

## Computer Adaptive Placement Test of Vocabulary and Grammar Components: Investigation of Test Anxiety in Persian/Farsi Non-Native Learners

Vol. 15, No. 4, Tome 82  
pp. 63-96  
September & October  
2024

Zari Saeedi<sup>1\*</sup> , Reza Morad Sahraei<sup>2</sup>  & Fatemeh Kelari<sup>3</sup> 

### Abstract

The research aims to compare anxiety of non-native (Iraqi) Persian language learners in computer adaptive placement test of Persian grammar and vocabulary with its linear computerized test counterpart. The main concern of this study was to scrutinize the probable relationship between linear and adaptive vocabulary and grammar placement tests and testees' anxiety (in case there is any)?" Explanatory mixed-method approach was used in the present study. The procedure for data collection was field research via the application of questionnaire and tests. The designed test was given to 48 Iraqi male and female students with the age range of 25-40 years from Persian language center of Tarbiat Modares University, with different language levels and in two forms of adaptive and linear computerized tests. According to positive upper and lower limits of average variance, the researcher came up to the conclusion that the average of anxiety component in linear computer test is more than the average 3. The findings revealed that test anxiety level of Iraqi students in adaptive test was lower than the linear computer test. These results help to encourage teachers and educational centers to design more computer adaptive tests.

**Keywords:** computer adaptive placement test, computerized test, Persian/Farsi language, assessment, test anxiety

<sup>1</sup> Corresponding Author: Associate Professor and Faculty Member of Allameh Tabataba'i University, Tehran, Iran; Email: [saeediz240@yahoo.com](mailto:saeediz240@yahoo.com), ORCID ID: <https://orcid.org/0000-0003-2165-8891>

<sup>2</sup> Associate Professor and Faculty Member of Allameh Tabataba'i University, Tehran, Iran, ORCID ID: <https://orcid.org/0000-0001-9953-0789>

<sup>3</sup> MA Student of Teaching Persian to Non-Persians, Allameh Tabataba'i University, Tehran, Iran. ORCID ID: <https://orcid.org/0009-0004-6397-2365>

## 1. Introduction

The inefficiency of the language level test cannot be ignored at all. Even if a learner is misplaced and placed in a course that does not suit his needs, it does not diminish the importance of this issue. Inconsistency in the content or at least in the framework of tests will be considered a big weakness in education. Lack of valid placement test means lack of correct identification; and misdiagnosing the learner's level means assessing the needs, determining the wrong goal and curriculum, as well as wasting his and the teacher's energy, and ultimately discouraging the learner. On the one hand, this issue will also lead to increased anxiety in the language learner. On the other hand, according to the world of communication and the ever-increasing progress in technology, more efficient and appropriate methods should be used for test-making or test implementation. In this regard, it should be kept in mind that the quality of the test does not sacrifice its appearance and shape.

### Research Question(s)

The main research question of this study can be formulated as follows:

What is the relationship between linear computerized test and computer adaptive placement test of Farsi vocabulary and grammar on the testee's anxiety level, if there is a relationship?

Following this question, the present hypothesis is proposed:

The level of test anxiety in the computer adaptive placement test is lower than in the equivalent linear computerized test.

## 2. Literature Review

Measurement is the process by which a better understanding of existing knowledge is gained. Quantitative and qualitative features are involved in this process. Measurement can be as simple as the teacher's personal judgment based on his classroom observations of language learners in the

class, or it can be as complex as a standard test lasting several hours (Birjandi & Mosallinejad, 2012). In the teaching-learning environment, assessment is always needed to measure the efficiency or quality and efficiency of the teaching or learning process (Masters, 2014). This important process, which is generally called assessment, is not only done after teaching but can also be done before teaching or during the teaching process. Assessment and testing are two inseparable and complementary pillars of language teaching. The relationship between teaching and testing is so closely interdependent that it is not possible to only address one of them without paying attention to the other (Zandi et al., 2018).

Until recently, test designers and teachers used traditional paper-and-pencil tests to assess and measure the achievement of cognitive learning goals (Weiss & Kingsbury, 1984). The use of computers in the field of language testing and assessment dates back to 1935 when correctional officers used computers to quickly calculate test scores in the United States (Abukhzam, 2013). Computerized tests opened a new window in future tests to measure the language skills of language learners (Fleming & Hiple, 2004).

There are different types of computer-based tests. The two main types of computer-based tests are adaptive computer tests and direct computer tests. Direct (linear) computerized tests refer to the use of a computer for a test that is the same as a paper-and-pencil test format. In this way, this test has the same questions as the paper-and-pencil test, and the items are displayed in the same order as the paper-and-pencil version. Most computer-based tests are designed based on the classical test theory and extensive studies have been conducted on them (Dunn et al., 1972; Elwood, 1969; Kiely et al., 1986).

However, there will undoubtedly always be those for whom the computer is a mystery. Therefore, computers and how to use them in the language test should be introduced and taught for their needs. Anxiety is a reaction caused by internal feelings in response to external threats (Casbarro, 2005). Test

anxiety is an important issue that arises during tests. This multidimensional structure has two main factors: the cognitive dimension and the emotional dimension. The cognitive dimension deals with mental activities around the test situation and leads to worry and unrelated or negative thoughts along with feelings of discomfort. The emotional dimension is related to the psychological component, which includes anxiety, physical reaction, and arousal (Cassady, 2004; Zeidner, 2007).

### **3. Methodology**

To evaluate the correctness of the proposed hypotheses, the integrated research method was used. The method of data collection is in the field and using questionnaires and tests. In this way, firstly, the vocabulary and grammar test questions were designed in computer format for male and female Persian learners aged 25-40 with Iraqi nationality at the Persian Language Center for non-Persian speakers of Tarbiat Modares University, Shahid Beheshti University and Imam Khomeini International University, at different language levels. In the form of a computerized test, the characteristics of the test were determined. Then, 48 male and female foreign students of Iraqi nationality, who had participated in the Farsi course of Tarbiat Modares Language Center and were studying at the university at the time of the research, participated in both tests and passed the test. After the test, all language learners completed the test anxiety questionnaire.

### **4. Results**

To answer the question of the present research, the Donald Powers (1999) questionnaire was used after the tests. Examining the research data showed that the average score of the anxiety component in the adaptive test is lower than the average. On the other hand, the average score of the anxiety component in the computer test is more than average. In this way, the

average scores of anxiety in the computer test are higher than the adaptive test.

Also, the examination of the scores of both tests showed that the mean scores of anxiety in computerized and adaptive test have a significant difference. Therefore, the average score of computerized test anxiety is significantly higher than the average score of adaptive test. As a result, the level of test anxiety in adaptive computer test is lower than the linear computerized test. These findings are in line with Henning's (1991) research. However, it is contrary to the research findings of Hooper and Hannafin (1989) that computer tests cause more anxiety in testees.

These results help to encourage teachers and educational centers to design more computer adaptive tests. It also shows the effect of these tests in reducing the anxiety of language learners compared to paper-and-pencil tests or direct computer tests. Therefore, by using adaptive computer tests, language learners can take the test according to their level at home or in the institution with less anxiety, and the test results will be more reliable.





دوماهنامه بین‌المللی

د ۱۵، ش ۴ (پیاپی ۸۲)، مهر و آبان ۱۴۰۳، صص ۶۳-۹۶

مقاله پژوهشی

<http://dorl.net/dor/20.1001.1.23223081.1401.0.0.142.4>

## آزمون انطباقی رایانه‌ای تعیین سطح مؤلفه‌های واژه و دستور: بررسی میزان اضطراب فارسی‌آموزان

زری سعیدی<sup>۱\*</sup>، رضامراد صحرایی<sup>۲</sup>، فاطمه کلاری<sup>۳</sup>

۱. دانشیار آموزش زبان انگلیسی، دانشگاه علامه طباطبائی، تهران، ایران

۲. دانشیار زبان‌شناسی، دانشگاه علامه طباطبائی، تهران، ایران

۳. دانشجوی کارشناسی ارشد آموزش زبان فارسی به غیر فارسی‌زبانان، دانشگاه علامه طباطبائی، تهران، ایران

تاریخ پذیرش: ۱۴۰۰/۱۲/۲۶

تاریخ دریافت: ۱۴۰۰/۰۸/۱۸

### چکیده

هدف از انجام پژوهش حاضر، مقایسه میزان اضطراب آزمون فارسی‌آموزان خارجی (عراقی) در آزمون تعیین سطح دانش زبان فارسی (دستور و واژه) به دو صورت رایانه‌ای مستقیم و آزمون انطباقی رایانه‌ای همتاست. به این منظور، این سؤال مطرح شد که رابطه بین آزمون‌های تعیین سطح مؤلفه‌های دستور و واژه فارسی انطباقی رایانه‌ای و رایانه‌محور بر میزان اضطراب آزمودنی‌ها، در صورت وجود ارتباط، چیست؟ در این پژوهش از روش تلفیقی استفاده شد. شیوه گردآوری داده به صورت میدانی و با استفاده از پرسش‌نامه و آزمون بود. به این ترتیب، آزمون طراحی شده پس از تعیین ویژگی‌های آزمون از ۴۸ نفر از فارسی‌آموزان دختر و پسر ۲۵ - ۴۰ ساله با ملیت عراقی مرکز آموزش زبان فارسی به غیرفارسی‌زبانان دانشگاه تربیت مدرس، در سطوح زبانی مختلف در قالب آزمون رایانه‌ای مستقیم و انطباقی اجرا شد. با توجه به مثبت بودن حد بالا و حد پایین فاصله اطمینان اختلاف میانگین، درمی‌یابیم که میانگین نمره مؤلفه اضطراب در آزمون رایانه‌ای بیش از حد متوسط ۳ است. در نتیجه میزان اضطراب فارسی‌آموزان عراقی در آزمون انطباقی رایانه‌ای کمتر از آزمون رایانه‌ای مستقیم همتاست.

واژه‌های کلیدی: آزمون انطباقی تعیین سطح، آزمون رایانه‌ای، زبان فارسی، سنجش.

E-mail: saeedi.za@atu.ac.ir

\* نویسنده مسئول مقاله:

## ۱. مقدمه

بنا به تجربه آموزش زبان انگلیسی در آموزشگاه‌ها و مؤسسات مختلف، همواره شاهد زبان‌آموزانی بوده‌ایم که سطح دانش زبانی آن‌ها متناسب با ترم و دوره آموزشی کنونی‌شان نبود. این مشکل حتی در آموزشگاه‌هایی که بد طولایی در امر آموزش زبان انگلیسی داشتند، اتفاق می‌افتاد. علی‌رغم فعالیت‌ها و پژوهش‌های انجام‌شده در حوزه آموزش زبان انگلیسی در اقصی نقاط دنیا و استفاده از سیستم‌های ارتباطی و فناوری مدرن و همچنین روش‌های نوآورانه در آموزش زبان، همچنان خلأهایی در آزمون‌سازی و تعیین سطح دانش زبانی زبان‌آموزان به چشم می‌خورد. حقیقتاً زبان فارسی نیز با مشکلاتی این‌چنینی در آزمون‌سازی و سیستم تعیین سطح دانش زبان‌آموزان مواجه است. ناکارآمدی آزمون تعیین سطح زبان به هیچ عنوان قابل اغماض نیست. حتی اگر یک زبان‌آموز اشتباه تعیین سطح شود و در دوره‌ای که متناسب با نیازهای وی نیست قرار گیرد، از اهمیت این مسئله نمی‌کاهد. عدم تجانس و یکپارچگی در محتوا یا حداقل در چارچوب آزمون‌ها ضعف بزرگی در امر آموزش به حساب خواهد آمد. نبود آزمون تعیین سطح دقیق یعنی عدم شناسایی صحیح؛ و تشخیص نادرست سطح زبان‌آموز به معنای نیازسنجی، تعیین هدف و برنامه درسی اشتباه و همچنین هدررفت انرژی وی و مدرس و نهایتاً دلسردی زبان‌آموز است. این مسئله همچنین به افزایش اضطراب در زبان‌آموز منجر خواهد شد.

از طرفی، رشد سریع اطلاعات و فناوری ارتباطات در آموزش و یادگیری به تغییر الگوی سیستم ارزشیابی از آزمون مداد - کاغذی به آزمون‌های رایانه‌محور منجر شده است. ارزیابی تکنولوژی‌محور فرصت‌هایی برای سنجش شکل پیچیده دانش فراهم می‌کند. ارزیابی و مشارکت از طریق روش‌های سنتی ممکن نیست. از طرفی، سؤالات و آیتم‌های آزمون انطباقی رایانه‌ای با هدف سنجش سطوح مشخصی از توانایی آزمودنی طراحی شده است. این آزمون‌ها قادرند تا نتایج دقیق‌تر و قابل اعتمادتری را نسبت به آزمون‌های سنتی در اختیار معلم قرار دهند.

از آنجا که آزمون‌های رایانه‌ای در ۱۰ - ۱۵ سال اخیر رشد فزاینده‌ای داشته است، تعداد پژوهش‌های انجام‌شده در حوزه آزمون‌سازی، بالاخص مقایسه آزمون‌های مداد - کاغذی با آزمون‌های رایانه‌محور افزایش چشمگیری یافته است. علی‌رغم فعالیت‌ها و پژوهش‌های



انجام شده در حوزه آموزش زبان انگلیسی در اقصی نقاط دنیا و استفاده از سیستم‌های ارتباطی و فناوری مدرن و همچنین روش‌های نوآورانه در آموزش زبان، همچنان خلأهایی در آزمون‌سازی و تعیین سطح دانش زبانی زبان‌آموزان به چشم می‌خورد. حقیقتاً زبان فارسی نیز با مشکلاتی این‌چنینی در آزمون‌سازی و سیستم تعیین سطح دانش زبانی زبان‌آموزان مواجه است.

از طرفی، با توجه به دنیای ارتباطات و پیشرفت روزافزون در فناوری و تکنولوژی، پیوسته باید روش‌های کارآمدتر و مناسب‌تری را برای آزمون‌سازی و یا اجرای آزمون به‌کار برد. در این راستا، این نکته را باید در نظر داشت که کیفیت آزمون فدای ظاهر و شکل آن نشود. آزمون‌های مداد - کاغذی به دلیل ویژگی‌های منحصر به فرد خود از دیرباز تاکنون همواره مورد توجه طراحان آزمون بوده‌اند. با افزایش دسترسی و استفاده از رایانه‌های شخصی در مراکز آموزشی، بازار آزمون‌های رایانه‌محور همچنان رو به توسعه است. از جهت دیگر، آزمون‌های رایانه‌ای، مخصوصاً آزمون انطباقی رایانه‌ای، مزایای زیادی دارد که طراحان سؤال را به استفاده از آن ترغیب می‌کند. در پژوهش حاضر در صدد پاسخ به این پرسش هستیم:

**سؤال پژوهش:** رابطه بین آزمون‌های تعیین سطح مؤلفه‌های دستور و واژه فارسی انطباقی رایانه‌ای و رایانه‌محور بر میزان اضطراب آزمودنی‌ها، در صورت وجود ارتباط، چیست؟

به دنبال این پرسش، فرضیه مطرح است:

**فرضیه پژوهش:** میزان اضطراب آزمون در آزمون انطباقی رایانه‌ای تعیین سطح مؤلفه‌های دستور و واژه فارسی کم‌تر از آزمون رایانه‌محور معادل است.

به منظور ارزیابی صحت فرضیه‌های مطرح شده از روش پژوهش تلفیقی استفاده شد. شیوه گردآوری داده به صورت میدانی و با استفاده از پرسش‌نامه و آزمون است. به این ترتیب، ابتدا سؤالات آزمون واژه و دستور طراحی شده در قالب رایانه برای فارسی‌آموزان دختر و پسر ۲۵ - ۴۰ ساله با ملیت عراقی مرکز آموزش زبان فارسی به غیرفارسی‌زبانان دانشگاه تربیت مدرس، دانشگاه شهید بهشتی و دانشگاه بین‌المللی امام خمینی، در سطوح زبانی مختلف در قالب آزمون رایانه‌ای اجرا و ویژگی‌های آزمون تعیین شد. سپس ۴۸ دانشجوی خارجی دختر و پسر با ملیت عراقی که در دوره فارسی‌آموزی مرکز زبان تربیت مدرس شرکت کرده بودند و

در زمان پژوهش در دانشگاه مشغول تحصیل بودند، در هر دو آزمون شرکت کردند و امتحان دادند. پس از آزمون، همه زبان‌آموزان پرسش‌نامه اضطراب مربوط به آزمون‌ها را تکمیل کردند.

## ۲. پیشینه تحقیق

سنجش فرایندی است که طی آن درک بهتری از دانش کنونی فراگیر به دست می‌آید. ویژگی‌های کمی و کیفی در این فرایند دخیل هستند. سنجش می‌تواند به سادگی قضاوت شخصی مدرس مبنی بر مشاهدات کلاسی وی از زبان‌آموزان در کلاس باشد و یا به پیچیدگی یک آزمون استاندارد چندساعته باشد (بیرجندی و مصلی‌نژاد، ۱۳۸۲). در محیط یاددهی - یادگیری، همیشه به پیمانه‌ای نیاز است تا میزان بازدهی یا کیفیت و کارآمدی آموزش یا فرایند یادگیری را اندازه بگیرد (Masters, 2014). این فرایند مهم را که به‌طور کلی به آن ارزیابی<sup>۱</sup> می‌گویند، تنها بعد از آموزش صورت نمی‌گیرد بلکه می‌تواند پیش از آموزش یا حین فرایند آموزش نیز انجام شود. به‌طور تخصصی‌تر، مفاهیم آزمون، سنجش، و ارزشیابی همچنان در فعالیت‌های آموزشی در جهان غالب هستند. گاهی اوقات این مفاهیم به اشتباه مترادف یکدیگر برداشت می‌شوند (Kizlik, 2012).

به‌طور کلی، ارزیابی یکی از ارکان اساسی و مهم آموزشی است که معلمان و کارمندان آموزش با استفاده از آن سطح مهارت یا دانش زبان‌آموزان را می‌سنجند (Taras, 2005). ارزیابی به معلمان کمک می‌کند تا نقاط قوت و ضعف زبان‌آموزان را دریابند و به آن‌ها انگیزه بدهند. افزون بر این، ارزیابی بازخوردهای کارآمدی درباره فراگیری زبان‌آموزان در اختیار معلمان قرار می‌دهد (Stiggins, 2009). این روش معلمان را قادر می‌سازد تا به ارزشیابی یادگیری بپردازند و سپس از آن اطلاعات در جهت ارتقای دانش زبان‌آموزان استفاده کنند. یکی از اشکال ارزیابی، آزمون‌ها هستند. آزمون روش‌هایی به‌منظور سنجش میزان یادگیری زبان‌آموز در نقطه خاصی از زمان است و غالباً شامل جمع‌آوری اطلاعات به شکل داده‌های عددی است (Richards, 1994). آزمون، ابزار مناسب و دقیق در فرایند اندازه‌گیری است. در واقع آزمون به هر نوع ابزار یا شیوه‌ای اشاره می‌کند که جهت اندازه‌گیری عملکرد یا توانایی به کار می‌رود.

آزمون‌ها اشکال گوناگونی دارند و برای اهداف مختلف به کار می‌روند. به عبارت ساده، آزمون روشی برای اندازه‌گیری توانایی یا دانش شخص در یک زمینه معین است. آزمون در درجه اول یک روش است، این روش نوعی ابزار است که شامل مجموعه‌ای از تکنیک‌ها، راه‌های انجام کار و مواد آزمودنی می‌شود. آزمون‌ها به گونه‌ای انتخاب می‌شوند که در آن آزمودنی یا آزمون‌گیرنده یا هر دوی آن‌ها فعالیت‌های خاصی را انجام می‌دهند (Douglas, 1980, p.262).

در نیم قرن اخیر، یکی از حوزه‌های مهم در زمینه زبان‌آموزی که روز به روز به میزان اثربخشی و اهمیت آن در گستره‌های مرتبط با زبان توجه شده، آزمون‌سازی است (قنسولی، ۱۳۸۹). ارزشیابی و آزمون دو رکن جدایی‌ناپذیر و کامل‌کننده آموزش زبان هستند. ارتباط آموزش و آزمون به قدری وابسته به هم و تنگاتنگ است که تنها پرداختن به یکی از آن‌ها بدون توجه به دیگری امکان‌پذیر نیست. مسلماً یکی از ابعاد مهم و تأثیرگذار آزمون‌ها، قلمرو سیاسی، فرهنگی و شکل‌دهی هویت ملی هر کشور است. آموزش و آزمون زبان فارسی نیز مشمول این قاعده می‌شود. دانشجویان خارجی که به منظور تحصیل به ایران می‌آیند، سطح زبانی یکسانی ندارند. بنابراین برای سنجش سطح زبانی دانشجویان خارجی نیازمند ملاک و معیار زبانی معتبری هستیم تا بتوانیم براساس آن اقدام به راهنمایی تحصیلی کنیم. علاوه بر این، باید به گونه‌ای عمل کرد که هزینه صرف‌شده برای داوطلبان یادگیری زبان فارسی، به نحو مطلوبی به نتیجه برسد و بتواند ضامن یادگیری زبان فارسی باشد (زندى و همکاران، ۱۳۹۸).

تا چندی پیش، طراحان آزمون و معلمان از آزمون‌های مداد - کاغذی و سنتی به منظور ارزیابی و سنجش میزان دستیابی به اهداف یادگیری شناختی استفاده می‌کردند (Weiss & Kingsbury, 1984). استفاده از رایانه در حوزه آزمون‌سازی زبان و ارزیابی به سال ۱۹۳۵ زمانی که مصححان از رایانه برای محاسبه سریع نمرات آزمون در آمریکا استفاده می‌کردند، برمی‌گردد (Abukhzam, 2013). آزمون‌های رایانه‌ای پنجره جدیدی در آزمون‌سازی‌های آتی برای سنجش مهارت زبانی زبان‌آموزان گشود (Fleming & Hipple, 2004). حوزه<sup>۲</sup> (۱۹۹۴، ص. ۳۸) سنجش زبان به کمک رایانه را این‌گونه تعریف می‌کند: «روش یکپارچه‌ای که در آن عملکرد زبانی با کمک رایانه آشکار و ارزیابی می‌شود». اولین آزمونی که در آن از رایانه استفاده شد، آزمون نگارش بود که در آن آزمودنی‌ها از رایانه به‌عنوان واژه‌پرداز و لغت‌نامه

استفاده می‌کردند (Larson & Madsen, 1985).

آزمون‌های رایانه‌محور انواع مختلفی دارد. دو نوع اصلی آزمون رایانه‌محور، آزمون انطباقی رایانه‌ای و آزمون رایانه‌ای مستقیم هستند. آزمون‌های رایانه‌ای مستقیم (خطی) به استفاده از رایانه برای آزمونی دقیقاً مشابه با فرمت آزمون مداد - کاغذی اطلاق می‌شود. به این ترتیب، این آزمون سؤالاتی مشابه با آزمون مداد - کاغذی دارد و آیتم‌ها دقیقاً با همان ترتیب نسخه آزمون مداد - کاغذی نمایش داده می‌شوند. بیشتر آزمون‌های رایانه‌محور براساس نظریه کلاسیک آزمون طراحی شده‌اند و مطالعات گسترده‌ای در مورد آن‌ها انجام شده است (Dunn et al., 1972; Elwood, 1969; Kiely et al., 1986).

اولین آزمون انطباقی طراحی شده، آزمون هوش بینه - سیمون<sup>۲</sup> است که در سال ۱۹۰۵ اجرا شد (Van der Linden, 2010). آزمون انطباقی رایانه‌ای شکل پیچیده‌ای از آزمون است که در آن رایانه با توجه به پاسخ آزمودنی به سؤالات پیشین به‌طور پویا آیتم‌هایی را برای سنجش ارائه می‌کند. به عبارتی، در آزمون انطباقی، ابتدا یک سؤال با درجه دشواری متوسط ارائه می‌شود. در صورتی که پاسخ آزمودنی درست باشد، سؤال بعدی دشوارتر خواهد بود و اگر آزمودنی پاسخ اشتباه بدهد، سؤال آسان‌تری نشان داده می‌شود (Carlson, 1994). در صورت پاسخ درست به سؤال، یک سؤال با همان درجه دشواری به‌عنوان سؤال نشانگر جایگزین می‌شود. سؤال نشانگر، سؤالی است که سطح توانایی فراگیر را مشخص می‌کند (Kline, 2005). بیشتر سیستم‌های ارائه‌شده برای آزمون‌های انطباقی براساس نظریه پرسش - پاسخ طراحی شده‌اند و اساساً اکثر آزمون‌های انطباقی برای سنجش پیشرفت و توانایی آزمودنی‌ها طراحی شده‌اند (Ju, 1993). افزون بر این، آزمون انطباقی تعداد سؤالات زیادی را ذخیره می‌کند، که به آن بانک سؤال می‌گویند. سؤالات از این بانک با روش پرسش - پاسخ انتخاب می‌شوند که منجر به افزایش روایی و پایایی آزمون می‌شود (Kirschner et al., 2016).

سنجش زبان به کمک آزمون انطباقی رایانه‌ای ماهیت حوزه ارزیابی زبان را با مزایای بالقوه‌ای که دارد، دستخوش تغییرات قرار داده است. برای نمونه، آزمودنی در آزمون‌های انطباقی هیچ‌گونه محدودیت زمانی ندارد. این ویژگی در کاهش اضطراب آزمودنی در زمان

امتحان بسیار مؤثر است (Henning, 1991).

با این همه، بدون شک همیشه کسانی هستند که رایانه برای آن‌ها مانند رمز و رازی سری است. بنابراین باید برای نیازهای آن‌ها رایانه و طرز استفاده از آن در آزمون زبان معرفی و آموزش داده شود. اضطراب واکنشی است که با احساسات درونی در پاسخ به خطرات خارجی ایجاد می‌شود (Casbarro, 2005). اضطراب (یا در برخی مواقع عصبانیت) روی افراد مختلف تأثیرات متفاوتی می‌گذارد. در برخی، اضطراب به گیجی و عدم اطمینان حین آزمون منجر می‌شود. در حالی که در برخی دیگر، ممکن است حتی به تمرکز و دقت بیشتر و داشتن عملکردی بهتر منجر شود. در هر دو حالت، ما همچنان نمی‌دانیم که استفاده از رایانه در آزمون‌سازی و اضطراب ناشی از ناآشنایی با رایانه چه تأثیری بر عملکرد آزمودنی خواهد گذاشت. این مسئله‌ای است که در استفاده از آزمون‌سازی الکترونیکی رفته رفته کم‌رنگ‌تر خواهد شد (Tylor et al., 1998).

اضطراب آزمون مسئله مهمی است که در موقعیت سنجش و آزمون در آزمودنی ایجاد می‌شود. این سازه چند بعدی دو عامل اصلی دارد: بُعد شناختی و بُعد عاطفی. بعد شناختی به فعالیت‌های ذهنی حول موقعیت آزمون می‌پردازد و به نگرانی و افکار نامرتب یا منفی همراه با احساس ناراحتی منجر می‌شود. بعد احساسی به مولفه روان‌شناختی مربوط می‌شود که شامل تشویش، واکنش بدنی و برانگیختگی است (Cassady, 2004; Zeidner, 2007).

وارد، هوپر و هانافین<sup>۹</sup> (1989) در پژوهشی نشان دادند اگرچه میزان اضطراب آزمودنی در آزمون‌های رایانه‌ای بیشتر است، عملکرد آزمودنی در هر دو نوع آزمون رایانه‌ای و مداد - کاغذی کاملاً مشابه است.

همچنین اولئا و همکاران<sup>۱۰</sup> (2000) در پژوهش دیگر، به بررسی تأثیر مرور سؤالات بر عملکرد آزمودنی در هر دو نوع آزمون انطباقی و تعداد ثابت سؤالات و رابطه آن با اضطراب آزمون پرداخت. در این پژوهش، آزمودنی‌ها به‌طور تصادفی در یکی از چهار گروه (تعداد سؤال ثابت با مرور، تعداد سؤال ثابت بدون مرور، انطباقی با مرور، انطباقی بدون مرور) قرار گرفتند و میزان اضطراب آن‌ها در آزمون اندازه‌گیری شد. نتایج حاکی از آن است که سطح اضطراب با مرور سؤالات رابطه منفی دارد. به عبارتی، اضطراب آزمودنی در آزمون‌های با

مرور بسیار کم‌تر از آزمون‌های بدون مرور است.

علاوه بر این، مطالعات انجام‌شده در مورد جنسیت و رابطه آن با میزان اضطراب آزمودنی در آزمون انطباقی نشان می‌دهد هیچ رابطه معناداری بین جنسیت و میزان اضطراب در این نوع آزمون‌ها وجود ندارد (Fritts & Marszaleck, 2010). این نتیجه غافلگیرکننده است، چرا که پژوهش‌ها نشان می‌دهد جنسیت در میزان اضطراب آزمودنی در آزمون‌های سنتی تأثیر زیادی دارد (Hembree, 1988).

همانند بسیاری از کشورهای در حال توسعه دیگر، ایرانیان نیز باید خود را با تکنولوژی جدیدی که بر پایه برنامه‌های رایانه‌ای است، سازگار کنند. تمام جوانب زندگی مدرن بشر، از جمله آموزش، درگیر رایانه است. رایانه‌ها معمولاً در کلاس‌ها، سمینارها و کنفرانس‌های ایران به کار می‌روند. با این حال، به دلیل دسترسی محدود و هزینه بالا در فرایند ارزشیابی جایگاه خاصی ندارند (رضایی و گلشن، ۲۰۱۵).

در باب طراحی آزمون‌های انطباقی رایانه‌ای زبان فارسی مخصوصاً آزمون تعیین سطح غیرفارسی‌زبانان پژوهش‌چندانی انجام نشده است. یکی از پژوهش‌های انجام‌شده در این حوزه، رساله‌ای با عنوان «طراحی آزمون‌های انطباقی رایانه‌ای زبان فارسی برای سنجش سطح زبانی غیرفارسی‌زبانان» در سال ۱۳۹۳ توسط فاطمه محمودی برای دریافت درجه کارشناسی ارشد رشته آموزش زبان فارسی به غیر فارسی‌زبانان است. محمودی در پژوهش خود به بررسی میزان کارآمدی آزمون انطباقی رایانه‌ای، نحوه عملکرد آزمودنی در این آزمون زبان‌آموزان نسبت به ویژگی‌های آزمون انطباقی پرداخته است. به این منظور از دو آزمون خواندن، مداد - کاغذی و انطباقی رایانه‌ای، استفاده کرد. در رابطه با کارآمدی آزمون‌های انطباقی وی نتیجه گرفت میزان کارآمدی آزمون انطباقی طراحی‌شده در تعیین سطح زبان‌آموزان غیرفارسی‌زبان، مناسب و رضایت‌بخش می‌باشد و برخلاف آزمون مداد - کاغذی، در پاسخ به سؤالات آزمون انطباقی، امکان حدس و گمان برای آزمودنی عملاً وجود ندارد. در بررسی نحوه عملکرد آزمودنی محمودی نتیجه گرفت «برآورد نمره واقعی آزمودنی‌ها از طریق آزمون انطباقی، نسبت به آزمون مداد - کاغذی دقیق‌تر است» (محمودی، ۱۳۹۳، ص. ۷۸). نتایج به‌دست‌آمده در رابطه با نگرش زبان‌آموزان نشان می‌دهد که «نگرش زبان‌آموزان پس از شرکت در آزمون‌های انطباقی مثبت است و نسبت به این آزمون رضایت

خاطر بیشتری دارند» (محمودی، ۱۳۹۳، ص. ۷۹).

با توجه به محدود پژوهش‌های انجام‌شده در حوزه آزمون‌سازی و بالخص طراحی آزمون انطباقی زبان فارسی، درمی‌یابیم اکثر پژوهش‌ها مرتبط با آزمون بسندگی و مهارتی و از نوع مداد - کاغذی است و توجه کمی به آزمون‌های انطباقی رایانه‌ای به‌ویژه آزمون تعیین سطح شده است. علی‌هذا اهمیت و تأکید بر انجام پژوهش در این حوزه برجسته‌تر و جدی‌تر می‌شود.

### ۳. چارچوب نظری

در طراحی آزمون انطباقی رایانه‌ای روش‌های مختلفی وجود دارد. به‌منظور طراحی آزمون حاضر از روش لیلی (2007) استفاده شده است. مراحل ساخت آزمون در این روش شامل انتخاب سؤال، مدل محاسبه، قوانین توقف، اعتبارسنجی آزمون، بانک سؤال، تنظیم سؤال، انتخاب سؤال، اجرای سؤال، نمردهی و ویژگی‌های آن و گزارش نتیجه آزمون می‌شود (Lilley, 2007).

به طور کلی، انتخاب سؤال با استفاده از نظریه پرسش - پاسخ انجام می‌شود. نظریه پرسش - پاسخ از پاسخ آزمودنی (درست یا نادرست) به سؤال استفاده می‌کند تا سطح توانایی آزمودنی را تخمین بزند و بر اساس پاسخ آزمودنی سؤال را انتخاب کند. انتخاب سؤال اول مورد خاصی از فرایند انتخاب سؤال است. به‌طور کلی، در آغاز آزمون، اطلاعات کمی در مورد آزمودنی وجود دارد (در برخی موارد، هیچ اطلاعاتی وجود ندارد). در صورتی که اطلاعات پیش‌زمینه از آزمودنی وجود داشته باشد، سؤال اول بر اساس آن اطلاعات انتخاب می‌شود (Lilley, 2007).

مرحله بعدی، تعیین مدل محاسبه‌ای است که به آن‌ها مدل‌های راش می‌گویند. مدل‌های راش در واقع روی‌آوردی ریاضی برای آزمون این فرضیه است که اندازه‌های مربوط به معنا و واحد یک سازه را می‌توان از ابزاری که برای آن خصیصه تهیه شده است به‌دست آورد. وقتی داده‌ها با این مدل‌ها برازش پیدا می‌کند، به معنای آن است که ابزار اندازه‌گیری و اندازه‌ها در یک واحد فاصله‌ای مشترک مقیاس‌بندی شده‌اند و می‌توانند در انواع یا شکل‌های مختلف آن ابزار و نیز در بین نمونه‌های مختلف یک جامعه ثابت باقی بمانند (Wright, & Stone, 1979). این مدل‌ها

شامل مدل محاسبه‌ای یک پارامتری<sup>۷</sup>، دو پارامتری<sup>۸</sup> و سه پارامتری<sup>۹</sup> هستند. مدل محاسبه‌ای یک پارامتری (1PL)، ساده‌ترین و برجسته‌ترین مدل نظریه‌ی پرسش - پاسخ است. فرض اصلی این مدل این است که امکان مشاهده‌ی مستقیم ویژگی‌های مورد نظر وجود ندارد. به همین سبب، این ویژگی‌ها را ویژگی‌های نهفته (پنهان)<sup>۱۰</sup> می‌نامند. با این حال، با استفاده از پاسخ‌های فرد به سؤالات، مخصوصاً در قالب آزمون، می‌توان ویژگی‌های نهفته هر فرد را نتیجه‌گیری کرد. در مدل دو فاکتوری راش، امتیاز پاسخ درست برابر با ۱ و امتیاز پاسخ نادرست برابر با صفر است. این مدل که در آن پاسخ‌ها به دو طبقه (بلی - خیر، درست - نادرست) تقسیم می‌شوند، شناخته‌شده‌ترین و رایج‌ترین مدل راش و دارای تابع ساده‌ی منطقی است. مدل محاسبه‌ای دو پارامتری (2PL) از توانایی آزمودنی و دو پارامتر به‌منظور پیش‌بینی احتمال پاسخ استفاده می‌کند. مدل محاسبه‌ای سه پارامتری (3PL)، پارامتر سوم را به مدل دو پارامتری اضافه می‌کند. این پارامتر سوم، حدس زدن است. همانطور که از نام آن پیداست، این پارامتر احتمال پاسخ درست آزمودنی بدون دانش لازم به سؤال از طریق حدس پاسخ را نشان می‌دهد (Lilley, 2007).

لحظه‌ای که آزمون به پایان می‌رسد، لحظه‌ی بسیار مهمی در آزمون انطباقی است. اگر آزمون خیلی کوتاه باشد، تعیین سطح توانایی دقیق نخواهد بود و در صورتی که آزمون خیلی طولانی باشد، ممکن است آزمودنی خسته شود و اشتباهاتی از روی بی‌دقتی داشته باشد و در نتیجه، پایین‌تر از سطح توانایی خود عمل کند. افزون بر این، منابع و زمان ازدست می‌رود. دو راه برای اتمام آزمون انطباقی وجود دارد؛ براساس نحوه‌ی اتمام آزمون دو نوع آزمون با طول ثابت<sup>۱۱</sup> و طول متغیر<sup>۱۲</sup> وجود دارد. آزمون با طول ثابت زمانی به پایان می‌رسد که آزمودنی حداکثر تعداد سؤالات را پاسخ داده است و یا مقدار زمان مشخصی گذشته باشد. در مقابل، آزمون طول متغیر از مدل نظریه‌ی پرسش - پاسخ برای اتمام آزمون استفاده می‌کند (Lilley, 2007).

خلاصه‌ی آمار سؤالات و آزمودنی در آزمون انطباقی به‌طور چشمگیری با آزمون‌های مداد - کاغذی متفاوت است. اولاً، هر سؤال تنها توسط تعداد کمی از آزمودنی‌ها پاسخ داده می‌شود. بنابراین، تعداد پاسخ‌های از دست‌رفته ۹۰ درصد یا حتی بیشتر است. این مسئله باعث می‌شود



تحلیل کلاسیک سؤالات ناکارآمد باشد. دوماً، تمام سؤالات به گونه‌ای اجرا می‌شوند که میزان موفقیت در تمام سؤالات حدود ۵۰ درصد (یا ۶۰ درصد، ۷۰ درصد، و غیره) باشد. لذا بررسی مقدار  $p$  ناکارآمد است. سوماً، تقریباً تمام سؤالاتی که برای آزمودنی ارائه شده است، با هدفی است. این بدان معناست که تقریباً هیچ شانسی برای یافتن الگوهایی حدس زدن وجود ندارد و حتی شناخت تفاوت در ضریب تمایز سؤالات نیز دشوار می‌شود. با این حال، تحلیل سؤال و اعتبارسنجی آزمون همچنان می‌تواند انجام شود. علی‌رغم میزان بالای داده‌های از دست‌رفته، سؤالات و آزمودنی‌ها شبکه‌ای به هم مرتبط می‌سازند. زمانی که سؤالات و افراد براساس توانایی و دشواری مرتب می‌شوند، مقیاس‌نگاری از پاسخ‌ها در قالب ماتریس توزیعی بلوک - مورب نمایش داده می‌شود (Guttman, 1994). این مقیاس‌نگار پایه‌ای برای تخمین میزان دشواری سؤالات و توانایی‌های آزمودنی در موقعیت راش فراهم می‌کند.

بانک سؤالات در مرکز سیستم‌های اجرایی آزمون انطباقی قرار دارد. این بانک شامل مجموعه‌ای از سؤالاتی است که تنها بعضی از آن سؤالات به هر آزمودنی در طول آزمون نشان داده می‌شود. ویژگی‌های مهم بانک سؤال شامل اندازه، ترکیب، ساختار، و ماهیت مدل پرسش - پاسخ است که بر اساس آن سؤالات تنظیم می‌شوند. طراحی آزمون شامل تعیین بانک سؤالات است. انواع سؤالات موجود در بانک سؤال، ترکیب آن را مشخص می‌کند. فرمت، محتوا، و ویژگی‌های عملکردی سؤالات همگی به ترکیب بانک سؤال مربوط می‌شود. روش انتخاب سؤال به صورت دوره‌ای مشابه لیست خرید در دست است. بیشترین کارآمدی آزمون انطباقی زمانی است که سؤالات آن دائماً در انبار موجود باشند. ساختار بانک سؤال به اینکه آیا و چگونه هر سؤال به سؤال دیگر مرتبط است، مربوط می‌شود. سؤالات می‌توانند مانند یک شیء منفرد و جدا در بانک سؤال بمانند، انتخاب و ارائه شوند و یا اینکه به دسته‌های از پیش تعیین‌شده‌ای تقسیم شوند که به‌عنوان یک واحد انتخاب و ارائه می‌شوند. نهایتاً، تمام سؤالات بانک باید تحت مدل پرسش - پاسخ تنظیم شوند. آزمون‌های انطباقی در پشت صحنه، به اجرای مدل نظریه پرسش - پاسخ تکیه می‌کنند تا دو نقش کلیدی را ایفا کنند. اولین نقش ایجاد نمراتی است که بین آزمودنی‌هایی که آزمون‌های مختلفی داده‌اند، قابل قیاس است. دومین نقش شناسایی مناسب‌ترین سؤالات برای آن دسته از آزمودنی‌هایی است که سطح عملکرد آن‌ها زودتر در آزمون مشخص شده است (Lilley, 2007).

جنبه کلیدی دیگر آزمون انطباقی رایانه‌ای درجه‌بندی یا تنظیم بانک سؤالات است. مدل‌های مذکور نظریه پرسش - پاسخ بین یک تا سه پارامتر دارند؛ دشواری، تمایز، و حدس. پیشنهاد می‌شود بانک سؤالات تا جایی که امکان دارد بزرگ باشد و سؤالات با سطوح دشواری متفاوتی در بانک سؤالات وجود داشته باشد (Lilley, 2007).

در ارتباط با مرحله انتخاب سؤال می‌توان گفت که اکثر سیستم‌های آزمون انطباقی رایانه‌ای سؤالات را به‌منظور بهترین سنجش و اندازه‌گیری بهینه سطح دانش آزمودنی انتخاب می‌کنند، به همین دلیل محدودیت‌ها و قوانینی برای انتخاب این سؤالات وجود دارد. ماهیت این قوانین و تعاریف متعدد از «بهترین» است که به تفاوت‌هایی در سیستم‌های آزمون‌سازی انطباقی منجر می‌شود. این تفاوت‌ها می‌توانند در سه مقوله اصلی قرار گیرند. اولین مقوله این است که آیا سؤالات به صورت منفرد و یکی یکی از بانک سؤالات انتخاب می‌شوند و یا در مجموعه‌هایی از پیش تعیین‌شده قرار دارند. در حالت اول، چرخه انتخاب - اجرا - نمره‌دهی تا زمانی که امتحان ادامه دارد، برای هر سؤال، تکرار می‌شود. در حالت دوم، این چرخه به تعداد کم‌تری تکرار می‌شود و در هر چرخه چند سؤال انتخاب می‌شوند. یک سؤال چند مرحله‌ای<sup>۱۳</sup> با یک سؤال نسبتاً سخت آغاز می‌شود. بر اساس عملکرد آزمودنی در تمام این چرخه، یکی از دسته آزمون‌های مرحله دوم که سطوح دشواری آنان کاملاً متفاوت است، برای آزمودنی‌ها انتخاب می‌شود. پس از آن مراحل بعدی نیز دنبال می‌شوند. با این حال، کارآمدی این نوع آزمون از مرحله دوم یا سوم به بعد خیلی کم می‌شود. خواه سؤالات به صورت منفرد انتخاب شوند خواه به صورت دسته‌ای، سؤالات برای کمک به سنجش یک یا چند هدف در آزمون انتخاب می‌شوند. رایج‌ترین نوع این انتخاب، بیشینه دقت است که با آن هر آزمودنی سنجیده می‌شود (Lilley, 2007).

اجرای سؤالات نیز بخش مهمی از مراحل تهیه آزمون است. ویژگی‌های روان‌سنجی نمرات آزمون می‌تواند بر کیفیت و ویژگی‌های واسطه یا نرم‌افزاری که سؤالات را ارائه می‌دهد و پاسخ‌ها را جمع‌آوری می‌کند، تأثیر بگذارد. در حالت ایدئال، این واسطه و نرم‌افزار باید تأثیر خنثی بر آزمودنی داشته باشد. این ابزار نباید به افزایش یا کاهش میزان عملکرد با توجه به سازه‌های قابل اندازه‌گیری منجر شود. در این حالت، طراحان نرم‌افزار باید خط مشخصی بین

ابزار واسطه که استفاده از آن برای آزمودنی‌ها آسان است اما در کارایی نقص دارند، و ابزارهایی که کارایی‌های بسیاری دارند، اما برای آزمودنی گیج‌کننده و دشوار هستند، ترسیم کنند. خطا در هر یک از این موارد به اشکال در روند سنجش و اندازه‌گیری منجر می‌شود. سؤالات آزمون باید تا جای ممکن سریع، راحت و واضح نمایش داده شوند. پیشرفت در حوزه فناوری رایانه به این تلاش‌ها کمک می‌کند. رشد چشمگیر آزمون‌های اینترنتی بازتاب راحتی استفاده از آن است، چراکه سؤالات آزمون به راحتی در قالب HTML یا کد مناسب ارائه می‌شوند. درحقیقت، مشکل صفحه‌پردازان طراحی ساده بدون حواس‌پرتی‌های غیرضروری است. طراحی کلید «کمک»<sup>۴</sup>، لینک‌ها و تصاویر متحرک، با هدف کمک به عملکرد بهینه آزمودنی، بسیار آسان است. با این همه، این ویژگی‌ها ممکن است به حواس‌پرتی، خستگی، یا هدر رفتن زمان منجر شود (Lilley, 2007).

رایج‌ترین متد برای نمره‌دهی آزمون‌های انطباقی از طریق تخمین‌های نظریه پرسش - پاسخ است. این نمرات این مزیت را دارند که در آن نمره آزمودنی‌ها در آزمون‌های خیلی متفاوت قابل مقایسه است. با این حال، معمولاً در مقیاسی که برای اهداف گزارشی نامناسب است، ارائه می‌شود (تخمین مهارتی معمولاً در گستره‌ای بین ۵ - تا +۵، با میانگین در حدود صفر است). به این ترتیب، اکثر آزمون‌های انطباقی بین نمرات موقتی که در حین آزمون به‌دست می‌آید و سؤال بعدی را انتخاب می‌کند، و نمرات پایانی که در آخر گزارش می‌شوند، تمایز قائل می‌شوند (Lilley, 2007).

یکی از مشهودترین تفاوت‌های آزمون انطباقی و مداد - کاغذی در بازخورد و گزارش است. گزارش آزمون انطباقی بدون هیچ ابهامی می‌تواند مشخص کند آزمودنی بر چه حوزه‌هایی تسلط دارد، چه مواردی را باید یاد بگیرد، و مجموعه بعدی سؤالاتی که باید نمایش داده شود، چیست. این ویژگی هم به معلم و هم به زبان‌آموز کمک می‌کند تا روی تلاش‌های کارآمدتر تمرکز کنند. آزمون انطباقی خیلی سریع نشان می‌دهد کدام زبان‌آموزان به توجه بیشتری نیاز دارند، و روی چه مواردی باید تمرکز بیشتری کرد (Lilley, 2007).

#### ۴. روش تحقیق

بخش طراحی آزمون تعیین سطح مؤلفه دستور و واژه فارسی انطباقی رایانه‌ای در پژوهش حاضر از نوع تولیدی/ کاربردی بوده و تحلیل‌های صورت‌گرفته در رابطه با آزمون و نظرات شرکت‌کنندگان به صورت کمی صورت گرفته است. به‌طور کلی، طرح پژوهش حاضر از نوع طرح پژوهش تلفیقی توضیحی<sup>۱۰</sup> است و شامل دو مرحله می‌شود.

در مرحله اول به مطالعه دقیق سؤالات موجود (سؤالات آزمون تعیین سطح مرکز آموزش زبان فارسی به غیرفارسی‌زبانان علامه طباطبائی و دانشگاه تربیت مدرس)، مطالعه استاندارد مرجع آموزش زبان فارسی به غیرفارسی‌زبانان، طراحی سؤالات اولیه آزمون و اصلاح صورت سؤالات با توجه به نظرات کارشناسان آنفا و فارسی‌آموزان عراقی پرداختیم.

در مرحله دوم این پژوهش روش‌های پژوهشی کمی استفاده شد. ابتدا با کمک ۳۳۲ فارسی‌آموز عراقی شامل ۱۱۲ فارسی‌آموز عراقی سطح مقدماتی، ۱۱۰ فارسی‌آموز عراقی سطح میانی و ۱۱۰ فارسی‌آموز عراقی سطح پیشرفته به تعیین آلفای کرونباخ آزمون و سپس به سنجش ۴۸ فارسی‌آموز عراقی با استفاده از دو آزمون رایانه‌ای مستقیم و انطباقی رایانه‌ای پرداختیم. در نهایت از پرسش‌نامه اضطراب آزمون به منظور سنجش میزان اضطراب فارسی‌آموزان عراقی به هر دو آزمون استفاده شد. هنگام استفاده از پرسش‌نامه به منظور افزایش روایی و پایایی نتایج پژوهش، از روش سه‌گانه<sup>۱۱</sup> برای سنجش میزان اضطراب زبان‌آموزان استفاده شد. به این ترتیب، ابتدا پیش از اجرای آزمون رایانه‌ای مستقیم و انطباقی رایانه‌ای، یک بار پرسش‌نامه اضطراب آزمون توزیع شد تا زبان‌آموزانی که به‌طور ذاتی اضطراب زیادی حین آزمون دارند شناسایی شوند و میزان اضطراب آن‌ها در نتیجه آزمون خلی ایجاد نکند. در مرحله بعد، پس از اجرای آزمون رایانه‌ای مستقیم و سپس آزمون انطباقی رایانه‌ای مجدد پرسش‌نامه اضطراب آزمون به زبان‌آموزان داده شد.

با نهایی شدن سؤالات آزمون، آزمون رایانه‌ای مستقیم و انطباقی رایانه‌ای مربوطه نیز طراحی شد. به‌دلیل همه‌گیری کووید-۱۹ امکان برگزاری آزمون به‌صورت مستقیم نبود. به همین دلیل هر دو آزمون به‌صورت آنلاین برگزار شد. با این حال، برای جلوگیری از تقلب و هر گونه مشکل احتمالی، از پلتفرم بیگ بلو باتن برای برگزاری آزمون استفاده شد. در این مرحله،

وب کم همه زبان آموزان روشن بود و چهره آنها قابل مشاهده بود. هم‌زمان با حضور در پلتفرم، آزمون‌های مربوطه نیز گرفته شد. پس از هر آزمون لینک پرسش‌نامه‌های اضطراب مربوط به هر آزمون ارائه شد و از آنها خواسته شد تا همان لحظه به پرسش‌نامه پاسخ دهند. گفتنی است به دلیل محدودیت‌های زمانی و ظرفیت محدود پلتفرم، و به منظور اجرای دقیق و صحیح آزمون و جلوگیری از هر گونه مشکل و تقلب، ۴۸ فارسی‌آموز عراقی در دسته‌های شش نفره امتحان دادند. به عبارتی، در هر نوبت تنها از شش فارسی‌آموز عراقی همه آزمون‌ها و پرسش‌نامه گرفته می‌شد.

#### ۴-۱. جامعه و نمونه پژوهش

به منظور طراحی آزمون انطباقی رایانه‌ای مورد نظر در بخش‌های مختلف طراحی آزمون، از چند گروه فارسی‌آموز عراقی زن و مرد، با تعداد متفاوت بهره گرفتیم که در این بخش به طور مفصل آنها را شرح می‌دهیم:

الف) بخش اول نمونه آماری، سه گروه ۱۰ نفره از فارسی‌آموزان عراقی زن و مرد سطح مقدماتی، میانی و پیشرفته از مرکز آموزش زبان فارسی دانشگاه تربیت مدرس جهت تعیین میزان شفافیت و وضوح سؤالات هستند. گفتنی است هر سه گروه در سطح خود نه بسیار ضعیف و نه بسیار قوی بودند. سعی شد تنها از زبان‌آموزان با دانش حد وسط برای تعیین وضوح و رفع ابهامات سؤالات استفاده شود.

ب) بخش دوم نمونه آماری به طور کلی از ۳۳۲ فارسی‌آموز عراقی زن و مرد شامل ۱۱۲ نفر فارسی‌آموز عراقی زن و مرد سطح مقدماتی، ۱۱۰ نفر فارسی‌آموز عراقی سطح میانی و ۱۱۰ نفر فارسی‌آموز عراقی سطح پیشرفته از دانشگاه‌های تربیت مدرس، شهید بهشتی و دانشگاه بین‌المللی امام خمینی به منظور تعیین ضریب تمیز، ضریب دشواری و آلفای کرونباخ آزمون می‌شوند.

ج) و در نهایت بخش سوم نمونه آماری شامل ۴۸ نفر از فارسی‌آموزان زن و مرد دانشگاه تربیت مدرس با ملیت عراقی که در دوره فارسی‌آموزی مرکز زبان تربیت مدرس شرکت کرده‌اند و هم‌اکنون در دانشگاه مشغول تحصیل هستند، است. این گروه در آزمون‌های نهایی شرکت کرده و در آخر به پرسش‌نامه مربوطه پاسخ دادند.

#### ۴-۲. ابزارها و بیکره پژوهش

۱. آزمون چهارگزینه‌ای سنجش مؤلفه‌های واژه و دستور با ۳۹۳ سؤال که پس از رفع ابهامات، اعمال نظر کارشناسان آرفا، تعیین ضریب تمییز و دشواری آزمون، غربال شد و تعداد آن به ۳۰۴ سؤال کاهش یافت. سؤالات نهایی در پیوست الف و ب و پ قابل مشاهده است.
۲. آزمون رایانه‌ای مستقیم با ۳۰۴ سؤال که در نرم‌افزار آی اسپرینگ طراحی شد. هدف از این آزمون تعیین سطح دانش زبانی آزمودنی در سه سطح مقدماتی، میانی و پیشرفته است. در این آزمون ۳۰۴ سؤال به‌طور تصادفی به فارسی‌آموزان نشان داده می‌شود. با این حال، همه فارسی‌آموزان به ۳۰۴ سؤال واحد و کاملاً یکسان پاسخ می‌دهند. برای جلوگیری از تقلب، ترتیب سؤالات به‌صورت تصادفی قرار داده شده است.
۳. نسخه انطباقی رایانه‌ای آزمون با بانک سؤال ۳۰۴ سؤال که در ابتدا یک سؤال با سطح دشواری متوسط نشان می‌دهد. سپس با توجه به الگوریتم‌های تعریف‌شده و درستی یا نادرستی پاسخ آزمودنی، سطح وی را در سه سطح مقدماتی، میانی و پیشرفته تعیین می‌کند. گفتنی است به‌دلیل همه‌گیری ویروس کووید - ۱۹ این آزمون با قابلیت اجرای آنلاین طراحی شد. نسخه آنلاین این آزمون در نرم‌افزار لارا اول به کمک پلتفرم پی اچ پی طراحی شد. این سنجش از یافته‌ها و نتایج مدل‌های نظریه پرسش - پاسخ در تحلیل سؤال‌های آزمون‌ها متأثر است.
۴. برای سنجش میزان اضطراب آزمودنی‌ها از پرسش‌نامه اضطراب پاورز<sup>۱۷</sup> (۱۹۹۹) شامل ۲۰ سؤال با ۵ مقیاس خیلی زیاد، زیاد، متوسط، کم، خیلی کم استفاده شد. از آنجا که به‌طور معمول میزان اضطراب بعضی آزمودنی‌ها حین آزمون بسیار زیاد است، به‌منظور جلوگیری از تأثیر این اضطراب بر نتایج آزمون، پرسش‌نامه اضطراب یک بار پیش از شروع آزمون، یک بار پس از آزمون رایانه‌ای و یک بار پس از آزمون انطباقی رایانه‌ای توزیع شد. به این ترتیب، با استفاده از پرسش‌نامه اضطراب پیش از آزمون توانستیم زبان‌آموزانی که به‌طور کلی استرس و اضطراب زیادی حین آزمون دارند شناسایی کنیم. در ادامه، پس از آزمون رایانه‌ای مستقیم و پس از آزمون انطباقی رایانه‌ای پرسش‌نامه اضطراب مجدداً توزیع و میزان اضطراب آزمودنی نسبت به هر دو آزمون سنجیده شد. نسخه فارسی پرسش‌نامه اضطراب پاورز (۱۹۹۹) در پیوست الف آمده است.

#### ۳-۴. روایی و پایایی آزمون

به‌منظور طراحی آزمون‌های رایانه‌ای مستقیم و انطباقی رایانه‌ای در ابتدا باید سؤالات آن طراحی می‌شد. به این ترتیب با توجه به محتوای آموزشی در هر سطح (مقدماتی - میانی - پیشرفته) و با استفاده از استاندارد مرجع آموزش زبان فارسی (صحرایی و مرصوص، ۱۳۹۵) ۳۹۳ سؤال طراحی شد. این سؤالات شامل ۱۸۵ سؤال سطح مقدماتی، ۱۰۸ سؤال سطح میانی، و ۱۰۰ سؤال سطح پیشرفته می‌شود. پس از طراحی سؤالات، با استفاده از روش دلفای، این سؤالات خام به ۵ کارشناس آموزش زبان فارسی به غیرفارسی‌زبانان داده و اصلاح شد. پس از اصلاحات نهایی تعداد سؤالات برابر با ۳۲۴ سؤال شامل ۱۲۷ سؤال مقدماتی، ۹۸ سؤال سطح میانی، و ۹۹ سؤال سطح پیشرفته شد. علاوه بر این، صورت برخی از سؤالات تغییر کرد. سپس به‌منظور اصلاح مجدد سؤالات و اطمینان از میزان شفافیت و وضوح سؤالات و نبود ابهام در آن‌ها از (۳۰ نفر) سه گروه ۱۰ نفره از زبان‌آموزان در سه سطح مقدماتی (۱۰ نفر)، میانی (۱۰ نفر) و پیشرفته (۱۰ نفر) امتحان گرفته شد. به هر گروه صرفاً سؤالات مربوط به سطح خود داده شد. در انتها نظر زبان‌آموزان درباره هر سؤال پرسیده شد و ابهامات و مشکلاتی که در صورت سؤالات یا گزینه‌ها وجود داشت، برطرف شد.

پس از اعمال اصلاحات مذکور، تعیین ضریب تشخیص و دشواری و آلفای کرونباخ آزمون مستلزم وجود تعداد زیادی فارسی‌آموز عراقی بود. به همین منظور، به کمک همکاران در این حوزه توانستیم علاوه بر فارسی‌آموزان عراقی در دانشگاه تربیت مدرس، از فارسی‌آموزان عراقی دانشگاه بین‌المللی امام خمینی و شهید بهشتی نیز امتحان بگیریم و مؤلفه‌های آزمون را تعیین کنیم. پس از انجام محاسبات لازم در این مرحله و غربال سؤالات، ۳۰۴ سؤال شامل ۱۲۰ سؤال سطح مقدماتی، ۸۸ سؤال سطح میانی و ۹۶ سؤال سطح پیشرفته باقی ماند. این تعداد سؤالات به‌عنوان سؤال نهایی آزمون انتخاب شد و ضریب دشواری آنان بین ۰/۳۷ تا ۰/۶۳ بود. همچنین آلفای کرونباخ آزمون سطح مقدماتی برابر ۰/۸۳، سطح میانی برابر ۰/۷۷ و سطح پیشرفته برابر ۰/۸۵ است.

#### ۴-۴. روش تحلیل داده‌ها

به‌منظور پاسخ به پرسش پژوهش و به‌منظور مقایسه نمرات حاصل از پرسشنامه

شاخص‌های اضطراب، از میانگین نمرات به دست آمده از هر پرسش‌نامه (مجموع نمرات تقسیم بر تعداد سؤالات پرسش‌نامه) استفاده خواهد شد.

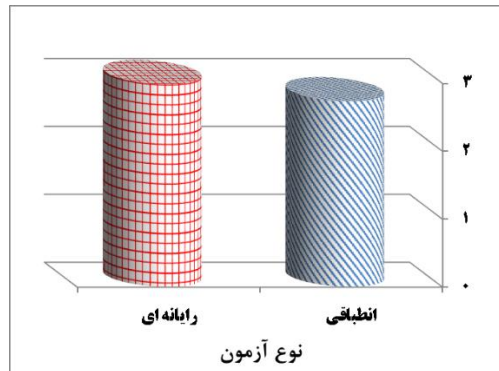
### ۵. تحلیل داده‌ها

جدول ۱ نشان می‌دهد که میانگین نمرات پرسش‌نامه اضطراب در آزمون رایانه‌ای بیشتر از آزمون انطباقی محاسبه شده است و بیشترین مقدار پراکندگی نیز مربوط به آزمون انطباقی است.

جدول ۱: شاخص‌های توصیفی نمرات پرسش‌نامه اضطراب

Table 1: Descriptive Scores Indices of Anxiety Questionnaire

متغیر	نوع آزمون	تعداد	دامنه	حداقل	حداکثر	میانگین	انحراف معیار	واریانس
اضطراب	آزمون انطباقی	۴۸	۳/۲۰	۱/۴۰	۴/۶	۲/۸۳۳۳	۰/۸۴۴۰۲	۰/۷۱۲
	آزمون رایانه‌ای	۴۸	۲/۷۵	۱/۸۰	۴/۵۵	۲/۹۸۵۴	۰/۷۱۲۶۵	۰/۵۰۸



نمودار ۱: میانگین نمرات اضطراب به تفکیک نوع آزمون

Graph 1: Anxiety Score Average with Test Type



نمودار ۱ مشخص می‌کند که به‌طور شهودی نیز میانگین نمرات اضطراب در آزمون رایانه‌ای بیشتر است که در بخش بعد با استفاده از آزمون آماری مناسب این مسئله نیز مورد بررسی قرار خواهد گرفت.

با توجه به اینکه نمرات برای هر متغیر در دو نوع آزمون (انطباقی و رایانه‌ای) محاسبه شده‌اند، برای بررسی فرضیه پژوهش از آزمون پارامتری  $t$  با نمونه‌های جفتی ( $t$  زوجی) استفاده می‌شود. اما قبل از استفاده از این آزمون لازم است از نرمال بودن میانگین نمرات مؤلفه‌ها در دو نوع آزمون اطمینان پیدا کنیم. بنابراین با استفاده از آزمون ناپارامتری کولموگروف اسمیرنوف، نرمال بودن نمرات بررسی می‌شوند. نتایج این آزمون در جدول ۲ آمده است.

جدول ۲: آزمون ناپارامتری برای نرمال بودن مؤلفه‌ها

Table 2: Kolmogorov - Smirnov Test for Factors Normal Being

سطح معنی‌داری تقریبی (دو طرفه)	آماره Z کولموگروف - اسمیرنوف	بیشترین اختلاف			تعداد	مؤلفه‌ها
		منفی	مثبت	مطلق		
۰/۹۰۷	۰/۵۶۵	- /۰۷۲	/۰۸۲	/۰۸۲	۴۸	اضطراب آزمون انطباقی
۰/۹۰۵	۰/۵۶۷	- /۰۶۰	/۰۸۲	/۰۸۲	۴۸	اضطراب آزمون رایانه‌ای

چون سطوح معنی‌داری در جدول ۲ (ستون آخر جدول) به جز برای عملکرد آزمون انطباقی از مقدار خطای آزمون یعنی ۰/۰۵ بیشتر است، پس فرض نرمال بودن میانگین نمره مؤلفه‌های پژوهش در سطح خطای ۰/۰۵ رد نمی‌شود. بنابراین برای آزمون مقایسه میانگین نمرات هر شاخص می‌توان از آزمون پارامتری  $t$  زوجی استفاده کرد. در ادامه به بررسی فرضیه‌های تحقیق پرداخته می‌شود.

جدول ۱ و نمودار ۱ مشخص کرد که میانگین نمرات اضطراب در آزمون رایانه‌ای بیش از آزمون انطباقی محاسبه شده است. با توجه به سؤال دوم پژوهش (رابطه بین آزمون‌های تعیین

سطح مؤلفه‌های دستور و واژه فارسی انطباقی رایانه‌ای و رایانه‌محور بر میزان اضطراب آزمودنی‌ها، در صورت وجود ارتباط، چیست؟) اکنون فرضیه دوم مورد بررسی قرار می‌گیرد. فرضیه پژوهش: میزان اضطراب آزمون (در صورت وجود اضطراب) در آزمون انطباقی رایانه‌ای کمتر از آزمون رایانه‌محور است.

جدول ۳: نتایج آزمون t مربوط به میانگین نمره مؤلفه اضطراب در آزمون رایانه‌ای

**Table 3:** The Results of t - Test Related to Score Average of Anxiety Factor in Computerized Test

مقدار آزمون = ۳						مؤلفه
فاصله اطمینان ۹۵ درصد برای اختلاف میانگین		اختلاف میانگین	سطح معنی‌داری	درجه آزادی	مقدار آماره آزمون	
حد بالا	حد پایین					
۰/۲۰۷۸	۰/۰۵۲۲	۰/۱۳	۰/۰۰۲	۴۴	۳/۳۶۷	اضطراب آزمون رایانه‌ای

جدول ۱ و نمودار ۱ نشان داد که میانگین نمرات اضطراب در آزمون رایانه‌ای بیشتر از آزمون انطباقی است. به‌منظور بررسی فرضیه تحقیق از آزمون t زوجی استفاده می‌شود.

جدول ۴ خلاصه نتایج بررسی فرض‌های فوق را نشان می‌دهد.

جدول ۴: مقایسه میانگین اضطراب در آزمون‌های رایانه‌ای و انطباقی

**Table 4:** Anxiety Average Comparison in Adaptive and Linear Testing

سطح معناداری	درجه آزادی	آماره آزمون t	اختلاف زوج‌ها				متغیر	
			فاصله اطمینان ۹۵ درصدی		خطای استاندارد میانگین	انحراف معیار		میانگین
			حد بالا	حد پایین				
۰/۰۰۱	۴۷	- ۳/۹۰۸	- /۱۲۲۳۳	- /۳۸۱۸۴	/۰۶۴۵۰	/۴۴۶۸۵	- /۲۵۲۰۸	اختلاف نمرات اضطراب رایانه‌ای و انطباقی

با توجه به جدول ۴ چون سطح معنی‌داری آزمون (مقدار ۰/۰۰۱) کمتر از خطای آزمون (مقدار ۰/۰۵) محاسبه شده است، پس میانگین نمرات اضطراب رایانه‌ای و انطباقی اختلاف معناداری دارند. اکنون با توجه به منفی بودن حد بالا و حد پایین فاصله اطمینان ۹۵ درصدی برای اختلاف میانگین، نتیجه می‌شود که اختلاف میانگین بین نمرات اضطراب رایانه‌ای و انطباقی منفی است (البته این استنباط با توجه به اختلاف میانگین شاخص (مقدار ۰/۲۵۲۰۸ -) که مقداری منفی است نیز امکان‌پذیر است). بنابراین میانگین نمرات اضطراب رایانه‌ای به‌طور معنی‌داری بیش از میانگین نمرات آزمون انطباقی است. در نتیجه فرضیه دوم مبنی بر اینکه میزان اضطراب آزمون در آزمون انطباقی رایانه‌ای کمتر از آزمون رایانه‌محور است، مورد تأیید قرار خواهد گرفت.

## ۶. نتیجه

در ارتباط با سؤال پژوهش (رابطه بین آزمون‌های تعیین سطح مؤلفه‌های دستور و واژه فارسی انطباقی رایانه‌ای و رایانه‌محور بر میزان اضطراب آزمودنی‌ها، در صورت وجود ارتباط، چیست؟)، از پرسش‌نامه دنالد پاورز (۱۹۹۹) در پیوست الف استفاده شد.

با استفاده از آزمون پارامتری  $t$  زوجی برای تحلیل پرسش‌نامه و با توجه به جدول ۳ سطح معنی‌داری آزمون یعنی ۰/۰۰۱ از خطای آزمون یعنی مقدار ۰/۰۵ کوچکتر است. بنابراین میانگین نمره مؤلفه اضطراب در آزمون انطباقی با عدد ۳ اختلاف معناداری دارد. اکنون با توجه به منفی بودن حد بالا و حد پایین فاصله اطمینان اختلاف میانگین، درمی‌یابیم که میانگین نمره مؤلفه اضطراب در آزمون انطباقی کمتر از ۳ است (البته این استنباط با توجه به اختلاف میانگین این مؤلفه (۰/۰۷۱۱۱ -) که مقداری منفی است نیز امکان‌پذیر است). بنابراین میانگین نمره مؤلفه اضطراب در آزمون انطباقی کمتر از حد متوسط است.

همچنین، با توجه به جدول ۴ چون سطح معنی‌داری آزمون یعنی ۰/۰۰۲ از خطای آزمون یعنی مقدار ۰/۰۵ کوچکتر است، بنابراین میانگین نمره مؤلفه اضطراب در آزمون رایانه‌ای نیز با عدد ۳ اختلاف معناداری دارد. اکنون با توجه به مثبت بودن حد بالا و حد پایین فاصله اطمینان اختلاف میانگین، درمی‌یابیم که میانگین نمره مؤلفه اضطراب در آزمون رایانه‌ای بیش از حد متوسط ۳ است (البته این استنباط با توجه به اختلاف میانگین این مؤلفه (۰/۱۳) که مقداری مثبت

است نیز امکان‌پذیر است). بنابراین میانگین نمره مؤلفه اضطراب در آزمون رایانه‌ای بیش از حد متوسط است. به این ترتیب، میانگین نمرات اضطراب در آزمون رایانه‌ای بیشتر از آزمون انطباقی است.

با توجه به جدول ۴ سطح معنی‌داری آزمون (مقدار ۰/۰۰۱) کمتر از خطای آزمون (مقدار ۰/۰۵) محاسبه شده است. بنابراین میانگین نمرات اضطراب رایانه‌ای و انطباقی اختلاف معناداری دارند. اکنون با توجه به منفی بودن حد بالا و حد پایین فاصله اطمینان ۹۵ درصدی برای اختلاف میانگین، نتیجه می‌شود که اختلاف میانگین بین نمرات اضطراب رایانه‌ای و انطباقی منفی است (البته این استنباط با توجه به اختلاف میانگین شاخص (مقدار ۰/۲۵۲۰۸ -) که مقداری منفی است نیز امکان‌پذیر است). بنابراین میانگین نمرات اضطراب رایانه‌ای به‌طور معنی‌داری بیش از میانگین نمرات آزمون انطباقی می‌باشد. در نتیجه فرضیه دوم مبنی بر اینکه میزان اضطراب آزمون در آزمون انطباقی رایانه‌ای کمتر از آزمون رایانه‌محور است، مورد قرار خواهد گرفت. این یافته‌ها با پژوهش هنینگ<sup>۱۸</sup> (۱۹۹۱) در یک سو قرار دارد. اما با یافته‌های پژوهش هوپر و هانافین (۱۹۸۹) مبنی بر اینکه آزمون‌های رایانه‌ای اضطراب بیشتری در آزمودنی ایجاد می‌کنند، مغایر است.

این نتایج به تشویق مدرسان و مراکز آموزشی به طراحی آزمون انطباقی رایانه‌ای بیشتر کمک می‌کند. همچنین تأثیر این آزمون‌ها در میزان کاهش اضطراب زبان‌آموزان نسبت به آزمون‌های مداد - کاغذی و یا آزمون‌های رایانه‌ای مستقیم را نشان می‌دهد. بنابراین با استفاده از آزمون‌های انطباقی رایانه‌ای زبان‌آموزان می‌توانند در خانه و یا مؤسسه با اضطراب و نگرانی کمتر آزمون متناسب با سطح خود را بدهند و نتایج آزمون به میزان بیشتری پایایی خواهد داشت.

## ۷. پی‌نوشت‌ها

1. assessment
2. José Noijons
3. Binet & Simon
4. response theory
5. Ward, T., Hooper, S., & Hannafin, K.

6. Olea et. al
7. one parametric logistic model
8. two parametric logistic model
9. three parametric logistic model
10. latent
11. fixed- length test
12. variable - length test
13. multistage question (MST)
14. help
15. mixed - explanatory research design
16. triangulation
17. powers
18. Henning

#### ۸. منابع

- زندی، ب.، روشن، ب.، وکیلی فرد، ا.ر.، و گلپور، ل. (۱۳۹۸). طراحی و اعتباربخشی آزمون استاندارد زبان فارسی (بر پایه چهار مهارت زبانی). *جستارهای زبانی*، ۱۰ (۵۴)، ۱۳۹ - ۱۷۱.
- صحرایی، ر.م.، و مرصوص، ف. (۱۳۹۵). *استاندارد مرجع آموزش زبان فارسی*. تهران: نشر دانشگاه علامه طباطبائی.
- قنصولی، ب. (۱۳۸۹). طراحی و روایی سنجی آزمون بسندگی زبان فارسی. پژوهش زبان‌های خارجی، ۵۷، ۱۱۵-۱۲۹.
- محمودی، ف. (۱۳۹۳). *طراحی آزمون‌های انطباقی رایانه‌ای زبان فارسی برای سنجش سطح زبانی غیر فارسی‌زبانان*. پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبائی.

#### References

- Abukhzam, O. M. (2013). Tracing the journey of the use of computer technology in the field of second and foreign language assessment & testing. *International Refereed & Indexed Journal of English Language & Translation Studies*, 1(1), 87–99.
- Birjandi, P., Bagheridust, E., & Mosalanejad, P. (2004). *Language testing a concise collection for graduate applicants*. Shahid Mahdavi Pub.

- Brown, H. D. (1980). *Principles of language learning and teaching*. Pearson Publication.
- Casbarro, H. (2005). *Test anxiety and what you can do about it: A practical guide for teachers, parents, and kids*. Dude Publishing.
- Cassady, J. C. (2004). The influence of cognitive test anxiety across the learning - testing cycle. *Learning and Instruction*, 14, 569–592. <http://dx.doi.org/10.1016/j.learninstruc.2004.09.002>.
- Dunn, T. G., Lushene, R. E., & O'neil, H. F. (1972). Complete automation of the MMPI and a study of its response in latencies. *Journal of Consulting and Clinical Psychology*, 39, 381–387.
- Elwood, D. L. (1969). Automation of psychological testing. *American Psychologist*, 24, 287–289.
- Fleming, S., & Hiple, D. (2004). Distance education to distributed learning: Multiple formats and technologies in language instruction. *CALICO Journal*, 22(1), 63–82.
- Fritts, B. E., & Marszalek, J. (2010). Computerized adaptive testing, anxiety levels, and gender differences. *Social Psychology of Education*, 13(3), 441–458. <http://dx.doi.org/10.1007/s11218-010-9113-3>
- Ghonsouli, B. (2010). Development and validation of a Persian proficiency test. *Pazhuhesh - E Zabanha - Ye Khareji*, 57, 115–129[In Persian].
- Guttman, L. (1944) A basis for scaling qualitative data. *American Sociological Review*, 9, 139–150. <http://dx.doi.org/10.2307/2086306>.
- Hembree, R. (1988). Correlates, causes, effects and treatment of text anxiety. *Review of Educational Research*, 58, 47–77.
- Henning, G. T. (1991). Validating an item bank in a computer-assisted or computer - adaptive test: Using item response theory for the process of validating CATS. In P. Dunkel (Ed.), *Computer-assisted language learning and testing: Research issues*

and practice, (pp. 209–222). Newbury House.

- Ju, G., & Fon, N. (1993). *A computer - based Chinese edition of DAT mechanical reasoning: Comparing computer - based and paper - pencil formats of a timed pictorial tests in Taiwan*. University of Illinois at Urbana - Champaign.
- Kiely, G. L., Zara, A. R., & Weiss, D. J. (1986). *Equivalency of computer and paper - and - pencil Armed Services Vocational Aptitude Battery tests*. Brooks Air Force Base, TX: Air Force Human Resources Laboratory.
- Kirschner, P. A., Kirschner, F., & Janssen, J. (2014). The collaboration principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (Second edition) (pp. 547–575). Cambridge University Press.
- Kizlik, R. J. (2012) Measurement, assessment, and evaluation in education. Retrieved on October 2019. [Online]. Available: <https://www.adprima.com/measurement.htm>].
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. The Guilford Press.
- Larson, J. W., & Madsen, H. S. (1985). Computer - adaptive language testing: Moving beyond computer - assisted testing. *CALICO Journal*, 2(3), 32–6.
- Lilley, M. (2007). *The development and application of computer - adaptive testing in a higher education environment*. Unpublished doctoral dissertation, Hertfordshire University.
- Linden, W. J., & Glas, A. W. (2000). *Computerized Adaptive Testing: Theory and Practice*. Kluwer.
- Madsen, H. S. (1991). Computer-adaptive test of listening and reading comprehension: The Brigham Young University approach. In P. Dunkel (Ed.), *Computer-assisted language learning and testing: Research issues and practice*, (pp. 237–257). Newbury House.
- Masters, G. (2014). Getting to the essence of assessment. Retrieved on October

2019. [Online]. Available: <https://rd.acer.org/article/getting-to-the-essence-of-assessment>.

- Mmahmoudi, F. (2014). *Designing computerized adaptive tests for evaluating the language level of non-Persian speakers*. M.A thesis: Allameh Tabataba'i University. [In Persian].
- Noijons, J. (1994). Testing computer assisted language tests: Towards a checklist for CALT. *CALICO Journal*, 12(1), 37–58.
- Olea, J.V., Revuelta, J., Ximenez, M.C., & Abad, F. J. (2000). Psychometric and psychological effects of review on computerized fixed and adaptive tests. *Psicología*, 21, 157–173.
- Rezaie M., & Golshan, M. (2015). Computer adaptive test (CAT): Advantages and limitations. *Int J Educ Investig*, 2(5), 128–37.
- Richards, J. C. (1994). Teacher thinking and foreign language teaching. *MEXTESOL Journal*, 18(2), 7–13.
- Sahraie, R. M., & Marsus, F. (2017). *Persian language teaching standard reference*. Allameh Tabataba'i University Publication [In Persian].
- Stiggins, R. (2009). Essential formative assessment competencies for teachers and school leaders. In H.L. Andrade & G. J Cizek (Eds.), *Handbook of formative assessment*, (pp. 233–250). Routledge.
- Taras, M. (2005). Assessment: Summative and formative: Some theoretical reflections. *British Journal of Educational Studies*, 53(4), 466 – 478.
- Van der Linden, W., & Hambleton, R. (1997). *Handbook of modern item response theory*. Springer - Verlag.
- Ward, T., Hooper, S., & Hannafin, K. (1989). The effect of computerized tests on the performance and attitudes of college students. *Journal of Educational Computing Research*, 327–333. <https://doi.org/10.2190/4UID-VQRM-J70DJEQF>.



- Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21, 361–375. <https://doi.org/10.1111/j.1745-3984.1984.tb01040.x>.
- Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Mesa Press.
- Zandi, B., Rovshan, B., Vakilifard, A., & Golpour, L. (2020). planning and validating Persian language standard test (On the Basis of Four Language Skills). *Language Related Research*, 10(6), 139–171. [In Persian].
- Zeidner, M. (1998). *Test anxiety: The state of the art*. Plenum Press.
- Zeidner, M. (2007). Test anxiety in educational contexts: Concepts, findings, and future directions. In M. Zeidner, P. A. Schutz, & R. Elsevier Academic Press. <http://dx.doi.org/10.1016/B978-012372545-5/50011-3>

ضمیمه الف. پرسشنامه اضطراب آزمون دنالد پاورز (1991)

Appendix A. Donal Powers Test Anxiety Questionnaire (1991)

خیلی کم	کم	متوسط	زیاد	خیلی زیاد	سوالات
					۱. هنگام آزمون بسیار تحت فشار و مردد هستم.
					۲. هنگامی که آزمون می‌دهم، خیلی نگران و ناراحت هستم.
					۳. فکر کردن به نمره‌ای که می‌گیرم باعث می‌شود هنگام امتحان نتوانم درست تمرکز کنم.
					۴. هنگام امتحان هنگ می‌کنم و گیج می‌شوم.
					۵. هنگام آزمون به اینکه آیا نمره قبولی را می‌گیرم یا نه فکر می‌کنم.
					۶. هرچه بیشتر برای امتحان تلاش می‌کنم، هنگام امتحان بیشتر گیج می‌شوم.

خیلی کم	کم	متوسط	زیاد	خیلی زیاد	سوالات
					۷. فکر کردن به عملکرد ضعیفم هنگام امتحان باعث اختلال در تمرکز می‌شود.
					۸. هنگام امتحان بسیار عصبی هستم.
					۹. حتی زمانی که خیلی برای امتحان آماده هستم، باز هم احساس نگرانی و استرس دارم.
					۱۰. دقیقاً قبل از گرفتن نمره بسیار نگران و مضطرب هستم.
					۱۱. در طول آزمون بسیار نگران هستم.
					۱۲. فکر می‌کنم امتحانات من را اذیت نمی‌کنند.
					۱۳. هنگام آزمون به قدری اضطراب دارم که دل‌درد می‌گیرم.
					۱۴. زمانی که دارم آزمون می‌دهم انگار دارم با خودم رقابت می‌کنم.
					۱۵. هنگامی که آزمون می‌دهم بسیار استرس دارم و وحشت‌زده هستم.
					۱۶. قبل از آزمون بسیار نگران هستم.
					۱۷. قبل از آزمون به شکست در امتحان و پیامدهای آن فکر می‌کنم.
					۱۸. هنگام آزمون قلبم به شدت می‌تپد.
					۱۹. بعد از اتمام آزمون، سعی می‌کنم دیگر نگران آزمون نباشم، اما نمی‌توانم.
					۲۰. در طول آزمون به قدری مضطرب و نگران می‌شوم که حتی مطالبی که می‌دانم فراموش می‌کنم.