

کاربرد سنجش تشخیصی شناختی به منظور تحلیل سؤال‌های درک مطلب کنکور کارشناسی ارشد رشته زبان انگلیسی

ریحانه رحیمی^۱، جلیل یونسی^۲، مهران مکرمی^۳

تاریخ دریافت: ۹۷/۰۲/۱۶

تاریخ پذیرش: ۹۷/۰۵/۲۵

چکیده

هدف پژوهش حاضر تعیین مهارت‌های پایه آزمون خواندن و درک مطلب و تحلیل سؤال‌های آزمون و درنهایت تعیین نقاط ضعف و قوت آزمودنی‌ها از طریق مدل‌های سنجش تشخیصی شناختی با تحلیل بازبازش بود. داده‌های مورد بررسی در این پژوهش شامل پاسخنامه ۳۰۰۰ نفر از شرکت‌کنندگان در آزمون سراسری کارشناسی ارشد رشته زبان انگلیسی بود که از بین ۱۶۰۴۴ نفر شرکت‌کننده به تصادف انتخاب شده بود. در ابتدا با کمک نظر متخصصین یک ماتریس Q که شامل روابط بین ۲۰ سؤال آزمون خواندن با ۷ مهارت پایه بود تشکیل گردید و سپس با به کارگیری مدل G-DINA تجزیه و تحلیل نهایی انجام شد. درنهایت با استفاده از روش پسینی مورد انتظار در چهارچوب مدل G-DINA احتمال تسلط هر فرد بر هر یک از مهارت‌ها با کمک بسته CDM در محیط نرم‌افزار R محاسبه شد و با تعیین نقطه برش ۰/۶ برای حد تسلط، افراد مورد بررسی در هر یک از مهارت‌ها به دو گروه مسلط و غیر مسلط تقسیم شدند. یافته‌های حاصل از پژوهش نشان داد که در هر یک از مهارت‌ها بیشتر افراد به حد تسلط نرسیده‌اند، البته وضعیت افراد در مهارت SUM (تجزیه و تحلیل و ارزیابی اهمیت نسبی اطلاعات موجود در متن با کمک تفاوت قائل شدن بین ایده اصلی و جزئیات) نسبت به سایر مهارت‌ها بهتر بود و حدود نیمی از افراد در این مهارت به حد تسلط رسیده بودند. نتیجه‌ی نهایی گویای این مطلب بود که به‌طور کلی وضعیت آزمودنی‌ها در کاربرد راهبردهای شناختی برای رسیدن به پاسخ صحیح نامطلوب است و باید این راهبردها در سرفصل‌های درسی مهارت خواندن و درک مطلب گنجانده شود.

۱. دانشجوی دکتری سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران. (نویسنده مسئول)

reyhanerahimi1367@gmail.com

۲. دانشیار گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران.

۳. دانشجوی دکتری سنجش و اندازه‌گیری، دانشگاه تهران، تهران، ایران.

واژگان کلیدی: برآورد پسینی مورد انتظار، تحلیل باز برازش، سنجش تشخیصی شناختی، ماتریس Q، مدل G-DINA

مقدمه

زبان شناسان ۴ مهارت شنیدن^۱، گفتن^۲، خواندن^۳ و نوشتن^۴ را از مهارت‌های اصلی یادگیری زبان معرفی می‌کنند. از میان این ۴ مهارت، مهارت درک مطلب و خواندن یکی از مهارت‌های اساسی برای یادگیری زبان‌های خارجی است؛ به طوری که آلدerson^۵ (۲۰۰۰) ادعا می‌کند در یادگیری زبان انگلیسی بسیاری از مباحث و سؤال‌های مرتبط با یادگیری به مهارت خواندن و درک مطلب مربوط می‌شود. کدا^۶ (۲۰۰۵) معتقد است که خواندن و درک مطلب فرایند پیچیده‌ای است که آزمودنی برای رسیدن به نتیجه مطلوب باید بین متن و ذهن خود ارتباط برقرار کند و استراتژی‌های مختلفی را به کاربرد. پژوهش‌های صورت گرفته، عموماً مهارت درک مطلب را فرایندی تعاملی و اجتماعی - شناختی معرفی می‌کنند (برنهاردت^۷، ۱۹۹۳). در دورانی که نظریه رفتارگرایی بر سیستم‌های آموزشی حاکم بود، شیوه‌های آموزش و سنجش زبان نیز تحت تأثیر این نظریه بود؛ اما با ظهور نظریه شناختی در روان‌شناسی و تأثیر زبان‌شناسی گشتاری - زایشی بر علم زبان‌شناسی، زبان‌شناسی رفتاری رو به افول گذاشت. در این سیستم یادگیری، دیگر نتیجه تکرار و تقلید نیست؛ بلکه یادگیری با فرایندهای ذهنی پیچیده مرتبط است. با ایجاد این تحولات، شیوه‌های آموزش زبان و به دنبال آن سنجش آن تغییر کرد و منجر به طراحی آزمون‌های جدیدی همچون IELTTS^۸، OTESOL^۹ گردید (فرهادی، ۱۳۸۱). از جمله آزمون‌هایی که در زمینه سنجش زبان در کشور ما برگزار می‌شود می‌توان به آزمون و آزمون‌های ورودی دانشگاه در سه مقطع کارشناسی، کارشناسی ارشد و دکتری رشته زبان انگلیسی اشاره کرد. با توجه به اینکه

1. Listening
2. Speaking
3. Reading
4. Writing
5. Alderson
6. Koda
7. Bernhardt
8. International English Language Testing Service
9. of English for Speaker of Other Languages Ontario Test

آزمون‌های ورودی دانشگاه‌ها به منظور جذب دانشجو با هدف پرورش نیروی متخصص انجام می‌گیرد، دارای جایگاه ویژه‌ای است. با وجود اهمیت این آزمون‌ها و هزینه‌های اقتصادی که برای داوطلبان و دولت دربردارد و نیز اثرات مثبت و منفی که بر داوطلبان و نظام آموزشی دارد، به طور شایسته و همه جانبه‌ای اطلاعات حاصل از آن مورد تجزیه و تحلیل قرار نمی‌گیرد. به عبارت دیگر نتایج یک آزمون در صورتی که به طور مناسبی تفسیر شوند می‌تواند اطلاعات مفیدی درباره پیشرفت تحصیلی شاگردان در اختیار مجریان و مدیران آموزشی قرار دهد و از این طریق در زمینه آموزش صحیح به آنان کمک کند. نمره‌های حاصل از یک آزمون این ظرفیت را دارند که به مدرسان و آزمودنی‌ها در شناسایی نقاط قوت و ضعف آنان و اقدام به فعالیت‌های جبرانی یاری دهند (هومن، ۱۳۶۹). تاکر^۱ (۲۰۱۰) سنجش مطلوب را سنجشی می‌داند که در کنار استانداردهای روانسنجی دقیق، بازخوردهایی راجع به اینکه آزمودنی‌ها چگونه یاد گرفته‌اند و چه ویژگی‌هایی برای رسیدن به هدف نیاز دارند را فراهم کند. از بین انواع آزمون‌های پیشرفت تحصیلی، آزمون‌هایی که با هدف تشخیص طراحی شده‌اند از این ویژگی برخوردار هستند. در واقع آزمون‌های تشخیصی برای تشخیص نقاط ضعف و قدرت آزمودنی‌ها در مهارت‌های اساسی به کار می‌روند و با تجزیه و تحلیل عناصر مختلف پیشرفت و مهارت‌های اصلی یادگیری می‌توان به شکل مؤثری از این آزمون‌ها در زمینه بهبود یادگیری بهره برد (شریفی، ۱۳۹۰). با اهمیت یافتن سنجش تشخیصی از یک سو و با مطرح شدن نظریه‌های شناختی از سوی دیگر و همچنین با فعالیت‌های اسنو و لهمان^۲ (۱۹۸۹) در زمینه ترکیب اصول روان‌شناسی شناختی با شیوه‌های اندازه‌گیری، رویکرد جدیدی در سنجش تحت عنوان سنجش شناختی تشخیصی^۳ (CDA) پدید آمد.

سنجش تشخیصی شناختی نوعی از سنجش آموزشی است که برخلاف مدل‌های سنتی سنجش این توانایی را دارد که ساختارهای دانش و فرایندهای مهارت را در آزمودنی‌ها اندازه بگیرد و اطلاعاتی راجع به نقاط ضعف و قدرت آنان فراهم کند. مدل‌های تشخیصی شناختی^۴ (CDMs) که در رویکرد سنجش تشخیصی شناختی پیشنهاد می‌گردند؛ مدل‌های

1. Toker
2. Snow & Lohman
3. Cognitive Diagnostic Assessment
4. Cognitive Diagnostic Models

ریاضی را با مدل‌های روانسنجی معروف مثل نظریه سؤال پاسخ ترکیب می‌کنند تا از این طریق هم خواسته‌های شناختی سؤال و هم سطح توانایی آزمودنی‌ها را نشان دهند (لیتون و گریل، ۲۰۰۷). مدل‌های تشخیصی شناختی مانند تحلیل عامل تأییدی در ذات تأییدی هستند. در این مدل‌ها متغیر مکنون که همان مهارت‌ها هستند، به وسیله سؤال‌های موجود در ماتریس Q اندازه گرفته می‌شوند. ماتریس Q فرضیه‌هایی درباره مهارت‌های لازم برای پاسخ درست هستند (لی، ۲۰۱۱) و معمولاً عناصر این ماتریس بر اساس قضاوت متخصصان قبل از تجزیه و تحلیل مدل تشخیصی شناختی تعیین می‌شود و در تجزیه و تحلیل‌های آماری درستی این روابط بررسی می‌شود. در مجموع مدل‌های تشخیصی شناختی به دنبال آن هستند، که با بررسی فرایندهای ذهنی اطلاعات دقیقی را در زمینه درک و فهم آزمودنی‌ها از حیطه-ای که مورد سنجش قرار گرفته است، ارائه کنند. اهمیت و کارایی این نوع از ارزیابی در بررسی حیطه‌ها و موضوعاتی بیشتر نمایان می‌شود؛ که آزمودنی برای پاسخگویی به سؤال‌های آن حوزه نیاز به مهارت‌های مختلف و استفاده از فرایندهای عالی ذهنی دارد. از همین روی بیشتر پژوهش‌های صورت گرفته در زمینه کاربرد مدل‌های تشخیصی شناختی به آزمون‌های ریاضیات و زبان (به ویژه مهارت درک مطلب) که نیازمند مهارت‌های شناختی هستند، یافته است. پژوهش‌های تشخیصی شناختی که تاکنون در داخل کشور انجام شده است نیز در زمینه آزمون‌های ریاضیات (مینایی، ۱۳۹۱؛ افضل‌ی، ۱۳۹۳؛ محسن پور، ۱۳۹۳)، علوم تجربی (کبیری، ۱۳۹۲) و زبان انگلیسی (مقدم، ۱۳۹۴) بوده است.

اصلی‌ترین کاربرد مدل‌های تشخیصی شناختی در زمینه ساخت آزمون‌های تشخیصی است؛ یعنی اینکه پژوهشگر با هدف تشخیص نقاط ضعف و قوت آزمودنی‌ها و نارسایی‌های موجود در سیستم آموزشی اقدام به طراحی یک آزمون تشخیصی کرده و بعد از اجرای آزمون طراحی شده داده‌های حاصل از آن را با مدل تشخیصی مناسب مورد تجزیه و تحلیل قرار می‌دهد. این تنها کاربرد این مدل‌ها نیست؛ بلکه داده‌های موجود که از آزمون‌های غیر تشخیصی به دست آمده‌اند را نیز می‌توان با مدل‌های تشخیصی شناختی بررسی کرد. در روش دوم که به تحلیل بازبرازش معروف است، از داده‌های آزمون‌های غیر تشخیصی با هدف تشخیص استفاده می‌شود. با توجه به این قابلیت مدل‌های تشخیصی شناختی، می‌توان

-
1. Leighton & Gierl
 2. Li

داده‌های حاصل از هر آزمونی که قبلاً اجرا شده است - حتی اگر هدف تشخیصی نیز نداشته باشد - را در چارچوب تشخیصی شناختی بررسی نمود. به عنوان مثال آزمون تافل^۱ بارها توسط پژوهشگران مختلف با کمک تحلیل بازبرازش مورد ارزیابی واقع شده است. کسای^۲ (۱۹۹۷)، باک^۳ و همکاران (۱۹۹۷) و اسکات^۴ (۱۹۹۸) با مدل فضای قاعده، ون داویر^۵ (۲۰۰۵) با مدل GDM، جانگ^۶ (۲۰۰۹) با مدل RUM، لی و سوواکی^۶ (۲۰۰۹) با مدل GDM و RUM و یی^۷ (۲۰۱۳) با مدل‌های LCDM DINA DINO و ACDM آزمون تافل را مورد مطالعه قرار داده‌اند؛ ولی هر کدام تعداد مهارت‌های متفاوتی را برای پاسخ به سؤال‌ها تعیین کردند. در مجموع بر سر تعداد مهارت‌های زیربنایی مهارت خواندن و درک مطلب اختلاف نظر وجود دارد؛ به عنوان مثال مانی^۸ (۱۹۸۱) ۱۹ مهارت پایه برای خواندن تعریف می‌کند، در مقابل برخی از صاحب‌نظران دیگر تنها دو مهارت لغت و رمزگشایی را به عنوان مهارت زیربنایی خواندن در نظر گرفته‌اند (جانسون و رینولدز^۹، ۱۹۴۰، پتی و کوکریل^{۱۰}، ۱۹۷۴، ورنون^{۱۱}، ۱۹۶۲). در رابطه با کاربرد مدل‌های تشخیصی شناختی در حیطه آزمون‌های پیشرفت تحصیلی زبان می‌توان به پژوهش اسوتینا، گرین و تاتسوکا^{۱۲} (۲۰۱۲) بر روی آزمون ورودی دبیرستان، پژوهش ژانگ^{۱۳} (۲۰۱۳) بر روی آزمون خواندن مدرسه و مقدم (۱۳۹۴) بر روی کنکور دکتری رشته زبان انگلیسی اشاره کرد.

اختلاف نظر دیگری که بر سر آزمون‌های خواندن و درک مطلب وجود دارد، جبرانی بودن یا نبودن مهارت‌های زیربنایی است؛ به این معنا که آیا فقدان یک مهارت از طریق وجود سایر مهارت‌ها قابل جبران هست یا نه. از همین روی تعدادی مدل اشباع‌شده یا کلی که به نسبت به مدل‌های جبرانی و غیر جبرانی انعطاف پذیرترند، به وجود آمده‌اند. در پژوهشی

1. TOEFL
2. Kasai
3. Buck
4. Scott
5. von Davier
6. Lee & Sawaki
7. Yi
8. Munby
9. Jnson & Reynolds
10. Pettit & Cockriel
11. Vernon
12. Svetina, Gorin & Tatsuka
13. Zhang

که توسط لی، هانترو و لی^۱ (۲۰۱۵) بر روی آزمون خواندن ملیب انجام شد دو مدل جبرانی "ACDM, DINO" با دو مدل غیر جبرانی "RRUM, DINA" و یک مدل اشباع شده "G-DINA" با یکدیگر مقایسه شدند. نتیجه این بود که مدل‌های ACDM و G-DINA برآزش بهتری با داده‌ها نسبت به سایر مدل‌ها داشتند. در پژوهشی دیگر که توسط راوند و رویچ^۲ (۲۰۱۵) بر روی آزمون خواندن درک مطلب ورودی دانشگاه‌های ایران صورت گرفت، شاخص‌های برآزش نسبی مدل G-DINA با دو مدل ACDM و DINA مورد مقایسه قرار گرفت؛ نتیجه این مقایسه برتری مدل G-DINA بر دو مدل دیگر را تأیید کرد. در نهایت، با توجه به اختلاف نظر موجود در مورد جبرانی یا غیرجبرانی بودن مهارت‌های زیربنایی درک مطلب و با در نظر گرفتن نتایج پژوهش‌هایی که از مقایسه بین مدل‌های تشخیصی شناختی مختلف به دست آمده است؛ مدل کلی G-DINA برای تجزیه و تحلیل داده‌های پژوهش حاضر استفاده شد.

با جمع‌بندی مطالب ذکر شده، هدف از پژوهش پیش‌رو، بررسی سؤال‌های درک مطلب کنکور کارشناسی ارشد رشته زبان انگلیسی در چارچوب سنجش تشخیصی و با کمک مدل کلی G-DINA است؛ تا از این طریق ضمن تعیین مهارت‌های شناختی مورد نیاز برای پاسخ به این سؤال‌ها وضعیت تسلط شرکت کنندگان و نقاط ضعف و قوت آن‌ها را مشخص گردد. با عنایت به این نکته که کنکور کارشناسی ارشد در بین دو مقطع تحصیلی کارشناسی و کارشناسی ارشد قرار دارد، بنابراین با نتایج حاصل از بررسی آن می‌توان برای هر دو مقطع برنامه‌ریزی نمود. به عبارت دیگر با بررسی کنکور کارشناسی ارشد در چارچوب مدل‌های تشخیصی شناختی، از یک طرف می‌توان به نقاط ضعف آزمودنی‌ها به ویژه در حوزه‌هایی که اکثر افراد در آن دچار مشکل هستند پی برد و از این طریق با برنامه‌ریزی شایسته و مناسب برای مقطع کارشناسی به سایر دانشجویانی که وارد این مقطع می‌شوند کمک کرد، به گونه‌ای که نارسایی‌های موجود در دانشجویان بعدی برطرف شود. از طرف دیگر می‌توان به برنامه‌ریزی در مقطع کارشناسی ارشد کمک کرد و با بازخورد دادن به داوطلبان و نیز مراجع سیاست‌گذار در حوزه تعیین سرفصل‌های درسی برای کمک به رفع نقاط ضعف این افراد اقدام کرد.

-
1. Li, Hunter & Lei
 2. Ravavd & Robitzsch

روش پژوهش

با توجه به اینکه پژوهش حاضر از تکنیک سنجش تشخیصی شناختی برای تحلیل بخشی از داده‌های کنکور کارشناسی ارشد رشته زبان که قبلاً (سال ۱۳۹۴) اجرا شده است؛ استفاده کرده است جزء تحقیقات ثانویه^۱ به حساب می‌آید. از طرف دیگر به دلیل اینکه سنجش‌های تشخیصی شناختی به دنبال تشخیص نقاط ضعف و قوت آزمودنی‌ها برای بهبود یادگیری و بالا بردن کیفیت تدریس و تدوین برنامه مناسب هستند؛ این پژوهش از نظر هدف در طبقه تحقیقات کاربردی قرار می‌گیرد.

جامعه آماری پژوهش حاضر شامل کلیه داوطلبانی می‌شود که در آزمون سراسری کارشناسی ارشد سال ۱۳۹۴ رشته زبان انگلیسی شرکت کرده‌اند. طبق آمار دریافتی از سازمان سنجش و آموزش کشور تعداد شرکت‌کنندگان رشته زبان در سال ۱۳۹۴، ۱۶۰۴۴ نفر هستند که از این تعداد ۱۲۰۷۶ نفر زن و ۳۹۶۸ نفر مرد می‌باشند. تعداد کل افراد گروه نمونه که پاسخنامه آن‌ها، مورد بررسی قرار گرفت، ۳۰۰۰ نفر می‌باشد. از این ۳۰۰۰ نفر افراد گروه نمونه، ۲۰۵۸ نفر زن و ۹۴۲ نفرشان مرد هستند. با توجه به اینکه نمونه مورد نظر بنا به درخواست پژوهشگر از بانک داده‌های سازمان سنجش و آموزش کشور تهیه شد و پژوهشگر هیچ دخالتی در امر انتخاب نداشت و سازمان سنجش به صورت تصادفی داده‌های درخواستی پژوهشگر را از بین کلیه داده‌ها موجود انتخاب کرده است؛ می‌توان روش نمونه‌گیری پژوهش را نمونه‌گیری تصادفی ساده محسوب کرد.

ابزار اندازه‌گیری مورد استفاده در پژوهش حاضر آزمون ورودی دوره کارشناسی ارشد ناپیوسته مجموعه زبان انگلیسی سال ۱۳۹۴ که توسط سازمان سنجش و آموزش کشور طراحی شده است، می‌باشد؛ که از بین ۲۴۰ سؤال موجود در این دفترچه ۲۰ سؤال قسمت زبان عمومی که مربوط به مهارت درک مطلب است برای مطالعه در نظر گرفته شده است (سؤال‌های ۴۱ تا ۶۰ دفترچه عمومی). روش نمره‌گذاری سؤال‌های آزمون به صورت صفر و یک (دو ارزشی) است. در صورتی که فرد در پاسخ به یک سؤال گزینه صحیح را انتخاب کرده باشد، نمره یک و اگر گزینه‌های اشتباه را انتخاب کرده یا اصلاً پاسخ نداده باشد نمره صفر می‌گیرد؛ بنابراین نمره کل داوطلبی که موفق شود به همه سؤال‌ها پاسخ صحیح بدهد

۱. استفاده از روش‌های آماری جدید جهت بررسی مجدد داده‌هایی که قبلاً گردآوری شده‌اند.

برابر با نمره بیست و نمره کل داوطلبی که نتواند به هیچ یک از این بیست سؤال، پاسخ صحیح بدهد برابر با نمره صفر است.

نتایج پژوهش

هر تحلیل از نوع سنجش تشخیصی شناختی به دو دسته مختلف از داده‌ها نیاز دارد. دسته اول شامل پاسخنامه داوطلبانی است که در آزمون شرکت کرده‌اند و دسته دوم شامل مدل شناختی که در قالب ماتریس Q تدوین می‌شود. ماتریس Q شامل رابطه بین مهارت‌های شناختی پایه و سؤال‌ها است؛ به این صورت که هر سطر متعلق به یک سؤال و هر ستون متعلق به یک مهارت است. اگر در محل تلاقی یک سؤال با یک مهارت عدد صفر قرار گرفته باشد، به معنی عدم رابطه بین مهارت و سؤال است؛ درحالی که عدد یک حاکی از الزام وجود آن مهارت شناختی برای پاسخ به سؤال موردنظر است. به‌منظور تدوین ماتریس Q در این پژوهش، ابتدا ادبیات نظری و پیشینه تجربی موجود در زمینه خواندن و درک مطلب بررسی شد و با مشورت گرفتن از متخصصان حوزه سنجش زبان که در تدریس دروس مربوط به خواندن و درک مطلب سابقه داشتند؛ ۹ مهارت پیشنهادی جانگ (۲۰۰۹) به عنوان مهارت‌های پایه آزمون مورد بررسی در نظر گرفته شد. انتخاب این ۹ مهارت به این خاطر بود که از نظر متخصصانی که با آن‌ها مشورت شد، مهارت‌های پیشنهادی جانگ نسبت به مهارت‌های پیشنهادی افراد دیگر توانایی بالاتری برای پوشش تمامی سؤال‌های آزمون را دارند. ۹ مهارت موردنظر همراه با تعریف هر یک از آن‌ها در جدول ۱ ارائه شده است.

جدول ۱. مهارت‌های پایه و تعریف هر یک از آن‌ها

ردیف	مهارت	تعریف
۱	CDV	استنباط معنی یک کلمه یا عبارت با جستجو و تجزیه و تحلیل متن با کمک سرخ‌هایی که در متن موجود است.
۲	CIV	تشخیص معنی کلمه بیرون از متن با کمک دانش قبلی.
۳	SSL	ادراک روابط منطقی بین قسمت‌های مختلف متن با کمک انسجام واژگانی و گرامری موجود در جملات متوالی.
۴	TEI	خواندن سریع جملات سراسر یک پاراگراف برای معنا کردن تحت الفظی جملات یک متن.
۵	TIM	خواندن گزینشی یک پاراگراف یا همه پاراگراف‌ها برای تشخیص عقیده اصلی که بیان شده بر اساس اطلاعات ضمنی در متن.

خواندن سطحی یک پاراگراف و استنتاج‌های گزاره‌ای درباره استدلال‌ها یا هدف نویسنده با مراجعه به اطلاعات بیان‌شده ضمنی یا دانش استقرایی.	INF	۶
خواندن دقیق یا مؤثر به منظور تعیین مکان اطلاعات مناسب در متن و تعیین اطلاعات درست یا نادرست.	NEG	۷
تجزیه و تحلیل و ارزیابی اهمیت نسبی اطلاعات موجود در متن با کمک تفاوت قائل شدن بین ایده اصلی و جزئیات.	SUM	۸
تشخیص استدلال‌ها و مغایرت‌های اصلی مزبور به ساختار بیانی موجود در متن شامل روابطی مثل مقایسه/تضاد، علت/معلول یا استدلال‌های جایگزین و ترسیم آن‌ها در چارچوب ذهنی.	MCF	۹

۹ مهارت جدول بالا همراه با سؤال‌های آزمون در اختیار سه متخصص سنجش زبان با سابقه تدریس درک مطلب و خواندن قرار گرفت. از این سه متخصص درخواست شد، که تعیین کنند برای پاسخ به هریک از ۲۰ سؤال آزمون کدام یک از این مهارت‌ها لازم است. طبق نظر هریک از این متخصصان یک ماتریس Q تشکیل شد، سپس این سه ماتریس بر اساس همپوشانی‌هایی که داشتند در همدیگر ادغام شده و یک ماتریس نهایی به دست آمد. دوباره ماتریس Q نهایی به هریک از این سه متخصص داده شد و توسط آن‌ها مورد تأیید قرار گرفت. در ماتریس نهایی دو مهارت CIV (تشخیص معنی با کمک دانش قبلی) و TEI (خواندن سریع جملات سراسر یک پاراگراف برای معنا کردن تحت الفظی جملات یک متن) حذف شدند. دلیل اصلی حذف این مهارت‌ها این بود که برای قابل‌اعتماد بودن اطلاعات تشخیصی بهتر است هر ویژگی حداقل به سه سؤال مرتبط شود (هارتز، روسوس و استوت، ۲۰۰۲)؛ ولی این دو مهارت توسط هر متخصص فقط به یک یا دو سؤال مرتبط شدند. از طرفی خرد شدن زیاد مهارت‌ها به حفظ انسجام یک آزمون لطمه زده و منجر به بی‌ثباتی و عدم اطمینان طبقه‌بندی آزمودنی‌ها می‌شود (لی و سواکی، ۲۰۰۹)؛ بنابراین ماتریسی شامل ۲۰ سطر (تعداد سؤال‌ها) و هفت ستون (تعداد مهارت‌ها) تشکیل شد. در جدول ۲ رابطه بین هر سؤال و هر مهارت نشان داده شده است.

جدول ۲: رابطه سؤال‌ها و مهارت‌ها

شماره سؤال	CDV	SSL	TIM	INF	NEG	SUM	MCF
۱	۰	۰	۰	۱	۰	۱	۰
۲	۱	۰	۰	۰	۰	۰	۱
۳	۱	۱	۰	۰	۰	۰	۰
۴	۱	۱	۰	۰	۰	۰	۰
۵	۰	۱	۱	۱	۰	۰	۱
۶	۱	۱	۰	۰	۱	۰	۰
۷	۰	۱	۰	۱	۰	۰	۰
۸	۰	۱	۱	۱	۰	۱	۱
۹	۰	۰	۰	۱	۱	۰	۰
۱۰	۱	۱	۰	۰	۱	۰	۰
۱۱	۰	۰	۱	۱	۰	۰	۰
۱۲	۰	۰	۱	۱	۰	۱	۰
۱۳	۰	۰	۱	۱	۱	۰	۰
۱۴	۰	۱	۱	۰	۰	۰	۰
۱۵	۰	۱	۱	۰	۰	۰	۰
۱۶	۰	۱	۱	۰	۰	۰	۰
۱۷	۰	۱	۱	۰	۰	۱	۰
۱۸	۱	۰	۰	۰	۰	۰	۰
۱۹	۱	۰	۱	۱	۰	۰	۰
۲۰	۰	۰	۱	۰	۱	۰	۱

بعد از تدوین ماتریس Q تجزیه و تحلیل‌های کمی با کمک نرم‌افزار R و پکیج‌های CDM و G-DINA انجام شد. ابتدا روایی ماتریس Q بررسی و تأیید شد. سپس برازش مدل G-DINA با داده‌های پژوهش از طریق شاخص ریشه دوم استاندارد میانگین مجذورات باقی‌مانده^۱ (SRMSR) بررسی گردید. شاخص SRMSR که برابر با متوسط اندازه باقی‌مانده‌های بین کوواریانس مشاهده‌شده و کوواریانس پیش‌بینی شده است، در صورت داشتن ارزشی کوچک‌تر از ۰/۰۵ نشان‌دهنده برازش خوب مدل با داده‌هاست (مای

1. The Standardized Root Mean Square Residual

دیا- اولیوارس^۱، ۲۰۱۳). شاخص SRMSR محاسبه شده برابر با ۰/۰۲۶ است؛ لذا مدل G-DINA با داده‌ها برازش دارد.

جدول ۳. ماتریس همبستگی بین مهارت‌ها

مهارت	CDV	SSL	TIM	INF	NEG	SUM	MCF
CDV	۱						
SSL	۰/۰۵	۱					
TIM	۰/۶۱	۰/۶۴	۱				
INF	۰/۶۱	۰/۶۲	۰/۰۵	۱			
NEG	۰/۵۹	۰/۳۲	۰/۸۵	۰/۲۶	۱		
SUM	۰/۸۷	۰/۶۴	۰/۹۰	۰/۱۱	۰/۶۶	۱	
MCF	۰/۱۸	۰/۸۷	۰/۴۶	۰/۷۹	۰/۱۲	۰/۱۸	۱

در جدول بالا همبستگی بین مهارت‌ها گزارش شده است. بر طبق اطلاعات جدول بیشترین همبستگی بین مهارت SUM و مهارت TIM ($r=0/90$)، مهارت SUM با مهارت-های CDV ($r=0/87$) و مهارت TIM و مهارت NEG ($r=0/85$) و نیز بین مهارت MCF با مهارت SSL ($r=0/87$) است.

قبل از برآورد پارامترهای مربوط به هر سؤال، برازش سؤال‌ها با مدل از طریق شاخص میانگین مجذور خطای تقریب (RMSEA) محاسبه و در جدول ۳ گزارش شده است.

جدول ۴. شاخص برازش سؤال

شماره سؤال	RMSEA	شماره سؤال	RMSEA
۱	۰/۰۲۰	۱۱	۰/۰۳
۲	۰/۰۳	۱۲	۰/۰۲
۳	۰/۰۳	۱۳	۰/۰۳
۴	۰/۰۳	۱۴	۰/۰۲
۵	۰/۰۲	۱۵	۰/۰۳
۶	۰/۰۳	۱۶	۰/۰۳
۷	۰/۰۲	۱۷	۰/۰۲
۸	۰/۰۳	۱۸	۰/۰۴
۹	۰/۰۳	۱۹	۰/۰۱
۱۰	۰/۰۴	۲۰	۰/۰۲

زمانی که شاخص RMSEA سؤالی کوچک‌تر از ۰/۰۵ باشد، آن سؤال برازش خوبی با مدل دارد (کانینا-هینت، راپ و ویلهلم، ۲۰۰۹). لذا همان‌گونه که در جدول بالا مشاهده می‌شود، میزان RMSEA همه سؤال‌ها کمتر از ۰/۰۵ است و در نتیجه همه سؤال‌ها برازش خوبی با مدل G-DINA دارند.

حدس و لغزش دو پارامتر مربوط به سؤال هستند که درصد افرادی را که پاسخی خلاف انتظار داده‌اند را نشان می‌دهند. به عبارت دیگر پارامتر حدس یعنی احتمال پاسخ صحیح در حالی که فرد به هیچ‌یک از مهارت‌های لازم برای پاسخ‌گویی نرسیده است؛ و در مقابل پارامتر لغزش یعنی احتمال اینکه فرد در صورت داشتن مهارت لازم برای پاسخ به سؤال، در پاسخگویی به آن شکست بخورد. در جدول ۴ پارامترهای حدس و لغزش و خطای برآورد مربوط به هر یک آورده شده است.

جدول ۵. پارامترهای حدس و لغزش برای مدل G-DINA

شماره سؤال	حدس	خطای برآورد	لغزش	خطای برآورد
۱	۰/۰۹	۰/۰۱	۰/۲۶	۰/۰۳
۲	۰/۰۲	۰/۰۰۷	۰/۴۱	۰/۰۳
۳	۰/۰۶	۰/۰۱	۰/۶۵	۰/۰۲
۴	۰/۰۹	۰/۰۱	۰/۶۵	۰/۰۲
۵	۰/۰۱	۰/۰۱	۰/۴۹	۰/۰۳
۶	۰/۱۱	۰/۰۲	۰/۲۹	۰/۰۳
۷	۰/۰۱	۰/۰۰۸	۰/۲۴	۰/۰۳
۸	۰/۰۰۲	۰/۰۰۶	۰/۸۳	۰/۰۳
۹	۰/۰۲	۰/۰۰۸	۰/۴۸	۰/۰۳
۱۰	۰/۰۹	۰/۰۲	۰/۲۴	۰/۰۳
۱۱	۰/۰۸	۰/۰۱	۰/۱۶	۰/۰۲
۱۲	۰/۰۰۹	۰/۰۰۵	۰/۵۳	۰/۰۳
۱۳	۰/۰۱	۰/۰۰۵	۰/۸۰	۰/۰۲
۱۴	۰/۰۳	۰/۰۰۷	۰/۵۸	۰/۰۲
۱۵	۰/۰۲	۰/۰۰۶	۰/۷۵	۰/۰۹
۱۶	۰/۰۰۷	۰/۰۰۴	۰/۸۹	۰/۰۱

شماره سؤال	حدس	خطای برآورد	لغزش	خطای برآورد
۱۷	۰/۰۰۴	۰/۰۰۳	۰/۹۰	۰/۰۲
۱۸	۰/۰۶	۰/۰۰۸	۰/۷۷	۰/۰۱
۱۹	۰/۰۲	۰/۰۰۶	۰/۳۵	۰/۰۳
۲۰	۰/۰۶	۰/۰۱	۰/۲۹	۰/۰۳

با توجه به اطلاعات جدول بالا میزان پارامتر حدس همه سؤال‌ها بسیار کم است، یعنی احتمال حدس و پاسخ صحیح دادن به سؤال‌ها، در صورت نداشتن مهارت لازم بسیار ناچیز است. در مقابل میزان پارامتر لغزش، به‌ویژه در مورد سؤال‌های ۱۶ و ۱۷ زیاد است. در جدول ۶ پارامترهای برآورد شده در چارچوب مدل G-DINA ارائه شده است. در این مدل برای هر سؤال یک پارامتر منفرد ارائه نمی‌شود؛ بلکه بر اساس الگوهای تسلط بر مهارت‌های مرتبط با سؤال احتمال پاسخ به سؤال برآورد می‌گردد. برای مثال سؤال اول با دو مهارت (SUM و INF) مرتبط است و ۴ حالت برای این سؤال متصور است: حالت اول زمانی رخ می‌دهد که شخص به هیچ‌یک از مهارت‌ها تسلط نیافته است (A00)، حالت دوم زمانی رخ می‌دهد که شخص فقط بر مهارت اول تسلط یافته است (A10)، حالت سوم زمانی رخ می‌دهد که شخص فقط بر مهارت دوم تسلط یافته است (A01) و حالت چهارم زمانی رخ می‌دهد که شخص بر هر دو مهارت تسلط یافته است (A11). مدل G-DINA بر اساس هر یک از این چهار حالت یک احتمال پاسخ به سؤال ارائه می‌دهد، یعنی در مجموع برای سؤال یک ۴ احتمال مختلف محاسبه می‌شود.

جدول ۶. برآورد پارامتر سؤال‌ها بر اساس مدل G-DINA

سؤال	مهارت مربوط به هر سؤال	توزیع مهارت	احتمال	خطای برآورد
۱	INF-SUM	A00	۰/۰۹	۰/۰۱
		A10	۰/۳۶	۰/۰۵
		A01	۰/۴۶	۰/۰۳
		A11	۰/۷۳	۰/۰۳
۲	CDV-MCF	A00	۰/۰۲	۰/۰۰۷
		A10	۰/۱۵	۰/۰۲
		A01	۰/۲۲	۰/۰۳
		A11	۰/۵۹	۰/۰۳
۳	CDV-SSL	A00	۰/۰۶	۰/۰۱
		A10	۰/۲۹	۰/۰۳

خطای برآورد	احتمال	توزیع مهارت	مهارت مربوط به هر سؤال	سؤال		
۰/۰۳	۰/۲۲	A01	CDV-SSL	۴		
۰/۰۲	۰/۳۵	A11				
۰/۰۱	۰/۰۹	A00				
۰/۰۳	۰/۲۵	A10				
۰/۰۳	۰/۲۶	A01				
۰/۰۲	۰/۳۵	A11				
۰/۰۱	۰/۰۱	A0000				
۰/۳۳	۰/۹۹	A1000				
۰/۰۴	۰/۲۴	A0100				
۰/۰۸	۰/۳۴	A0010				
۰/۲۶	۰/۹۹	A0001				
۰/۰۶	۰/۲۴	A1100				
۰/۲۲	۰/۰۰۰۱	A1010	SSL-TIM-INF-MCF	۵		
۰/۰۷	۰/۱۴	A1001				
۰/۱۰	۰/۴۵	A0110				
۰/۰۵	۰/۰۳	A0101				
۰/۱۹	۰/۰۰۰۱	A0011				
۰/۱۴	۰/۰۰۲	A1110				
۰/۰۶	۰/۰۹	A1011				
۰/۱۸	۰/۷۲	A1101				
۰/۰۹	۰/۲۹	A0111				
۰/۰۳	۰/۵۱	A1111				
۰/۰۲	۰/۱۱	A000			CDV-SSL-NEG	۶
۰/۰۴	۰/۴۰	A100				
۰/۱۲	۰/۹۹	A010				
۰/۰۷	۰/۴۷	A001				
۰/۱۱	۰/۰۰۰۱	A110				
۰/۰۶	۰/۱۲	A101				
۰/۰۵	۰/۲۷	A011				
۰/۰۳	۰/۷۱	A111				
۰/۰۰۸	۰/۰۱	A00	SSL-INF	۷		
۰/۰۳	۰/۴۶	A10				
۰/۰۴	۰/۰۰۰۱	A01				

سؤال	مهارت مربوط به هر سؤال	توزیع مهارت	احتمال	خطای برآورد
		A11	۰/۷۶	۰/۰۳
		A00000	۰/۰۰۲	۰/۰۰۶
		A10000	۰/۰۳	۰/۲۱
		A01000	۰/۵۷	۰/۱۷
		A00100	۰/۰۰۰۱	۰/۲۶
		A00010	۰/۰۰۰۱	۰/۱۰
		A00001	۰/۰۰۰۱	۰/۱۹
		A11000	۰/۳۱	۰/۳۵
		A10100	۰/۰۰۰۱	۰/۴۹
		A10010	۰/۹۸	۰/۴۳
		A10001	۰/۹۹	۰/۳۷
		A01100	۰/۱۱	۰/۰۹
		A01010	۰/۱۹	۰/۰۵
		A01001	۰/۰۰۰۲	۰/۰۸
		A00110	۰/۱۴	۰/۰۶
		A00101	۰/۴۳	۰/۱۵
		A00011	۰/۰۰۰۱	۰/۱۴
		A11100	۰/۰۰۰۱	۰/۲۸
		A11010	۰/۱۷	۰/۰۶
		A11001	۰/۰۳	۰/۱۴
		A10110	۰/۷۲	۰/۲۰
		A10101	۰/۹۹	۰/۹۳
		A10011	۰/۰۰۰۱	۰/۰۷
		A01110	۰/۶۳	۰/۲۷
		A01101	۰/۰۰۰۱	۰/۰۸
		A01011	۰/۰۰۰۱	۰/۴۲
		A00111	۰/۱۱	۰/۱۸
		A11110	۰/۸۱	۰/۴۰
		A11101	۰/۳۰	۰/۱۲
		A11011	۰/۲۶	۰/۱۰
		A10111	۰/۳۴	۰/۱۵
		A01111	۰/۰۱	۰/۴۶
		A11111	۰/۱۷	۰/۰۳

SSL-TIM-INF-SUM-MCF

۸

سؤال	مهارت مربوط به هر سؤال	توزیع مهارت	احتمال	خطای برآورد
۹	INF-NEG	A00	۰/۰۲	۰/۰۰۸
		A10	۰/۰۰۲	۰/۰۳
		A01	۰/۱۲	۰/۰۲
		A11	۰/۵۲	۰/۰۳
۱۰	CDV-SSL-NEG	A000	۰/۰۹	۰/۰۲
		A100	۰/۰۸	۰/۰۴
		A010	۰/۴۰	۰/۰۷
		A001	۰/۶۲	۰/۰۷
		A110	۰/۵۷	۰/۰۸
		A101	۰/۵۴	۰/۰۷
		A011	۰/۲۱	۰/۰۵
		A111	۰/۷۶	۰/۰۳
		A00	۰/۰۸	۰/۰۱
		A10	۰/۳۶	۰/۰۲
		A01	۰/۲۹	۰/۰۵
۱۲	TIM-INF-SUM	A11	۰/۸۴	۰/۰۲
		A000	۰/۰۰۹	۰/۰۰۵
		A100	۰/۰۰۴	۰/۰۴
		A010	۰/۹۹	۰/۲۴
		A001	۰/۰۰۱	۰/۰۴
		A110	۰/۰۴	۰/۰۴
		A101	۰/۲۱	۰/۰۳
		A011	۰/۰۰۱	۰/۰۴
۱۳	TIM-INF-NEG	A111	۰/۴۷	۰/۰۳
		A000	۰/۰۱	۰/۰۰۵
		A100	۰/۲۵	۰/۰۴
		A010	۰/۰۱	۰/۰۲
		A001	۰/۰۳	۰/۰۲
		A110	۰/۰۰۱	۰/۰۳
		A101	۰/۰۰۱	۰/۰۳
		A011	۰/۴۴	۰/۰۸
۱۴	SSL-TIM	A111	۰/۲۰	۰/۰۲
		A00	۰/۰۳	۰/۰۰۷

سؤال	مهارت مربوط به هر سؤال	توزیع مهارت	احتمال	خطای برآورد
		A10	۰/۰۲	۰/۰۲
		A01	۰/۲۰	۰/۰۳
		A11	۰/۴۲	۰/۰۲
		A00	۰/۰۲	۰/۰۰۶
	SSL-TIM	A10	۰/۱۰	۰/۰۳
۱۵		A01	۰/۲۵	۰/۰۵
		A11	۰/۲۵	۰/۰۲
		A00	۰/۰۰۷	۰/۰۰۴
	SSL-TIM	A10	۰/۰۰۹	۰/۰۱
۱۶		A01	۰/۱۷	۰/۰۲
		A11	۰/۱۱	۰/۰۱
		A000	۰/۰۰۴	۰/۰۰۳
		A100	۰/۰۰۰۱	۰/۰۶
		A010	۰/۰۰۰۲	۰/۰۳
	SSL-TIM-SUM	A001	۰/۰۰۰۱	۰/۰۳
۱۷		A110	۰/۰۷	۰/۰۵
		A101	۰/۰۰۰۱	۰/۰۳
		A011	۰/۳۴	۰/۰۴
		A111	۰/۱۰	۰/۰۲
	CDV	A0	۰/۰۶	۰/۰۰۸
۱۸		A1	۰/۲۳	۰/۰۱
		A000	۰/۰۲	۰/۰۰۶
		A100	۰/۰۰۰۱	۰/۰۷
		A010	۰/۵۲	۰/۰۵
	CDV-TIM-INF	A001	۰/۶۶	۰/۱۳
۱۹		A110	۰/۲۰	۰/۰۳
		A101	۰/۰۰۷	۰/۰۲
		A011	۰/۰۰۰۱	۰/۰۷
		A111	۰/۰۳	۰/۰۳
		A000	۰/۰۶	۰/۰۱
	TIM-NEG-MCF	A100	۰/۰۸	۰/۰۵
۲۰		A010	۰/۱۳	۰/۰۴
		A001	۰/۰۸	۰/۰۵

سؤال	مهارت مربوط به هر سؤال	توزیع مهارت	احتمال	خطای برآورد
		A110	۰/۳۷	۰/۰۴
		A101	۰/۹۹	۰/۰۹
		A011	۰/۲۷	۰/۰۶
		A111	۰/۷۱	۰/۰۳

در پایان نیز وضعیت تسلط آزمودنی‌ها^۵ که هدف اصلی هر سنجش تشخیصی شناختی است- بر هر یک از مهارت‌های شناختی مورد بررسی قرار گرفته است. به منظور نیل به این هدف ابتدا احتمال تسلط هر فرد در هر مهارت با روش پسینی مورد انتظار^۱ (EAP) محاسبه و نقطه برش ۰/۶ به عنوان معیار حد تسلط در نظر گرفته شد. نقطه برش ۰/۶ با در نظر گرفتن پژوهش‌های قبلی در حوزه خواندن و دشوار بودن آزمون تعیین شد؛ بنابراین اگر آزمودنی در یک مهارت احتمال تسلط بالای ۰/۶ داشت به عنوان یک فرد مسلط در آن صفت محسوب شده و در غیر این صورت (احتمال کمتر از ۰/۶) فرض بر این است که در آن صفت به حد تسلط نرسیده است. بعد از طبقه‌بندی افراد در دو گروه مسلط و غیر مسلط فراوانی و درصد فراوانی افراد در هر یک از این دو گروه و برای هر مهارت به طور جداگانه محاسبه و در جدول ۷ گزارش شده است.

جدول ۷. فراوانی و درصد فراوانی افراد مسلط و غیر مسلط در هر مهارت

مهارت	احتمال حاشیه‌ای تسلط	مسلط		غیر مسلط	
		فراوانی	درصد	فراوانی	درصد
CDV	۰/۴۴	۱۰۰۱	۳۳/۴	۱۹۹۹	۶۶/۶
SSL	۰/۳۷	۹۴۰	۳۱/۳	۲۰۶۰	۶۸/۷
TIM	۰/۵۲	۱۳۸۳	۴۶/۱	۱۶۱۷	۵۳/۹
INF	۰/۲۳	۴۲۷	۱۵/۷	۲۵۲۸	۸۴/۳
NEG	۰/۴۲	۱۰۲۰	۳۴	۱۹۸۰	۶۶
SUM	۰/۵۵	۱۴۹۳	۴۹/۸	۱۵۰۷	۵۰/۲
MCF	۰/۳۱	۷۱۸	۲۳/۹	۲۲۸۲	۷۶/۱

با توجه به فراوانی‌های نشان داده شده در جدول بالا نزدیک نیمی از افراد گروه نمونه در دو مهارت SUM (۴۹/۸٪) و TIM (۴۶/۱٪) به تسلط رسیده‌اند؛ اما تعداد بسیار کمی از افراد در دو مهارت INF (۱۵/۷٪) و MCF (۲۳/۹٪) به تسلط رسیده‌اند. همچنین حدود یک

1. Expected A Poteriori

سوم افراد گروه نمونه در مهارت‌های (CDV (۳۳/۴٪)، (SSL (۳۱/۳٪) و (NEG (۳۴٪) به تسلط رسیده‌اند؛ بنابراین بیشترین ضعف آزمودنی‌ها در دو مهارت INF (خواندن سطحی یک پاراگراف و استنتاج‌های گزاره‌ای درباره استدلال‌ها یا هدف نویسنده با مراجعه به اطلاعات بیان‌شده ضمنی یا دانش استقرایی) و MCF (تشخیص استدلال‌ها و مغایرت‌های اصلی مزبور به ساختار بیانی موجود در متن شامل روابطی مثل مقایسه/ تضاد، علت/ معلول یا استدلال‌های جایگزین و ترسیم آن‌ها در چارچوب ذهنی) است.

بحث و نتیجه‌گیری

به منظور سنجش تشخیصی شناختی ۲۰ سؤال آزمون مورد مطالعه، مانند هر سنجش تشخیصی شناختی دیگری به دو مدل شناختی و روان‌سنجی نیاز بود. در جهت طراحی مدل شناختی در قالب ماتریس Q، ضمن مطالعه ادبیات نظری و جمع‌آوری پیشینه تجربی از نظر متخصصان سنجش زبان نیز کمک گرفته شد و در نهایت هفت خرده مهارت به عنوان مهارت پایه این آزمون تعیین شدند. به‌طور خلاصه این هفت مهارت شامل پی بردن به معنای واژگان از طریق تجزیه و تحلیل متن، درک روابط بین قسمت‌های مختلف متن و استفاده از دانش نحوی، خواندن گزینشی متن برای پی بردن به عقیده اصلی و خواندن سطحی برای پی بردن به استدلال‌ها و هدف نویسنده، خواندن دقیق متن برای تشخیص اطلاعات درست و نادرست، ارزیابی اهمیت نسبی اطلاعات موجود در متن به منظور یافتن ایده اصلی و در نهایت تشخیص استدلال‌ها و مغایرت‌ها و ترسیم یک نقشه ذهنی می‌شود.

با در نظر گرفتن اختلاف‌هایی که بر سر جبرانی یا غیرجبرانی بودن خرده مهارت‌های درک مطلب وجود دارد و با توجه به پژوهش‌هایی نظیر لی، هانت و لی (۲۰۱۵) و راوند و رویچ (۲۰۱۵) که نتیجه آن‌ها برتری مدل G-DINA را نسبت به مدل‌های مورد مقایسه تأیید کرده‌اند، مدل G-DINA برای تجزیه و تحلیل اطلاعات کمی انتخاب شد. مدل G-DINA یک مدل کلی است. در مدل‌های کلی برای احتمال هر سؤال 2^k (K برابر است با تعداد مهارت‌های مرتبط با سؤال) احتمال مختلف وجود دارد؛ مثلاً اگر سؤالی با دو مهارت مرتبط باشد ۴ احتمال مختلف بر طبق تسلط بر هر یک از مهارت‌ها برای آن سؤال متصور است. به عبارت دیگر در این مدل‌ها بر اساس تسلط افراد بر هر یک از مهارت‌های مرتبط با سؤال، احتمال پاسخگویی سؤال برآورد می‌گردد. با وجود شواهد نظری و تجربی که بر

مناسب بودن مدل G-DINA برای آزمون خواندن دلالت داشت؛ از طریق تجزیه و تحلیل آماری نیز برازش این مدل بررسی گردید. شاخص SRMSR برآورد شده نشان داد که مدل با داده‌ها برازش دارد. همچنین شاخص RMSEA برای هر سؤال محاسبه شد که نتیجه آن حاکی از برازش هر ۲۰ سؤال با مدل بود.

پارامترهای حدس و لغزش سؤال‌ها نشان داد در صورت فقدان خرده مهارت‌های لازم برای هر سؤال احتمال حدس زدن پاسخ صحیح بسیار کم است؛ یعنی ارزش پارامتر حدس سؤال‌ها ناچیز است که این امر با نمره منفی داشتن پاسخ اشتباه در آزمون مذکور همخوانی دارد. در واقع افراد به‌طور آگاهانه از حدس زدن پاسخ سؤال‌ها پرهیز کرده‌اند. در مقابل احتمال شکست افراد در صورت تسلط داشتن بر مهارت‌ها زیاد است. همچنین میزان احتمالات برآورد شده برای پاسخ به هر سؤال نیز پایین بود. همه این موارد، شواهدی دال بر دشوار بودن آزمون مورد بررسی هستند.

نتیجه بررسی وضعیت آزمودنی‌ها نشان داد بیشترین احتمال تسلط برای مهارت SUM تجزیه و تحلیل و ارزیابی اهمیت نسبی اطلاعات موجود در متن با کمک تفاوت قائل شدن بین ایده اصلی و جزئیات) و TIM (خواندن گزینشی یک پارگراف یا همه پاراگراف‌ها برای تشخیص عقیده اصلی که بیان شده بر اساس اطلاعات ضمنی در متن) است؛ حدود نیمی از افراد به عنوان افراد مسلط در این دو صفت شناخته شدند. در ۳ مهارت CDV (استنباط معنی یک کلمه یا عبارت با جستجو و تجزیه و تحلیل متن با کمک سرخ‌هایی که در متن موجود است)، SSL (ادراک روابط منطقی بین قسمت‌های مختلف متن با کمک انسجام واژگانی و گرامری موجود در جملات متوالی) و NEG (خواندن دقیق یا مؤثر به‌منظور تعیین مکان اطلاعات مناسب در متن و تعیین اطلاعات درست یا نادرست) حدود یک سوم افراد به‌عنوان افراد مسلط طبقه‌بندی شدند. در دو مهارت INF (خواندن سطحی یک پاراگراف و استنتاج‌های گزاره‌ای درباره استدلال‌ها یا هدف نویسنده با مراجعه به اطلاعات بیان‌شده ضمنی یا دانش استقرایی) و MCF (تشخیص استدلال‌ها و مغایرت‌های اصلی مزبور به ساختار بیانی موجود در متن شامل روابطی مثل مقایسه/ تضاد، علت/ معلول یا استدلال‌های جایگزین و ترسیم آن‌ها در چارچوب ذهنی) تعداد بسیار کمی از افراد به تسلط رسیده بودند.

در پژوهشی که توسط جانگک (۲۰۰۹) بر روی آزمون تافل انجام شد و مهارت‌های موجود در این پژوهش نیز از همان مطالعه استخراج شده است، مهارت SUM نسبت به سایر مهارت‌ها بیشترین افراد مسلط را به خود اختصاص داد، همچنین در پژوهش جانگک ۳۸٪ افراد در مهارت NEG به تسلط رسیدند که در این پژوهش میزان ۳۴٪ برای این مهارت به دست آمد با این تفاوت که در پژوهش جانگک مهارت NEG ضعیف‌ترین مهارت از نظر میزان افراد مسلط بود. تشابه دیگر این دو پژوهش این بود که مهارت MCF در هر دو آن‌ها نسبت به سایر مهارت‌ها وضعیت مطلوبی از نظر میزان افراد تسلط یافته نداشت؛ با این تفاوت که در پژوهش جانگک درصد افراد مسلط برای مهارت MCF نسبت به این پژوهش بیشتر بود. به طور کلی وضعیت تسلط افراد بر مهارت‌ها در پژوهش جانگک نسبت به پژوهش حاضر مطلوب‌تر بود. در پژوهش لی (۲۰۱۱) بر روی آزمون MELAB، ۴ مهارت پایه معرفی شد. ۴ مهارت پیشنهادی لی را می‌توان با مهارت‌های CDV، SSL، TIM و SUM معادل دانست. پژوهش لی از این جهت با پژوهش حاضر شباهت دارد که تعداد افراد مسلط بر هر یک از مهارت کم است؛ یعنی بیشتر آزمودنی‌ها در مهارت‌های مورد نظر به حد تسلط نرسیده‌اند.

با در نظر گرفتن وضعیت آزمودنی‌ها در دو مهارت SUM و TIM که نسبت به سایر مهارت‌ها بهتر است و تقریباً نیمی از افراد در این دو مهارت به حد تسلط رسیده‌اند؛ می‌توان ادعا کرد که آزمودنی‌ها به طور نسبی در تشخیص ایده اصلی متن موفق بوده‌اند؛ اما در سایر مهارت‌ها عملکرد آزمودنی‌ها مطلوب نبوده است. به ویژه در دو مهارت INF و MCF دارای عملکرد بسیار ضعیفی بودند؛ به عبارتی آزمودنی‌ها در مقایسه و پی‌بردن به روابط علت و معلولی و تشخیص تضادها و سپس تشکیل یک نقشه ذهنی ناتوان بوده و نتوانسته‌اند به اهداف نویسنده به درستی پی ببرند. در واقع افراد در خواندن سریع و داشتن سرعت عمل در رسیدن به پاسخ با کمک سرنخ‌ها و پی‌بردن به نکاتی ریزی همچون روابط علی یا تضادها ضعیف هستند. در مجموع یافته‌ها نشان داد که آزمودنی‌ها در استفاده از راهبردها و استراتژی‌های موجود به منظور یافتن پاسخ ناموفق عمل کرده‌اند؛ یعنی اینکه آزمودنی‌ها در مهارت‌های پایه درک مطلب و خواندن دارای ضعف جدی بوده و نیاز به آموزش ویژه در این زمینه دارند؛ بنابراین توصیه می‌شود که یک دوره کلاس‌های آموزشی توسط متخصصین حوزه سنجش زبان طراحی و مهارت‌های مورد نظر به آزمودنی‌ها آموزش داده شود و در نهایت اثربخشی

این دوره بر روی عملکرد افراد مورد بررسی قرار گیرد؛ زیرا این امکان وجود دارد که خیلی از افراد از وجود چنین راهبردهایی برای عملکرد بهتر بی‌اطلاع باشند. علاوه بر این طراحی یک آزمون خواندن و درک مطلب متناسب با مهارت‌های شناختی مطرح شده در این پژوهش، پیشنهاد می‌شود. به این صورت که ابتدا یک مدل شناختی طراحی و سپس بر اساس آن سؤال‌های آزمون طراحی شود؛ زیرا یکی از مشکلات عمده رویکرد تحلیل بازبازش، ناکافی بودن سؤال‌های مرتبط با یک مهارت و از طرفی وجود تعداد زیادی سؤال که یک مهارت خاص را می‌سنجند، است. همان‌طور که در پژوهش حاضر نیز ناکافی بودن سؤال‌های مرتبط با دو مهارت منجر به حذف این مهارت‌ها شد. در واقع زمانی که آزمون بر اساس اهداف شناختی طراحی نشده باشد پژوهشگر در رسیدن به اهداف شناختی و پی‌بردن به ضعف‌های شناختی آزمودنی‌ها با مشکل روبرو خواهد شد.

منابع

- افضلی، افشین. (۱۳۹۳). *تدوین و ارزشیابی مدل تشخیصی شناختی (CDM) ریاضیات پایه اول دبیرستان با استفاده از روش سلسله مراتبی صفات (AHM)*. پایان‌نامه دکتری: دانشگاه علامه طباطبایی.
- ثرن‌دایک، رابرت ال. (۱۳۶۹). *روانسنجی کاربردی*. ترجمه حیدرعلی هومن. تهران: انتشارات دانشگاه تهران.
- شریفی، حسن پاشا. (۱۳۹۰). *اصول روان‌سنجی و روان‌آزمایی*. تهران: رشد.
- فرهادی، حسین. (۱۳۸۱). «نقدی بر آزمون‌های کنکور کارشناسی ارشد رشته زبان انگلیسی» پژوهش‌های زبان خارجی، ۱۳، ۷۹-۱۰۶.
- کیبیری، مسعود (۱۳۹۳). *کاربرد سنجش تشخیص به منظور تعیین مهارت‌های کسب شده علوم تجربی در دانش‌آموزان سال سوم راهنمایی ایران بر اساس داده‌های تیمز ۲۰۱۱*. پایان‌نامه دکتری: دانشگاه تهران
- محسن‌پور، مریم (۱۳۹۳). *طراحی و ساخت آزمون شناختی- تشخیصی سواد ریاضی کاربردی دانش‌آموزان سال اول دبیرستان و سنجش اثربخشی بسته آموزشی جبرانی مبتنی بر آن*. پایان‌نامه دکتری: دانشگاه تهران.

مینایی، اصغر. (۱۳۹۱). مدل‌پردازی تشخیصی شناختی (CDM) سؤال‌های ریاضیات تیمز ۲۰۰۷ در دانش‌آموزان پایه هشتم ایران با استفاده از مدل یکپارچه با پارامتر پردازی مجدد (RUM) و مقایسه مهارت‌های ریاضی دانش‌آموزان دختر و پسر. پایان‌نامه دکتري: دانشگاه علامه طباطبائی.

مقدم، اعظم. (۱۳۹۴). کاربرد مدل‌های تشخیصی شناختی به منظور تعیین مهارت‌های زیربنایی عملکرد داوطلبان در آزمون ورودی زبان انگلیسی عمومی دوره دکتري. پایان‌نامه دکتري: دانشگاه علامه طباطبائی.

- Alderson, J. C. (2005). *Assessing reading*. Ernst Klett Sprachen.
- Bernhardt, E. B. (1993). *Reading development in a second language: Theoretical, empirical, & classroom perspectives*. Ablex Publishing Corporation, 355 Chestnut St., Norwood, NJ (paperback: ISBN-0-89391-734-6; cloth: ISBN-0-89391-675-7).
- Buck, G., Tatsuoaka, K., & Kostin, I. (1997). The Subskills of Reading: Rule-space Analysis of a Multiple-choice Test of Second Language Reading Comprehension. *Language Learning*, 47(3), 423-466.
- Davier, M. (2005). A general diagnostic model applied to language testing data. *ETS Research Report Series*, 2005(2).
- Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion Model application to LanguEdge assessment. *Language Testing*, 26(1), 031-73.
- Johnson, D. M., & Reynolds, F. A. (1940). A factor analysis of verbal ability. *The Psychological Record*, 4, 181.
- Kasai, M. (1997). *Application of the rule space model to the reading comprehension section of the test of English as a foreign language (TOEFL)* (Doctoral dissertation, University of Illinois at Urbana-Champaign).
- Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge University Press.
- Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2009). A practical illustration of multidimensional diagnostic skills profiling: Comparing results from confirmatory factor analysis and diagnostic classification models. *Studies in Educational Evaluation*, 35(2), 64-70.
- Lee, Y. W., & Sawaki, Y. (2009). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly*, 6(3), 239-263.
- Leighton, J., & Gierl, M. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.
- Li, H. (2011). A cognitive diagnostic analysis of the MELAB reading test. *Spaan Fellow Working Papers in Second or Foreign Language Assessment*, 9, 17-46.

- Li, H., Hunter, C. V., & Lei, P. W. (2016). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*, 33(3), 391-409.
- Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models. *Measurement: Interdisciplinary Research and Perspectives*, 11 (3), 71-101.
- Munby, J. (1981). *Communicative syllabus design: A sociolinguistic model for designing the content of purpose-specific language programmes*. Cambridge University Press.
- Pettit, N. T., & Cockriel, I. W. (1974). A factor study of the literal reading comprehension test and the inferential reading comprehension test. *Journal of Literacy Research*, 6(1), 63-75.
- Ravand, H., & Robitzsch, A. (2015). Cognitive Diagnostic Modeling Using R. *Practical Assessment, Research & Evaluation*, 20.
- Scott, H. S. (1998). *Cognitive diagnostic perspectives of a second language reading test* (Doctoral dissertation, University of Illinois at Urbana-Champaign).
- Snow, R. E., & Lohman, D. F. (1989). *Implications of cognitive psychology for educational measurement*. American Council on Education.
- Svetina, D., Gorin, J. S., & Tatsuoka, K. K. (2011). Defining and comparing the reading comprehension construct: A cognitive-psychometric modeling approach. *International Journal of Testing*, 11(1), 1-23.
- Toker, T. (2010). Cognitive Diagnostic Assessment of Timss-2007 Mathematics items For 8th Graders in Turkey.
- Vernon, P. E. (1962). The determinants of reading comprehension. *Educational and Psychological Measurement*.
- Yi, Y. (2013). *Implementing a cognitive diagnostic assessment in an institutional test: a new networking model in language testing and experiment with a new psychometric model and task type* (Doctoral dissertation, University of Illinois at Urbana-Champaign).
- Zhang, J. (2013). *Relationships between missing responses and skill mastery profiles of cognitive diagnostic assessment* (Doctoral dissertation, University of Toronto).