibid) فرض می‌کند. جامع‌ترین دسته‌بندی در خصوص تعاریف ارائه‌شده از روانی و ابعاد مختلف آن، از آن Tavakoli and Hunter (2018) است. آن دو بر اساس داده‌های کمی و کیفی به دست آمده از ۸۴ معلم زبان دوم در انگلستان به این نتیجه رسیدند که چهار رویکرد نسبت به روانی در بین آن‌ها وجود دارد: یکی رویکرد بسیار کلان، که روانی را با کلماتی چون تسلط^۹ و بسندگی عمومی^{۱۰} تعریف می‌کند؛ دیگری رویکرد کلان، که روانی را مساوی با توانایی صحبت کردن به زبان دوم^{۱۱} می‌داند؛ سوم رویکرد خرد، که در آن روانی با آسانی^{۱۲}، سلاست^{۱۳} و تداوم^{۱۴} هم‌معنی است و متمایز از دقت و پیچیدگی است و درنهایت، رویکرد بسیار خرد، که دارای سه وجه سرعت، شکست و اصلاح است.

این رویکرد آخر (= بسیار خرد) همان است که پیش‌تر و در بخش «مقدمه و بیان مسئله» ذکرش رفت و در ادامه، بیشتر به آن پرداخته می‌شود. در واقع، این رویکرد مبنای انجام این

-
1. speech-language pathology
 2. speech rate
 3. length and positioning of silent pauses
 4. frequency and distribution of filled pauses
 5. self-correction
 6. oral proficiency
 7. correctness
 8. relevance
 9. mastery
 10. general proficiency
 11. L2 speaking ability
 12. ease
 13. flow
 14. continuity

پژوهش است و برای انتخاب آن دو دلیل وجود داشت: یکی این که این رویکرد ریزبینه‌ترین رویکرد در میان چهار رویکرد پیش‌گفته است و دیگر این که این رویکرد دیدی عینی^۱ و تجربه‌پذیر به روانی دارد و اجزایی را برای آن برمی‌شمرد که همگی آن‌ها به لحاظ علمی و تجربی قابل اندازه‌گیری هستند.

پیرامون رویکرد بسیار خرد نسبت به روانی که در «پیشینه پژوهش» به آن اشاره شد، باید دانست که در حال حاضر، یکی از تغییر جهت‌های عمده‌ای که مبحث روانی در دو دهه گذشته شاهد آن بوده است، آگاه‌شدن پژوهشگران از نیاز به اندازه‌گیری روانی از سه منظر سرعت، شکست و اصلاح است (Tavakoli & Wright, 2020, p. 48). بر اساس این سه منظر، این رویکرد از سه نوع روانی نام می‌برد: روانی سرعت^۲، روانی شکست^۳ و روانی اصلاح^۴.

منظور از روانی سرعت، مجموعه اندازه‌هایی^۵ است که سلاست و تداوم گفتار را ترسیم می‌کنند. روانی شکست به معنی اندازه‌هایی است که مکث‌ها و سکوت‌های شکننده سلاست گفتار را منعکس می‌کنند و روانی اصلاح نیز شامل اندازه‌هایی است که فرایندهای نظارتی و اصلاحی همچون تکرار و بازسازی^۶ را نشان می‌دهند (Tavakoli & Wright, 2020, p. 47).

متغیرهایی که در هر نوع روانی محاسبه می‌شوند، تنوع قابل ملاحظه‌ای دارند. پژوهش حاضر تنها ناظر به روانی اصلاح است که رایج‌ترین متغیرهای آن عبارت‌اند از: خوداصلاحی، خطای شروع و تکرار ناقص و کامل (اعم از واژه، گروه و جمله). خوداصلاحی به معنی اصلاح خطاهای زبان‌آموز توسط خود آن‌هاست (Pishghadam et al., 2011, p. 958 as cited in Sultana, 2009, p. 11). در خوداصلاحی، گوینده متوجه اشتباه خود در تلفظ، دستور، انتخاب واژه و مانند این‌ها می‌شود و فوراً آن را اصلاح می‌کند (Pishghadam et al., 2011, p. 958, as cited in Lam, 2006, p. 144). در پژوهش حاضر، هر جا که آزمون‌دهنده فارسی‌آموز متوجه اشتباه خود می‌شد و فوراً آن را اصلاح می‌کرد، یک مورد خوداصلاحی در نظر گرفته می‌شد.

1. objective
2. speed fluency
3. breakdown fluency
4. repair fluency
5. measures
6. reformulation

در خطای شروع، فرد نطق خود را شروع و سپس آن را، گویی فکر جدیدی به ذهنش رسیده باشد، پیش از پایان یافتن، ناگهان متوقف می‌کند و نطق جدیدی آغاز می‌کند (Fox Tree, 1995؛ مؤسسه آواشناسی و پردازش گفتار، بی تا، بند اول). برای اشاره به اطلاعات ناتمام و اطلاعات جدید به ترتیب از اصطلاح‌های خطای شروع و شروع دوباره^۱ استفاده می‌شود (Fox Tree, 1995, p. 710). در این پژوهش، هر جا که آزمون‌دهنده فارسی آموز نطق خود را شروع و آن را پیش از تمام شدن، ناگهان متوقف و نطقی جدید و متفاوت یا نامرتب با نطق پیشین را آغاز می‌کرد، یک مورد خطای شروع در نظر گرفته می‌شد. تکرار نیز در اینجا معنی و تعریف متفاوتی نسبت به تعریف عام آن ندارد و کم‌وبیش به همان معنی است که همگان می‌دانیم. این اصطلاح به معنی گفتن مجدد و بلافاصله واژه‌ها یا گروه‌هاست (Houston, 2016; Skehan et al., 2016). تنها چیزی که باید در نظر داشت، این است که فارسی آموز به‌طور طبیعی، گفته خود را گاهی ناقص و گاهی کامل تکرار می‌کند. شایان ذکر است پژوهش حاضر، از میان این سه متغیر، دو متغیر خطای شروع و تکرار را مورد بررسی قرار می‌دهد.

این پژوهش، با توجه به هدف و پرسش خود، رایج‌ترین مؤلفه‌های قابل محاسبه در بخشی از روانی اصلاح، یعنی تکرار واژه، گروه و جمله (اعم از ناقص و کامل) و خطای شروع، بر اساس Tavakoli and Wright (2020) را برای اندازه‌گیری برگزیده است. این پژوهش بر آن است تا با پیاده‌سازی داده‌های شفاهی آزمونگر-آزمون‌دهنده در شرایط آزمونی و برچسب‌زنی آن‌ها، یکی از سه وجه کلان روانی، یعنی اصلاح، و اجزای سازنده آن را در سطوح مختلف اندازه‌گیری کند و بدین ترتیب، بخش کوچکی از مقیاس آرمانی پیش‌گفته برای ارزیابی گفتار را فراهم کند. بر این اساس، می‌توان گفت هدف غایی این پژوهش و پژوهش‌های مشابهی که هر کدام گوشه‌ای از این جورچین را کامل کنند، تسهیل فرایند ارزیابی گفتار فارسی آموزان خارجی و نیز هرچه عادلانه‌تر کردن تصمیماتی است که بر پایه نتایج آزمون‌ها گرفته می‌شوند. بر اساس چارچوب نظری اتخاذشده، این پژوهش برای پاسخگویی به دو پرسش زیر انجام می‌شود:

۱. خطاهای شروع تا چه اندازه در میان آزمون‌دهندگان سطوح مختلف فارسی آموزی تغییر می‌کند؟

۲. بسامد تکرارهای ناقص و کامل (اعم از تکرار جمله، گروه و واژه) در میان آزمون‌دهندگان سطوح مختلف فارسی‌آموزی تا چه اندازه تغییر می‌کند؟

روش

در این پژوهش، گفتار تعداد ۲۳ آزمون‌دهنده فارسی‌آموز به کمک مقیاس ارزیابی گفتار بنیاد سعدی تعیین سطح شد. در این مرحله، گفتارهایی انتخاب شد که پژوهشگر در مورد سطح زبانی آن‌ها بیشترین اطمینان را داشت؛ یعنی با اطمینان (و در نتیجه سرعت) بسیار بیشتری درباره سطح بسندگی به نتیجه می‌رسید.

از مجموع ۲۳ گفتار ارزیابی شده، ۸ گفتار در سطح پیش‌میان، ۶ گفتار در سطح میانی، ۶ گفتار در سطح فوق‌میان و ۳ گفتار در سطح پیشرفته قرار گرفتند. مدت زمان گفتارها در هر سطح برابر با ۱۱ و در مجموع ۴۴ دقیقه است. همچنین، ۳ عدد از این گفتارها طی آزمون حضوری و ۲۰ عدد طی آزمون اینترنتی ضبط شده‌اند. در ضمن، ۸ تن از آزمون‌دهندگان مرد و ۱۵ تن زن بودند. پس از تعیین سطح، تمامی گفتارها با تمام جزئیات (اعم از واژه‌های تکراری و ناقص، خطاها و پرکننده‌ها) پیاده‌سازی و در پیکره بین‌المللی بنیاد سعدی به همراه فراداده‌ها درج شد.

ضمناً جهت پاسخ‌گویی به پرسش‌ها، تعداد ۷ برجسب روانی به قرار زیر، پس از تعریف در پیکره بین‌المللی بنیاد سعدی، بر روی گفتارهای درج‌شده زده شد. این برجسب‌ها از Tavakoli and Wright (2020) برگرفته شدند. یک برجسب مربوط به خطای شروع و ۶ برجسب متعلق به انواع تکرار بود.

۱. خطای شروع
۲. تکرار ناقص واژه
۳. تکرار ناقص گروه
۴. تکرار ناقص جمله
۵. تکرار کامل واژه
۶. تکرار کامل گروه
۷. تکرار کامل جمله

شایان‌ذکر است آنچه در Tavakoli and Wright (2020) آمده است، صرفاً تفکیک میان تکرار ناقص و کامل به‌طور کلی است؛ اما پژوهشگر با کمی دقت در داده‌های گفتاری

خود دریافت که این تکرار ناقص و کامل باید در میان سطوح نحوی نیز تفکیک شود؛ بنابراین، تقسیم‌بندی انواع تکرار به قسمی که در فهرست بالا مشاهده می‌شود، در آمد. پیکره بین‌المللی بنیاد سعدی حاوی ۶۱۳ نمونه گفتار فارسی آموزان خارجی است اما ویژگی‌هایی که در ادامه ذکر می‌شود، موجب شد نمونه‌گیری از آن‌ها کاملاً تصادفی نباشد: یکی این که کیفیت گفت‌وگوهای ضبط‌شده همگی بالا نبود و پژوهشگر مجبور بود بعضی از آن‌ها را نمونه‌گیری نکند؛ دیگر این که به‌طور طبیعی تعداد قابل توجهی از آن‌ها در سطوح میانی و پایین‌تر جای داشتند و ممکن نبود تعداد زیادی از آن‌ها در نمونه قرار بگیرد؛ سوم این که پژوهشگر سعی داشت تا آنجا که امکان دارد، تنوع ملیتی فارسی آموزان شرکت‌کننده در این پژوهش را بالا ببرد درحالی که پراکندگی فارسی آموزان در جهان، و بنابراین پراکندگی جغرافیایی فارسی آموزان شرکت‌کننده در این مصاحبه‌ها یکسان نیست؛ چهارم آن که پژوهشگر متغیرهای جنسیت و زبان اول را کنترل نمی‌کرد (یعنی قصد داشت از لحاظ جنسیت و زبان اول در میان آزمون‌دهندگان فارسی آموز مورد مطالعه خود تنوع ایجاد کند) و درنهایت، این که پژوهشگر، به دلیل دچاربودن به همان مسئله‌ای که انجام این پژوهش را ضروری می‌نماید (= عدم وجود مقیاس ارزیابی دقیق و پیکره‌بنیاد برای گفتار و به‌طور کلی تولیدات زبانی فارسی آموزان خارجی)، نمی‌توانست از نتیجه ارزیابی خود از گفتار تمام فارسی آموزان جامعه آماری‌اش مطمئن باشد. گفتنی است که به‌طور طبیعی و در عمل، ارزیابی بعضی از تولیدات زبانی، اعم از گفتار یا نوشتار، با دشواری کمتری همراه می‌شود بدین معنی که ارزیاب در قضاوت خود درباره آن‌ها به نتیجه مطمئن‌تری می‌رسد. غالباً دلیل این عدم دشواری آن است که مؤلفه‌های موردنظر، در فرایند یادگیری به‌صورت هماهنگ رشد کرده‌اند.

همچنین پیکره بنیاد سعدی حاوی گفتارهایی از سطوح مختلف بسندگی است اما سطوح نوآموز، مقدماتی و ماهر مندرج در مقیاس ارزیابی گفتار بنیاد سعدی، برای بررسی در این پژوهش انتخاب نشدند؛ اولی به این دلیل بدیهی که طبیعتاً به‌هیچ‌وجه تولید زبانی ندارد؛ دومی به این دلیل که در آن، گفتار همچنان بسیار پراکنده، خطاها بسیار زیاد و پیام ناتمام، گاهی نامفهوم و عمدتاً غیرقابل درک است و حتی مهارت فارسی آموز در شنیدن نیز به‌عنوان مهارتی ادراکی به‌اندازه‌ای پایین است که منظور آزمونگر را به‌زحمت می‌تواند بفهمد و سومی نیز به این دلیل که هم شمار فارسی آموزانی که در ایران بزرگ نشده باشند (یعنی زبان

فارسی زبان اولشان نباشد) و در عین حال، به آن دست یافته باشند، بسیار کم است و هم این که فرد در این سطح به نوعی زبان‌ور محسوب می‌شود و توانایی زبانی او، به گواه مقیاس ارزیابی مورد استفاده پژوهشگر، اختلاف و تفاوت بسیار اندکی با فارسی‌گویان بومی دارد و بنابراین نمی‌تواند در این پژوهش موضوع و مورد مطالعه باشد زیرا پژوهش حاضر روانی گفتار را در فارسی‌آموزان مورد بررسی قرار می‌دهد نه فارسی‌گویان؛ بنابراین، پژوهش حاضر به اندازه‌گیری روانی در ۴ سطح قرار گرفته در میان دو سطح نوآموز و مقدماتی از یک سو و سطح ماهر از سویی دیگر، یعنی به ترتیب سطوح پیش‌میانی، میانی، فوق‌میانی و پیشرفته، پرداخت.

آزمون گفتار بنیاد سعدی و مقیاس ارزیابی گفتار بنیاد سعدی ابزارهای این پژوهش بودند. آزمون گفتار بنیاد سعدی حاوی پرسش‌هایی دربارهٔ تجارب جهان‌شمول و زیستهٔ بشری، همچون موسیقی، خانواده، روابط انسانی، تحصیل، تعطیلات، بازی و ورزش، کتاب و مطالعه، تاریخ، وسایل نقلیه، ترافیک، غذا، فیلم، مهمانی، هدیه، تلویزیون، سفر و مانند این‌ها بود و بنابراین، صرفاً توانایی عمومی زبانی فارسی‌آموزان را می‌سنجید. همچنین آزمونگر در مقام پرسشگر و آزمون‌دهنده در مقام پاسخگو ظاهر می‌شدند و بنابراین، تقریباً تمام گفتار آزمون‌دهنده شامل جملات غیرپرسی ساده و مرکب بود. در این آزمون، عملکرد گفتاری آزمون‌دهندگان فارسی‌آموز بر اساس مقیاس ارزیابی بنیاد سعدی به تفکیک چهار جنبهٔ گفتار که در زیر آمده است، مورد ارزیابی قرار گرفت:

۱. روانی و پیوستگی

۲. منبع واژگانی

۳. دقت و غنای دستوری

۴. تلفظ

به هر یک از این چهار مؤلفه، بر اساس مقیاس ارزیابی گفتار بنیاد سعدی، یکی از سطوح ۱ تا ۷ داده شد و میانگین سطوح چهار معیار، پس از گردشدن به سمت پایین، به عنوان سطح فرد در آزمون گفتار در نظر گرفته شد.

در این پژوهش، همان‌طور که گفته شد، تعداد هفت متغیر روانی گفتار مورد بررسی قرار گرفتند. نام کامل فارسی و انگلیسی این متغیرها در جدول زیر قابل مشاهده است:

جدول ۱. نام فارسی و انگلیسی متغیرهای پژوهش

شماره	نام فارسی پارامتر	نام انگلیسی پارامتر
۱	تکرار ناقص واژه	partial-word-repetition
۲	تکرار ناقص گروه	partial-phrase-repetition
۳	تکرار ناقص جمله	partial-sentence-repetition
۴	تکرار کامل واژه	complete-word-repetition
۵	تکرار کامل جمله	complete-sentence-repetition
۶	تکرار کامل گروه	complete-phrase-repetition
۷	خطای شروع	false-start

سطوح زبانی موردبررسی در پژوهش مطابق «چارچوب مرجع بنیاد سعدی» (صحرائی و مرصوص، ۱۳۹۵) به شرح جدول زیر هستند. در پژوهش حاضر، گفتار فارسی آموزان چهار سطح پیش‌میانی، میانی، فوق‌میانی و پیشرفته بررسی شدند.

جدول ۱. نام سطوح بسندگی چارچوب مرجع بنیاد سعدی و وضعیت آن‌ها در پژوهش حاضر

شماره	نام سطح	موجود در پژوهش حاضر؟
۱	نوآموز	نیست
۲	مقدماتی	نیست
۳	پیش‌میانی	هست
۴	میانی	هست
۵	فوق‌میانی	هست
۶	پیشرفته	هست
۷	ماهر	نیست

در این پژوهش، هر یک از پارامترهای پژوهش را در چهار سطح زبانی بررسی کردیم؛ بنابراین، پارامترهای نامبرده، متغیرهای وابسته و سطوح زبانی متغیرهای مستقل پژوهش محسوب می‌شدند.

نتایج

در این پژوهش، گفتار ۲۳ فارسی‌آموز از سطوح مختلف بسندگی بررسی شد. تعداد متون فارسی‌آموزان هر سطح و میانگین تعداد واژه‌ها و کاراکترهای هر سطح و همچنین مجموع تعداد واژه در هر سطح به شرح جدول زیر است.

جدول ۲. اطلاعات آماری اولیه متغیرها

شماره	نام سطح	تعداد متن	میانگین تعداد واژه	میانگین تعداد کاراکتر	مجموع تعداد واژه	مدت زمان
۱	پیش‌میانی	۸	۱۰۹,۵	۴۸۱,۵	۸۷۶	۱۱ دقیقه
۲	میانی	۶	۱۷۲,۶	۷۶۱,۶	۱۰۳۶	۱۱ دقیقه
۳	فوق‌میانی	۶	۲۰۷	۹۴۳,۳	۱۲۴۲	۱۱ دقیقه
۴	پیشرفته	۳	۴۱۷	۱۹۸۱,۶	۱۲۵۱	۱۱ دقیقه

جدول زیر تعداد، میانگین و انحراف معیار هر یک از پارامترهای پژوهش را در هر سطح نشان می‌دهد.

جدول ۳. تعداد، میانگین و انحراف معیار هر متغیر

پارامتر	سطح	تعداد	میانگین	انحراف معیار
تکرار ناقص واژه	پیش‌میانی	۲۱	۲,۶۲۵	۳,۳۷۸
	میانی	۱۲	۲	۲,۶۰۸
	فوق‌میانی	۹	۱,۵	۱,۳۷۸
	پیشرفته	۵	۱,۶۶۷	۱,۵۲۸
تکرار ناقص گروه	پیش‌میانی	۱۲	۱,۵	۱,۵۱۲
	میانی	۸	۱,۳۳۳	۱,۹۶۶
	فوق‌میانی	۵	۰,۸۳۳	۰,۹۸۳
	پیشرفته	۱۰	۳,۳۳۳	۴,۹۳۳
تکرار ناقص جمله	پیش‌میانی	۰	۰	۰
	میانی	۰	۰	۰
	فوق‌میانی	۰	۰	۰
	پیشرفته	۲	۰,۶۶۷	۱,۱۵۵
تکرار کامل واژه	پیش‌میانی	۱۶	۲	۳,۱۶۲
	میانی	۱۴	۲,۳۳۳	۲,۵۸۲
	فوق‌میانی	۴	۰,۶۶۷	۰,۵۱۶
	پیشرفته	۵	۱,۶۶۷	۲,۰۸۲
تکرار کامل جمله	پیش‌میانی	۱	۰,۱۲۵	۰,۳۵۴
	میانی	۱	۰,۱۶۷	۰,۴۰۸
	فوق‌میانی	۱	۰,۱۶۷	۰,۴۰۸
	پیشرفته	۰	۰	۰
تکرار کامل گروه	پیش‌میانی	۰	۰	۰

پارامتر	سطح	تعداد	میانگین	انحراف معیار
	میانی	۰	۰	۰
	فوق میانی	۰	۰	۰
	پیشرفته	۰	۰	۰
خطای شروع	پیش میانی	۳	۰,۳۷۵	۰,۷۴۴
	میانی	۲	۰,۳۳۳	۰,۵۱۶
	فوق میانی	۵	۰,۸۳۳	۱,۶۰۲
	پیشرفته	۶	۲	۱,۷۳۲

جدول ۴ را می‌توان به صورت جدول ۵ خلاصه کرد:

جدول ۵. تغییرات وضعیت هر متغیر در طول سطوح بسندگی

شماره	پارامتر	تغییرات وضعیت در سطوح بسندگی
۱	تکرار ناقص واژه	پیش میانی < میانی < فوق میانی < پیشرفته
۲	تکرار ناقص گروه	پیش میانی < میانی < فوق میانی > پیشرفته
۳	تکرار ناقص جمله	پیش میانی = میانی = فوق میانی > پیشرفته
۴	تکرار کامل واژه	پیش میانی < میانی < فوق میانی > پیشرفته
۵	تکرار کامل جمله	پیش میانی = میانی = فوق میانی < پیشرفته
۶	تکرار کامل گروه	پیش میانی = میانی = فوق میانی = پیشرفته
۷	خطای شروع	پیش میانی < میانی > فوق میانی > پیشرفته

طبق گفته Strunk and Mwavita (2022) بررسی یک متغیر میان چند گروه از روش آنوای تک‌راهه^۱ امکان‌پذیر است. در واقع آزمون تی^۲ نیز حالت خاصی از آنوا است (هر دو آزمون از نوع مدل خطی عمومی^۳ هستند). انجام آزمون آنوای پنج پیش فرض دارد:

۱. نوع متغیر وابسته باید فاصله‌ای^۴ یا نسبی^۵ باشند. چون پارامترهای این پژوهش در هر متن شمارش شده‌اند، نوع آن‌ها فاصله‌ای است.

1. One Way ANOVA
2. t-test
3. General Linear Model
4. interval
5. ratio

۲. متغیر وابسته باید توزیع نرمال^۱ داشته باشد. بررسی توزیع نرمال در این پژوهش با آزمون شاپیرو-ویلک^۲ انجام می‌شود. در این آزمون سطح معنی‌داری کمتر از ۰,۰۵ به معنی عدم توزیع نرمال داده‌هاست.

۳. مشاهدات باید مستقل باشند. چون در این پژوهش، مصاحبه با هر فارسی‌آموز جداگانه انجام شده است، مشاهدات بر یکدیگر تأثیری ندارند و مستقل از هم هستند.

۴. انتخاب داده‌ها باید تصادفی باشد. در این پژوهش، از میان افراد مختلفی که در آزمون گفتار شرکت کرده‌اند، چند فرد به شکل تصادفی تعدادی انتخاب شده‌اند.

۵. واریانس داده‌ها باید همگن^۳ باشد. این موضوع با استفاده از آزمون لوین^۴ بررسی می‌شود. فرض صفر^۵ این آزمون برابری واریانس‌ها در میان گروه‌هاست؛ بنابراین اگر سطح معنی‌داری آزمون کمتر از ۰,۰۵ باشد به معنی عدم همگنی واریانس داده‌هاست.

در این پژوهش تحلیل‌های آماری با استفاده از نرم‌افزار جامووی^۶ نسخه ۲,۲,۵ انجام شده‌اند (The jamovi project, 2021; Fox & Weisberg, 2022; R Core Team, 2021).

نتایج بررسی توزیع نرمال داده‌ها به شرح جدول زیر است. مشخص است که در همه‌ی موارد سطح معنی‌داری کمتر از ۰,۰۵ است و بنابراین توزیع این داده‌ها را نمی‌توان نرمال فرض کرد. در مورد تکرار کامل گروه چون هیچ موردی مشاهده نشد، این آزمون و آزمون‌های بعدی انجام نمی‌شوند.

جدول ۵. نتیجه بررسی توزیع نرمال

پارامتر	آماره (W)	سطح معنی‌داری
تکرار ناقص واژه	۰,۸۴۵	۰,۰۰۲
تکرار ناقص گروه	۰,۹۱۲	۰,۰۴۵
تکرار ناقص جمله	۰,۴۶۸	۰,۰۰۱

1. normal distribution
2. Shapiro-Wilk
3. homogeneity
4. Levene's Test
5. Null hypothesis
6. jamovi

پارامتر	آماره (W)	سطح معنی‌داری
تکرار کامل واژه	۰,۸۳۸	۰,۰۰۲
تکرار کامل جمله	۰,۵۱۸	۰,۰۰۱
خطای شروع	۰,۷۹۵	۰,۰۰۱

جدول زیر نتایج بررسی همگنی واریانس داده‌ها را نشان می‌دهد. در دو مورد شرط همگنی نقض شده است و در سایر موارد این شرط برقرار است.

جدول ۷. نتیجه بررسی همگنی واریانس

پارامتر	آماره (F)	سطح معنی‌داری
تکرار ناقص واژه	۰,۸۶۳	۰,۴۷۷
تکرار ناقص گروه	۷,۰۵	۰,۰۰۲
تکرار ناقص جمله	۴۴,۰۵۸	۰,۰۰۱
تکرار کامل واژه	۲,۵۷۸	۰,۰۸۴
تکرار کامل جمله	۰,۹۵۲	۰,۴۳۶
خطای شروع	۲,۱۹۶	۰,۱۲۲

مشخص است که در هیچ موردی نمی‌توان از آزمون آنووا استفاده کرد و باید از معادل ناپارامتری این آزمون یعنی آزمون کراسکال-والیس^۱ استفاده کرد چون در تمامی موارد یک یا هر دو شرط توزیع نرمال و همگنی واریانسها نقض شده است. جدول زیر نتایج اجرای آزمون ناپارامتری کراسکال-والیس برای پارامترها را نشان می‌دهد. سطح معنی‌داری در تمامی پارامترها بیشتر از ۰,۰۵ است و نمی‌توان تفاوت بین گروه‌ها در این پارامترها را معنی‌دار تلقی کرد.

جدول ۶. نتیجه آزمون کراسکال-والیس

پارامتر	آماره (χ ²)	سطح معنی‌داری
تکرار ناقص واژه	۰,۱۲۵	۰,۹۸۹
تکرار ناقص گروه	۰,۷۸۲	۰,۸۵۴
تکرار ناقص جمله	۶,۶۶۷	۰,۰۸۳
تکرار کامل واژه	۱,۳۱۴	۰,۷۲۶
تکرار کامل جمله	۰,۵۶۵	۰,۹۰۴
خطای شروع	۵,۲۳۷	۰,۱۵۵

1. Kruskal-Wallis

بحث

پژوهش حاضر با هدف اندازه‌گیری بخشی از سازه روانی اصلاح (که به همراه روانی سرعت و روانی شکست، یک انگاره کامل از روانی گفتار ارائه می‌دهند)، انجام شد و درصدد پاسخگویی به این پرسش بود که دو متغیر خطای شروع و تکرار چگونه در طول سطوح بسندگی زبان فارسی تغییر می‌کنند. این پژوهش (به همراه دیگر پژوهش‌هایی که ابعاد دیگری از گفتار، همچون دستور، واژه، تلفظ، روانی و پیوستگی و مانند این‌ها، را بررسی و در سطوح مختلف بسندگی اندازه‌گیری کنند)، در نهایت، مواد خام تهیه یک مقیاس ارزیابی پایا برای گفتار فارسی‌آموزان خارجی را فراهم خواهد نمود. یافته‌های این پژوهش وضعیت رشد روانی (یا در واقع، کاهش ناروانی) در طول سطوح بسندگی را روشن می‌کنند و این که آیا هیچ کدام از آن‌ها می‌توانند یک سطح را از سطحی دیگر متمایز کنند یا خیر. اینک بر اساس ارقام جدول ۸ (که نشان‌دهنده نوع تفاوت میان سطوح بسندگی در خصوص متغیرهای این پژوهش هستند)، یافته‌های پژوهش را مرور می‌کنیم.

خطای شروع در میان سطوح بسندگی مورد مطالعه در این پژوهش، همواره در حال تغییر است. سطوح پیش‌میانی و میانی با سطوح فوق‌میانی و پیشرفته تفاوت آشکاری دارند. دو سطح نخست با یکدیگر و دو سطح دوم هم با یکدیگر نزدیکی قابل ملاحظه‌ای دارند؛ یعنی این متغیر تنها دو سطح نخست را از دو سطح دوم متمایز می‌کند نه تک‌تک سطوح را از یکدیگر. ممکن است دلیل این امر نوع پرسش‌های مطرح‌شده در آزمون باشد، شاید محدودتر کردن موضوع مورد سؤال در آزمون، نتایج دیگری به دست دهد؛ گو این که تمامی موضوعات مطرح‌شده در آزمون عمومی و جهان‌شمول بودند. همچنین، اجرای آزمون ناپارامتری کراسکال-والیس نشان داد در خصوص متغیر خطای شروع، تفاوت معنی‌داری میان آزمودنی‌های سطوح مختلف وجود ندارد.

متغیر تکرار خود به شش نوع تکرار ناقص واژه، تکرار ناقص گروه، تکرار ناقص جمله، تکرار کامل واژه، تکرار کامل گروه و تکرار کامل جمله تقسیم می‌شود که در ادامه، به طور جداگانه به هر یک از این انواع می‌پردازیم جز تکرار کامل گروه که اساساً هیچ رخدادی از آن در گفتارهای مورد بررسی یافت نشد.

تکرار ناقص واژه تنها متغیر این پژوهش است که از سطح پیش‌میانی تا پیشرفته به طور پیوسته رو به کاهش بوده است. در این میان، سطح پیش‌میانی با همه سطوح بالاتر از خود

آشکارا تفاوت دارد؛ اما سه سطح میانی، فوق میانی و پیشرفته، نسبتاً اختلاف کمتری با هم دارند. ممکن است این امر جنبه روانی داشته باشد و آن این که با افزایش سطح زبانی، فرد به طور طبیعی اعتماد بیشتری به دانش واژگانی خود پیدا می‌کند و کمتر دچار خطا در کاربرد واژه‌ها می‌شود. در ضمن، آزمون کراسکال-والیس تفاوت معنی‌داری میان سطوح مختلف نشان نداد.

در خصوص تکرار ناقص گروه، بیشترین نزدیکی و شباهت میان سطوح پیش میانی و پیشرفته از یک سو و سطوح میانی و فوق میانی از سوی دیگر دیده می‌شود هرچند این متغیر از ابتدا تا سطح فوق میانی با سرعت یکسانی رو به کاهش بوده است. احتمال دارد دلیل افزایش دوبرابری آن در سطح پیشرفته نسبت به سطح فوق میانی این باشد که آزمون دهنده با توجه به توانایی زبانی خود، در تلاش است مفاهیم پیچیده‌تری را بیان کند. به هر حال، آزمون کراسکال-والیس از این لحاظ، تفاوت معنی‌داری میان سطوح مختلف نشان نداد.

تکرار ناقص جمله تنها در سطح پیشرفته دارای رخداد بوده است. پس این سطح تفاوت قابل ملاحظه‌ای با سطوح پیش از خود دارد. هرچند این تفاوت به گواه آزمون کراسکال-والیس معنی‌دار نیست. احتمالاً این امر به دلیل مشابهی که در مورد تکرار ناقص گروه عنوان شد، رخ داده باشد.

در خصوص متغیر تکرار کامل واژه، تمایز و تفاوت فاحشی میان دو سطح پیش میانی و میانی از یک سو و دو سطح فوق میانی و پیشرفته از سوی دیگر برقرار است. این متغیر نیز مانند متغیر خطای شروع تنها دو سطح نخست را از دو سطح دوم متمایز می‌کند نه تک تک سطوح را از یکدیگر. ممکن است علت این امر همان علتی باشد که در خصوص خطای شروع ذکر شد. همچنین آزمون کراسکال-والیس مؤید معنی‌دار نبودن تفاوت میان سطوح بسندگی از لحاظ متغیر تکرار کامل واژه است.

در نهایت، تکرار کامل جمله تا سطح فوق میانی ثابت بوده و سپس کاهش پیدا کرده است؛ بنابراین سطح پیشرفته با دیگر سطوح از این نظر متفاوت است و آزمون کراسکال-والیس هم تفاوت میان سطوح را معنی‌دار تشخیص نداد. پس باید گفت هیچ یک از متغیرهای پژوهش حاضر تفاوت معنی‌داری میان سطوح ایجاد نمی‌کنند.

آنچه در باب یافته‌های این پژوهش گفته شد، قابل مقایسه با چهار پژوهش جدید است که هم از لحاظ شرایط جمع‌آوری داده‌ها (=شرایط آزمونی)، هم (بخشی از) متغیرهای

موردبررسی و هم تنوع زبان و ملیت آزمودنی‌ها شباهت بسیاری با یکدیگر و با پژوهش حاضر دارند.

دو عدد از این پژوهش‌ها در Tavakoli و همکاران (2017) و Tavakoli و همکاران (2020) است که هر دو با اندازه‌گیری عملکردهای گفتاری ۳۲ آزمون‌دهنده در تکلیف‌های مختلف آزمون گفتار ایتیس و در طول سطوح بسندگی A2 تا C1 در چارچوب مشترک اروپایی (مجموعاً چهار سطح)، به این نتیجه رسیدند که متغیرهای روانی اصلاح و از جمله تکرار و خطای شروع پایداری بالایی در طول سطوح نداشتند به این صورت که در سطوح A2 و B1 شاهد بیشترین میزان کاربرد متغیرها بودند. تحلیل آن‌ها این بود که گویندگان سطح A2 بین هر واژه مکث می‌کردند تا میزان نیاز به اصلاح را به حداقل برسانند و گویندگان سطح B1 هنگام تولید نطق‌های طولانی‌تر اصلاحات بیشتری به کار می‌بردند. سپس میزان این استفاده، در سطوح B2 و C1 معتدل‌تر می‌شد. جالب آن‌که تنوع تکلیف‌های آزمونی در این پژوهش، تأثیری در نتایج مذکور ایجاد نکرد.

علاوه‌براین، Tavakoli و همکاران (forthcoming) با تکرار پژوهش Tavakoli و همکاران (۲۰۱۷)، این بار در خصوص ۶۰ شرکت‌کننده در آزمون تیپ^۱، نشان دادند متغیرهای اصلاح و از جمله تکرار و خطای شروع تمایزی بین سطوح ایجاد نمی‌کردند و هیچ کدام از متغیرهای اندازه‌گیری‌شده در دو سطح B2 و C1 را از هم منفک نمی‌کردند (as cited in Tavakoli & Wright, 2020, P. 115). شایان‌ذکر است تفاوت پژوهش اخیر با دو پژوهش پیش‌گفته یکی در تعداد آزمون‌دهنده‌ها (به ترتیب ۶۰ در مقابل ۳۲) و دیگری در حالت گفتار (به ترتیب گفتگوی حضوری در مقابل رایانه واسط غیرهمزمان) بود. همچنین، Tavakoli و همکاران (2023) با بررسی عملکرد ۵۶ آزمون‌دهنده در تکلیف تک‌گویی آزمون تیپ در چهار سطح بسندگی که قابل‌مقایسه با سطوح ۵، ۵٫۵، ۶٫۵ و ۷٫۵ در نظام نمره‌دهی آزمون آیلتس هستند، به این نتیجه رسیدند که هیچ تمایزی میان سطوح بسندگی از منظر این متغیرها وجود ندارد. لازم به ذکر است Tavakoli و همکاران (2023) مجموع متغیرهای سازه اصلاح را اندازه‌گیری کرده‌اند نه هر کدام را جداگانه.

بدین ترتیب، مشاهده می‌شود که یافته‌های پژوهش حاضر مشابه و مؤید پژوهش‌های مرتبط پیشین که همگی در رابطه با تولیدات گفتاری انگلیسی‌آموزان بودند، هستند. این بدان معنی است که در خصوص هیچ‌کدام از متغیرها تفاوت معنی‌داری بین سطوح بسندگی

1. TEEP (Test of English for Educational Purposes)

موردنظر در این پژوهش، یعنی سطوح پیش‌میانی، میانی، فوق‌میانی و پیشرفته مشاهده نشد؛ بنابراین، سطح فارسی‌آموزان تأثیری بر میزان استفاده آن‌ها از این متغیرها ندارد. در پایان، لازم به ذکر است سایر منابعی که پژوهشگر به بررسی و مطالعه آن‌ها پرداخت، اگرچه تکرار و دیگر متغیرهای سازه اصلاح را مدنظر داشتند اما در آن‌ها رابطه این متغیرها با متغیرهای دیگری همچون تکرار تکلیف، سختی تکلیف، ارزیابی ارزیابان و مانند این‌ها سنجیده شده بود و تعداد پژوهش‌هایی که نوع و میزان ارتباطشان با این پژوهش، تعیین‌کننده یا حتی صرفاً روشنگر و راهگشا باشد، اندک بود.

نتایج حاصل از این پژوهش بعضی از متغیرهایی که لزوماً ارتباط مستقیمی با سطح بسندگی ندارند را بر ما روشن کرد. زین پس می‌دانیم که دست‌کم در شرایطی مانند شرایط حاکم بر این پژوهش، فرض کاهش یا افزایش معنی‌دار متغیرهای تکرار و خطای شروع با تغییر سطح بسندگی، فرضی نادرست است و با مشاهده این متغیرها در گفتار فارسی‌آموزان خارجی نمی‌توانیم در خصوص سطح بسندگی آن‌ها به قطعیت برسیم. این امر وقتی اهمیت خود را نشان می‌دهد که بدانیم ارزیابی مبتنی بر متغیرهایی بی‌ارتباط با سطح بسندگی، مخصوصاً در آزمون‌های مقیاس بالا (همچون آمفا) ممکن است چه تأثیرات جدی‌ای بر نتایج نهایی آزمون بگذارد.

با وجود این که این پژوهش به پرسش مهمی پاسخ داد، هنوز می‌توان با کنترل کردن یا افزودن بعضی متغیرها به یافته‌های دقیق‌تری رسید. برای نمونه، می‌توان شرایط روانی یا محیطی حاکم بر آزمون را کنترل کرد یا تنها یک موضوع برای پرسیدن در آزمون (مثلاً هدیه‌دادن) برگزید. همچنین، حال که می‌دانیم چه متغیرهایی متأثر از سطح بسندگی فارسی‌آموزان نیستند، باید بفهمیم چه متغیرهایی با تغییر سطح بسندگی به‌طور معنی‌داری تغییر می‌کنند تا بتوانیم به مقیاس ارزیابی آرمانی مدنظر خود دست یابیم.

نتیجه‌گیری

هدف از انجام این پژوهش اندازه‌گیری بخشی از سازه روانی اصلاح به‌عنوان یکی از سه وجه انگاره روانی گفتار (Skehan, 2003; 2009) بود که روانی را دارای سه وجه سرعت، شکست و اصلاح می‌داند. بررسی گفتار ۲۳ آزمون‌دهنده فارسی‌آموز در چهار سطح پیش‌میانی، میانی، فوق‌میانی و پیشرفته از این منظر نشان داد هیچ‌یک از هفت متغیر روانی اصلاح که در این پژوهش مدنظر بودند، یعنی خطای شروع، تکرار ناقص واژه، تکرار ناقص

گروه، تکرار ناقص جمله، تکرار کامل واژه، تکرار کامل گروه (که البته هیچ رخدادی نداشت) و تکرار کامل جمله نمی‌توانند سطوح بسندگی زبان فارسی را از یکدیگر متمایز کنند. اگرچه در دو مورد از آن‌ها (خطای شروع و تکرار کامل واژه) میان دو سطح نخست از یک سو و دو سطح دوم از سوی دیگر تفاوت بسامدی آشکاری ملاحظه شد اما در مجموع و بر اساس نتیجه آزمون کراسکال-والیس، تفاوت این متغیرها میان تک تک سطوح معنی‌دار نبود.

این نتیجه‌گیری دست‌کم برای دو گروه از فعالان حوزه آموزش زبان فارسی کاربرد دارد. یکی برای کسانی که به‌واسطه فعالیت در این حوزه، نیاز دارند هرازگاهی به دلایل مختلف، گفتار فارسی آموزان را ارزیابی (=تعیین سطح) کنند. این افراد نیاز ندارند در ارزیابی خود به این امر دقت کنند که آزمون‌دهنده تا چه اندازه مرتکب خطای شروع می‌شود یا واژه‌ها و جمله‌ها را تکرار می‌کند. دوم برای گروه‌ها یا مؤسسه‌هایی که قصد تهیه یک مقیاس ارزیابی گفتار را دارند. این گروه‌ها یا مؤسسات می‌توانند فارغ از متغیرهایی که در این پژوهش مورد بررسی قرار گرفت، دست به تهیه مقیاس بزنند و برای این کار، به دنبال متغیرهای دیگری باشند که سطوح بسندگی را از یکدیگر متمایز کنند.

این پژوهش مانند بسیاری از پژوهش‌های دیگر دچار محدودیت‌هایی بود. مهم‌ترین محدودیت این پژوهش نبود مقیاس ارزیابی مناسب برای تعیین سطح یا همان ارزیابی گفتارها بود. در واقع، این پژوهش خود گامی در جهت تهیه یک مقیاس ارزیابی پایا بود. برای کاهش اثر این محدودیت، تنها گفتارهایی را انتخاب شد که کمترین تردید را در خصوص نتیجه ارزیابی (=سطح تعیین‌شده نهایی) در ذهن ایجاد می‌کردند. دیگر محدودیت این پژوهش این بود که حتی یک نمونه مشابه آن درباره فارسی آموزان خارجی انجام نشده بود تا پژوهشگران با مراجعه به آن بتوانند تصور روشن‌تری از آنچه باید در جریان پژوهش انجام دهند، کسب کنند. حتی تعداد پژوهش‌های خارجی مشابه پژوهش حاضر نیز اندک بود و این کار را برای پژوهشگران سخت‌تر می‌کرد. هرچند اندک پژوهش‌های موجود (سه عدد) هم با این که تنها بخشی از آن‌ها به تکرار و خطای شروع اختصاص داشت، از نظر شرایط جمع‌آوری داده‌ها و تنوع زبانی و ملیتی آزمون‌دهنده‌ها شباهت‌های قابل‌توجهی با این پژوهش داشتند. در پایان، پیشنهاد می‌شود متغیرهای این پژوهش در شرایط پژوهشی دیگری از جمله در شرایط غیرآزمونی، با پرسش‌های آزمونی یکسان (مثلاً فقط درباره ترافیک از

آزمون‌دهندگان سؤال شود)، با کنترل متغیرهایی نظیر ملیت، جنسیت و سن، با مدت‌زمان بیشتری از گفتار به ازای هر سطح و با یک مقیاس ارزیابی دیگر، اندازه‌گیری شوند.

تعارض منافع

نویسندگان هیچ‌گونه تعارض منافی ندارند.

منابع

- جلیلی، سید اکبر. (۱۳۹۶). ارزیابی تولیدات نوشتاری فارسی‌آموزان سطح پیشرفته: طراحی یک چارچوب جزئی‌نگر. *پژوهش‌نامه آموزش زبان فارسی به غیرفارسی‌زبانان*، ۶(۱)، ۲۹-۵۶.
- رودمعجنی، محسن و قبول، احسان. (۱۳۹۹). اعتبارسنجی معیار نمره‌دهی مهارت صحبت‌کردن در آزمون جامع زبان فارسی فردوسی. *جستارهای زبانی*، ۱۱(۴)، ۱۸۳-۲۱۰.
<http://lrr.modares.ac.ir/article-14-36045-fa.html>
- زارع، سعیده، صفایی‌قلاتی، مهدی و یاراحمدزهی، ناهید. (۱۳۹۹). انتقال آوایی در گفتار فارسی فارسی‌آموزان چینی با تمرکز بر فضای واکه‌ای. *پژوهش‌نامه آموزش زبان فارسی به غیرفارسی‌زبانان*، ۹(۱)، ۱۴۹-۱۷۶. doi: 10.30479/JTPSOL.2020.14062.1490
- صادقی، وحید. (۱۳۹۹). انتقال آوایی در گفتار فارسی‌آموزان چینی و عربی: مطالعه موردی انفجاری‌های بدنه‌ای. *پژوهش‌نامه آموزش زبان فارسی به غیرفارسی‌زبانان*، ۹(۱)، ۳-۲۴. doi: 10.30479/JTPSOL.2020.10185.1428
- صحرائی، رضامراد و مرصوص، فائزه. (۱۳۹۵). *استاندارد مرجع آموزش زبان فارسی*. تهران، ایران: دانشگاه علامه طباطبایی.

References

- Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge, NY: Cambridge University Press.
- Buitrago, A. G. (2017). Collaborative and self-directed learning strategies to promote fluent EFL speakers. *English Language Teaching*, 10(5), 139-157. <https://doi.org/10.5539/elt.v10n5p139>
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of Language Testing*. Cambridge, England: Cambridge University Press.
- De Keyser, R. & Zamel, V. (1987). Longman dictionary of applied linguistics. *TESOL Quarterly*, 21(2), 351-353. <https://doi.org/10.2307/3586739>
- Estaji, M. & Kia Shemshaki, L. (2021). Representations and uses of conjunctions in Persian learners' academic writings: The predictive power of Saadi Foundation writing rubric. In A. Aghdassi (Ed.), *Perspectives on academic Persian* (pp. 219-238). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-75610-9_12

- Estaji, M. & Rajabi, M. (2019). The Use of Self-Repair Strategies in Classroom Conversations: Does the teacher's level of reflection make a difference?. *Applied Research on English Language*, 8(3), 423-448. <https://doi.org/10.22108/are.2019.114882.1404>
- Fillmore, C. J. (1979). On fluency. In C. J. Fillmore, D. Kempler and W. Wang (Eds.), *Individual Differences in Language Ability and Language Behavior* (pp. 85–101). New York, United States: Academic Press.
- Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, 34(6), 709-738. <https://doi.org/10.1006/jmla.1995.1032>
- Fox, B. & Jasperson, R. (1995). A syntactic exploration of repair in English conversation. In P. W. Davis (Ed.), *Alternative linguistics: Descriptive and theoretical modes* (pp. 77-134). Amsterdam, Netherlands: J. Benjamins.
- Fox, J. & Weisberg, S. (2022). *car: Companion to Applied Regression*. [R package]. Retrieved from <https://cran.r-project.org/package=car>
- Fulcher, G. & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Oxon, England: Routledge.
- Fulcher, G. (2010). *Practical language testing*. London, England: Hodder Education.
- Fulcher, G. (2012). Scoring performance tests. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 378-392). Oxon, England: Routledge.
- Houston, S. M. (2016). *A longitudinal study of the development of fluency of novice Japanese learners: Analysis using objective measures* [MA thesis, Purdue University]. Open Access Theses and Dissertations.
- Hughes, R. (2011). *Teaching and researching speaking*. Harlow, England: Pearson Education Limited.
- Institute of Phonetics and Speech Processing (n.d.). Ludwig Maximilian University of Munich. Retrieved June 10, 2021 from https://www.phonetik.uni-muenchen.de/forschung/Verbmobil/trllex_e_html/projects/false_start.html
- Jalili, S. A., (2017). Assessing advanced Persian language learners' written production: Developing a detailed rubric. *Journal of Teaching Persian to Speakers of Other Languages (JTPSOL)*, 6(1), 29-56. [In Persian]
- Lam, W. Y. K. (2006). Gauging the effects of ESL oral communication strategy teaching: A multi-method approach. *Electronic Journal of Foreign Language Teaching*, 3(2), 142-157.
- Lennon P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40(3), 387-417. <https://doi.org/10.1111/j.1467-1770.1990.tb00669.x>
- Möhle, D. (1984). A comparison of the second language speech production of different native speakers. In H.W. Dechert, D. Möhle and M. Raupach (Eds.), *Second Language Productions* (pp. 26–49). Tübingen, Germany: Gunter Narr Verlag.
- Pishghadam, R., Hashemi, M. R., & Norouz Kermanshahi, P. (2011). Self-correction among Iranian EFL learners: An investigation into their preferences for corrective feedback. *Journal of Language Teaching and Research*, 2(5), 957-962. <https://doi.org/10.4304/jltr.2.5.957-962>
- R Core Team (2021). *R: A Language and environment for statistical computing*. (Version 4.0) [Computer software]. Retrieved from <https://cran.r-project.org> (R packages retrieved from MRAN snapshot).
- Roudmajani, M., & Ghabool, E. (2020). The validity of speaking scoring rubric in Ferdowsi Persian proficiency test. *Language Related Research*, 11(4), 183-210. [In Persian] <http://dorl.net/dor/20.1001.1.23223081.1399.11.4.22.7>

- Sadeghi, V. (2020). Phonetic transfer in the production of Chinese and Arabic learners of Persian: A case study on Persian dorsal stops. *Journal of Teaching Persian to Speakers of Other Languages (JTPSOL)*, 9(1), 3-24. [In Persian] <https://doi.org/10.30479/jtpsol.2020.10185.1428>
- Sahraee, R. M., & Marsoos, F. (2016). *Standard reference for teaching Persian language*. Tehran, Iran: Allameh Tabataba'i University. [In Persian]
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36(1), 1-14. <https://doi.org/10.1017/S026144480200188X>
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510-532. <https://doi.org/10.1093/applin/amp047>
- Skehan, P., Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics*, 54(2), 97-111.
- Strunk, K. & Mwavita, M. (2022). *Design and analysis in educational research using jamovi: ANOVA designs*. London: England: Routledge.
- Sultana, A. (2009). Peer correction in ESL classroom. *BRAC University Journal*, 6(1), 11-19.
- 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. <https://doi.org/10.1177/1362168817708462>
- 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. 239-273). Amsterdam/Philadelphia: John Benjamins.
- Tavakoli, P. & Wright, C. (2020). *Second language speech fluency: From research to practice*. Cambridge, England: Cambridge University Press.
- Tavakoli, P., Kendon, G., Hunter, A.-M. & Slaght, J. (forthcoming). *TEEP Speaking test fluency rating scales across four levels of proficiency*. Internal Report, International Study and Language Institute, University of Reading.
- Tavakoli, P., Kendon, G., Mazhurnaya, S., & Ziomek, A. (2023). Assessment of fluency in the Test of English for Educational Purposes. *Language Testing*, 40(1), 1-23. <https://doi.org/10.1177/02655322231151384>
- 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. London, England: British Council.
- Tavakoli, P., Nakatsuhara, F., & Hunter, A.-M. (2020). Aspects of oral fluency across assessed levels of proficiency: Do 'fluency profiles' for different proficiency levels exist? *Modern Language Journal*, 104(1), 169-191. <https://doi.org/10.1111/modl.12620>
- The jamovi project (2021). *jamovi*. (Version 2.2) [Computer Software]. Retrieved from <https://www.jamovi.org>
- Ur, P. (2012). *A course in English language teaching*. Cambridge, England: Cambridge University Press.
- Zare, S., Safaie-qalati, M., & Yarahamdzei, N. (2020). Phonological transfer in Chinese Persian language learners: Vocalic space in focus. *Journal of Teaching Persian to Speakers of Other Languages (JTPSOL)*, 9(1), 149-176. [In Persian] <https://doi.org/10.30479/jtpsol.2020.14062.1490>