

The Study of the Content Validity Supportive Factors of the EPT Multiple-Choice Reading Comprehension Test: Based on Componential Perspective

Roshanak Rezaei¹ , Faramarz Azizmalayeri^{2*} , Abass Bayar³  &
Hossein Ahmadi⁴ 

Vol. 13, No. 4, Tome 70
pp. 727-762
September & October
2022

Received: 27 February 2021
Received in revised form: 23 July 2021
Accepted: 29 August 2021

Abstract

The validity supporting factors of the EPT reading comprehension test, a standard language proficiency test held monthly in Iran, were investigated in this research paper. To this end, each of the test questions was assigned to one of the known effective linguistic and non-linguistic components used by different readers in comprehending a reading text. Our ability to read in a second language, at first, depends on our familiarity with linguistic knowledge, which is the words and grammar of the target language, and under the influence of our non-linguistic knowledge; that if the assessment of reading comprehension ability is unfairly mixed with our "non-linguistic knowledge", the result will not reflect the content validity of the test. Since high-stake tests like EPT are designed for examinees with different experiences and other intrapersonal cognitive abilities, balancing the content of the test will lead us to more favorable results. 85 questions from 22 expository texts were randomly selected, and 150 female students graduated from the theoretical fields of humanities, experimental, and mathematics from Malayer city participated in the test. The data were analyzed using Pearson's correlation statistical analysis, Heterotrait-monotrait ratio, Exploratory Factor Analysis, and the Partial Least Square model. The results showed that non-linguistic knowledge has a significant and more than expected effect on the performance of the testees. Teachers, lecturers, and institutes implementing and designing second language reading comprehension standardized tests will use the results of this research.

Keywords: English proficiency test; CEFR mid-band proficiency level; linguistic components; non-linguistic components.

1. Ph.D. candidate, Department of English Language, Malayer Branch, Islamic Azad University, Malayer, Iran; ORCID: 0000-0001-7428-7655
2. Corresponding author: Assistant Professor of TEFL, Department of English Language, Malayer Branch, Islamic Azad University, Malayer, Iran
Email: Faramarz.amalayeri@hotmail.com; ORCID: 0000-0001-8907-2411
3. Assistant Professor of TEFL, Department of English Language, Malayer Branch, Islamic Azad University, Malayer, Iran; ORCID: 0000-0002-5625-0717
4. Assistant Professor of TEFL, Department of English Language, Malayer Branch, Islamic Azad University, Malayer, Iran; ORCID: 0000-0002-9448-5760

1. Introduction

To determine the level of proficiency in a foreign language, several language tests are frequently held each year. Language proficiency tests do not consider the method used by language learners; their emphasis is generally on the past language experience of the examinees and what they have learned until the time of the test (Jafarpour, 2015). Among the valid language tests held in Iran is the EPT, which has been held almost every month by the Islamic Azad University since 2012. EPT is composed of three parts: vocabulary, grammar, and comprehension. The importance of this exam is in the graduation of the students of this university at the PhD level.

In terms of test structure, EPT can be compared with TOEFL PBT. The only difference lies in the lack of listening and writing sections. This test consists of 100 questions; it does not have a negative score, and the time allotted is set on 140 minutes. The reading comprehension section has 35 questions, usually consisting of 4 separate expository texts on different fields. 15 questions are in the format of cloze test, and the rest is passage-based multiple choice questions.

The importance of evaluation, as stated by Black and William (1998), is undeniable both for teachers and learners. Therefore, a test content validity can guarantee the appropriateness of the result interpretations (Linn & Gronlund, 2000). Test validity can be interpreted as an agreement between the test score and the quality it is intended to measure (Kaplan & Saccuzzo, 2001). Thus, it can be concluded that any investigation and analysis on the effectiveness of the supporting content validity factors of a proficiency test can predict the adequacy of the level of linguistic knowledge as opposed to non-linguistic knowledge of the testees.

According to Bachman's language proficiency model (1990), strategic competence of language learners is based on their language competence and their knowledge structures, or their knowledge of the world (Alderson, Clapham, & Wall, 2005), and the knowledge they have acquired while

learning their first language (Hedge, 2003). Here, it can be implied that theoretically, in order to be successful in reading comprehension tasks, in addition to language competence, one needs to be equipped with other reader-oriented variables such as synthesizing, inferring, and evaluating information (Alderson, 2000).

As Nunan (1991) has stated, reading comprehension in a foreign language is different from that in one's first language. A word of caution hence should be mentioned in this regard. The assessment of this language ability should be run under a more controlled condition in terms of the testee's knowledge of the world. Nunan (1991) believed in the interaction of language competence and world knowledge when doing a comprehension task. Reading comprehension in a foreign language, as the result, needs separate sub-skills that are required to be taught (Alderson, 2000).

Considering the above mentioned premises, foreign language reading comprehension tests for mid-band language proficiency levels are expected to make use of these different source-based variables equally. This means that despite the inherent difficulty to assess this language ability, language-based questions should be paid the same attention as nonlanguage-based questions all over the test as possibly as they can (Weir, 2005).

This parity is the motivation of the present study. Collecting intermediate passages employed in different EPT administrations, the study aimed at determining the load of different linguistic (text-dependent) and non-linguistic (reader-dependent) components of the test based on the componential perspective theories. The list of components is based on Farhadi and Hesami's (2005, p. 35) list of variables in their article, *construct validity of L2 reading comprehension skills*. The following research question is going to be answered:

Which group of components, linguistic versus non-linguistic components, has the strongest correlations with the value score obtained in the measurement of reading competence based on the gathered data?

2. Methodology

2.1. Subjects

A total number of 150 12th-grade female students took part in the present study. They were from three different fields of study (58 from experimental sciences, 48 from mathematical sciences, and 44 from humanities) from 3 language institutes in Malayer. They were selected by cluster random method. The subjects were between 18 to 22 years old. ($M=19.5$, $SD=1.51$). These students were intermediate language learners and got a score range between 51 and 70 out of 120 in the Oxford English Placement Test, so they were all placed at the CEFR level of B1.2 to B2.1.

2.2. Educational tools and materials

The material used in the present study consisted of 22 randomly-selected reading comprehension passages adapted from the EPT test. The test items underwent a close specification process to assign each item to the component it was intended to measure. The linguistic components understood vocabulary, using context, identifying facts and details, summarizing concepts, sequencing events, distinguishing between facts and opinions, and discovering main idea. For non-linguistic components, understanding point of view, determining author's purpose, drawing conclusions, identifying inferences, understanding cause and effect, using prior knowledge, identifying figurative language, and visualizing ideas were utilized.

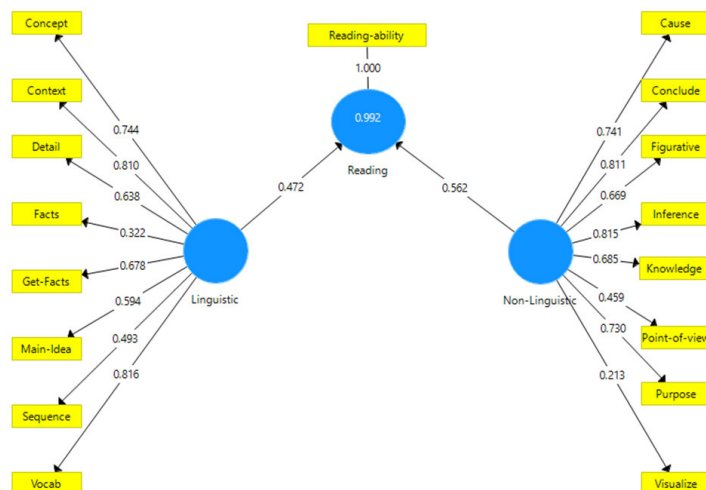
3. Results

The exploratory factor analysis (EFA) was run in order to get a maximum dispersion of factor loadings within factors of linguistic and nonlinguistic components. For a smaller number of variables highly onto each factor resulting in more interpretable clusters of factors, principal axis factoring and

varimax rotation methods were applied. SPSS results showed that the components of reading ability measured four constructs with an accuracy of 49.50 percent. From the results, scanning, and literal and reorganizational processing played an important role in comprehending second language intermediate texts of EPT.

The model of the standardized regression weights between each component and the overall linguistic and nonlinguistic constructs displayed in figure 1 indicates that the linguistic components of reading comprehension test had moderate and significant contributions to their construct, while nonlinguistic components of reading comprehension test had a high and significant contribution to their construct.

Figure. 1
The relationships between components of constructs
(standardized regression weights)





دوماهنامه بین‌المللی

د ۱۳، ش ۴ (پیاپی ۷۰)، مهر و آبان ۱۴۰۱، صص ۷۲۷-۷۲۲

مقاله پژوهشی

<http://dorl.net/dor/20.1001.1.23223081.1401.0.0.89.1>

مطالعه عوامل حمایت‌کننده روایی محتوا در آزمون چهارگزینه‌ای درک مطلب EPT: از دیدگاه مؤلفه‌محور

روشنک رضایی^۱، فرامرز عزیزملایی^{۲*}، عباس بیات^۳، حسین احمدی^۴

۱. دانشجوی دکتری آموزش زبان انگلیسی، دانشگاه آزاد اسلامی، ملایر، ایران.

۲. استادیار گروه آموزش زبان انگلیسی، دانشگاه آزاد اسلامی، ملایر، ایران.

۳. استادیار گروه آموزش زبان انگلیسی، دانشگاه آزاد اسلامی، ملایر، ایران.

۴. استادیار گروه آموزش زبان انگلیسی، دانشگاه آزاد اسلامی، ملایر، ایران.

تاریخ پذیرش: ۱۴۰۰/۰۶/۰۷

تاریخ دریافت: ۱۳۹۹/۱۲/۰۹

چکیده

در این پژوهش، به بررسی عوامل حمایت‌کننده روایی قسمتی از بخش درک مطلب آزمون استاندارد ای‌پی‌تی که هر ماه در ایران برگزار می‌شود می‌پردازیم و با اختصاص هر یک از سؤالات آن، به یکی از مؤلفه‌های مؤثرشناخته‌شده در درک مطلب، آن را تحلیل می‌کنیم. توانایی ما در خواندن در زبان دوم، در ابتدا، به میزان آشنایی ما با دانش متنی که همان کلمات و دستور زبان هدف است بستگی دارد؛ و سپس تحت تأثیر دانش غیرمتنی ماست؛ که اگر ارزیابی این توانایی با «دانش غیرمتنی» ما بیش از میزان هنجار آن آمیخته شود، نتیجه حاصله اعتبار استاندارد سؤال را تأمین نخواهد کرد. از آنجایی که آزمون‌های سرنوشت‌سازی^۱ مانند EPT برای آزمون‌دهندگان با تجربیات و دیگر توانایی‌های شناختی متفاوت فردی طراحی می‌شوند، متوازن ساختن این مداخله ما را به نتایج مطلوب‌تری خواهد رساند. ۸۵ سؤال از ۲۲ متن علمی توضیحی به‌طور تصادفی انتخاب شدند، و ۱۵۰ آزمون‌ور خانم فارغ‌التحصیل رشته‌های نظری علوم انسانی، تجربی، و ریاضی از شهرستان ملایر در آزمون شرکت کردند. داده‌ها با استفاده از تحلیل آماری همبستگی پیرسون، نسبت خصیصه متفاوت - خصیصه یکسان، تحلیل آماری اکتشافی، و مدل تعدیل جزئی بررسی شدند. نتایج حاصله نشان داد که دانش غیرمتنی در آزمون، تأثیر معنی‌دار و بیش از حد انتظاری نشان می‌دهد. معلمان، مدرسان، و مؤسسه‌های مجری طراحی آزمون‌های استاندارد در زبان دوم، از نتایج تحقیق حاضر استفاده خواهند کرد.

واژه‌های کلیدی: آزمون درک مطلب ای‌پی‌تی^۲، سطح متوسط زبانی بر مبنای سیفر^۳، مؤلفه‌های زبانی^۴، مؤلفه‌های غیرزبانی^۵.

۱. مقدمه

برای تعیین میزان توانایی افراد در زبان دوم، آزمون‌های متعددی برگزار می‌شود که از جمله آن‌ها آزمون‌های مهارت^۱ است. آزمون مهارت بر شیوه زبان‌آموزی آزمون‌وران توجهی ندارد و تأکید آن به‌طور کلی بر گذشته افراد و آنچه تا زمان آزمون فراگرفته‌اند، متمرکز است (جعفرپور، ۱۳۷۵). ازجمله آزمون‌های زبانی معتبر که در ایران برگزار می‌شود، آزمون مهارت زبانی ای‌پی‌تی است، که از سال ۱۳۹۲ تقریباً هر ماه توسط دانشگاه آزاد اسلامی در مراکز اکثر استان‌ها برگزار می‌شود. این آزمون مانند سایر آزمون‌های مشابه، مانند MSRT، MHLE و UTEPT، که فقط در ایران اعتبار دارند؛ شامل سه قسمت واژگان، دستور، و درک مطلب است. آزمون EPT در مقایسه با دو آزمون دیگر، MSRT و MHLE، که توسط وزارت علوم و وزارت بهداشت برگزار می‌شوند، فاقد بخش شنیداری است، و به همین دلیل علی‌رغم اینکه از درجه سختی آزمون TOEFL برخوردار است، نسبت به دیگر آزمون‌های مشابه ساده‌تر تلقی می‌شود. اهمیت این آزمون در فارغ‌التحصیلی دانشجویان این دانشگاه در مقطع دکتری است.

به‌لحاظ ساختار آزمون، می‌توان EPT را با TOEFL PBT مقایسه کرد. تنها تفاوت آن با آزمون TOEFL این است که بخش شنیداری و نگارش ندارد. این آزمون، متشکل از ۱۰۰ سؤال است؛ نمره منفی ندارد، و مدت پاسخ‌گویی به آن ۱۴۰ دقیقه است. نمره قبولی در این آزمون ۵۰ است. قسمت درک مطلب دارای ۳۵ سؤال است، که معمولاً شامل ۴ متن جداگانه با موضوعات توضیحی در زمینه‌های علمی متفاوت است. پانزده سؤال مربوط به cloze و بیست سؤال را درک مطلب چهارگزینه‌ای تشکیل می‌دهد.

اهمیت ارزشیابی، همانگونه که بلک و ویلیام^۲ (1998) بیان می‌کنند، آنجا نمودار می‌شود که نتیجه آن، نه‌تنها برای معلمان، بلکه برای یادگیرندگان مفید و راهگشاست و از آنجا که روایی یک ارزشیابی از کیفیت و مناسبت تفسیرهای نتایج‌بخش (Linn & Gronlund, 2000) حکایت می‌کند که از آن به صورت توافق بین نمره آزمون، و کیفیتی که قرار است اندازه بگیرد (Kaplan & Saccuzzo, 2001) می‌توان بهره جست، بررسی و تحلیل عوامل مؤثر در روایی قسمت درک مطلب آزمون موردنظر می‌تواند در پیش‌بینی بسندگی میزان دانش زبانی (متنی) و غیرزبانی (غیرمتنی) آزمون‌وران تأثیر بسزایی داشته باشد.

تعاریف دیگری توسط محققان مختلف از ارزشیابی داده شده است: جمع‌آوری منسجم اطلاعات (Palomba & Banta, 1999)؛ معیاری برای اصلاح یادگیری و آموزش (Black & William, 1998)؛ و ترکیب منسجم تعلیم و تعلم (Carr & Harris, 2001). آنچه از همه این تعاریف می‌توان برداشت کرد ماهیت ارزشیابی است که نقش آن را به عنوان عامل لاینفک آموزش، از طراحی و برنامه‌ریزی فعالیت‌های کلاسی گرفته تا تهیه برنامه درسی ملی، ضروری و انکارناپذیر می‌کند.

از سوی دیگر، هدف نهایی خواندن برای درک مطلب در آموزش زبان، رسیدن زبان‌آموز به استقلال در دریافت مطالب علمی از برون‌دادهای زبانی در قالب نوشتار است. طبق مدل توانش زبانی یکمن (1990) ترکیب توانش زبانی^۸ و دانش ساختاری (دانش عمومی جهان)^۹، (Alderson et al., 2005) که هم‌زمان با یادگیری زبان اول کسب کرده‌اند (Hedge, 2003)، به توانش راه‌کاری^{۱۰} زبان‌آموزان منجر می‌شود و زبان‌آموزان زبان دوم را به موفقیت در درک مطلب در سطوح مختلف می‌رساند. بنابراین می‌توان گفت که از لحاظ تئوری، متغیر خواندن برای درک مطلب در زبان اول، علاوه بر متغیرهای زبانی متنی، شامل دیگر متغیرهای خواننده‌محور شناختی از جمله توانایی ترکیب اطلاعات^{۱۱}، استنتاج از اطلاعات^{۱۲}، و ارزیابی اطلاعات^{۱۳} نیز خواهد شد (Alderson, 2000).

مهارت درک مطلب در زبان دوم متفاوت از این مهارت در زبان اول است (Nunan, 1991). بنابراین ارزشیابی این مهارت در زبان دوم، بهتر است تحت شرایط کنترل‌شده‌ای از دانش عمومی آزمون‌وران انجام گیرد. نونان (1991) معتقد است که کنش متقابل بین مؤلفه‌های زبانی گنجانده شده در متن و دیگر عوامل خارجی که نیاز به آموزش‌های متفاوتی خواهند داشت، مهارت خواندن در زبان دوم را تشکیل می‌دهد.

بنابراین درک مطلب تنها به محتوای زبانی متنی محدود نمی‌شود، بلکه به میزان آموزش و تجربیات گذشته زبان‌آموزان «متغیرهای غیرمتنی» نیز وابسته است (Alderson, 2000). آلدرسون (2000) به وضوح بیان می‌کند که: «آستانه زبانی (نمود پیدا کرده در متن) در آموزش زبان دوم باید قبل از اینکه توانایی زبان اول با آن مداخله کند، در نظر گرفته شود (p.121)». گرچه این ناخالصی را نمی‌توان به طور کامل زدود، با دقت بیشتر می‌توان در طراحی

سؤال‌ها برای سطح متوسط زبانی، از رجحان یکی از این مؤلفه‌ها (متغیرهای متن‌محور و متغیرهای خواننده‌محور) بر دیگری اجتناب کرد (Weir, 2005).

این توازن، انگیزه انجام پژوهش حاضر است. با جمع‌آوری متون سطح متوسط در معیار سیفر، از آزمون استاندارد ای‌پی‌تی که در ایران هر ماه اجرا می‌شود، این مطالعه قصد دارد بار مؤثر مؤلفه‌های زبانی (وابسته به متن) و غیرزبانی (وابسته به خواننده) متفاوتی را که براساس نظریه مؤلفه - محور و براساس لیست مؤلفه‌هایی که فرهادی و حسامی (2005, p.35) در مقاله خود با نام روایی سازه‌ای مهارت‌های درک‌مطلب در زبان دوم جمع‌آوری کرده‌اند بررسی کند. این مطالعه قصد دارد که به سؤال تحقیق زیر پاسخ دهد:

کدام گروه مؤلفه‌ها - گروه مؤلفه‌های زبانی در مقابل گروه مؤلفه‌های غیرزبانی - بیشترین همبستگی را با نمره کسب‌شده در ارزشیابی مهارت «خواندن برای درک‌مطلب» براساس داده‌های جمع‌آوری شده، داشته است؟

نتایج این تحقیق نکته مهمی را برای ارزشیابی مهارت نهفته خواندن آشکار خواهد کرد و چون به بررسی «روایی آزمون چندگزینه‌ای» یک آزمون استاندارد می‌پردازد، در پیش‌بینی راهکارهای مؤثر برای موفقیت در این آزمون سرنوشت‌ساز نیز سودمند خواهد بود، که طبیعتاً زبان‌آموزان و متخصصان در امر آموزش زبان از آن سود خواهند برد.

۲. پیشینه تحقیق

۲-۱. مبانی نظری

از آنجا که روایی مهم‌ترین ویژگی یک آزمون خوب است (Sahraei & Mamaghani, 2012; Walt & Steyn, 2008) بررسی آن، به‌خصوص برای موفقیت در آزمون‌های سرنوشت‌ساز ضروری است (Amirian et al., 2018). مسیک (1989) مفهوم روایی را به تفسیر نتایج حاصله از آزمون نسبت داد. او روایی سازه‌ای^{۱۴} را به‌عنوان عنصر اصلی روایی تفسیر نتایج آزمون‌ها مطرح کرد که روایی محتوایی و روایی سنج‌ای (ملاکی)، هر دو قسمت‌هایی از آن را تشکیل می‌دادند. در تقسیم‌بندی مسیک (1989) روایی سازه‌ای که براساس شواهد حاصل از آزمون به‌دست می‌آمد، کاربرد و ارتباط آزمون را با تغییری که برای اندازه‌گیری آن طراحی شده بود نشان می‌داد.

بنابراین، روایی درواقع زنجیره‌ای از استنتاج‌های وابسته به شواهد گرفته‌شده از آزمون‌هاست (Walt & Steyn, 2008). پژوهشگرانی چون بکمن (1990) و ویر (2007) با اشاره به ویژگی یکپارچگی روایی که توسط مسیک (1989) بیان شده بود، آن را به زیرطبقه‌های متفاوتی تحت عناوینی ازجمله «روایی نظریه‌پرداز»^{۱۰}، «روایی مضمون»^{۱۱}، «روایی نمره‌مدار»^{۱۲}، «روایی شناختی»^{۱۳} و «روایی خارجی»^{۱۴} تقسیم کردند. استین و والت (2008)، برای بررسی و تحلیل یک آزمون، رجوع به تئوری آزمون کلاسیک، مدل اندازه‌گیری رش، و تحلیل عامل نمرات حاصل از آزمون را توصیه کرده‌اند.

خواندن برای درک‌مطلب، یکی از مهم‌ترین مهارت‌های زبانی چهارگانه است که ارتباط مستقیمی با میزان سواد افراد دارد. طبق تعریف گروه مطالعه «رند» (2002)، فرایند هم‌زمان دریافت و ساخت معنا از کنش متقابل فرد با متن نوشته‌شده را، خواندن برای درک‌مطلب می‌نامند. این فرایند شامل سه زیرتوانش زبانی (Caldwell, 2008) مهارت درک‌مطلب؛ دانش پایه، انگیزه فرد برای خواندن، و ویژگی‌های متن، ازجمله میزان سختی آن است. طبق این نظریه برای داشتن فهم درست از یک متن، افراد علاوه بر دانش زبانی که مربوط به متغیرهای متن است، باید به دانش فراشناختی و فرازبانی نیز مجهز باشند؛ گرچه، معیاری برای بسندگی میزان هر کدام از این دانش‌ها هنوز هم مشخص نیست (Caldwell, 2008).

فرضیه‌های بسیاری برای تشریح ماهیت و طرز کار این مهارت اساسی در مغز ارائه شده است. این فرضیه‌ها را می‌توان به دو دسته فرایندمحور^{۲۰} و مؤلفه‌محور^{۲۱} تقسیم کرد. فرایندمحور بر عملکرد مغز و پردازش اطلاعات حین خواندن تأکید می‌کنند و در تلاش‌اند برای این فرایندهای ذهنی مدل‌های بهینه ارائه دهند (Parris & Hamilton, 2009). نظریه‌های مؤلفه‌محور به فرایند ذهنی توجهی ندارند، بلکه در جست‌وجوی حیطه‌های علمی و مهارت‌هایی هستند که در این فرایند ذهنی دخیل هستند. این فرضیه‌ها مؤلفه‌های متفاوتی را برای مهارت خواندن معرفی کرده‌اند (Hoover & Tunmer, 1993). پژوهشگرانی چون دیویس^{۲۲} (1968)، مانبی^{۲۳} (1978)، لانزر، ویت، و دالان^{۲۴} (1979)، کودی^{۲۵} (1979)، پرفتی^{۲۶} (1985)، گرب (1991)، فرهادی و حسامی (2005)، و وندرگرفت و گا^{۲۷} (2012) مجموعه‌های متفاوتی را از مؤلفه‌های گوناگون ارائه کرده‌اند. تقریباً در همه این مجموعه‌ها، مؤلفه‌های زبانی با ریزمؤلفه‌های متفاوت، و دانش

عمومی (تجربه‌های فردی) با ریزمؤلفه‌های متفاوت، به چشم می‌خورد. ژنگ^{۲۸} (2016)، نظریه‌های فرایندمحور را به سه گروه «جزء به کل»^{۲۹}، «کل به جزء»^{۳۰}، و «تعاملی»^{۳۱} که ترکیبی از دو مدل جزء به کل، و کل به جزء است، تقسیم‌بندی کرده است. مدل‌های جزء به کل تأکید بر متن دارند. در این مدل‌ها، خوانندگان متن را به صورت خطی پردازش می‌کنند، درحالی که مدل‌های کل به جزء، تأکید بر خواننده متن و دانش قبلی‌اش دارند. این پژوهشگران بر این باورند که «استنتاج» و «تجربیات قبلی»، دو عامل اساسی در فرایند خواندن به حساب می‌آیند. کینچ^{۳۲} (1998)، وندیک^{۳۳} (1977)، رامله‌هارت^{۳۴} (2004)، واستوویچ^{۳۵} (2000) به مدل تعاملی خواندن معتقدند. متن و دانش عمومی خواننده متن، هر دو از اهمیت یکسانی برخوردارند و رجحان یکی بر دیگری، در نظر این پژوهشگران، اعتبار ارزشیابی مهارت خواندن را تهدید خواهد کرد.

به دلیل کمبودن پژوهش در زمینه زبان‌ها و مهارت‌های آموزشی زبان دوم در حوزه بزرگسالان (August & Shanahan, 2010)، هنوز هیچ مدل تئوری جامعی، که بتواند فرایند درک مطلب را در هر دو زبان اول و دوم در بزرگسالان توضیح دهد، وجود ندارد (Pasquarell et al., 2012).

۱-۲. پیشینه مطالعات

باتوجه به مشکلات عدیده‌ای که فراگیران زبان پس از ورود به دوره دکتری در بحث مهارت‌های زبانی پیدا می‌کنند، و تهدید جدی که این امر بر روایی آزمون‌های سرنوشت‌ساز می‌گذارد (Amirian et al., 2018)، پرداختن و تحلیل سؤالات این نوع آزمون‌ها در بخش‌های مختلف زبانی، ضروری به نظر می‌رسد.

پژوهش‌های نسبتاً متعددی در زمینه روایی آزمون‌ها، به خصوص در سال‌های اخیر انجام شده است. تعدادی از این پژوهش‌ها، به بررسی جهت‌گیری آزمون‌های استاندارد و سرنوشت‌ساز در ایران تمرکز کرده‌اند. شهمیرزادی و همکاران (2020)، سطح توانایی آزمون‌وران را در مقیاس بزرگ، در آزمون ورودی زبان عمومی مقطع دکتری زبان‌انگلیسی مورد مطالعه قرار داده‌اند. این پژوهشگران با بررسی پاسخ‌نامه بیش از ۳۰۰۰ داوطلب این

آزمون، با کمک ماتریس کیو در مدل سنجش تشخیصی شناختی جی‌دینا^{۳۶}، و به‌وسیله پروتکل تفکر با صدای بلند، به مطالعه کنش افتراقی سؤال و خصیصه پرداختند، سوگیری در کنش افتراقی سؤالات آزمون مشاهده شد.

در همین راستا، همتی و سایرین (2016) عملکرد داوطلبان آزمون ورودی دانشگاه‌های ایران (کنکور)، را در سؤالات چهارگزینه‌ای بخش درک‌مطلب زبان‌انگلیسی، با به‌کارگیری مدل جی‌دینا در سنجش تشخیصی شناختی بررسی کردند. این محققان بر یکپارچگی توانش زبانی تأکید کردند. نتایج این تحقیق، مهارت در پنج مولفه درک‌مطلب را برای موفقیت آزمون‌دهندگان در بخش زبان انگلیسی آزمون ملی دانشگاه‌های ایران ضروری دانست، این درحالی است که ۵۰ درصد داوطلبان این آزمون، برای این پنج مولفه آمادگی نداشتند. این پنج مولفه عبارت بودند از: «استنتاج اطلاعات»، «استخراج اطلاعات متنی»، «حدس معنی کلمه ناآشنا با توجه به متن»، «شناخت مرجع ضمائر»، و «ارزیابی گزینه‌ها».

در نتایج به‌دست آمده از پژوهش‌های اخیر (Estaji & Zhaleh, 2020)، استفاده از کنش افتراقی سؤال برای تولید سؤالات بدون جهت‌گیری و عادلانه متناسب با پیشینه علمی، ضروری است. استاجی و ژاله (2020)، جهت اطمینان از اعتبار و عدالت سؤالات در آزمون‌های سرنوشت‌ساز، سؤالات درک‌مطلب آزمون ورودی دانشگاه‌ها، برای ورود به دوره کارشناسی‌ارشد گرایش‌های مختلف زبان‌انگلیسی را بررسی کردند. هدف از تحقیق، ارتباط بین رشته مورد مطالعه، و موفقیت در قسمت درک‌مطلب آزمون بود. مدل تئوری پاسخ سؤال استدلالی یک‌عامل - دوعامل^{۳۷}، برای بررسی عملکرد نمونه‌ای متشکل از ۳۵۸۸ آزمون‌ور استفاده شد. نتیجه این پژوهش نیز سوگیری در آزمون را، در سؤالات درک‌مطلب گزارش کرد. برخلاف نتایج بالا، یافته‌های روحانی و همکاران (2017)، تفاوت معنی‌داری را بین دانش زمینه‌ای ۵۸ دانشجوی دوره کارشناسی رشته زبان‌انگلیسی، و عملکرد آنان در آزمون درک‌مطلب گزارش نکرد. نتایج این تحقیق، دانش زمینه‌ای را تهدیدی برای اعتبار ساختاری آزمون ندانست.

مطالعات دیگر نیز هم‌بستگی معنی‌داری را بین دانش زمینه‌ای بالا و توانایی درک‌مطلب نشان دادند؛ این درحالی‌است که دانش زمینه‌ای کم، با توانایی درک‌مطلب هم‌بستگی کمی داشت

(et al., 2014 Adams)؛ این پژوهشگران دریافتند که علی‌رغم این حقیقت، عملکرد زبان‌آموزان با دانش زبانی بالا، به‌طور معنی‌داری بهتر از عملکرد زبان‌آموزان با دانش زبانی پایین‌تر است. شاید دلیل این مطلب، وجود آستانه زبانی در درک مطلب زبان دوم باشد (Grab, 2009)؛ نکته‌ای که تحقیق حاضر به آن پرداخته است.

در همین راستا، شیبانی و زراعت‌پیشه (2018)، با کمک مدل تحلیل اکتشافی، رابطه بین اسلوب سؤالات درک مطلب آزمون ورودی دانشگاه‌ها – سؤالات چهارگزینه‌ای در برابر سؤالات کلوز – و عملکرد آزمودنی‌ها را مطالعه کردند. نتایج این تحقیق، سوگیری را در سؤالات کلوز تأیید کرد که در جهت نقض تئوری یکپارچگی اعتبار آزمون مسیک (1989) بود.

همچنین، درباره ساخت آزمون درک مطلب در زبان اول (جباری و خادمی، ۱۳۹۲) و زبان دوم (Zandi et al., 2014) پژوهش شده است: عامل مشترک در این پژوهش‌ها، ضرورت مرور و بررسی سؤالات مورد استفاده در آزمون درک مطلب، و تبیین ارتباط هر سؤال با مولفه در نظر گرفته شده، متناسب با سطح زبانی یادگیرندگان، و هدف ارزش‌یابی مورد نظر بود.

رنجبران و علوی (2017) نیز، طراحی آزمون درک مطلب را براساس سنجش تشخیصی شناختی و مدل رام^{۳۸} تحقیق کرده‌اند. این محققان در بررسی‌ها دریافتند که مولفه «خلاصه کردن» در مقایسه با مولفه «استنتاج»، برای زبان‌آموزان چالش‌برانگیزتر است. گرچه، نظریه معروف فلچر (2009) مولفه «استنتاج» را چالش‌برانگیزترین سؤالات درک مطلب دانسته بود. براساس این تحقیق، درک دانش ضمنی نهفته در متن‌ها به سبب نیاز به دانش زمینه‌ای مرتبط، از جمله مؤلفه‌های دشوار در آزمون‌های درک مطلب است، و دریافتن اطلاعات متنی آشکار از جمله آسان‌ترین مؤلفه‌های زبانی است. آلدرسون (2000) علت این امر را سطح پایین شناختی این مؤلفه در نظریه بلوم (1956) می‌داند.

راج و همکاران (2019) تأثیر مؤلفه‌های زبانی را در پاسخ به فعالیت‌های متن‌های داستانی کم تا متوسط گزارش کرده‌اند. کولت (2006) از میان این مؤلفه‌ها، دانش واژگانی را مهم‌ترین عامل موفقیت زبان‌آموزان می‌داند. او معتقد است که دانش واژگانی به دو دسته «اشاره‌گر» یا راهنما و «ادراکی» یا مفهومی تقسیم می‌شود که درک دسته دوم (ادراکی) نیاز به دانش زمینه یا تجربیات قبلی دارد.

طبق جست‌وجوهای نگارندگان پژوهش حاضر، و باتوجه به مطالعات ذکرشده، مطالعه کم‌تری به بررسی مؤلفه‌های زبانی هدف در آزمون‌های چندگزینه‌ای برمبنای دانش زبانی و غیرزبانی پرداخته است. اعتبار قضاوت آزمون‌های درکمطلب باتوجه به تعادل استفاده از سؤالات مبتنی بر دانش زبانی و غیرزبانی، مطابق با سطح زبانی مخاطبان آزمون، قابل‌بررسی و بهبود است.

۳. روش‌شناسی تحقیق

۳-۱. آزمودنی‌ها

در این پژوهش، ۱۵۰ دانش‌آموز دختر پایه ۱۲ و فارغ‌التحصیل رشته‌های علوم تجربی، علوم ریاضی، و علوم انسانی (۵۸ نفر از رشته علوم تجربی، ۴۸ نفر علوم ریاضی، و ۴۴ نفر علوم انسانی) از سه مؤسسه زبان معتبر نشر، پیشگامان، و راستان در سطح شهرستان ملایر به روش تصادفی خوشه‌ای انتخاب شدند. سن آزمودنی‌ها ۱۸ تا ۲۲ سال بود ($M = 19.5$, $SD = 1.51$). این دانش‌آموزان و فارغ‌التحصیلان دارای مدرک زبانی متوسط بودند که در آزمون تعیین سطح آکسفورد پلیسمنت تست، دامنه نمره‌ای بین ۵۱ تا ۷۰ از ۱۲۰ داشتند، و طبق تقسیم‌بندی «سیفر» در محدوده زبانی گروه B1.2 تا B2.1 قرار گرفته بودند.

۳-۲. ابزارها و مواد آموزشی

۳-۲-۱. قسمت درکمطلب آزمون ای‌پی‌تی

آزمون EPT، ۱۰۰ نمره دارد که ۲۵ نمره آن مختص به واژگان، ۴۰ نمره دستور، و ۳۵ نمره درکمطلب است. زمان پاسخ به سؤالات به‌طور کلی ۱۴۰ دقیقه بوده، و نمره منفی گزارش نمی‌شود. حداقل نمره قبولی در این آزمون ۵۰ است.

از قسمت سؤالات چهارگزینه‌ای مربوط به متون توضیحی آزمون ای‌پی‌تی، در این پژوهش استفاده شد. از دوازده دوره آزمون ای‌پی‌تی، ۳۵ متن توضیحی، توسط هفت نفر از استادان رشته زبان که دارای مدرک دکتری در رشته آموزش زبان انگلیسی هستند، به‌عنوان سطح متوسط انتخاب شدند، و از بین این متون، ۲۲ متن به روش تصادفی ساده انتخاب شد.

متخصصان با کمک جدول تعیین مشخصات، سؤالات را کدگذاری کردند. هر سؤال با توجه به مهارتی که برای اندازه‌گیری آن طراحی شده بود، مشخص شد؛ و در یکی از دو دسته سؤالات زبانی و غیرزبانی قرار گرفت. آن دسته از سؤالات که تأکید بر معنی واژگان، پیدا کردن مراجع ضمیر، دستور زبان مفهومی مانند تشخیص جمله معلوم از مجهول، و استفاده از دانش درباره زبانی^{۳۹} جهت یافتن جزئیات و ایده اصلی متن داشت، زیرمجموعه دانش زبانی در نظر گرفته شدند؛ و سؤالاتی از قبیل ارزشیابی، تجزیه و تحلیل، دریافت معانی ضمنی کلمات، عبارات و اصطلاحات، دیدگاه و هدف نویسنده، و کاربرد اطلاعات، به مجموعه سؤالات غیرزبانی ارجاع شدند.

۲-۲-۳. نرم‌افزار CEFR Text Analyzer

نرم‌افزار مذکور براساس میانگین طول جملات تشکیل‌دهنده متون، به همراه میزان تواتر کلمات استفاده‌شده در آن‌ها، متون را به دسته‌های متوالی از آسان‌ترین A1 تا مشکل‌ترین C2 تقسیم می‌کند. برای تشخیص میزان دشواری کلمات نیز، نسبت تعداد واژگهای وابسته و آزاد در واحد کلمات محاسبه، و نتیجه مشخص می‌شود. متون مورد استفاده در پژوهش حاضر، در اولین مرحله، با نرم‌افزار تکست‌آنالایزر بررسی و متون مشخص‌شده با سطح B1 به استادان ارجاع شدند. از آنجایی که قضاوت نیروی انسانی تکمیل‌کننده فرایند تعیین سطح متون توسط ماشین است؛ پژوهش حاضر، جهت حصول اطمینان از سطح متون انتخابی از هر دو معیار استفاده کرد.

۳-۳. جمع‌آوری داده‌ها

این مطالعه در دو مرحله متوالی انجام شد. در مرحله اول، پس از نرم‌افزار، متتهای اقتباس‌شده از ۱۲ دوره آزمون آی‌پی‌تی، تحت بررسی گروهی از کارشناسان خبره آموزش زبان انگلیسی به عنوان زبان خارجی در پرسش‌نامه سی‌دی‌ای^{۴۰}، و با استفاده از طرح دلفی قرار گرفت، و ۲۲ متن از ۳۵ متن به طور تصادفی انتخاب شد. سپس، سؤالات رمزگذاری شدند. سؤالات، نهایتاً به هشت مولفه زبانی و هشت مولفه غیرزبانی گروه‌بندی شدند. سؤالاتی که براساس درک معنی کلمه، پیدا کردن مرجع ضمیر، جست‌وجوی مفاهیم آشکار در متن، پیدا کردن جزئیات و دریافت

اطلاعات، کوتاه کردن مفهوم جملات در قالب کلمات، درک توالی وقایع، و تمایز بین اطلاعات و ایده‌ها طراحی شده بودند، در گروه مولفه زبانی قرار گرفتند. و سؤالاتی که علاوه بر درک معنای زبانی، نیاز به فعالیت‌های شناختی عمومی از جمله استنباط، تجزیه و تحلیل، کاربرد، ارزیابی اطلاعات، و استفاده مستقیم از دانش قبلی داشتند، به عنوان گروه مولفه غیرزبانی شناخته شدند.

در مرحله دوم، ۱۵۰ آزمون‌ور خانم، در محدوده سنی ۱۸ تا ۲۲ سال، و با رشته‌های تحصیلی متفاوت علوم تجربی، علوم ریاضی، و علوم انسانی، به سؤالات پاسخ دادند. سپس نتایج جمع‌آوری شده با کمک تحلیل آماری همبستگی پیرسون، نسبت خصیصه متفاوت - خصیصه یکسان^{۴۱}، تحلیل آماری اکتشافی^{۴۲}، و مدل تعدیل جزئی^{۴۳} بررسی شدند. جمع‌آوری داده‌ها پس از یازده جلسه آزمون تکمیل شد. این آزمون‌های کوتاه مدت که هرکدام به طور متوسط شامل دو متن درک مطلب و سؤالاتی با دامنه ۶ تا ۱۰، در هر آزمون بودند، در فواصل منظم برگزار شدند و حدود سه ماه زمان بردند. زمان پاسخ برای هر سؤال، ۲ دقیقه و کل مهلت هر آزمون براساس تعداد سؤال‌ها متغیر بود. گفتنی است، که شرکت‌کننده‌ها از شرایط پژوهش اطلاعی نداشتند، و درواقع جهت بهبود توانایی درک مطلب خود، در آزمون‌های کوتاه مدت شرکت می‌کردند. از آنجا که مکان و زمان آزمون نیز بر نتیجه آن اثرگذار است، آزمون‌ها همگی در صبح و در مکان مؤسسات برگزار شدند.

۴. تجزیه و تحلیل داده‌ها

نمرات کلی آزمون، و خرده مؤلفه‌ها در نرم افزار ای بی ام اس پی.اس.اس^{۴۴}، نسخه ۲۶ ثبت شدند. سؤال تحقیق با کاربرد تحلیل آماری همبستگی پیرسون، نسبت خصیصه متفاوت - خصیصه یکسان، تحلیل آماری اکتشافی، و مدل تعدیل جزئی بررسی شد. نتایج این تحلیل‌ها در ذیل آورده شده است:

جدول ۱ نتایج همبستگی پیرسون را برای دو زیرمؤلفهٔ زبانی و غیرزبانی نشان می‌دهد. ماتریکس همبستگی به سه قسمت تقسیم شده، و هر بخش با رنگ‌های سفید، طوسی، و هاشور مشخص شده است. نواحی طوسی و هاشور، ضریب همبستگی بین زیرمؤلفه‌های زبانی، و غیرزبانی را نشان می‌دهد، در قسمت سفیدرنگ ضریب همبستگی بین دو زیرمؤلفه مشخص شده است. طبق نسبت خصیصهٔ متفاوت - خصیصهٔ یکسان، اگر زیرمؤلفه‌های مولفهٔ زبانی دارای اعتبار و اگر باشند، ضریب همبستگی بین آن‌ها، باید قوی‌تر از ضریب همبستگی آن‌ها، با زیرمؤلفه‌های مولفهٔ دیگر (غیرزبانی) باشد.

[illegible]

منطق متضمن نسبت خصیصه متفاوت - خصیصه یکسان، بر این اساس استوار است که میانگین زیرمؤلفه‌های اشتراک دو مؤلفه زبانی، و غیرزبانی (ناحیه سفیدرنگ)، کمتر از ریشه دوم ضریب همبستگی بین زیرمؤلفه‌های آن‌ها (نواحی طوسی و هاشور) باشد. میانگین ضریب همبستگی برای نواحی طوسی، هاشور، و سفید به ترتیب ۰.۳۲۵، ۰.۳۳۶، و ۰.۳۴۱ است. نسبت خصیصه متفاوت - خصیصه یکسان، به صورت زیر محاسبه می‌شود:

$$1.02 = \sqrt{0.325 * 0.336} / 0.341$$

طبق نظر گارسون (2016) و هیروسایرین (2017)، اگر این نسبت بیشتر از ۰.۹۰ باشد، آزمون اعتبار واگرا ندارد. به عبارت دیگر، ضریب همبستگی اشتراکی، بین دو مؤلفه مادر، بزرگ تر از همبستگی بین زیرمؤلفه‌هاست. این امر حاکی از این است که ضریب همبستگی بین زیرمؤلفه‌های زبانی و زیرمؤلفه‌های غیرزبانی، کمتر از ضریب همبستگی میانگین اشتراکی هر دو آن‌هاست.

۲-۴. تحلیل آماری اکتشافی

برای بررسی بسندگی نمونه سؤالات و مؤلفه‌ها و زیرمؤلفه‌های هدف برای تحلیل آماری اکتشافی، از روش چرخش واریمکس استفاده شد. جدول ۲ ماتریکس عامل مؤلفه‌های زبانی، و غیرزبانی مهارت خواندن را نشان می‌دهد. اولین دسته شامل چهار زیرمؤلفه زبانی: دریافت مرجع زبانی، دریافت اطلاعات، دانش واژگانی، و درک جزئیات، و سه زیرمؤلفه غیرزبانی: درک علت و معلولی جملات، تشخیص معنای ضمنی کلمات، و تشخیص هدف نویسنده است. دسته دوم، دو زیرمؤلفه زبانی توالی وقایع، و جملات کوتاه‌شده در قالب کلمات، و چهار زیرمؤلفه غیرزبانی: استنباط، درک دیدگاه، نتیجه‌گیری، و استفاده از دانش تجربی است. نتایج، همچنین، تشخیص تفاوت بین اطلاعات و ایده‌ها و تشخیص ایده اصلی متن را در دسته‌های دوم و سوم قرار می‌دهد. تصویرسازی اطلاعات با هیچ کدام از دسته‌ها هماهنگ نیست و دانش واژگانی در بیشتر دسته‌ها دیده می‌شود.

جدول ۲: ماتریکس عامل؛ مؤلفه‌های مهارت خواندن

Table 2: Rotated Factor Matrix^a: Components of Reading Ability

	عامل			
	1	2	3	4
دریافت مرجع ضمیر	.839	.372		
دریافت اطلاعات	.634			
درک علت و معلولی جملات	.580		.391	
تشخیص معنای ضمنی کلمات	.572		.462	
دانش واژگانی	.565	.464	.346	
درک جزئیات	.561			
تشخیص هدف نویسنده	.544	.307		
استنتاج		.738	.336	
درک دیدگاه		.654		
نتیجه‌گیری	.386	.626		
توالی وقایع		.517		
جملات کوتاه‌شده در قالب کلمات	.338	.476	.458	
استفاده از دانش تجربی (قبلی)	.369	.375		
تشخیص تفاوت بین اطلاعات و ایده‌ها			.463	
تصویرسازی اطلاعات				
تشخیص ایده اصلی متن		.329		.627

Extraction Method: Principal Axis Factoring.

Rotation Method: Varimax with Kaiser Normalization.

Rotation converged in 5 iterations.

۳-۴. مدل تعدیل جزئی

برای رسیدن به اهداف زیر، مدل تعدیل جزئی برای تحلیل داده‌های این مطالعه به‌کاررفت:

۱. دو گروه مؤلفه‌های زبانی و غیرزبانی تا چه میزانی در نمره‌های کل تأثیر گذاشته‌اند؛

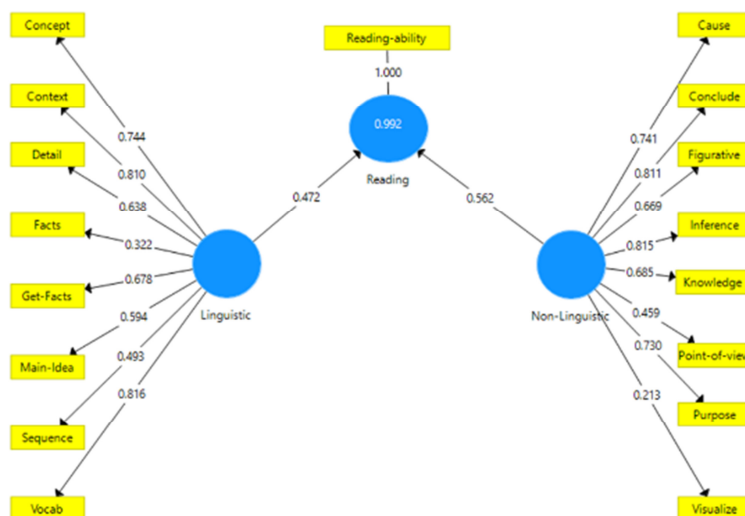
۲. تا چه حدی، دو گروه مؤلفه‌های زبانی و غیرزبانی، به‌صورت یک واحد کل در مهارت خواندن برای درک مطلب تأثیرگذار بوده‌اند.

مدل ۱ بار رگرسیون استاندارد^{۴۰}، بین هر مؤلفه و کل خرده‌مؤلفه‌های زبانی و غیرزبانی،

و مدل ۲ مشارکت هر دو گروه را در مهارت پنهان خواندن نمایش می‌دهد. گفتنی است که بار رگرسیون استاندارد (b)، با استفاده از معیار ضریب هم‌بستگی پیرسون نیز می‌تواند تفسیر شود. به این مفهوم که ضریب $0/1$ و کمتر از آن ضعیف، ضریب $0/3$ متوسط، و ضریب $0/5$ به بالا زیاد در نظر گرفته شود. مقدار تی و شاخص معادل احتمال آن، بیانگر معنادار بودن آماری بار رگرسیون استاندارد است؛ و در آخر، $97/5$ ضریب اطمینان، نشان‌دهنده معیار صحت تخمین نمونه آماری است. اگر محدوده این ضریب اطمینان، منفی یا برابر صفر شود بدین معناست که بار رگرسیون استاندارد، تصادفی به وجود آمده، و درواقع، می‌توانست صفر باشد.

براساس این نتایج و با توجه به جدول ۳ و مدل ۱، ریزمؤلفه‌های زبانی و غیرزبانی آزمون، مشارکت معنی‌داری در نمره سازه مورد سنجش – مهارت درک مطلب – داشته‌اند. به‌طور خلاصه، هیچ یک از ریز مؤلفه‌های زبانی، ضریب اطمینان منفی یا صفر ندارند، و تمامی رگرسیون‌ها، برابر، یا مساوی $0/30$ هستند.

اعداد ذکرشده درمورد ریزمؤلفه‌های غیرزبانی، نیز نشان می‌دهد که به‌طور متوسط این زیرمؤلفه‌ها، مشارکت معنادار و بالاتری در نمره‌های کل مهارت خواندن دارند. درمجموع، به‌جز خرده‌مهارت تصویرسازی اطلاعات، که با ضریب اطمینان منفی مشخص شده است، $0/005$ -، باقی زیرمؤلفه‌ها، مقدار رگرسیون $0/5$ و بالاتر دارند.



مدل ۱: ارتباط بین زیرمؤلفه‌ها (میزان رگرسیون استاندارد)

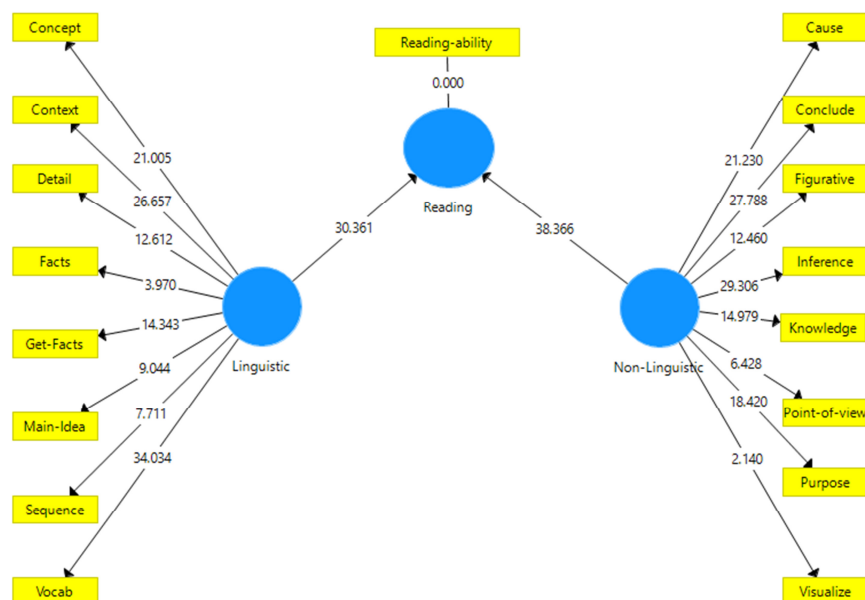
Model 1: Relationships between components of constructs (standardized regression weights)

جدول ۳: میزان رگرسیون استاندارد؛ مشارکت زیرمؤلفه‌های زبانی و غیرزبانی در مؤلفه‌های مادر

Table 3: Standardized Regression Weights; Contributions of Linguistic Components to their Constructs

	B	میانگین	انحراف معیار	ارزش تی	ارزش پی	2.5 %
علت و معلول < غیرزبانی -	0.741	0.739	0.038	19.561	0.000	0.654
مفاهیم < زبانی	0.744	0.745	0.035	21.458	0.000	0.667
نتیجه < غیرزبانی	0.811	0.810	0.030	27.248	0.000	0.746
متن < زبانی	0.810	0.812	0.030	26.812	0.000	0.747
جزئیات < زبانی	0.638	0.638	0.050	12.810	0.000	0.529
اطلاعات و ایده < زبانی	0.322	0.316	0.083	3.889	0.000	0.129
ضمنی < غیرزبانی	0.669	0.668	0.052	12.932	0.000	0.559
اطلاعات < زبانی	0.678	0.677	0.045	14.967	0.000	0.573
استنتاج < غیرزبانی	0.815	0.813	0.029	28.450	0.000	0.753
دانش قبل < غیرزبانی	0.685	0.686	0.045	15.390	0.000	0.590

2.5 %	ارزش پی	ارزش تی	انحراف معیار	میانگین	B	
0.445	0.000	9.132	0.065	0.593	0.594	ایده اصلی -> زبانی
0.306	0.000	6.354	0.072	0.460	0.459	دیدگاه -> غیرزبانی
0.644	0.000	18.525	0.039	0.730	0.730	هدف -> غیرزبانی
0.347	0.000	7.477	0.066	0.494	0.493	توالی -> زبانی
-0.005	0.046	2.001	0.106	0.212	0.213	تصویرسازی -> غیرزبانی
0.766	0.000	34.749	0.023	0.815	0.816	واژگان -> زبانی



مدل ۲: رابطه بین مؤلفه‌های مادر زبانی و غیرزبانی (ارزش تی)

Model 2: Relationships between components of constructs (t-values)

مدل ۲ ارتباط بین دو مولفه مادر، زبانی، و غیرزبانی با مهارت خواندن (ارزش تی) را نشان می‌دهد.

باتوجه به نتایج حاصله در جدول ۴ و مدل ۲، می‌توان نتیجه‌گیری کرد که مؤلفه‌های

غیرزبانی مشارکت معنی‌دار و بالایی در مهارت خواندن دارند ($b = 0.562$, $T = 38.36$, $P = 0.000$, $97\% \text{ CI}[0.530, 0.589]$)، درحالی‌که، مؤلفه‌های زبانی مشارکت معنی‌دار، اما متوسطی در مهارت خواندن برای درک مطلب دارند ($b = 0.472$, $t = 30.36$, $p = 0.000$, $97.5\% \text{ CI}[0.442, 0.504]$).

جدول ۴: میزان رگرسیون استاندارد؛ مشارکت مؤلفه‌های مادر زبانی و غیرزبانی در مهارت خواندن

Table 4: Standardized Regression Weights; Contributions of Linguistic and Non-Linguistic Components to Reading Ability

	B	میانگین	انحراف	ارزش	ارزش	2.5	97.5
		ن	معیار	تی	پی	%	%
مؤلفه زبانی > مهارت خواندن	0.47 2	0.473	0.016	30.361	0.000	0.44 2	0.504
مؤلفه غیرزبانی > مهارت خواندن	0.56 2	0.560	0.015	38.366	0.000	0.53 3	0.589

۵. بحث

سؤال تحقیق مبتنی بر بررسی و کشف ریزمؤلفه‌های گروه زبانی و گروه غیرزبانی بود که در نمره حاصله از آزمون مهارت خواندن، براساس داده‌های جمع‌آوری‌شده، بیشترین مشارکت را داشتند. به‌عنوان یک بررسی داده‌محور، تنها ریزمؤلفه‌هایی که در سؤالات آزمون گنجانده شده بودند، بررسی شدند.

در تحقیق حاضر، دانش زبانی آزمون‌وران به صورت‌های متفاوت در آزمون کدگذاری شده‌بود؛ برای مثال، مترادف و متضاد کلمات تحت عنوان «دانش واژگانی»، تقلیل جملات متنی در قالب کلمات منحصربه‌فرد در سؤالات آزمون، تحت عنوان «جملات کوتاه‌شده در قالب کلمات»، استفاده از کلمات توصیفی آشکار در سؤالات جهت بررسی فهم ظاهری جملات، با عنوان «تشخیص تفاوت بین اطلاعات و ایده‌ها»، پاسخ به سؤالاتی که در متن قابل ردیابی بودند، تحت عنوان «دریافت اطلاعات»، و سؤالاتی که پی‌گیر اسامی خاص یا اطلاعات بسیار جزئی در متن‌ها بودند، تحت عنوان «درک جزئیات»، از مؤلفه‌های این زیرتوانش زبانی بودند. سؤالاتی که روی دریافت مرجع ضمائر و درک انسجام متن تأکید می‌کردند، با نام‌های: «استفاده از متن»، و «توالی وقایع»؛ و سؤالاتی که ایده اصلی متن را هدف قرار داده بودند با نام

«تشخیص ایده اصلی متن» در مؤلفه‌های زبانی آزمون، گنجانده شده بودند.

بررسی‌ها نشان داد که «دانش واژگانی» به‌طور کلی، چه درمورد سؤالات وابسته به مهارت‌های زبانی، و چه درمورد سؤالات وابسته به مهارت‌های غیرزبانی، ارتباط تنگاتنگی با موفقیت در آزمون خواندن دارند. گرچه، پژوهشگرانی چون کولت (2006)، یافته این پژوهش را تأیید می‌کنند؛ گروهی دیگر از پژوهشگران، شروع از جزء به کل را به اندازه کافی مؤثر ندانسته‌اند (Zarei & Shakoori, 2014; Chehrzad & Ajideh, 2013; Brame & Biel, 2015).

نتایج مطالعه پژوهشگرانی چون پرفتی (2005)، لافر (1992)، شونن و همکاران (1998)، نیز با نتایج تحقیق حاضر همسوست. آن‌ها معتقدند دانش واژگانی درمقایسه با دیگر راهبردهای فراشناختی در درک مطلب اهمیت بیشتری دارد. همچنین، دیدگاه پیکره‌ای، و حافظه‌محور (Sinclair, 1991; Skehan, 1998) دستور زبان را جنبه‌ای از دانش واژگانی می‌داند که در پیوند با واژگان اطرافش در متن نقش ایفا می‌کند. یافته‌های تحقیق حاضر، در راستای این نظریه نیز است.

دانش غیرزبانی آزمون‌وران، با سؤالاتی مربوط به استنباط، نتیجه‌گیری، هدف نویسنده، دریافت ضمنی علت و معلولی، دانش و تجربیات قبلی، بیان دیدگاه و تصویر سازی اطلاعات، در این پژوهش سنجیده شد. از میان این مهارت‌ها، استنباط و نتیجه‌گیری بیشترین همبستگی را نشان دادند. تطابق یافته‌های رنجبران و علوی (2017)، و فلچر (2009) با یافته‌های این پژوهش قابل تأمل است. این مسئله شاید به این دلیل باشد که این دو خرده‌مؤلفه در تقسیم‌بندی بلوم (۱۹۵۶)، از جمله سطوح فکری بالاتر هستند که به درک عمیق‌تری از جملات نیازمندند. استنباط در مقایسه با سایر مؤلفه‌های غیرزبانی بیشترین قدرت پیش‌بینی را برای مهارت خواندن داشت. از سوی دیگر، در میان تمامی مؤلفه‌های زبانی و غیرزبانی، نتیجه‌گیری، بیشترین همبستگی را نشان داد. به دلیل چالش‌برانگیزتر بودن مهارت ترکیب اطلاعات در تقسیم بندی بلوم (۱۹۵۶)، می‌توان این برتری را توجیه کرد و به این نتیجه رسید که زبان‌آموزانی که مهارت عمومی بالاتری در ترکیب اطلاعات دارند، در پاسخ به سؤالات مربوط به نتیجه‌گیری‌های متنی موفق‌تر عمل می‌کنند.

سؤالاتی مثل: «به تصویر کشیدن اطلاعات» که نیاز به دانش غیرزبانی در سطح کاربرد، علاوه بر ترکیب داشتند، در نتایج آزمون، همبستگی معنی‌داری را نشان ندادند. درواقع، می‌توان

این‌گونه برداشت کرد که سؤالاتی از این‌دست، برای سطح متوسط شایستگی زبانی، پیش‌بینی‌کننده مهارت خواندن نیستند.

به‌طور کلی، نتایج حاصله، حکایت از برتری هم‌بستگی مؤلفه‌های غیرزبانی با مهارت خواندن دارد. درحالی‌که، طبق نظریه پژوهشگرانی نظیر ویر (2005)، و آلدرسون (2000) این برتری، ممکن است نشانه آلوده‌شدن آزمون با متغیرها و مؤلفه‌هایی فراتر از مؤلفه‌های مؤثر در مهارت خواندن در زبان دوم باشد.

نتایج تحقیقات پژوهشگرانی چون استاجی و ژاله (2020)، شیبانی و زراعت‌پیشه (2018)، شه‌میرزادی و همکاران (2020)، با نتایج پژوهش حاضر همسوست، درحالی‌که، در تضاد با یافته‌های پژوهش روحانی و همکاران (2017) است. البته، بررسی سؤالات درک‌مطلب آزمون سراسری که توسط همتی و همکاران (2016) انجام شده است، تناسب میزان استفاده از مؤلفه‌های غیرزبانی و زبانی را در محتوای آزمون تأیید می‌کند.

مهارت خواندن در زبان دوم، بنابراین باید با احتیاط مورد ارزیابی و سنجش قرار گیرد. این مطلب در طراحی تست‌های چهارگزینه‌ای استاندارد داخل کشور، مانند آزمون ای‌پی‌تی، ام‌اس‌آرتی، و آزمون ورودی دانشگاه‌ها (کنکور) بیشتر حائز اهمیت است؛ و به‌خصوص در طراحی آزمون‌هایی با مقیاس کوچک‌تر، مانند یک آزمون معلم‌ساز در سطح کلاسی، می‌تواند این اطمینان را بدهد که سؤالات از اعتبار لازم برای حصول نتیجه، قضاوت، و بازخورد مؤثر برخوردارند.

۶. نتیجه

این مطالعه، برای بررسی عوامل حمایت‌کننده روایی سازه‌ای آزمون چهارگزینه‌ای درک‌مطلب برای زبان‌آموزان سطح متوسط انجام شد، و هدف از آن بررسی توازن مؤلفه‌های زبانی و غیرزبانی سؤالات، جهت حصول اطمینان از عدالت آزمون بود. نتایج حاصله از این مطالعه به شرح زیر است:

ضریب هم‌بستگی پیرسون بین خرده‌مؤلفه‌های دو گروه: ۱. سؤالات مربوط به خرده‌مؤلفه‌های زبانی شامل دانش واژگانی، تقلیل جملات متن به واژگان در سؤالات، حقایق و ایده‌ها، ارجاعات محتوایی و متنی، توالی وقایع، و بررسی ایده اصلی متن، و ۲. سؤالات مربوط

به خرده‌مؤلفه‌های غیرزبانی، ازجمله درک ارتباط ضمنی علی و معلولی، گرفتن نتایج براساس شواهد، درک معانی ضمنی کلمات و اصطلاحات، هدف و گرایش نویسنده، تفسیر و ارزیابی اطلاعات، و تصویرسازی اطلاعات، محاسبه شد. علاوه بر این، ضریب همبستگی پیرسون بین دو گروه مؤلفه‌های زبانی و غیرزبانی نیز محاسبه شد. نتایج نشان دادند که ضریب همبستگی بین زیرمؤلفه‌های هر دو گروه زبانی و غیرزبانی کمتر از ضریب همبستگی میانگین آن‌هاست. در همین راستا، نسبت خصیصه متفاوت - خصیصه یکسان جهت حصول اطمینان از این‌که همگی این مؤلفه‌ها در هر دو گروه، یک متغیر پنهان - مهارت خواندن - را اندازه می‌گیرند، محاسبه شد. ضریب محاسبه‌شده حاکی از این بود که آزمون اعتبار واگرا ندارد؛ بنابراین، می‌توان نتیجه به‌دست‌آمده را به‌عنوان شاخصی برای سطح مهارت آزمون‌وران در خواندن در نظر گرفت.

در تحلیل آماری اکتشافی، رابطه خرده‌مؤلفه دانش واژگانی، از گروه مؤلفه‌های زبانی، با بیشتر سؤالات آزمون آشکار شد، و از گروه مؤلفه‌های غیرزبانی، خرده‌مؤلفه تصویرسازی اطلاعات، با هیچ‌یک از دسته‌های پیشنهادی هماهنگی نداشت.

باتوجه به این‌که پژوهش حاضر مبتنی بر یافتن عوامل مؤثر در روایی آزمون ساخته‌شده براساس نمونه سؤالات آزمون ای‌پی‌تی طراحی شده بود، از مدل تعدیل جزئی نیز برای بررسی میزان مشارکت دو گروه مؤلفه‌های زبانی و غیرزبانی، در سنجش مهارت خواندن استفاده شد. این انتخاب به‌دلیل کم بودن تعداد شرکت‌کننده‌ها در آزمون، در مقایسه با مدل‌سازی معادلات ساختاری^{۴۶} که برای آزمون‌هایی با مقیاس بزرگ استفاده می‌شود، صورت گرفت. نتایج نشان داد که گروه مؤلفه‌های غیرزبانی مشارکت معنی‌دار، اما بزرگ‌تری، نسبت به گروه مؤلفه‌های زبانی، دارند.

بیشتر پژوهش‌های پیشین به بررسی رابطه بین خرده‌مؤلفه‌های مهارت خواندن، با سنجش نهایی آن در زبان اول و دوم پرداخته‌اند (Ranjbaran & Alavi, 2017; Hemmati et al., 2016). پژوهش‌های دیگر یا درباره ساخت آزمون تشخیصی خواندن در زبان اول جهت دانش‌آموزان دوره ابتدایی، تحقیق کرده‌اند (جباری و خادمی، ۱۳۹۲) یا رابطه بین بررسی کیفی مشخصات سؤالات آزمون درک مطلب، قبل از اجرای آزمون را،

به‌عنوان شاهد اعتبار پیشینه^۷، بر بهبود اعتبار آزمون‌های درک‌مطلب در مقیاس‌های کوچک، تحت بررسی قرار داده‌اند (Zandi et al., 2014).

پژوهش حاضر در همین راستا، و با هدف ارائه راه‌کاری برای بهبود کیفیت سؤالات آزمون‌های درک‌مطلب، به‌ویژه، در مقیاس کوچک انجام شده است. یافته‌های این پژوهش می‌تواند در انتخاب متون و سؤالات مناسب و عادلانه برای آزمون‌های درک مطلب، جهت ارزشیابی تشخیصی و تکوینی، معلمان و مدرسان فعال در امر آموزش زبان دوم را یاری دهد. از دیدگاه نظری، مهارت خواندن در فرایند ساخت و طراحی آزمون از مؤلفه‌های متفاوتی بهره می‌برد، که ترکیب تلفیقی این سؤالات، به‌جهت نمردهی بی‌طرفانه، کار اجرای آزمون را سخت‌تر می‌کند. از آنجا که آزمون‌های چندگزینه‌ای از آسانی اجرا و عدالت نمردهی بالایی برخوردارند، سؤالاتی که برای سنجش این مهارت، در آزمون‌های استاندارد داخل کشور گنجانده می‌شوند، باید به روش‌های متعددی ویرایش شوند. به‌طوری که تهدید بیش‌برآورد^۸ (Weir, 2005)، اعتبار نتایج آزمون را تهدید نکند. نگرانی برای بهبود توانش زبانی باید به بهبود دانش واژگانی و دیگر دانش‌های زبان‌شناختی و فراشناختی و همچنین فرازبانی منجر شود؛ به‌طوری که زبان‌آموزان پس از تلاش برای ارتقای موارد ذکرشده، نگران دانش عمومی و تجربه‌های فردی خود برای موفقیت در آزمون‌های استاندارد در سطح متوسط نباشند.

پژوهش پیش‌رو محدودیت‌هایی نیز دارد: این پژوهش با تعداد کمی از آزمودنی‌ها، در مقایسه با دیگر پژوهش‌های مشابه انجام شده است؛ دسترسی به آزمودنی‌ها، در مقیاس بیشتر مانند شرکت‌کنندگان آزمون سراسری، جهت بررسی عدالت آزمون در قسمت درک مطلب زبان دوم، و همچنین بررسی اعتبار انواع دیگری از فعالیت‌ها در آزمون درک‌مطلب، مانند سؤالات جای‌خالی درون‌متنی، به نتایج معتبرتری خواهد انجامید. از دیگر محدودیت‌ها، در تفسیر نتایج این پژوهش، تک‌جنس بودن آزمودنی‌ران و یکسان بودن سطح زبانی آنان است. از آنجایی‌که تفاوت جنسیت و سطح زبانی ممکن است با چگونگی تجربیات قبلی افراد ارتباط مستقیم داشته باشد، پژوهش حاضر سعی در حذف این متغیرها کرده بود، که با دخیل کردن آن‌ها ممکن است نتایج متفاوتی حاصل شود. آزمون‌وران ممکن است برای جبران ضعف در دانش واژگانی و

دیگر خرده مؤلفه‌های زبانی، از راهکارهای منحصربه‌فردی بهره‌جویند. این نکته نیز، از دیگر محدودیت‌های پژوهش حاضر است که باید در استفاده از یافته‌های این پژوهش موردتوجه قرار گیرد.

۷. پی‌نوشت‌ها

1. High-stake tests
2. English Proficiency Test
3. B1.2- B2.1
4. Linguistic Components
5. Non- Linguistic Components
6. Language Proficiency Tests
7. Black & William
8. Language Competence
9. Knowledge Structures (Knowledge of the world)
10. Strategic Competence
11. Synthesis
12. Inferences
13. Evaluation
14. Construct Validity
15. Theory-based Validity
16. Context Validity
17. Scoring Validity
18. Cognitive Validity
19. External Validity
20. Process-Oriented
21. Component-Oriented
22. Davis
23. Munby
24. Lunzer, Waite, & Dolan
25. Coady
26. Perfetti
27. Vandergrift, & Goh
28. Zhang
29. Bottom-up
30. Top-down
31. Interactive
32. Kintsch
33. Van Dijk

34. Rumelhart
35. Stovich
36. Cognitive Diagnostic Assessment (G DINA) model
37. 1-Parameter and 2-Parameter Logistic Item Response Theory (IRT)
38. RUM
39. metalinguistic
40. Cognitive Diagnostic Assessment
41. Heterotrait-monotrait ratio (HTMT)
42. Exploratory Factor Analysis (EFA)
43. Partial Least Square Modeling (PLSPM)
44. IBM SPSS 26 2019
45. Standardized regression weights
46. Structural Equation Modeling
47. A priori evidence validation
48. Overestimation

۸ منابع

- امیریان، م. ر.، حشمتی‌فر، ز.، زارعیان، غ.، و داوودی، م. (۱۳۹۷). واکاوی کیفی عوامل مؤثر بر شرکت در دوره‌های آمادگی آزمون زبان انگلیسی وزارت علوم، تحقیقات و فناوری (MSRT): با رویکرد نظریه داده‌بنیاد. *پژوهش‌های زبان‌شناختی در زبان‌های خارجی*، ۸ (۱)، ۵۶-۲۷.
- جباری، س.، و خادمی، م. (۱۳۹۲). ساخت آزمون تشخیصی خواندن و درک مطلب جهت دانش‌آموزان دوره ابتدایی. *پژوهش‌های برنامه درسی*، ۳ (۲)، ۵۱-۳۳.
- جعفرپور، ع. (۱۳۷۵). *مقدمه‌ای بر آزمون‌شناسی زبان*. شیراز: انتشارات دانشگاه شیراز.
- صحرایی، ر. م.، و ممقانی، ه. (۱۳۹۱). ارزیابی روایی و پایایی آزمون زبان انگلیسی وزارت علوم، تحقیقات و فناوری. *اندازه‌گیری تربیتی*، ۳ (۱۰)، ۲۰-۱.

References

- Abdelaal, N. M., & Sase, A. S., (2014). Relationship between prior knowledge and reading comprehension. *Advances in Language and Literary Studies*, 5(6), 125-131.
- Adams, M. J., & Collins, A. (1977). *A schema-theoretic view of reading comprehension (Report No. 032)*. Center for the Study of Reading Technical

Report.

- Amirian, M. R., Heshmatifar, Z., Zareian, Gh., & Davoudi, M. (2018). Qualitative analysis of the factors affecting participation in English language preparation courses of the Ministry of Science, Research and Technology (MSRT): A data-driven approach. *Linguistic Research in Foreign Languages*, 8(1), 27-56. [In Persian]
- August, D., & Shanahan, T. (2010). Developing literacy in second-language learners: Report of the national literacy panel on language-minority children and youth. *Journal of Literacy Research*, 42(3), 341-348.
- Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.
- Alderson, J. C., Clapham, C., & Wall, D. (2005). *Language test construction and evaluation*. Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Black, P., & William, D. (1998). Assessment and Classroom Learning: Assessment in Education. *Assessment in Education*, 5, 7-74.
- Brame, C., & Biel, R. (2015). Test-enhanced learning: The potential for testing to promote greater learning in undergraduate science courses. *CBE Life Sciences Education*, 14(2), 1-12.
- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Longman University Press.
- Carr, J. F., & Harris, D. E. (2001). *Succeeding with standards linking curriculum, assessment, and action planning*. Association for Supervision and Curriculum Development.
- Caldwell, J. S. (2008). *Comprehension assessment: A classroom guide*. Guilford Publication Press.
- Chehrzad, M. H., & Ajideh, P. (2013). Effects of different response types on

Iranian EFL test-takers' performance. *Iranian Journal of Applied Language Studies*, 5(2), 29-50.

- Coady, J. (1979). A psycholinguistic model for the ESL reader. In R. MacKay, B. Barkman, & R. R. Jordan (Eds.), *Reading in a second language: Hypothesis, organization and practice* (5-12). Newbury House.
- Council of Europe. (2011). *Manual for language test development and examining: ALTE*. Strasburg. Language Policy Division.
- <https://www.coe.int/en/web/language-policy/home>
- Davis, A. (1968). *Language testing symposium: A psycholinguistic approach*. Oxford University Press.
- Estaji, M., & Zhaleh, K. (2020). Does field of study matter in academic performance: Differential item functioning analysis of a high-stakes test using one-parameter and two-parameter item response theory models. *Iranian Journal of English for Academic Purposes*, 9(3), 14-31.
- Farhadi, H., & Hessami, G. R. (2005). Construct validity of L2 reading comprehension skills. *International Journal of Applied Linguistics*, 8(2), 29-53.
- Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading*, 10, 323-330.
- Garson, G. D. (2016). *Partial least squares: Regression and structural equation models*. Statistical Associates Publishers.
- Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly*, 25, 375-406.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press.
- Hair, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage Publications.
- Hedge, T. (2003). *Teaching & learning in the language classroom*. OUP.
- Hemmati, S. J., Baghaei, P., & Bemani, M. (2016). Cognitive diagnostic modeling

of L2 reading comprehension ability: providing feedback on the reading performance of Iranian candidates for the university entrance examination. *International Journal of Language Testing*, 6(2), 92-100.

- Hoover, W. A., & Tunmer, W. E. (1993). The components of reading. In G. G. Thompson, W. E. Tunmer, & T. Nicholson (Eds.), *Reading acquisition processes* (1-19). Multilingual Matters.
- Jabbari, S., & Khademi, M. (2013). Creating a diagnostic test for reading comprehension of elementary school students. *Curriculum Research*, 3(2), 33-51. [In Persian]
- Jafarpour, A. J. (1996). *Introduction to Language Testing*. Shiraz University Press. [In Persian]
- Kaplan, R. M., & Saccuzzo, D. P. (2001). *Psychological testing: Principles, applications, and issues* (5th Ed.). Wadsworth.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Laufer, B. (1992). Reading in a foreign language: How does L2 lexical knowledge interact with the reader's general academic ability?. *Journal of Research in Reading*, 15(1), 95-103.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th Ed.). Merrill, Prentice Hall.
- Lunzer, E., Waite, M., & Dolan, T. (1979). Comprehension and comprehension test. In E. Lunzer & K. Gardner (Eds.), *The effective use of reading* (37-71). Heinemann Educational Books Ltd.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (13-103). Macmillan.
- Munby, J. (1978). *Communicative syllabus design*. Cambridge University Press.
- Nunan, D. (1991). *Language teaching methodology*. London: Prentice-Hall.

- Palomba, C., & Banta, T.W. (1999) *Assessment essentials: Planning, implementing, and improving assessment in higher education*. Jossey-Bass Publishers.
- Paris, S. G., & Hamilton, E. E. (2009). The development of children's reading comprehension. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (32-53). Routledge.
- Pasquarella, A., Gottardo, A., & Grant, A. (2012). Comparing factors related to reading comprehension in adolescents who speak English as first (L1) or second (L2) languages. *Scientific Studies of Reading*, 1(6), 1-29.
- Perfetti, C. (1985). *Reading ability*. New York: Oxford University Press.
- Perfetti, C., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skills. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: a handbook* (227-247). Blackwell.
- Perfetti, C., Landi, N., & Oakhill, J. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 8, 293-304.
- Quellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554-566.
- RAND Reading Study Group. (2002). *Reading for understanding: Towards an R&D program in reading comprehension*. OERI.
- https://www.rand.org/content/dam/rand/pubs/monograph_reports/2005/MR1465.pdf
- Ranjbaran, F., & Alavi, S. M. (2017). Developing a reading comprehension test for cognitive diagnostic assessment: A RUM analysis. *Studies in Educational Evaluation*, 55, 167-179.
- Rogde, K., Hagen, A. M., Melby-Lervag, M., & Lervag, A. (2019). The effect of linguistic comprehension instruction on generalized language and reading comprehension skills: A systematic review. *Campbell Systematic Reviews*, 15(4),

1-37.

- Roohani, A., Dayeri, K., & Farhang Ju, M. (2017). Role of background knowledge in Iranian EFL learners' reading comprehension test performance. *International Journal of Language Testing*, 7(1), 28-39.
- Rumelhart, D. (2004). Toward an interactive model of reading. In R. B. Rudell & N. J. Unrau (Eds.), *Theoretical models and processes of reading*. (1149-1179). International Reading Association.
- Sahraei, R., & Mamaghani, H. (2012). Evaluation of the validity and reliability of the English language test of the Ministry of Science, Research and Technology. *Educational Measurement Quarterly*, 3(10), 1-20. [In Persian]
- Schoonen, R., Hulstijn, J., & Bossers, B. (1998). Meracognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in grades 6, 8 and 10. *Language Learning*, 48(1), 71-106.
- Shirvan, M., Ghonsooly, B., & Fatemi, A. (2016). The Effectiveness of strategy-based instruction in teaching English as a second or foreign Language: A meta-analysis of experimental studies. *US-China Foreign Language*, 14(3), 163-181.
- Shahmirzadi, N., Siyyari, M., Marashi, H., & Geramipour, M. (2020). Test fairness analysis in reading comprehension PhD nationwide admission test items under CDA. *Journal of Foreign Language Research*, 10(1), 152-165.
- Sheybani, E., & Zeraatpishe, M. (2018). On the dimensionality of reading comprehension tests composed of text comprehension items and cloze test items. *International Journal of Language Testing*, 8(1), 12-26.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Stovich, K. E. (2000). *Progress in understanding reading: Scientific foundations*

and new frontiers. Guilford Press.

- Urquhart, A. H., & Weir, C. J. (1998). *Reading in a second language: Process, product, and practice*. Longman.
- Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.
- Van der Walt, J. L., & Steyn, H. S. (2008). The validation of language tests. *Stellenbosch Papers in Linguistics*, 38, 191-204.
- Van Dijk, T. A. (1977). *Text and context: Explorations in the semantics of text*. Longman.
- Weir, C. J. (2005). *Language test validation: An evidence-based approach*. Palgrave.
- Zandi, H., Keivanpanah, S., & Alavi, S. M. (2014). The effect of test specifications review on improving the quality of a test. *Iranian Journal of Language Teaching Research*, 2(1), 1-14.
- Zarei, A. A., & Shakoory Nia, S. (2014). The effect of vocabulary, syntax, and discourse-oriented activities on short and long-term L2 reading comprehension. *International Journal of Language and Linguistics*, 1(1), 30-39.
- Zhang, D. (2012). Vocabulary and grammar knowledge in second language reading comprehension: A structural equation modelling study. *The Modern Language Journal*, 96(4), 558-575.