



## Design and Validation of the First Persian Placement Test Based on the Educational Resources of Al-Mustafi (PBUH) Center for Teaching Persian to Non-Persian Speakers

**Zahra Samaee**

M.A. Student in TPNPS at Ferdowsi University of Mashhad, Instructor at Al-Mustafi Center for Teaching Persian to Non-Persian Speakers, Mashhad, Iran.

samaeez@gmail.com

**Atiyeh Kamyabi Gol\***

Corresponding Author, Associate Professor, Department of Persian Language and Literature and the Department of Linguistics, Ferdowsi University of Mashhad, Mashhad, Iran.

kamyabigol@um.ac.ir

**Mahmood Mehravaran**

Associate Professor, Department of Persian Language and Literature, Qom University, Qom, Iran.

mehravaran72m@gmail.com

### Abstract:

Testing has always been an inseparable part of teaching. Having a test to determine the level of proficiency for Persian language teaching non-Persian speakers (Azfa) centers is a necessity. To this purpose, the first placement test based on the Persian language teaching textbooks of Al-Mustafa was designed and validated. The purpose of the present research was to design a standard test to determine the language proficiency level of the Persian language learners (beginner, intermediate, and advanced) based on "Parsa" textbooks. The participants were 135 Persian learners from Al-Mustafi (PBUH) Center for Teaching Persian to Non-Persian Speakers from four different nationalities studying in Iran and Turkey. The participants were recruited based on convenience sampling. The designed placement test included vocabulary and grammar, listening, reading, writing and speaking. The item analysis of the test was calculated using item difficulty and item discrimination formulas and the results showed that both the difficulty and discrimination levels were acceptable. The face and content validities were analyzed and confirmed using a written interview with 20 teachers of Al-Mustafa Center for Teaching Persian to Non-Persian Speakers. The inter-rater reliability (Cohen's Kappa) between the evaluators' scores for the speaking and writing sections of the test also indicated an acceptable agreement. Finally, the reliability of the tests was calculated using KR-21, and the reliability results of all three tests were above 80%, which shows the suitability of the tests.

\***Cite this article:** Samaee, Zahra. Kamyabi Gol, Atiyeh. Mehravaran Mahmood. Design and Validation of the First Persian Placement Test Based on the Educational Resources of Al-Mustafi (PBUH) Center for Teaching Persian to Non-Persian Speakers. Vol. 12, No. 1 (Tome 25), April 2023,45-72.

DOI: 10.30479/JTPSOL.2022.17549.1602

Received on: 13/07/2022

Accepted on: 10/09/2022

© The Author(s).

Publisher: Imam Khomeini International University



## Introduction

Language testing is an inseparable part of language teaching. Through language assessment and evaluation, educators can determine whether the learner has successfully achieved the learning outcomes. A language test is an instrument which measures the quality and performance of a language program and educational resources. Tests are of different types and one of the most important tests is the placement test which helps educational institutions to place a learner in the right class or group based on the learner's language knowledge and proficiency (Farhady et al., 2014; Bachman & Palmer, 2010). There is no doubt that the more standard the test, the more successful the language program. Therefore, it is of the utmost importance to design and validate language proficiency tests to assign learners to the right classes (Farhady et al., 2014) and help remove their psychological barriers (Kim & Shin, 2006). The present study is focused on designing and validating a placement test based on the Al-Mustafi (PBUH) Center for Teaching Persian to Non-Persian Speakers' Persian textbooks. Al-Mustafi (PBUH) Center for Teaching Persian to Non-Persian Speakers was established around 30 years ago and has over 60 branches around the world. Past literature has focused on the design and validation of English placement tests (Akhidenor-Bamidele, 2019; Kadwa & Sheik, 2021; Kim & Shin, 2006; Nasem & Mohammad, 2018) and has highlighted the importance of using such tests to group learners. Persian proficiency tests (Abbassi, 2014; Ghonsooly, 2003; Zandi et al., 2020;) and placement tests (Sahraei & Rezapour, 2011) have also been the focus of some research, but to the best of the authors' knowledge, no study has been conducted on designing and validating a placement test for Al-Mustafi (PBUH) Center for Teaching Persian to Non-Persian Speakers. This was the research gap that motivated the authors to focus on designing such a test for the three language proficiency levels elementary, intermediate, and advanced. The research questions for the present research focused on the item analysis, reliability and validity of the designed placement test.

## Methodology

The designed tests contained 40 questions on vocabulary, grammar, reading, and listening. Also, two writing tasks were considered for the writing section. Finally, 10 oral questions were used for the speaking section of the test. Two raters scored the writing and speaking sections individually. Item analysis was used to analyze all multiple-choice items on the tests. Also, to check the inter-rater reliability for the speaking and writing sections of the test, Cohen's Kappa was used. SPSS version 26 was used for calculating the Kappa. The reliability of the tests was analyzed through KR-21 and teacher feedback (written interview) was collected for the face and content validities. After taking all teacher comments and suggestions on board, the test was ready to be administered. A total of 135 Persian language learner participants were recruited from the three language proficiency levels. The participants were from Pakistan, India, Azerbaijan, and Turkey. The test was administered both in Iran and Turkey. The 90 participants studying in Iran were all females and their age

range was 18-25 (SD=1.62) and the 45 participants studying in Turkey were 25 females and 15 males and their age range was 13-47 (SD= 9.15). The teacher participants who took part in a written interview included 15 females and 5 males. Their ages ranged between 25 and 58 (SD=7.92) with 4-18 years of teaching experience. All participants signed a written consent form.

## **Results and Discussion**

The internal consistency of the designed test items was measured through Cronbach's alpha. Cronbach's alpha reliability coefficient normally ranges between 0 and 1. The closer the coefficient is to 1, the more internally consistent the test items. The reliability coefficient results for the three placement test levels revealed 0.94 for the beginner level which was found to be highly reliable, 0.83 for the intermediate level which indicates good internal consistency, and 0.80 for the advanced level which again indicates good internal consistency for the test items. The Kuder-Richardson 21 formula was also used to measure the internal consistency of the test items. All KR-21 results for reliability showed acceptable results of over 0.7 (Dornyei, 2011; Farhady et al., 2014). Also, the inter-rater reliability results for the speaking and writing tasks of the tests were considered acceptable. The inter-reliability was measured to measure the degree of agreement among independent observers. This was used to limit the subjective scoring of each rater. These results ranged from 0.69-0.91. The face and content validity of the tests were conducted through a written interview and the results showed that all teachers approved both the face and content validities. The only concern some teachers had was regarding the test time. For the beginner level, 10% of the teachers, and five per cent of the teachers for the intermediate level test believed that the time should be increased by five minutes. The item analysis results also indicated that for the beginner level out of the 40 items, 39 items had acceptable difficulty levels and all 40 items enjoyed an acceptable discrimination level. For the intermediate level test, out of the 40 items, 32 items were considered to have acceptable difficulty levels and 8 items were considered difficult. 38 of the items also showed an acceptable discrimination level. Finally, for the advanced proficiency level, out of the 40 items, 38 items had an acceptable range and 2 items were considered to be simple. One reason for this simplicity might have been the shared interlingual lexicon since most advanced-level participants were from Pakistan. The discrimination level for 39 of the advanced-level test items was considered to be acceptable as well. All in all, it seems that the designed test was generally dominated by acceptable items. Also, most of the question items could successfully discriminate between high and low students. The overall face and content validities were also approved by the teachers and the reliability was approved through KR-21.

## Conclusion

The present study aimed to design and validate a placement test based on Al-Mustafi (PBUH) Center for Teaching Persian to Non-Persian Speakers' Persian textbooks. This was the first validated placement test to be designed for the Al-Mustafi (PBUH) Center for Teaching Persian to Non-Persian Speakers. The results revealed that the designed test for all three proficiency levels was considered acceptable. The internal consistency of the designed test items, measured through Cronbach's alpha and KR-21 were considered reliable and also Cohen's Kappa values showing the results for the inter-rater reliability measurement, ranging from 0.69-0.91, were considered very good to excellent. As for the validity of the placement tests, the results for the face validity and content validity of the tests, conducted through a written interview with teacher participants at all three proficiency levels (beginner, intermediate, and advanced), indicated an overall approval. The teachers' suggestions regarding the time limit on the test for each level were also taken into consideration. The item analysis results including both item difficulty and item discrimination also showed acceptable levels for the test items. In the end, the authors hope that the designed placement test could be used at Al-Mustafi (PBUH) Centers for Teaching Persian to Non-Persian Speakers worldwide and that this instrument could help both teachers and learners alike to have a better learning experience.

## Acknowledgment

This paper and the research behind it would not have been possible without the exceptional support from Al-Mustafi (PBUH) Center for Teaching Persian to Non-Persian Speakers in Istanbul and Mashhad.

## Conflict of Interest

No conflicts of interest are reported by the author for this article.

**Keywords:** Language learners corpora, Chinese learners of Persian, Category and Scale Grammar, Writing.



## طراحی و اعتبارسنجی اولین آزمون تعیین سطح زبان فارسی بر اساس منابع آموزشی جامعه‌المصطفی (ص) العالمیه (پژوهشی)

زهرا سمائی

کارشناسی ارشد آموزش زبان فارسی به غیر فارسی‌زبانان، دانشگاه فردوسی مشهد، مشهد، ایران.  
samaeez@gmail.com

عطیه کامیابی گل\*

نویسنده مسئول، دانشیار گروه زبان و ادبیات فارسی و گروه زبان‌شناسی، دانشگاه فردوسی مشهد، مشهد، ایران.  
kamyabigol@um.ac.ir

محمود مهرآوران

دانشیار گروه زبان و ادبیات فارسی، دانشگاه قم، ایران.

mehravaran72m@gmail.com

### چکیده

آزمون و آزمون‌سازی همواره بخش جدانشدنی آموزش زبان است. وجود یک آزمون تعیین سطح استاندارد برای مراکز آموزش زبان فارسی به غیرفارسی‌زبانان (آزفا) امری ضروری است. به این منظور برای اولین بار آزمون تعیین سطح متناسب با منابع آموزش زبان فارسی جامعه‌المصطفی (ص) طراحی و اعتبارسنجی شد. هدف پژوهش، طراحی آزمون استاندارد هنجار شده و علمی برای تعیین سطح زبان فارسی، در سه سطح مبتدی، میانی و پیشرفته بر اساس مجموعه کتاب‌های «پارسا» است. جامعه آماری این پژوهش ۱۳۵ فارسی‌آموز مراکز آموزش زبان فارسی جامعه‌المصطفی (ص) از چهار ملیت مختلف در ایران و ترکیه بودند که به صورت نمونه‌گیری در دسترس برای شرکت در پژوهش دعوت شدند. آزمون شامل بخش‌های واژگان و دستور، شنیدار، خواندار، نوشتار و گفتار است. تحلیل سؤال‌های آزمون با استفاده از فرمول‌های ضریب تمیز و ضریب دشواری محاسبه شد که نتایج نشان‌دهنده قدرت تشخیص مناسب و قابل قبول بودن ضریب دشواری ماده‌های آزمون بود. روایی صوری و محتوایی آزمون‌ها نیز با استفاده از مصاحبه کتبی با ۲۰ مدرس جامعه‌المصطفی (ص) و آکاوی و تأیید شد. در بخش شفاهی و سنجش مهارت گفتاری، پایایی بین نمرات دو ارزیاب، با ضریب کاپای کوهن محاسبه شد و نتایج نشان‌دهنده توافق قابل قبول بین آنها بود. در نهایت، پایایی آزمون‌ها با استفاده از فرمول کودر-ریچاردسون ۲۱ محاسبه شد که نتیجه پایایی هر سه آزمون بالای ۸۰٪ به دست آمد که مطلوب بودن آزمون‌ها را نشان می‌دهد.

### کلیدواژه‌ها

آزمون تعیین سطح زبان فارسی، پایایی، روایی، آزمون‌سازی زبان، مرکز آموزش زبان جامعه‌المصطفی (ص)

\* استناد: سمائی، زهرا. کامیابی گل، عطیه. مهرآوران، محمود. طراحی و اعتبارسنجی اولین آزمون تعیین سطح زبان فارسی بر اساس منابع آموزشی جامعه‌المصطفی (ص) العالمیه، سال دوازدهم، شماره اول (پیاپی ۲۵)، بهار و تابستان ۱۴۰۲، ۴۵-۷۲.

شناسه دیجیتال (DOI): 10.30479/JTPSOL.2022.17549.1602

تاریخ پذیرش نهایی مقاله: ۱۴۰۱/۰۶/۱۹

تاریخ دریافت مقاله: ۱۴۰۱/۰۴/۲۲

ناشر: دانشگاه بین‌المللی امام خمینی (ه)

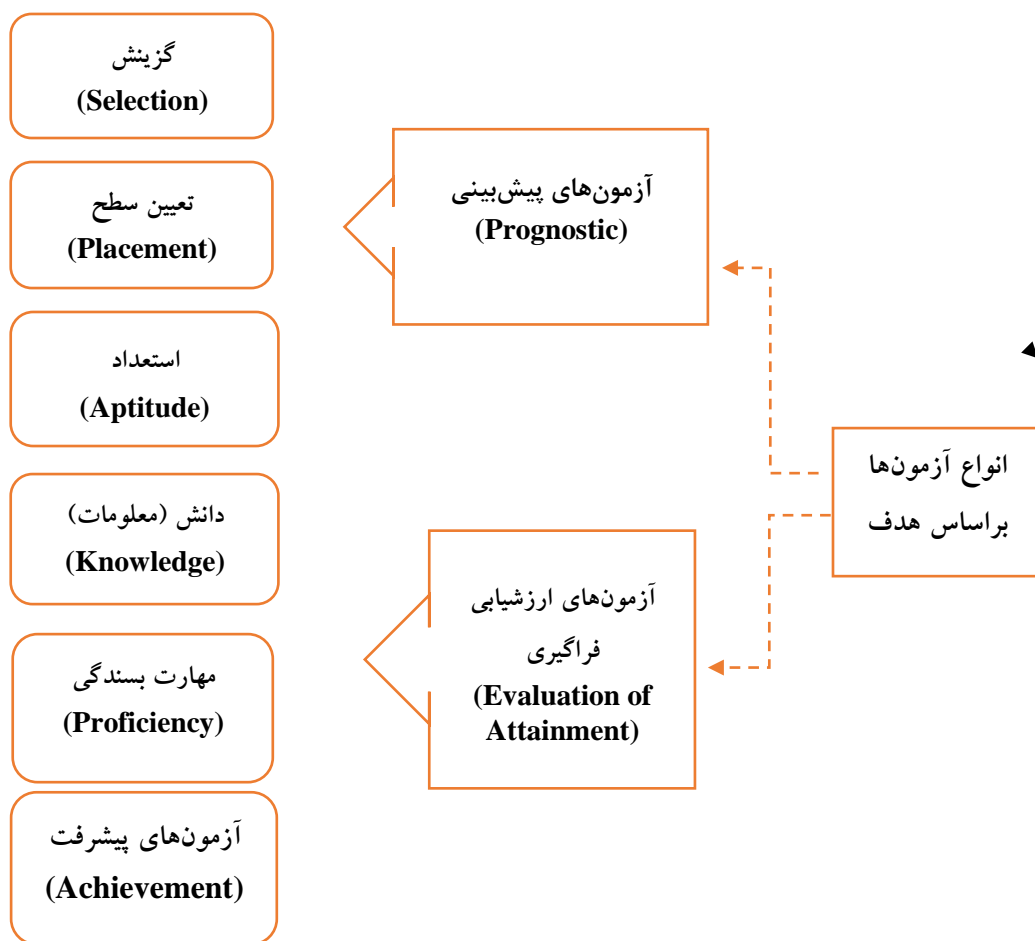
## ۱. مقدمه

یکی از موضوعاتی که در امر آموزش زبان فارسی اهمیّت شایانی دارد، آزمون و آزمون‌سازی است. آزمون‌سازی هر زبان، یکی از نقاط کلیدی در آموزش آن زبان است و بر اهداف آموزش و میزان موفقیت در دستیابی به آنها اثر می‌گذارد. Chastain (۱۹۸۸) معتقد است آموزش و یادگیری بدون ارزشیابی مکرر آموخته‌ها، بی‌ثمر و بی‌نتیجه خواهد بود. در امر آموزش زبان، آزمون‌ها با ارزیابی روند پیشرفت زبان‌آموزان، میزان یادگیری آنها را مشخص می‌کنند. آزمون‌ها به هدف‌گذاری بهتر و انتخاب درست‌تر مواد آموزشی مورد نیاز زبان‌آموزان نیز کمک می‌کنند. یکی دیگر از کاربردهای آزمون‌ها این است که از طریق تحلیل‌های انجام‌شده بر نتایج زبان‌آموزان، می‌توانند میزان موفقیت سیستم آموزش زبان و برنامه‌های آموزشی را آشکار کنند. از این‌رو، صاحب‌نظران آزمون‌سازی (Klein- Braley، 1984، Brown، 2010، Bachman، 1995، Bachman & Palmer، 2010) باور دارند آزمون‌سازی زبان امری بسیار پیچیده و چندوجهی است که در آن باید از زبان‌شناسی و فنون آموزش زبان نیز بهره جست.

آزمون‌ها انواع مختلفی دارند و یکی از مهم‌ترین آنها که آغازگر مسیر یادگیری است، آزمون تعیین سطح زبان است. آزمون تعیین سطح معمولاً در بدو ورود زبان‌آموزان به مراکز آموزش زبان و با هدف تعیین میزان توانایی و سطح آمادگی زبانی آنها برای یک دوره آموزش زبان انجام می‌گیرد (Farhady et al, 2014). به دلیل ضرورت دسته‌بندی زبان‌آموزان به گروه‌های همگن از نظر توانایی‌های زبانی، آزمون‌های تعیین سطح یک جزء بسیار مهم در اغلب برنامه‌های آموزش زبان به حساب می‌آیند (Brown, 2008, 2010). آزمون‌های تعیین سطح باید با روایی و پایایی بالا، زبان‌آموزان را در گروه‌های متجانس تقسیم‌بندی کنند تا هر گروه برنامه آموزشی مناسب خود را دریافت کند. فقدان آزمون‌های تعیین سطح و قرارگیری زبان‌آموزان با آمادگی‌های زبانی متفاوت در یک گروه، اثربخشی برنامه‌های آموزش زبان را به شدت کاهش می‌دهد. همچنین، در هنگام جذب زبان‌آموزان جدید و گروه‌بندی آنها بدون داشتن معیار مناسب و استاندارد، احتمال خطا و قرارگیری زبان‌آموز در دوره‌های نامناسب و مشکلات بعدی در امر آموزش زبان و موانع روان‌شناسی ایجادشده برای زبان‌آموز افزایش می‌یابد (Kim & Shin, 2006). بعضی از مراکز آرفا در ایران برای تعیین سطح زبانی زبان‌آموزان خود، آزمون طراحی کرده‌اند. جامعه‌المصطفی (ص) العالمیه به‌عنوان یکی از مراکز مهم آموزش زبان فارسی در دنیا آزمون تعیین سطح ندارد. بنابراین، این پژوهش بر آن شد تا براساس منابع آموزشی جامعه‌المصطفی (ص)، سه آزمون استاندارد و علمی برای تعیین سطح زبانی فارسی‌آموزان در سه سطح مبتدی، میانی و پیشرفته تهیه و ارائه کند و با توجه به گستردگی وسیع آموزش زبان فارسی این مجموعه در سطح بین‌المللی وجود آن بسیار ضروری است. پژوهش حاضر به دنبال پاسخ این پرسش‌های زیر است: ۱) آیا آزمون تعیین سطح طراحی‌شده دارای پایایی و روایی لازم است؟ ۲) آیا آزمون دارای ضریب دشواری و ضریب تمیز قابل قبول است؟ ۳) نظر مدرسان سطوح مختلف مرکز بین‌المللی آموزش زبان فارسی به غیرفارسی‌زبانان جامعه‌المصطفی (ص) درباره بخش‌های آزمون چیست؟

## ۲. چهارچوب نظری پژوهش

آزمون وسیله‌ای برای جمع‌آوری اطلاعات عددی درباره یک خصلت است. هدف از جمع‌آوری اطلاعات، تعیین میزان موجودیت آن خصلت است. منظور از نقش آزمون، هدفی است که آزمون به آن منظور طراحی شده است. از آنجاکه هدف آزمون به‌طور مستقیم در ساختار و کاربرد نتایج آن تأثیر می‌گذارد، لازم است نخست به شرح هدف‌هایی بپردازیم که آزمون‌ها برای آنها مورد استفاده قرار می‌گیرند. آزمون‌های زبان برای اهداف متفاوتی تهیه می‌شوند و کارکردهای مختلفی دارند (Birjandi & Mosallanejad, ۲۰۱۰؛ Jafarpur, 1997؛ Farhady et al, 2014).



شکل ۱. انواع آزمون‌ها، برگرفته‌شده از جعفرپور (۱۳۷۵)

## ۱.۲. انواع آزمون

آزمون‌ها برای اهداف متفاوتی به کار می‌روند و هدف آزمون به‌طور مستقیم در ساختار و کاربرد نتایج آن تأثیر می‌گذارد. طبق شکل (۱)، آزمون‌ها دو هدف عمده دارند و به دو دسته کلی آزمون‌های پیش‌بینی<sup>۱</sup> و آزمون‌های ارزشیابی فراگیری<sup>۲</sup> تقسیم می‌شوند، سپس هر بخش به سه بخش مجزا تقسیم می‌شود. دسته اول آزمون‌های پیش‌بینی، شامل آزمون‌های گزینش<sup>۳</sup>، تعیین سطح<sup>۴</sup> و آزمون استعداد<sup>۵</sup> می‌شود و دسته دوم آزمون‌های ارزشیابی فراگیری شامل آزمون‌های دانش معلومات<sup>۶</sup>، مهارت بسندگی<sup>۷</sup> و آزمون‌های پیشرفت<sup>۸</sup> می‌شود.

### ۱.۱.۲. آزمون‌های پیش‌بینی

هدف از آزمون‌های پیش‌بینی‌کننده، گرفتن تصمیمات منطقی درباره موفقیت‌های آینده داوطلبان براساس توانایی‌های فعلی‌شان است. این آزمون‌ها به‌منظور پیش‌بینی کردن جریان عملی در آینده درباره داوطلبان به کار برده می‌شوند.

### ۱.۱.۱.۲. آزمون‌های گزینشی

این آزمون‌ها برای فراهم‌ساختن اطلاعاتی است که براساس آن پذیرش یا عدم پذیرش داوطلبان در برنامه‌ای خاص تعیین می‌شود. برنامه لزوماً مرتبط با یک برنامه آموزشی نیست؛ مثل آزمون گرفتن گواهی‌نامه رانندگی (Farhady et al, 2014).

### ۱.۱.۲. آزمون‌های تعیین سطح

این آزمون‌ها به‌منظور تعیین مناسب‌ترین سطح آموزشی برای داوطلبان به کار می‌روند. در آزمون‌های تعیین سطح قبولی یا مردودی وجود ندارد. هدف آزمون‌های تعیین سطح صرفاً سنجیدن توانایی‌های داوطلب در دنبال کردن مسیر خاصی از یادگیری زبان است (Fulcher, 2013).

### ۱.۱.۳. آزمون‌های استعداد

آزمون‌های استعداد به‌منظور پیش‌بینی کردن موفقیت‌های داوطلبان در دستیابی به اهداف خاصی در آینده به کار می‌روند. به‌منظور شرکت در آزمون استعداد، لزومی ندارد داوطلب از محتوای آزمون دانش قبلی داشته باشد. این

۱. Prognostic Tests

۲. Evaluation of Attainment

۳. Selection

۴. Placement

۵. Aptitude

۶. Knowledge

۷. Proficiency

۸. Achievement



آزمون‌ها کاربردهای زیادی دارند؛ مثلاً برای تعیین رشته تحصیلی دانشجویان یا پیش‌بینی شغل مناسب برای داوطلبان مفید هستند (Fulcher, 2013).

## ۲.۱.۲. آزمون‌های ارزشیابی فراگیری

هدف از آزمون‌های فراگیری سنجش مهارت یا توانی است که آزمودنی تا زمان انجام آزمون به دست آورده است؛ به‌دیگرسخن، هدف آزمون فراگیری، تعیین توانایی آزمودنی در زمانی است که آزمون برگزار می‌شود، درحالی‌که هدف آزمون پیش‌بینی دستیابی به شواهدی برای پیشگویی توان یا موفقیت آزمودنی در زمان معینی پس از اجرای آزمون است. آزمون‌های فراگیری شامل آزمون‌های پیشرفت یا دستاورد، مهارت و دانش یا معلومات می‌شود.

### ۲.۱.۲.۱. آزمون پیشرفت

این آزمون برای سنجیدن میزان یادگیری دانشجو از مجموعه خاصی از مطالب آموزشی طرح می‌شوند و برپایه برنامه آموزشی ویژه‌ای تدوین می‌شود. این آزمون‌ها در محیط‌های آموزشی نقش حیاتی ایفا می‌کنند؛ مانند آزمون‌های میان‌ترم و پایان‌ترم (Farhady et al, 2014).

### ۲.۲.۱.۲. آزمون‌های مهارت

این نوع آزمون برای سنجیدن توانایی کلی زبان فراگیران به کار می‌روند. آزمون‌های مهارت برای سنجیدن میزان دانشی است که فراگیر در طی آموزش زبان کسب کرده است. میزان توانایی فراگیر در بخش‌های زبان و درجه‌ای که شخص قادر به کاربرد عملی از دانش زبانی‌اش است، طراحی می‌شوند (Bachman, 1995)؛ مثل آزمون انگلیسی تافل.

### ۲.۲.۱.۳. آزمون دانش

آزمون دانش، میزان آشنایی آزمودنی با شیوه اندیشه و آداب و رسوم و عادت‌های اهل زبان را می‌سنجد. آزمون‌های معلومات در مواقعی به کار می‌روند که محیط آموزش، زبانی غیر از زبان مادری فراگیران است. در این‌گونه موارد، زبان دوم به‌عنوان زبان آزمون‌ها برای سنجش معلومات داوطلبان، در مواردی غیر از خود زبان استفاده می‌شود (Brown, 2010)؛ به‌عنوان‌مثال، آزمون فیزیک یا روان‌شناسی که به زبان انگلیسی نوشته شده، به‌عنوان آزمون معلومات به حساب می‌آید.

## ۲.۲. پایایی آزمون

اولین گام در طراحی یک آزمون استاندارد این است که بتوانیم به‌وسیله آن، توانایی آزمون‌دهنده‌ها را به‌درستی اندازه‌گیری کنیم. پایایی به معنی درستی و همسانی نتایج است؛ آزمونی که در تکرار آن، نتایج همسانی به دست

آید، دارای پایایی لازم است (Jafarpur, 1997)؛ البته به‌طور قطع ابزاری یا آزمونی وجود ندارد که بدون خطا باشد و این یک حقیقت ساده است که در همهٔ آزمون‌ها، هنگام اندازه‌گیری و سنجش، میزانی از تناقض دیده می‌شود. بنابراین، اگر زبان‌آموزی یک آزمون خوب را برای بار دوم بدهد، احتمال دارد که نمرهٔ متفاوتی بگیرد (Douglas, 2014). پایایی یعنی تعیین خطای اندازه‌گیری‌ای که به‌وسیلهٔ آن می‌توان مشخص کرد نمرهٔ آزمودنی تا چه اندازه تحت تأثیر عوامل نامربوط و تصادفی<sup>۱</sup> قابل نوسان است (Bachman, 1995)؛ به‌بیانی دیگر، پایایی یک آزمون نشان‌گر آن است که تا چه اندازه تفاوت‌های فردی، واقعی یا آلوده به خطا هستند (Traub, 1994). میزان پایایی آزمون به ما نشان می‌دهد که چه مقدار از کل واریانس یک نمره نتیجهٔ خطاهای تصادفی و چه مقدار از آن، نشانهٔ توانایی یا دانش آزمودنی است (Jafarpur, 1997). روش‌های متعددی برای محاسبهٔ پایایی نمره‌ها وجود دارد که عبارتند از: پایایی بازآزمایی<sup>۲</sup>، پایایی هم‌تا<sup>۳</sup>، پایایی دونیمه‌کردن<sup>۴</sup>، پایایی کوردریچاردسون<sup>۵</sup>، پایایی نمره‌گذاران<sup>۶</sup>، پایایی آزمون‌های سرعت و پایایی براساس خطای استاندارد اندازه‌گیری<sup>۷</sup> (Jafarpur, 1997).

## ۲.۱. روش پایایی کوردر - ریچاردسون (KR-21 Method)

در روش کوردر<sup>۸</sup>، ریچاردسون<sup>۹</sup> آزمون تنها یک بار اجرا می‌شود. این روش نسبت به روش‌های دیگر مزیتی دارد. برخلاف روش‌های بازآزمایی، دونیمه‌کردن و موازی که ممکن است در اجرا خطاهایی وجود داشته باشند. این روش از یک فرمول کاملاً مشخص ساده آماری استفاده می‌کند و احتمال خطای ارزیابی را کاهش می‌دهد (Farhady et al, 2014). طراحان این فرمول که دو نفر به نام‌های کوردر و ریچاردسون، دو آماردان معروف بودند که برای بررسی همسانی درونی آزمون و تعیین پایایی آن دو فرمول را مورد استفاده قرار دادند (Traub, 1994؛ Glen, 2016). برای محاسبهٔ پایایی با این روش، به دو فاکتور میانگین نمرات و واریانس نیاز داریم. در فرمول KR-21 همبستگی بین نمرات مهم است (Farhady et al, 2014؛ Traub, 1994؛ Glen, 2016). محاسبهٔ پایایی مبتنی بر واریانس ماده‌هاست و فرض بر این است که ماده‌ها از اهمیت یکسانی برخوردار هستند و هیچ‌گونه وابستگی به هم ندارند. اگر ماده‌ها به نسبت مساوی، امتیاز نداشته باشند، پایایی به‌دست آمده، کمتر

1. Random

2. Test- Retest Reliability

3. Parallel- Form Reliability

4. Split- Half Reliability

5. Kuder- Richardson Reliability

6. Scorer Reliability

7. Standard Error of Measurement

8. Kuder

9. Richardson

از واقعیت است و اگر ماده‌ها واحدهای مستقل نباشند، پایایی به‌دست‌آمده، بیشتر از واقعیت است. نتیجه به‌دست‌آمده نشان می‌دهد پاسخ نهایی ضریب همبستگی، عددی بین صفر و یک است (Dorneyi, 2011). فرمول محاسباتی KR-21 به شکل زیر است:

$$(KR - 21)r = \left[ \frac{K}{K-1} \right] \left[ 1 - \frac{\bar{X}(K-\bar{X})}{KV} \right] =$$

K= تعداد سؤال‌های آزمون

$\bar{X}$ = میانگین نمره‌های آزمون

V= واریانس

### ۲.۲.۲. ضریب کاپای کوهن<sup>۱</sup>

ضریب کاپای کوهن یک معیار آماری برای به‌دست‌آوردن میزان پایایی نمرات بین ارزیاب‌های آزمون است. این ضریب در اصل معیاری برای توافق بین دو ارزیاب است که اولین بار در سال ۱۹۶۰ توسط جیکوب کوهن<sup>۲</sup> در مجله سنجش آموزش و روان‌شناسی منتشر شد. این ضریب شاخص مقدار بین صفر و یک دارد. صفر بیانگر عدم توافق بین دو ارزیاب و یک بیانگر توافق کامل بین دو ارزیاب در نمره‌دهی است. در اصل مقدار بالای ۰/۹ ضریب کاپای کوهن، بیانگر بالاترین میزان توافق بین ارزیابان است. مقدار مناسب و قابل قبول ضریب پایایی کوهن، رقمی بالای ۰/۶ است. فرمول زیر نحوه محاسبه ضریب کاپای کوهن را نشان می‌دهد (Carletta, 1996).

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

### ۲.۳. روایی آزمون

یکی از مهم‌ترین ویژگی‌های یک آزمون استاندارد، داشتن روایی است. روایی<sup>۳</sup> به درجه‌ای از کاربرد ویژه آزمون برای اندازه‌گیری آنچه را که می‌خواهیم بسنجد، اشاره می‌کند؛ به عبارت دیگر آزمونی روایی دارد که فقط آنچه را می‌خواهیم، بسنجد و نه چیز دیگری را. اگر آزمون واژگان که برای سنجش توانایی واژگانی آزمودنی‌ها طراحی شده است، توانایی دستوری آزمودنی‌ها را بسنجد، دارای روایی آزمون واژگان نیست؛ زیرا در ارتباط با کاربرد ویژه آن به کار نرفته است (Ghonsooly, 2003). بنابر آنچه گفته شد، آزمونی دارای روایی است که برای

1. Cohen Kappa Coefficient

2. Jacob Cohen

3. Validity

اندازه‌گیری آنچه مورد نظر است، کافی و مناسب باشد. روایی یک امر نسبی است. روایی به انواع مختلفی تقسیم می‌شود که عبارتند از: روایی محتوایی<sup>۱</sup>، روایی صوری<sup>۲</sup>، ملاکی (سنجه‌ای)<sup>۳</sup>، روایی سازه‌ای<sup>۴</sup> (Dornyei, 2011؛ Seif, 2011)، در بخش زیر به تعریف روایی محتوایی و صوری خواهیم پرداخت.

### ۲.۳.۱. روایی محتوایی

اینکه آزمون تا چه اندازه نمونه‌هایی از هدف‌های رفتاری<sup>۵</sup> مورد سنجش را در بر می‌گیرد، به روایی محتوایی آن مربوط می‌شود؛ یعنی محتوای آزمون تا چه اندازه مطالب مورد اندازه‌گیری را در بر دارد. برای تعیین موجودیت روایی محتوایی، باید محتوای آزمون بازبینی شود و میزان روایی آن با توجه به هدف آزمون، مشخص شود. روایی محتوایی را فقط از راه دآوری ذهنی تعیین می‌کنند. هیچ روش عینی برای تعیین این جنبه از روایی وجود ندارد (Jafarpur, 1997).

### ۲.۳.۲. روایی صوری<sup>۶</sup>

روایی صوری (ظاهری) به برداشت‌ها یا واکنش‌های آزمودنی‌ها، مدرسان، ناظران اجرای آزمون<sup>۷</sup>، مدیران برنامه آموزشی و متخصصان آزمون‌شناسی مربوط می‌شود. به این معنی که آزمون مورد بحث تا چه اندازه از نظر این افراد بهینه است. روایی صوری از اهمیت ویژه‌ای برخوردار است. اگر شکل ظاهری و نوع ماده‌های آزمونی، به نظر آزمودنی‌ها بهینه نباشد، در کارکرد آنها تأثیر منفی خواهد داشت و هدف آزمون‌گر از انجام آزمون که تعیین توانایی حقیقی اوست، برآورده نخواهد شد. روایی صوری شرطی لازم است؛ اما کافی نیست (Jafarpur, 1997).

### ۲.۳.۳. روایی سنجه‌ای<sup>۸</sup>

منظور از روایی سنجه‌ای، تعیین همبستگی بین آزمون و سنجه‌ای دیگر است. روایی سنجه‌ای بر دو نوع است: همزمان<sup>۹</sup> و پیشگو<sup>۱۰</sup>. در بررسی روایی سنجه‌ای همزمان، کارکرد آزمودنی در آزمون با سنجه‌ای دیگر مقایسه می‌شود که تقریباً همزمان برگزار شده است. در بررسی روایی سنجه‌ای پیشگو، کارکرد آزمودنی در آزمون مورد

1. Content- Related Validity

2. Face Validity

3. Criterion- Related Validity

4. Construct Validity

5. Behavioral Objectives

6. Face Validity

7. Proctors

8. Criterion- Related Validity

9. Concurrent

10. Predictive

بررسی با نتایج آزمونی مقایسه می‌شود که با فاصله زمانی برگزار شده و سنجه تلقی می‌شود. اعتبار روایی سنجه‌ای هر آزمون، به کیفیت سنجه‌ای بستگی دارد که آزمون با آن مقایسه می‌شود (Jafarpurm, 1997).

### ۴.۳.۲. روایی سازه‌ای<sup>۱</sup>

روایی سازه‌ای به این نکته توجه دارد که آزمون تهیه‌شده تا چه اندازه ویژگی یا توانایی یا سازه ویژه‌ای را اندازه‌گیری می‌کند. اندازه‌گیری بسندگی در یک زبان، علاقه به مطالعه و نمونه‌هایی از این دست، سازه دانسته می‌شوند. برای اینکه به آزمونی روایی سازه‌ای ببخشیم، لازم است نخست یافته‌های گسترده‌ای از منابع گوناگونی گردآوری شوند تا امکان داشته باشد سازه مورد نظر در قالب چهارچوبی تعریف شود که رابطه آن با دیگر سازه‌ها مشخص شده باشد. سپس آزمون چنان تهیه شود که بتواند آن سازه را اندازه‌گیری کند. آن‌گاه با توجه به پیش‌بینی‌های فرضیه باید نمره‌های ناشی از آزمون را با رابطه‌های مشخص‌شده در تعریف سازه از راه روش‌های بسیار پیچیده آماری؛ مانند واکاوی عوامل<sup>۲</sup> و روایی همگرایی-افتراقی<sup>۳</sup> مورد بررسی قرار داد، نت روایی سازه‌ای آزمون تأیید شود (Jafarpur, 1997)

### ۴.۲. تحلیل سؤال

در تحلیل آزمون به بررسی پاسخ‌های آزمودنی‌ها درباره تک‌تک پرسش‌های آزمون می‌پردازیم. در تحلیل سؤال، دو عامل ضریب دشواری و ضریب تشخیص مورد واکاوی قرار می‌گیرند.

### ۱.۴.۲. ضریب دشواری<sup>۴</sup>

ضریب دشواری سؤال به ما نشان می‌دهد که چند نفر از زبان‌آموزان می‌توانند یک سؤال را به‌خوبی پاسخ بدهند و برای همین برخی معتقدند که درجه دشواری باید به درجه سادگی تغییر نام دهد؛ زیرا ما به درصدی از زبان‌آموزان که پاسخ صحیح داده‌اند، اشاره می‌کنیم. بنابه تعریف، درصد کل آزمون‌شوندگانی که به یک سؤال جواب درست می‌دهند، ضریب دشواری آن سؤال است که با حرف P نشان داده می‌شود. اگر در تحلیل یک سؤال کلیه افراد یا کلیه برگه‌های امتحانی دخالت داشته باشند، برای محاسبه ضریب دشواری آن سؤال کافی است که تعداد کل افرادی که به آن سؤال جواب درست داده‌اند، بر تعداد کل آزمون‌شوندگان تقسیم کنیم و نتیجه را در ۱۰۰ ضرب نماییم. رقم حاصل ضریب دشواری سؤال است. مطابق فرمول زیر:

$$P=100 \frac{R}{T}$$

<sup>1</sup> Construct Validity

<sup>2</sup> Factor Analysis

<sup>3</sup> Convergent- Discriminant Analysis

<sup>4</sup> Item Difficulty

در این فرمول R معرف تعداد کسانی است که به سؤال پاسخ صحیح داده و T نشان دهنده تعداد کل آزمون‌شوندگان است (Seif, 2011). با تحلیل ضریب دشواری، بازه قابل قبولی که نشان بدهد سؤالات ما نه خیلی سخت و نه خیلی ساده است، به دست می‌آید. برای رسیدن به این مطلوب از فرمول ضریب دشواری استفاده می‌شود. اگر نتیجه به دست آمده بیشتر از ۸۰ درصد باشد، نشانگر ساده بودن سؤال است. نتیجه زیر ۳۰ درصد نیز بیانگر دشواری آن است و در نهایت نتیجه بین ۳۰ تا ۸۰ درصد، قابل قبول است (Farhady et al, 2014).

### ۲.۴.۲. ضریب تمیز<sup>۱</sup>

برخلاف ضریب دشواری که میزان آسان بودن یا دشوار بودن یک سؤال را برای گروه آزمودنی‌ها نشان می‌دهد، ضریب تمیز که با حرف D نشان داده می‌شود، قدرت سؤال را در تمایز گذاری یا تشخیص بین گروه قوی و گروه ضعیف آزمودنی‌ها مشخص می‌کند؛ یعنی معلوم می‌کند که سؤال تا چه اندازه می‌تواند گروه قوی را از گروه ضعیف جدا سازد. برای محاسبه ضریب تمیز یک سؤال از فرمول زیر استفاده می‌شود (سیف، ۲۰۱۱).

فرمول محاسبه ضریب تمیز سؤالات آزمون:

$$D = \frac{\text{انتخاب‌های درست گروه پایین} - \text{انتخاب‌های درست گروه بالا}}{\text{تعداد افراد یک گروه (پایین یا بالا)}}$$

در تحلیل ضریب تمیز، عدد به دست آمده اگر ۱۰ درصد یا کمتر از آن باشد، درجه تشخیص سؤال ضعیف و در نتیجه غیر قابل قبول است. اگر نتیجه از ۱۵ درصد به بالا باشد، نتیجه قابل قبول است. بنابراین سؤال مورد نظر توانسته زبان آموز قوی را از ضعیف تشخیص بدهد و اگر نتیجه بین ۱۰ تا ۱۵ درصد باشد، ضریب تمیز مرزی است. اگر عدد به دست آمده صفر یا منفی باشد، به این معناست که در آن سؤال گروه قوی از گروه ضعیف عملکرد پایین تری داشته است. این گونه سؤال‌ها دارای معایب اساسی هستند که باید به کلی کنار گذاشته شوند یا در آنها تجدید نظر اساسی صورت پذیرد.

### ۳. پیشینه پژوهش

آزمون‌شناسی زبان، علم نوپایی است و از آغاز پیدایش خود، همواره شاهد دگرگونی‌های سریع و نوین بوده است؛ اما به حق می‌توان ادعا کرد که در چند دهه اخیر، این علم آن‌چنان به سرعت دگرگون شده است که تغییرات را می‌توان روز به روز مشاهده کرد؛ طوری که مشاهده می‌شود از گذشته تا به امروز در حوزه آزمون‌سازی زبان و

1. Item Discrimination

ارزشیابی درسی، پژوهش‌ها و تحقیقات زیادی در داخل و خارج از ایران انجام شده و یا در حال انجام است. در ادامه تعدادی از پژوهش‌هایی که در این حوزه انجام شده‌اند، بررسی می‌شوند.

Ghonsooly (۲۰۱۰) نمونه‌ای از آزمون بسندگی در زبان فارسی را طراحی کرده است. او در این پژوهش، ضمن ساخت یک نمونه آزمون بسندگی زبان فارسی با تأکید بر آزمون شنیدن، خواندن و دستور، روایی و پایایی آزمون را سنجیده است. برای روایی‌سنجی آزمون از فرآیند دو مرحله‌ای شامل تحلیل داده‌های بازنگری سه دانشجوی دوره دکتری زبان و ادبیات فارسی و بهره‌گیری از نظریه بحث در سنجش روایی بهره گرفته است. پایایی آزمون نیز از طریق آلفای کرونباخ محاسبه شده است. نتایج این آزمون نشان داد که دانشجویان خارجی در بخش خواندن و درک مفاهیم بهترین عملکرد را داشتند و بخش دستور دشوارترین بخش برای آنها بود و پایایی آزمون ۰/۷۰ تعیین شد.

Sahraei and Jalili (۲۰۱۱) در پژوهشی مبانی تهیه و تدوین آزمون مهارتی فارسی (آمفا) را تحلیل و بررسی کرده‌اند. در این تحقیق، ابتدا اهمیت و ضرورت آزمون مهارتی (بسندگی) که با هدف ارزیابی پیشرفت زبانی برگزار می‌شود، تبیین شد. سپس نمونه‌هایی از آزمون‌های مهارتی در زبان‌های انگلیسی، عربی و ژاپنی معرفی شد. در نتیجه این پژوهش، قالب کلی یک آزمون برای سنجش و ارزیابی مهارت زبانی فارسی‌آموزان خارجی ارائه شد.

Abbasi (۲۰۱۴) یک نمونه آزمون بسندگی معیار زبان فارسی را همانند دو آزمون تافل و آیلتس، تهیه و تنظیم کرده است که فارسی‌آموز را در همه مهارت‌های زبانی می‌سنجد و سطح دانش کلی زبانی او را نشان می‌دهد. در این پژوهش، ارزیابی روایی و پایایی آزمون براساس ضریب همبستگی پیرسون و ضریب آلفای کرونباخ انجام شده است. نتیجه این پژوهش، ارائه نمونه‌ای از آزمون‌های بسندگی برای زبان فارسی است که طبق نظر نویسنده، می‌تواند ملاکی برای ثبت‌نام غیرفارسی‌زبانان در دانشگاه‌ها و مراکز آموزشی فارسی زبان در ایران و خارج از ایران باشد.

Zandi et al (۲۰۱۱) نمونه‌ای از آزمون بسندگی زبان فارسی برپایه چهار مهارت زبانی درک شنیداری، درک خوانداری، گفت‌وگو و نگارش طراحی کرده‌اند. جامعه آماری این پژوهش، ۱۳۰ نفر از فارسی‌آموزان مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام‌خمينی (ره) بودند. برای محاسبه پایایی از فرمول آلفای کرونباخ استفاده شده است. میزان ضریب آلفای کرونباخ ۰/۷ به دست آمد که نشان‌دهنده پایابودن نتایج آزمون است. روایی آزمون براساس سنجش روایی محتوایی و ظاهری و همین‌طور سنجش روایی سازه‌ای بررسی شده است. پس از بررسی آزمون طراحی‌شده، نتایج آماری نشان‌دهنده این موضوع بوده است که این آزمون، ضریب دشواری و ضریب تشخیص بهینه‌ای دارد و دارای پایایی و روایی قابل قبولی نیز هست.

Sahraei and Rezapur (۲۰۱۱) در پژوهشی به بررسی ارزیابی و پایایی آزمون تعیین سطح زبان فارسی برای غیر فارسی‌زبانان بنیاد سعدی پرداخته‌اند. آزمون مورد نظر دارای سه بخش اصلی دانش زبانی (دستور، واژه و خواندن)، نگارش و صحبت کردن است. آنها برای محاسبه پایایی از روش ضریب کرونباخ و برای بررسی روایی از روایی سازه‌ای بهره برده‌اند. نتایج حاصل از این پژوهش نشان می‌دهد که پایایی آزمون ۰/۷۰ محاسبه شد که قابل قبول بود؛ اما آزمون از روایی مناسب و قابل قبولی برخوردار نبود و زبان‌آموزان در بخش مهارت‌های تولیدی ضعیف عمل کرده بودند و این موضوع بیانگر این است که در تدریس آنها، فقط به دانش محض زبانی پرداخته شده است و به دانش روندی توجه نشده است.

Samaee and Kamyabi Gol (۲۰۲۰) به ارزیابی یکی از آزمون‌های جامع مرکز آموزش زبان و معارف اسلامی (مدرسه المهدی (ع)) جامعه المصطفی (ص) پرداخته‌اند. آزمونی که در واقع نشان می‌دهد که فارسی‌آموز پس از طی دوره جامع آموزش فارسی از نظر مهارت زبانی در چه سطحی قرار دارد. در پژوهش آنها، بررسی روایی از طریق تحلیل سؤالی انجام شد و نشان داد که سؤالات آزمون از سطح دشواری مناسبی برخوردار هستند؛ اما ضریب تشخیص بالایی ندارند و نیاز است به این قسمت بیشتر توجه شود. آنها پایایی این آزمون را نیز به روش کودر- ریچاردسون ۲۱ سنجیدند و مشخص شد که آزمون از پایایی متوسطی (۰,۷۳) برخوردار بود.

Kim and Shin (۲۰۰۶) آزمون تعیین سطحی برای ارزیابی زبان‌آموزان انگلیسی طراحی کردند. سپس ۲۹ شرکت‌کننده از سه سطح مشخص مبتدی، متوسط و پیشرفته در آزمون شرکت کردند. در آخر نتایج سطح‌بندی آزمون تعیین سطح با سطح اصلی هر زبان‌آموز مقایسه شد. طبق این مطالعه، نتایج آزمون طراحی شده با سطح اصلی هر زبان‌آموز مطابقت داشت و به خوبی شرکت‌کننده‌ها را در دسته‌های همگن گروه‌بندی کرد. با توجه به اینکه تعداد شرکت‌کننده‌ها در این مطالعه پایین است، نتایج مطالعه به بررسی‌های تکمیلی نیاز دارد.

Kusiak-Pisowacka (۲۰۱۵) با استفاده از نرم‌افزار ایتمن به طراحی یک آزمون تعیین سطح کلی برای زبان‌های خارجی پرداختند. این آزمون ابتدا برای زبان‌آموزان هلندی برگزار شد و سپس نتایج آن با کمک روش‌های آماری مورد بررسی قرار گرفت. بنابر نتایج مطالعه، این تست از اعتبار بالایی برخوردار است؛ اما با توجه به اینکه تنها برای زبان‌آموزان هلندی بررسی شده است، نیاز به پژوهش‌های تعقیبی دارد.

Akhidenor-Bamidele (۲۰۱۹) به بررسی تأثیر استفاده تکنولوژی و آزمون‌های تعیین سطح آنلاین در جهت مشخص کردن سطح زبان‌آموزان و ثبت‌نام آنها در دوره آموزش زبان انگلیسی پرداخته است. تأثیر تکنولوژی در دنیای امروز و به خصوص بر فراگیران زبان، آن را به یک ابزار یادگیری زبان تبدیل کرده است. این پدیده می‌تواند به علت نیاز مبرم فاصله‌گرفتن از روش‌های آموزش سنتی؛ مانند تخته و گچ و روی آوردن به روش‌های مشارکتی تر نسبت به این سیستم باشد. همچنین، او در این پژوهش، عملکرد زبان‌آموزان در آزمون‌های پایان دوره را نیز تحلیل و بررسی کرده است و مشخص کرده که چه میزان آزمون‌های تعیین سطح آنلاین در کسب نتیجه در پایان دوره، مؤثر بوده است.



Nasem and Mohammad (۲۰۱۸) پژوهشی درباره بررسی اثرگذاری آزمون‌های تعیین سطح آنلاین دانشگاه آکسفورد از دیدگاه مدرسان مؤسسه زبان انگلیسی در دانشگاه شاه عبدالعزیز انجام دادند. آزمون‌های تعیین سطح زبان انگلیسی یک جز بسیار مهم از برنامه سالانه آمادگی PYPS است که زبان‌آموزان را در سطوح مطابق با میزان مهارتشان طبقه‌بندی می‌کند. در این تحقیق سه مرحله جمع‌آوری اطلاعات تبیین شد: (۱) مصاحبه‌های رودررو (۲) پرسشنامه و (۳) مصاحبه‌های تلفنی. مدرسانی که با این آزمون آشنا بودند، اشاره کردند که این آزمون تعیین سطح، آزمون جامع بوده و زبان‌آموزان را به‌درستی سطح‌بندی کرده است.

Kadwa and Sheik (۲۰۲۱) به بررسی آزمون تعیین سطح زبان انگلیسی آکسفورد در یکی از دانشگاه‌های عربستان سعودی پرداخته‌اند. در این پژوهش زبان‌آموزان با دو روش کمی (آزمون تعیین سطح) و کیفی (مصاحبه و آزمون شفاهی) آزموده شده‌اند و سپس نتایج با هم مقایسه شده است. طبق نتایج به‌دست‌آمده، آزمون تعیین سطح دانشگاه آکسفورد اعتبار بالایی داشت؛ اما سطح زبان‌آموزانی را که در آزمون نمره کمتر از ۳۰٪ گرفتند، با دقت مشخص نمی‌کرد. با توجه به آنکه دسته‌بندی اصولی زبان‌آموزان ضعیف‌تر، از اهمیت بیشتری نیز برخوردار است، طراحی آزمون‌های بهتر را پیشنهاد داده‌اند.

با توجه به مطالعات انجام‌شده، این سؤالات درباره اعتبارسنجی پژوهش مطرح است که آیا ضریب دشواری و ضریب تمیز آزمون‌های کتبی تعیین سطح در سه سطح مبتدی، میانی و پیشرفته، قابل قبول است؟ آیا آزمون‌های تعیین سطح سه‌گانه این پژوهش (مبتدی، میانی، پیشرفته) دارای پایایی و روایی لازم هستند؟

#### ۴. روش‌شناسی پژوهش

طرح پژوهش حاضر از نوع کاربردی بوده و انجام آن با استفاده از روش میدانی صورت گرفته است. به‌منظور تجزیه و تحلیل داده‌های تحقیق، از شیوه توصیفی-تحلیلی بهره برده شده است. نخست مطالعات لازم در حوزه آزمون‌های تعیین سطح داخلی و خارجی صورت گرفت و سپس سه آزمون در سه سطح مبتدی، میانی و پیشرفته براساس منابع آموزشی مراکز جامعه‌المصطفی<sup>(ص)</sup> العالمیه طراحی شد. آزمون براساس دوره ۱۳ جلدی پارسا<sup>۱</sup> طراحی شده است. داده‌های پژوهش از طریق برگزاری آزمون در قالب آزمون تعیین سطح در دو مرکز از مراکز آموزش زبان فارسی جامعه‌المصطفی<sup>(ص)</sup> جمع‌آوری شد.

#### ۴.۱. شرکت‌کنندگان

جامعه آماری این پژوهش، فارسی‌آموزان مراکز آموزش زبان فارسی و معارف اسلامی جامعه‌المصطفی<sup>(ص)</sup> در ایران و ترکیه و همچنین مدرسان این مراکز هستند. کل جامعه آماری فارسی‌آموزان شرکت‌کننده در دو مرکز، ۱۳۵

۱. مجموعه پارسا: کتاب‌های آموزش زبان فارسی جامعه‌المصطفی<sup>(ص)</sup> (با نام‌های زبان فارسی عمومی، خواندن، نوشتن، شنیدن).

نفر بود. در مرکز آموزش زبان فارسی نمایندگی مشهد (تمهیدیه خواهران)، ۹۰ نفر فارسی‌آموز خانم اهل کشورهای هند و پاکستان و آذربایجان حضور داشتند. در ترکیه نمایندگی المصطفی (ص) در شهر استانبول، ۲۵ نفر خانم و ۱۵ آقا و در مجموع ۴۵ نفر اهل ترکیه حضور داشتند. بازه سنی فارسی‌آموزان در ایران بین ۱۸ تا ۲۵ سال با میانگین سنی ۲۱/۸۵ با انحراف از معیار ۱/۶۲۱ ثبت شد. فارسی‌آموزان در ترکیه در گروه سنی بین ۱۳ تا ۴۷ سال قرار داشتند که میانگین سنی آنها ۲۹/۴۹ با انحراف از معیار ۹/۱۵۹ محاسبه شد. تمامی داده‌ها در سال ۱۴۰۰ جمع‌آوری شد.

از بین مدرّسان جامعه‌المصطفی (ص)، تعداد ۲۰ مدرس در بخش مصاحبه کتبی برای بررسی روایی صوری و محتوایی آزمون شرکت داشتند. از این تعداد، تعداد پنج نفر از مدرّسان مرد و ۱۵ نفر را مدرّسان خانم تشکیل دادند. بازه سنی مدرّسان بین ۲۵ تا ۵۸ سال با درجه انحراف معیار ۷/۹۲ بود. با سابقه‌ترین مدرّس ۱۸ سال تجربه تدریس داشت و کم‌تجربه‌ترین مدرس دارای چهار سال سابقه تدریس بود. تعداد ۱۳ نفر معادل ۶۵ درصد از مدرّسان فارغ‌التحصیل رشته آزا و زبان و ادبیات فارسی بودند و تعداد هفت نفر معادل ۳۵ درصد از مدرّسان رشته‌هایی غیرمرتبط با آموزش زبان فارسی بودند. این مدرّسان دوره‌های دانش‌افزایی روش‌های تدریس را گذرانده و امتیاز لازم برای ورود به حوزه تدریس در جامعه‌المصطفی (ص) را کسب کرده بودند.

#### ۲.۴. ابزارهای پژوهش

اولین و اصلی‌ترین ابزار پژوهش که از آن بهره برده شده است، آزمون تعیین سطح زبان فارسی طراحی شده در این پژوهش است. دومین ابزار مصاحبه کتبی نیمه ساخت‌یافته با مدرّسان محترم جامعه‌المصطفی (ص) است که در راستای بررسی روایی محتوایی آزمون، نظرشان درباره سؤالات آزمون و نمرات کسب‌شده توسط فارسی‌آموزان پرسیده شد.

#### ۵. ارائه و واکاوی پژوهش، محتوا و ساختار آزمون

آزمون تعیین سطح طراحی شده شامل دو بخش کتبی و شفاهی (به صورت مصاحبه) است. آزمون کتبی شامل ۴۰ سوال از بخش‌های واژگان، دستور زبان، خواندن، شنیدن و همچنین مهارت نوشتار با دو تمرین نوشتاری کوتاه و بلند می‌شود. آزمون شفاهی شامل ۱۰ سؤال درباره سنجش مهارت گفتاری، تعاملی و ارتباطی فارسی‌آموز است. این آزمون در سه سطح مبتدی، میانی و پیشرفته و در سه مرحله برگزار شده است. دامنه واژگانی سطوح زبانی مختلف براساس میزان واژگان به کاررفته در مجموعه کتاب‌های پارسا است. نوع سؤالات و مدت زمان لازم برای هر قسمت در جدول شماره ۱ نمایش داده شده است.

جدول ۱. ساختار آزمون تعیین سطح طراحی شده

گفت‌وگو ۱۰ دقیقه	مقولهٔ واژگانی و مقولهٔ دستور ۱۵ دقیقه	نوشتن ۴۰ دقیقه	خواندن و درک مطلب ۲۰ دقیقه	شنیدن ۱۵ دقیقه	ماده‌های آزمون سطوح زبانی
<ul style="list-style-type: none"> <li>● معرفی خود</li> <li>● بیان ملیت، شغل، تحصیلات، سن</li> </ul>	۱۰ سؤال تصویری برای واژگان ۱۰ سؤال دستور	یک موضوع انشا: انشای توضیحی با ۵۰ کلمه	دو متن کوتاه و متوسط ۱۰ سؤال (چندگزینه‌ای)	دوگفت‌وگوی کوتاه و متوسط ۱۰ سؤال (جای خالی چندگزینه‌ای)	سطح مبتدی
<ul style="list-style-type: none"> <li>● معرفی یک دوست با بیان ویژگی‌های اخلاقی و شخصیتی.</li> <li>● بحث دربارهٔ یک موضوع خاص مانند گردشگری</li> </ul>	۱۰ سؤال واژگان ۱۰ سؤال دستور	دو موضوع انشا: انشای اول توضیحی ۱۰۰ کلمه انشای دوم، نوشتن نامه‌ای به دوست	دو متن متوسط و بلند ۱۰ سؤال (چندگزینه‌ای)	دوگفت‌وگوی متوسط و بلند ۱۰ سؤال (جای خالی چند گزینه‌ای، درست و نادرست)	سطح میانی
<ul style="list-style-type: none"> <li>● توصیف یک شخصیت علمی</li> <li>● سؤالات استنباطی، مقایسه‌ای، نتیجه‌گیری</li> </ul>	۱۰ سؤال واژگان ۱۰ سؤال دستور	دو موضوع انشا: انشای اول، با موضوع توصیفی ۱۵۰ کلمه انشای دوم، نوشتن نامه‌ای اداری به رئیس دانشگاه	دو متن بلند ۱۰ سؤال (چندگزینه‌ای)	بیان خاطره/ مصاحبه با موضوع اجتماعی ۱۰ سؤال (جای خالی چندگزینه‌ای، درست و نادرست، چندگزینه‌ای استنباطی)	سطح پیشرفته

بنابر جدول (۱)، در بخش شفاهی آزمون از هر فارسی‌آموز به صورت انفرادی خواسته شد تا دربارهٔ موضوعی خاص مانند گردشگری صحبت کنند. در این آزمون شفاهی از شیوهٔ ارزیابی مستقیم استفاده شد. دو آزمون‌گر مهارت گفتار، آزمون‌دهی‌ها را در گفت‌وگویی که صورت پذیرفت، به صورت منفصل<sup>۱</sup> و تحلیلی<sup>۲</sup> و با توجه به اجزای<sup>۳</sup> گفتار ارزشیابی کردند. اجزای گفتار شامل تلفظ صحیح، ساختار، شامل تطبیق فعل و فاعل و بیان صحیح زمان افعال، کاربرد صحیح واژگان، درک کلام، لحن و آهنگ کلام و رسایی و روانی بیان بود که در مجموع ۲۰ نمره داشت و هر کدام از آزمون‌گرها به صورت جداگانه نمره‌ای را برای آزمون‌دهی ثبت کردند.

1. Discrete  
 2. Analytic  
 3. Components

## ۵. ۱. برگزاری آزمون

پس از طراحی آزمون تعیین سطح، پرسش‌های آزمون بررسی شد. برای بررسی روایی محتوایی، نمونه‌سؤالات طراحی شده در هر سطح، در اختیار مدرسانی از همان سطح قرار گرفت و بازخوردهای مدرسان به صورت مصاحبه کتبی نیمه‌ساخت‌یافته<sup>۱</sup> مورد بررسی قرار گرفت. در این نوع مصاحبه پرسش‌های از قبل تعیین شده به همراه برخی پرسش‌های آزاد و فضای خالی برای توضیح بیشتر در اختیار مدرسان قرار گرفت و بازخوردهای ایشان بررسی شد. آزمون‌ها در سه مرحله و در دو مرکز آموزش زبان جامعه‌المصطفی (ص)<sup>۲</sup> در ایران و ترکیه اجرا شد. شرایط آزمون طبق استانداردهای موجود در هر مرکز شامل مکانی با نور کافی، سکوت لازم، میز و صندلی راحت برای نشستن، دمای مناسب و غیره تنظیم شد (Bachman, 1995؛ Farhady et al, 2014؛ Fukcher, 2013؛ Cohen, 1994؛ Hughes, 2012). قبل از آزمون با فارسی‌آموزان صحبت شد و سعی گردید تا با برقراری ارتباط صمیمی از میزان اضطراب آنها کاسته شود (Hughes, 2012؛ Fulcher, 2013). مدت زمان آزمون ۱۰۰ دقیقه بود. بعد از اجرای آزمون و تصحیح برگه‌ها، طبق فرمول ضریب دشواری و ضریب تمیز سؤال‌ها محاسبه شد. در بحث روایی، روایی محتوایی نیز از طریق نظرسنجی از مدرسان هر سطح که از قبل بررسی شده بود، براساس نظرات مدرسان، اصلاحات واژگانی و دستوری پیشنهادی ایشان در نسخه نهایی آزمون اعمال شد. پایایی درونی آزمون نیز با فرمول کودر ریچاردسون (KR-21) اندازه‌گیری شد. در ضمن پایایی بین نمرات آزمون‌گرهای بخش شفاهی نیز با استفاده از کاپای کوهن محاسبه شد.

## ۶. تحلیل داده‌های پژوهش

در محاسبه پایایی درونی آزمون‌ها از فرمول کودر- ریچاردسون ۲۱ استفاده شد. در این روش محاسبه پایایی، اگر نتیجه به دست آمده بالای ۷۰ درصد باشد، به این معناست که آزمون از پایایی خوبی برخوردار است.

جدول ۲. نتایج پایایی آزمون‌های تعیین سطح

سطوح آزمون	تعداد آزمودنی	تعداد سؤالات	بیشینه نمره	کمینه نمره	میانگین	واریانس	انحراف معیار	نتایج پایایی KR21
مبتدی	۴۵	۴۰	۳۹	۱۸	۳۰/۶۶	۴۵/۸۶	۶/۷۷	۰/۹۴
میانی	۴۵	۴۰	۳۹	۱۴	۲۸/۵۱	۴۲/۶۱	۶/۵۲	۰/۸۳
پیشرفته	۴۵	۴۰	۳۸	۱۹	۲۸/۸۰	۳۱/۴	۵/۶۴	۰/۸۰

با توجه به اطلاعات جدول شماره (۲)، میزان پایایی آزمون سطح مبتدی، ۹۴ درصد است که پایایی عالی آزمون را در سطح مبتدی نشان می‌دهد و سؤالات از سازگاری درونی و همگنی بالایی برخوردار هستند (Traub, 1994؛ Glen, 2016؛ Dornyei, 2011). میزان پایایی در آزمون‌های سطح میانی، ۸۳ درصد و در آزمون

۱. Semi-Structured

سطح پیشرفته ۸۰ درصد به دست آمد که نشان‌دهنده پایایی خوب و قابل قبول آزمون‌هاست. پایایی کل آزمون نیز ۸۵ درصد محاسبه شد. پایایی بین نمرات آزمون‌گرها در سنجش مهارت‌های گفتار و نوشتار که با ضریب کاپای کوهن محاسبه شد، بیانگر همبستگی بین ارزیابان است، در جدول (۳) قرار گرفته است.

جدول ۳. نتایج پایایی (کاپای کوهن) نمرات ارزیابان برای مهارت‌های گفتار و نوشتار

آزمون‌ها	سطح مبتدی	سطح میانی	سطح پیشرفته
مهارت گفتاری	۰/۷۸	۰/۹۱	۰/۷۰
مهارت نوشتاری	۰/۶۹	۰/۷۰	۰/۷۱

بنابر داده‌های جدول شماره (۳)، ضریب کاپای کوهن بین نمرات مهارت گفتاری که دو آزمون‌گر به فارسی‌آموزان داده‌اند، در نمرات آزمون سطح مبتدی ۰/۷۸ محاسبه شد که قابل قبول ارزیابی می‌شود. در نمرات آزمون گفتار سطح میانی ۰/۹۱ به دست آمده که خیلی خوب است. در نمرات آزمون گفتار سطح پیشرفته، عدد ۰/۷۰ محاسبه شده است و نتیجه قابل قبول است. این رابطه درباره نمرات مهارت نوشتاری نیز بررسی و محاسبه شد. ضریب کاپای کوهن در نمرات آزمون نوشتاری سطح مبتدی عدد ۰/۶۹ است که پایایی متوسط ارزیابی می‌گردد. در نمرات نوشتاری سطح میانی عدد ۰/۷۰ و در نمرات نوشتاری سطح پیشرفته عدد ۰/۷۱ به دست آمده است و این رابطه قابل قبول ارزیابی می‌شود.

#### ۶.۱. تحلیل روایی صوری و محتوایی آزمون‌ها از نظر مدرّسان جامعه‌المصطفی<sup>(ص)</sup>

روایی صوری به برداشتها یا واکنش‌های آزمودنی‌ها، مدرّسان، ناظران اجرای آزمون، مدیران برنامه آموزشی و متخصصان آزمون‌شناسی مربوط می‌شود؛ به این معنی که تا چه اندازه آزمون مورد بحث، به نظر این افراد بهینه است (Jafarpur, 1997). اگرچه کیفیت به نظر می‌رسد، جنبه عینی ندارد، اما به دلیل نقشی که ظاهر آزمون می‌تواند در کارکرد آزمودنی‌ها داشته باشد، روایی صوری و ظاهری از اهمیت ویژه‌ای برخوردار است. اگر شکل ظاهری و نوع ماده‌های آزمون به نظر آزمودنی‌ها بهینه نباشد، این برداشت ممکن است در کارکرد آنها تأثیر منفی بگذارد و هدف آزمون‌گر از انجام آزمون، برآورده نخواهد شد. روایی صوری آزمون به‌تنهایی، روایی آزمون را تضمین نمی‌کند، شرط لازم است؛ ولی کافی نیست. بنابراین، علاوه بر روایی صوری، روایی محتوایی نیز در نظر گرفته شد. از آنجایی که هیچ روش عینی و آماری برای تعیین روایی محتوایی موجود نیست، برای دستیابی به این امر مهم و برای به‌دست آوردن روایی صوری و محتوایی آزمون‌ها، ابزار مصاحبه کتبی به کار گرفته شد تا نظر متخصصان آزمون را که همان مدرّسان مراکز جامعه‌المصطفی<sup>(ص)</sup> بودند، درباره محتوای آزمون جویا شویم. اینکه آزمون تا چه اندازه نمونه‌هایی از هدف‌های رفتاری<sup>۱</sup> مورد سنجش را در بر می‌گیرد، به روایی محتوایی آزمون

۱. Behavioral Objectives

مربوط می‌شود (Jafarpur, 1997). پس از فراخوان مدرسان برای همکاری در پژوهش، تعداد ۲۰ نفر از مدرسان جامعه‌المصطفی (ص) در این مصاحبه شرکت کردند. این مدرسان به دلیل تجربه نسبتاً بالایی که در تدریس محتوای آموزشی آن مجموعه داشتند (بین ۴-۱۸ سال)، نظرات راهبردی و مفیدی ارائه دادند. سؤالات مصاحبه و نتیجه آن در جدول (۴) آمده است.

جدول ۴. نتایج نظرسنجی مدرسان درباره آزمون‌های تعیین سطح

شماره سؤال	سؤالات مصاحبه کتبی	سطح آزمون	ضعیف	متوسط	خوب	خیلی خوب
۱	تناسب سطح دشواری واژگان آزمون‌ها با منبع آموزشی مورد نظر	مبتدی	۰	(/۱۰) ۱	(/۵۰) ۱۰	(/۴۰) ۸
		میانی	۰	(/۱۵) ۳	(/۳۵) ۷	(/۵۰) ۱۰
		پیشرفته	۰	۰	(/۴۰) ۸	(/۶۰) ۱۲
۲	تناسب سطح دشواری سؤالات ساختاری با سطح آزمودنی‌ها	مبتدی	۰	(/۱۰) ۲	(/۵۵) ۱۱	(/۳۵) ۷
		میانی	۰	(/۱۵) ۳	(/۷۰) ۱۴	(/۱۵) ۳
		پیشرفته	۰	(/۱۰) ۲	(/۶۰) ۱۲	(/۳۰) ۶
۳	کیفیت و گویا بودن تصاویر آزمون سطح مبتدی	مبتدی	۰	۰	(/۲۰) ۴	(/۸۰) ۱۶
۴	تعداد و ترتیب سؤالات آزمون‌ها	مبتدی	۰	۰	(/۶۰) ۱۲	(/۴۰) ۸
		میانی	۰	(/۱۵) ۳	(/۲۵) ۵	(/۶۰) ۱۲
		پیشرفته	۰	(/۱۰) ۲	(/۲۰) ۴	(/۷۰) ۱۴
۵	میزان زمان در نظر گرفته شده برای پاسخ‌گویی به سؤالات. (ضعیف=کم/ متوسط=تقریباً کافی/ خوب=کافی/ خیلی خوب = زیاد)	مبتدی	۲ (/۱۰)	(/۳۰) ۶	(/۶۰) ۱۲	۰
		میانی	(/۵) ۱	(/۲۵) ۵	(/۷۰) ۱۴	۰
		پیشرفته	۰	(/۲۰) ۴	(/۸۰) ۱۶	۰
۶	تعداد و طول متن‌ها در بخش مهارت خواندن	مبتدی	۰	(/۱۰) ۲	(/۳۰) ۶	(/۶۰) ۱۲
		میانی	۰	(/۱۵) ۳	(/۱۵) ۳	(/۷۰) ۱۴
		پیشرفته	۰	(/۵) ۱	(/۱۵) ۳	(/۸۰) ۱۶
۸	ارزیابی کلی محتوای آزمون	مبتدی	۰	(/۵) ۱	(/۲۰) ۴	(/۷۵) ۱۵
		میانی	۰	(/۱۰) ۱	(/۲۰) ۴	(/۷۰) ۱۴
		پیشرفته	۰	۰	(/۱۵) ۳	(/۸۵) ۱۷

براساس جدول شماره (۴)، نتایج کسب‌شده از نظرسنجی مدرّسان جامعه‌المصطفی<sup>(ص)</sup>، بیانگر میزان روایی صوری و محتوایی آزمون‌های تعیین سطح طراحی شده است. این آمار توصیفی نشان می‌دهد که محتوای آزمون‌ها تا چه اندازه مطالب مورد اندازه‌گیری را در بر داشته است. برای تعیین موجودیت روایی محتوایی، هر سه آزمون توسط مدرّسان جامعه‌المصطفی<sup>(ص)</sup> بررسی و با در نظر گرفتن اهداف آزمون‌ها به سؤال‌های مصاحبه پاسخ داده شد. کافی یا ناکافی بودن میزان زمان در نظر گرفته شده (پیشنهادی طراحان آزمون) برای پاسخ‌گویی به سؤالات، در سؤال ۵ مصاحبه بیان شد. بنابر نتایج مندرج در جدول (۴)، در سطح مبتدی تعداد ۱۲ نفر معادل ۶۰٪ از مدرّسان، مدت زمان آزمون را کافی و تعداد ۶ نفر معادل ۳۰٪ از آنها این زمان را تقریباً کافی ارزیابی کرده‌اند. حدود ۱۰٪ از مدرّسان به اندازه تعداد ۲ نفر، این مقدار زمان را ناکافی دانسته‌اند. آنها اشاره کردند «آزمودنی‌ها در بخش خواندن به حدود ۵ دقیقه بیشتر، زمان نیاز دارند. به نظر ایشان در بخش شنیدن نیز لازم است، زمان بیشتری حدود ۵ دقیقه به آزمودنی‌ها داده شود». در ارزیابی آزمون سطح میانی، تعداد ۱۴ نفر معادل ۷۰٪ از مدرّسان مدت زمان داده شده را کافی دانسته‌اند. تعداد ۵ نفر معادل ۲۵٪ این زمان را تقریباً کافی ارزیابی کردند و تعداد ۱ نفر معادل ۵٪ از مدرّسان معتقد بودند: «در بخش خواندن، زمان کم است و باید ۵ دقیقه به زمان خواندن اضافه شود». در ارزیابی سطح پیشرفته، تعداد ۱۶ نفر معادل ۸۰٪ از مدرّسان، زمان آزمون را کافی ارزیابی کردند و تعداد اندکی از آنها؛ یعنی حدود ۴ نفر معادل ۲۰٪ این زمان را تقریباً کافی دانسته‌اند. در این سطح هیچ مصاحبه‌شونده‌ای زمان را ناکافی ندانسته است. در کل سطوح گویه زیاد بودن زمان انتخاب نشده است. بنابراین، کسب این نتایج به این معناست که زمان آزمون برای برخی آزمون‌ها، نیاز به بازنگری داشت که قبل از اجرای آزمون تغییرات اعمال شد. برخی پیشنهاد‌های مدرّسان درباره جایگزینی برخی واژگان و دستور در آزمون‌ها نیز بررسی و اصلاحات اعمال شد.

## ۲.۶. تحلیل سؤالات آزمون‌ها

هدف از تحلیل سؤال‌های آزمون، بررسی تک‌تک سؤال‌ها و تعیین میزان دقت و نارسایی‌های آنهاست. در تحلیل سؤال‌های آزمون، نقاط قوت و ضعف یک آزمون و کیفیت همه سؤال‌های آن از طریق محاسبه ضریب دشواری و ضریب تمیز مشخص می‌شد. در این تحقیق، هر آزمون دارای ۴۰ سؤال کتبی است که برای به‌دست‌آوردن ضریب دشواری و ضریب تمیز تک‌تک سؤال‌ها محاسبه و بررسی شد که نتایج آن در جدول شماره (۵) قرار گرفته است. از مجموع ۴۰ سؤال آزمون سطح مبتدی، ۳ درصد (یک سؤال) ضریب دشواری کمتر از ۳۰٪ داشته و نشان‌دهنده دشواری بودن سؤال است. ضریب دشواری ۹۷ درصد از سؤالات (۳۹ سؤال) بین ۳۰-۸۰ درصد به دست آمده است که قابل قبول بودن آنها را نشان می‌دهد. این آزمون سؤال ساده و آسان نداشته است. از مجموع ۴۰ سؤال، ۵ سؤال معادل ۱۳٪ ضریب تشخیص لب مرز داشتند و قدرت تشخیص کمتر دارد. زمانی که قدرت تشخیص سؤال عددی بین ۱۰٪ تا ۱۵٪ باشد، به معنای لب مرز است. ضریب تشخیص تعداد ۳۵ سؤال معادل

۸۷٪ بالای ۱۵٪ درصد بوده است و این نشان از اینست که سؤالات آزمون قدرت تشخیص قابل قبولی داشته‌اند. هیچ سؤالی نیز در آزمون زیر ۱۰٪ که نشانگر ضعیف بودن قدرت تشخیص است، وجود نداشت.

جدول ۵. نتایج ضریب دشواری و ضریب تمیز سؤالات آزمون

سطوح آزمون	ضریب دشواری			ضریب تمیز		
	ساده	دشواری	قابل قبول	ضعیف	لب مرز	قوی
سطح مبتدی	-	سؤال ۱ (/۳)	سؤال ۳۹ (/۹۷)	-	سؤال ۵ (/۱۳)	سؤال ۳۵ (/۸۷)
سطح میانی	-	سؤال ۸ (/۱۸)	سؤال ۳۲ (/۸۲)	سؤال ۲ (/۵)	سؤال ۴ (/۱۰)	سؤال ۳۴ (/۸۵)
سطح پیشرفته	سؤال ۲ (/۵)	-	سؤال ۳۸ (/۹۵)	سؤال ۱ (/۳)	سؤال ۳ (/۵)	سؤال ۳۶ (/۹۲)

از مجموع ۴۰ سؤال سطح میانی، تعداد ۸ سؤال معادل ۱۸٪ ضریب دشواری زیر ۳۰٪ را کسب کرده‌اند و این به این معناست که ۱۸٪ از سؤالات این آزمون سطح نسبتاً دشواری داشته است. ضریب دشواری تعداد ۳۲ سؤال معادل ۸۲٪ بین ۳۰٪ تا ۸۰٪ درصد به دست آمده که نشان‌گر سطح دشواری قابل قبولی است. این آزمون سؤال ساده نداشته است. در محاسبه ضریب تمیز، از مجموع ۴۰ سؤال آزمون سطح میانی، تعداد ۳۴ سؤال معادل ۸۵٪، ضریب تشخیص بالای ۱۵٪ درصد را کسب کرده‌اند و این امر نشان‌دهنده این است که این سؤالات قدرت تشخیص قابل قبولی داشته‌اند. تعداد ۴ سؤال معادل ۱۰٪، ضریب تمیز بین ۱۰٪ تا ۱۵٪ را دریافت کرده‌اند. بنابراین، این سؤالات دارای درجه تمیز لب مرز هستند. تعداد ۲ سؤال معادل ۵٪ هم درجه تمیز زیر ۱۰٪ را کسب کرده‌اند که بیانگر قدرت تشخیص ضعیف است. پس از بررسی نتیجه، این سؤالاتی که ضریب تمیز پایینی داشتند، از آزمون نهایی حذف و با سؤال بهتری جایگزین شدند.

از مجموع ۴۰ سؤال سطح پیشرفته، تعداد ۳۸ سؤال معادل ۹۵٪، ضریب دشواری بین ۳۰٪ تا ۸۰٪ را کسب کرده‌اند و این امر بیان‌گر سطح دشواری نرمال و قابل قبولی است. تعداد ۲ سؤال معادل ۵٪ ضریب دشواری بالای ۸۰٪ را دریافت کرده‌اند که میزان سؤالات ساده را نشان می‌دهد. این آزمون سؤال دشوار نداشته است. دلیل احتمالی ساده بودن این سؤالات را می‌توان اشتراکات زبانی بین زبان فارسی و زبان اردو دانست. برخی از سازه‌های زبانی، مانند واژه‌ها یا ضرب‌المثل‌ها یا اصطلاحاتی که در دو زبان مشترک هستند، برای فارسی‌آموزان ساده به نظر می‌رسد. تعداد ۳۶ سؤال معادل ۹۲٪، ضریب تشخیص بالای ۱۵٪ درصد را کسب کرده‌اند و این امر نشان می‌دهد که این سؤالات قدرت تشخیص قابل قبولی داشته‌اند. تعداد ۳ سؤال معادل ۵٪، ضریب تشخیص بین ۱۰٪ تا ۱۵٪ را دریافت کرده‌اند. بنابراین، این سؤالات دارای درجه تمیز لب مرز هستند. تعداد ۱ سؤال معادل ۳٪ هم درجه تمیز زیر ۱۰٪ را کسب کرده است که بیانگر قدرت تشخیص ضعیف می‌باشد. سؤالاتی که درجه دشواری پایینی دارند و آسان هستند، معمولاً ضریب تشخیص لب مرز دارند.



سؤال پژوهش درباره میزان قدرت تمیز و دشواری ماده‌های آزمون تعیین سطح بود که فرضیه قابل قبول نبودن ضریب تمیز و ضریب دشواری را پیش‌بینی کرده بود؛ اما پس از محاسبه دقیق ضریب تمیز و ضریب دشواری ماده‌های آزمون، نتایج تحلیل داده‌ها نشان داد که ماده‌های آزمون از قدرت تشخیص و میزان دشواری قابل قبولی برخوردار بودند.

## ۷. نتیجه‌گیری و پیشنهادهای پژوهشی

نتیجه این پژوهش، تهیه و تنظیم و ارائه سه آزمون در سه سطح مبتدی، میانی و پیشرفته براساس مجموعه کتاب‌های آموزشی «پارسا» است که در مراکز مختلف جامعه‌المصطفی<sup>(ص)</sup> تدریس می‌شوند. هر آزمون تمام مراحل آزمون‌سازی را با موفقیت پشت سر گذاشته است. ضریب دشواری، ضریب تمیز، پایایی و روایی صوری و محتوایی آنها به‌طور مجزا و به شیوه‌ای علمی محاسبه شده و مناسب و مطلوب تشخیص داده شده‌اند.

اولین پرسش پژوهش درباره پایایی و روایی آزمون طراحی شده بود و با مشاهده نتایج آماری و داده‌پردازی‌های این دو امر مهم، مشخص شد که آزمون دارای روایی و پایایی مناسب و استاندارد است. در این پژوهش نتایج محاسبه پایایی نشان داد که پایایی آزمون سطح مبتدی ۰/۹۴، پایایی آزمون سطح میانی ۰/۸۳ و پایایی آزمون سطح پیشرفته ۰/۸۰ و پایایی کل آزمون ۰/۸۵ است. این نتایج بیانگر این بود که پایایی در آزمون سطح مبتدی بسیار عالی، در آزمون‌های سطح میانی و پیشرفته خوب و قابل قبول بوده است. نتایج این پژوهش در مقایسه با پژوهش‌های Sahraei and Rezapur (2011) و Zandi et al. (2011) و Ghonsooly (2010) که پایایی هر سه ۰/۷۰ محاسبه شده بود و همچنین با پژوهش Samaee and Kamyabi Gol (۲۰۲۰) که پایایی ۰/۷۳ گزارش شده بود، دارای پایایی مطلوب‌تری است. میزان پایایی حداقل قابل قبول در فرمول‌های آلفای کرونباخ و KR21، ۰/۷۰ است.

در پژوهش حاضر برای محاسبه روایی، از روایی محتوایی و صوری با به‌کارگیری مصاحبه نیمه‌ساخت‌یافته استفاده شده و نتایج روایی، مطلوب ارزیابی شده است. نتایج روایی در پژوهش Sahraei and Rezapur (۲۰۱۱) قابل قبول نبود و زبان‌آموزان در بخش مهارت‌های تولیدی ضعیف عمل کرده بودند؛ اما پژوهش Zandi et al. (۲۰۱۱) از روایی قابل قبول برخوردار بوده است. پرسش دوم پژوهش درباره میزان ضریب دشواری و ضریب تمیز ماده‌های آزمون بود که نتایج در کل نشان داد که این آزمون‌ها از ضریب دشواری مطلوب و قابل قبولی برخوردار هستند. در محاسبه میانگین ضریب دشواری در آزمون سطح پیشرفته مشخص شد که این آزمون پرسش دشوار نداشته است و حدود ۵ درصد از پرسش‌ها دارای ضریب دشواری ۸۰ درصد بوده و ساده تشخیص داده شدند. دلیل احتمالی ساده‌بودن این پرسش‌ها را می‌توان اشتراکات زبانی بین زبان فارسی و زبان اردو دانست. برخی از سازه‌های زبانی؛ مانند واژه‌ها یا ضرب‌المثل‌ها یا اصطلاحاتی که در دو زبان مشترک هستند، برای فارسی‌آموزان با ملیت پاکستانی و هندی ساده به نظر می‌رسید. همچنین، پس از محاسبه ضریب تمیز ماده‌های

آزمون، نتایج تحلیل داده‌ها نشان داد که ماده‌های آزمون از قدرت تشخیص قابل قبولی برخوردار بودند. این نتایج در مقایسه با پژوهش زندی و همکاران (۲۰۱۱) قابل تأمل بودند. در پژوهش آنها میانگین ضریب دشواری پرسش‌های مهارت خواندن مناسب و بهینه بود؛ اما میانگین ضریب دشواری پرسش‌ها در بخش مهارت شنیدن آسان ارزیابی شد که این نتیجه می‌تواند متأثر از ملیت تاجیک بعضی از فارسی‌آموزان و مدت زمان یکساله یادگیری زبان فارسی آنها بوده باشد. آخرین پرسش پژوهش دربارهٔ نظر مدرسان سطوح سه‌گانه (مبتدی، میانی و پیشرفته) مرکز بین‌المللی آموزش زبان فارسی به غیرفارسی‌زبانان جامعه‌المصطفی (ص) دربارهٔ آزمون تعیین سطح طراحی شده بود که تقریباً تمامی مدرسان سطح مبتدی (۹۰٪)، سطح دشواری واژگان و دستور را متناسب با سطح دشواری و دستور واژگان کتاب‌های پارسا تشخیص دادند. در بررسی مدت زمان آزمون، حدود ۹۰ درصد مدرسان مدت زمان آزمون را مناسب ارزیابی کردند. از نظر مدرسان سطح میانی، تمام بخش‌های آزمون مورد قبول بود. از نظر مدرسان سطح پیشرفته نیز شکل ظاهری آزمون‌ها و نوع و محتوای ماده‌های آزمون، بهینه بودند. پس از اعتبارسنجی علمی آزمون‌ها، درنهایت، می‌توان چنین نتیجه گرفت که آزمون طراحی شده مطابق با معیارهای استاندارد، طراحی شده است.

شایان ذکر است که نتیجه کاربردی این پژوهش برای مراکز جامعه‌المصطفی (ص) بسیار ضروری است و می‌توان از آنها در تمام مراکز آموزش زبان فارسی جامعه‌المصطفی (ص) در داخل و خارج از کشور استفاده کرد. یکی از کاربردهای این آزمون در انجام تعیین سطح آن دسته از علاقه‌مندان زبان فارسی است که با توجه به شرایط کرونایی، به‌صورت خودخوان با زبان فارسی آشنا شده‌اند و حالا علاقه‌مند به حضور در کلاس‌های مراکز داخلی و خارجی جامعه‌المصطفی (ص) هستند. توضیح آنکه گاهی این فارسی‌آموزان به کمک نرم‌افزارهای مرتبط و با بهره‌گیری از امکانات فضای مجازی (مودل‌ها، وبسایت‌ها و شبکه‌های اجتماعی)، کتاب‌های آموزشی زبان جامعه‌المصطفی (ص) را به‌صورت خودخوان فرامی‌گیرند و داوطلب ورود به یکی از این مراکز در داخل یا خارج می‌شوند. در چنین حالتی، این آزمون‌ها می‌توانند هر داوطلب را در هر سطح با توجه به متون جامعه‌المصطفی (ص) بسنجد و مشخص کند. در نتیجه کار گروه‌بندی را برای کارشناسان و مدیران آموزشی آسان می‌کنند. از طرفی، تعیین سطح درست به فارسی‌آموزان نیز انگیزه کافی برای ادامه‌دادن مسیر فارسی‌آموزی را می‌دهد. با توجه به مبحث مهم آزمون‌سازی و نوپا بودن این علم که موضوع اصلی این پژوهش است، می‌توان موضوعات زیر را به‌عنوان پیشنهادهایی برای پژوهش‌های آینده مطرح کرد:

- ۱) مقایسه و بررسی آزمون تعیین سطح جامعه‌المصطفی (ص) با آزمون‌های تعیین سطح مراکز دیگر.
- ۲) بررسی نقش اشتراکات زبان مادری زبان آموز با زبان فارسی و تأثیر آن در موفقیت فارسی‌آموز در آزمون تعیین سطح.
- ۳) بررسی متغیر ملیت در نتیجه به‌دست‌آمده از برگزاری آزمون تعیین سطح.
- ۴) تحلیل و بررسی تأثیر و نقش محیط دوم در میزان موفقیت آزمودنی‌ها در آزمون تعیین سطح.

این مقاله مستخرج از طرح پژوهشی با عنوان طراحی آزمون تعیین سطح براساس منابع آموزشی زبان فارسی جامعه‌المصطفی<sup>(ص)</sup> و کد ۱۰۲۲۲۶ و با سفارش جامعه‌المصطفی<sup>(ص)</sup> العالمیه و با نظارت دانشگاه فردوسی مشهد صورت پذیرفته است.

## فهرست منابع:

- جعفریور، عبدالجواد. (۱۳۷۵). *مقدمه‌ای بر آزمون‌شناسی زبان*. شیراز: انتشارات دانشگاه شیراز.
- داگلاس، دن. (۲۰۱۴). *درآمدی بر آزمون سازی زبان*. ترجمه هدی صحرایی و دیگران. (۱۳۹۷). تهران: انتشارات نشر نویسه پاریسی.
- زندى، بهمن؛ روشن، بلقیس؛ وکیلی فرد، امیررضا و گلپور، لیلان. (۱۳۹۸). طراحی و اعتباربخشی آزمون استاندارد زبان فارسی برپایه چهار مهارت زبانی. *جستارهای زبانی* ۱۰ (۶)، ۱۳۹-۱۷۱.
- سمائی، زهرا و کامیابی گل، عطیه. (۱۳۹۹). اعتبار سنجی آزمون نهایی زبان فارسی مدرسه المهدی<sup>(عج)</sup> جامعه المصطفی<sup>(ص)</sup> العالمیه. *همایش ملی الزامات آموزش زبان وادبیات فارسی در مدارس*. مراکز زبان آموزی، دانشگاه‌ها و فضای مجازی، دانشگاه فردوسی مشهد.
- سیف، علی‌اکبر. (۱۳۹۰). *اندازه‌گیری، سنجش و ارزشیابی آموزشی*. تهران: نشر دوران.
- صحرایی، رضامراد و جلیلی، سیداکبر. (۱۳۹۱). مبانی تهیه و تدوین آزمون مهارتی فارسی (آمفا). *پژوهش‌نامه آموزش زبان فارسی به غیرفارسی‌زبانان*، ۱ (۱)، ۱۲۴-۱۵۰.
- صحرایی، رضامراد و رضاپور، غلامحسین. (۱۳۹۸). ارزیابی روایی و پایایی آزمون تعیین سطح زبان فارسی برای غیرفارسی‌زبانان بنیاد سعدی. *اندازه‌گیری تربیتی*، ۹ (۳۷)، ۹۱-۱۱۱.
- عباسی، زهرا. (۱۳۹۳). تهیه و تدوین آزمون بسندگی زبان فارسی. *نخستین همایش زبان فارسی*، تهران، دانشگاه تربیت مدرس.
- قنسولی، بهزاد. (۱۳۸۲). *فرهنگ واژگان آزمون‌سازی (توصیفی-توضیحی)*. مشهد: انتشارات تابران.
- قنسولی، بهزاد. (۱۳۸۹). طراحی و روایی سنجی آزمون بسندگی زبان فارسی. *پژوهش‌های زبان‌های خارجی*، ۵۷، ۱۲۹-۱۱۵.

## References:

- Abbasi, Z. (2014). *Preparing and compiling the Persian language proficiency test*. Tarbiat Modares University. [ In Persian]
- Akhidenor-Bamidele, A. (2019). The roles of online placement test in English language teaching. *KnE Social Sciences*, 3 (24), 1-9. <https://doi.org/10.18502/kss.v3i24.5163>
- Bachman, L. F. (1995). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford University Press.
- Birjandi, P., & Mosallanejad, P., (2010), *An overview of testing and assessment: An overview of testing and assessment*. Sepahan Publication.
- Brown, D. (2010). *Language assessment: Principles and classroom practices*. Pearson Education.
- Brown, J. D. (2008). *Testing in language programs*. USA: McGraw-Hill.
- Carletta, J. (1996). Assessing agreement on classification tasks: The Kappa statistic. *Computational Linguistics*, 22 (2), 249-254.
- Chastain, K. (1988). *Developing second language skills*. Harcourt Brace Jovanovich Publishers.

- Cohen, A. D.** (1994). *Assessing language ability in the classroom*. Heinle & Heinle.
- Dornyei, Z.** (2011). *Research methods in applied linguistics*. Oxford University Press.
- Douglas, D.** (2014). *Understanding language testing*. Translators: Sahraei, Rezamorad; Salimi, Hoda; Marsus, Faeze. Tehran. Neveeseh Parsi Press. [In Persian]
- Farhady, H. & Ja'farpur, A. & Birjandi, P.** (2014). *Testing language skills: From theory to practice*. SAMT.
- Fulcher, G.** (2013). *Practical language testing*. Routledge.
- Ghonsooli, B.** (2003). *A dictionary of language testing*. Tabaran Press. [In Persian]
- Ghonsooli, B.** (2010). Designing and validation of Persian language proficiency test. *Foreign Language Research*, 57, 115-129. [In Persian]
- Glen, S.** (2016). Kuder-Richardson 20 (KR-20) & 21 (KR-21). Retrieved from *statisticsHowTo.com: Elementary Statistics for the rest of us!* <https://www.statisticshowto.com/kuder-richardson/> on July 05, 2022.
- Hughes, A.** (2012). *Testing for language teachers*. Cambridge University Press.
- Jafarpur, A.** (1997). *An Introduction to language Testing*. Shiraz University Press. [In Persian]
- Kadwa, M. S., & Sheik, A.** (2021). A critical review of the Oxford q-skills placement test at a Saudi Arabian University. *International Journal of Linguistics, Literature and Translation*, 4 (1), 198-209.
- Kim, H. J., & Shin, H. W.** (2006). A reading and writing placement test: Design, evaluation, and analysis. *Studies in Applied Linguistics and TESOL*, 6 (2), 68-80.
- Klein-Braley, C.** (1984). Advance Prediction of Difficulty with C-Tests. In T. Culhane, C. Klein-Braley, and D. K. Stevenson (Eds.), *Practice and problems in language testing* (pp. 97-112). University of Essex.
- Kusiak-Pisowacka, M.** (2015). The use of the IteMan program in designing a general foreign language placement test. *International Journal of Scientific and Research Publications*, 5 (12), 50-68.
- Nasem, H. & Mohammad, A.** (2018). *Teachers' perceptions on the effectiveness of the Oxford Online Placement Test at King Abdulaziz University* (Doctoral dissertation, Durham University) .
- Sahraee, R., Rezapour, G.** (2019). Evaluation of reliability and validity of Persian for non-Persian speakers placement tests of Sa'adi Foundation. *Quarterly of Educational Measurement*, 10 (37), 91-111. doi: 10.22054/jem.2020.39125.1892 [In Persian]
- Sahraei, R., & Jalili, A.** (2011). Skill based Persian language test on the basis of 4 main language skills. *Journal of Teaching Persian to Speakers of Other Languages (JTPSOL)*, 1 (1), 124-150. [In Persian]
- Samaei, Z., & Kamyabi Gol, A.** (2020). *Validation of the final Persian language exam of Al-Mahdi School (AS) of Al-Mustafa University*. National conference on the requirements of teaching Persian language and literature in schools, language learning centers, universities and virtual space. Ferdowsi University of Mashhad. [In Persian]
- Seif, A. A.** (2011). *Pedagogical measurement, examination and evaluation*. Davaran Publ. [In Persian]
- Traub, R.E.** (1994). *Reliability for the social sciences: Theory and applications*. Sage.
- Zandi, B., Rovshan, B., Vakilifard, A., & Golpour L.** (2011) Planning and validating Persian language standard test (On the basis of four language skills). *LRR.*, 10 (6), 139-171. [In Persian]