فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی
سال دهم، شماره ۲۹، بهار ۱۳۹۹، صص۲۲۷-۲۰۵

Journal of Educational Measurement & Evaluation Studies
Vol.10, No. 29, Spring 2020

انتخاب مدل تشخیصی شناختی مناسب برای آزمون‌های درک مطلب (مورد مطالعه: آزمون کارشناسی ارشد رشته زبان انگلیسی)

ریحانه رحیمی*
علی دلاور**
جلیل یونسی***
زهره‌سادات نصیری****

تاریخ دریافت مقاله: ۹۷/۰۸/۱۹
تاریخ پذیرش مقاله: ۹۷/۱۲/۱۸

چکیده

سنجش تشخیصی شناختی، نوعی از سنجش آموزشی است که با به‌کارگیری مدل‌های روان‌سنجی به تعیین و تشخیص نارسایی‌های یادگیری کمک می‌کند. با توجه به اینکه مدل‌های زیادی در این زمینه وجود دارد؛ انتخاب یک مدل تشخیصی شناختی مناسب برای تحلیل داده‌ها اهمیت زیادی خواهد داشت. هدف از اجرای پژوهش حاضر، تعیین مدل مناسب برای آزمون‌های خواندن و درک مطلب است. داده‌های مورد بررسی در این پژوهش شامل پاسخنامه ۳۰۰۰ نفر از شرکت‌کنندگان در آزمون سراسری کارشناسی ارشد رشته زبان انگلیسی بوده است که از بین ١٦٠٤٤ شرکت‌کننده به صورت تصادفی انتخاب شده‌اند. روش کار به این صورت بود که پس از تشکیل ماتریس Q مدل کلی (G-DINA و LCDM)، دو مدل جبرانی (ACDM و DINO) و دو مدل غیر جبرانی (RRUMو DINA) برای مقایسه در نظر گرفته شدند. ازنظر شاخص برازش ماکزیمم خی-دو هیچ‌یک از مدل‌ها برازش نداشتند؛ ولی در مقابل نتایج شاخص SRMSR نشان داد که همه مدل‌ها برازش قابل قبولی با داده‌ها دارند. در مقایسه بین شاخص‌های برازش نسبی مشخص شد که مدل‌های G-DINA و ACDM برازش بهتر و مدل‌های DINO و DINA برازش ضعیف‌تری نسبت به سایر مدل‌ها دارند. در بررسی شاخص برازش سؤال نیز فقط مدل‌های G-DINA و LCDM با همه سؤال‌ها برازش مطلوبی نشان دادند.

واژگان کلیدی: سنجش تشخیصی شناختی، ماتریس Q، مدل کلی، مدل جبرانی، مدل غیرجبرانی

---

*دانشجوی دکتری سنجش و اندازه گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، ایران (نویسنده مسئول: reyhanerahimi1367@gmail.com)

**استاد گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، ایران

***دانشیار گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، ایران

****کارشناسی ارشد سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، ایران

مقدمه

سنجش تشخیصی شناختی[1] نوعی از سنجش آموزشی است که برخلاف مدل‌های سنتی سنجش این توانایی را دارد که ساختارهای دانش و فرایندهای مهارت[2] را در آزمودنی‌ها اندازه بگیرد و اطلاعاتی راجع به نقاط ضعف و قدرت آنان فراهم کند. سنجش تشخیصی شناختی، رویکردی بین‌رشته‌ای دارد؛ به عبارتی خط اتصال بین روان‌شناسی شناختی و تحلیل‌های آماری است. در این نوع از سنجش، رابطه فرایندهای روانی و استراتژی‌های زیر بنایی عملکرد در یک آزمون با پاسخ‌های ارائه‌شده برای آن سؤال‌ها از طریق تجزیه‌وتحلیل‌های آماری پیچیده، بررسی می‌شود. به‌طور کلی، اندازه‌گیری‌های روان‌شناختی از تعدادی مدل‌های ریاضی برای انجام فرایند اندازه‌گیری کمک می‌گیرند. بر اساس این مدل‌ها اندازه‌های عددی متغیر مستقل با یکدیگر ترکیب شده و اندازه‌های مربوط به متغیر وابسته را پیش‌بینی می‌کنند (امبرتسون و رایس[3] ۲۰۰۰؛ ترجمه شریفی، فرزاد، حبیبی و ایزانلو، ۱۳۸۸). به‌این‌ترتیب سنجش تشخیصی شناختی نیز مجموعه‌ای از مدل‌های ریاضی با عنوان مدل‌های تشخیصی شناختی[4] (CDMs) را برای مطالعات پژوهشی خود تدوین کرده است. کار اصلی این مدل‌ها تبدیل نظریه‌های شناختی به مدل‌های احتمالاتی است (سینهارا و المونده[5] ۲۰۰۷). به سخن دیگر، مدل‌های تشخیصی شناختی (CDMs) که در رویکرد سنجش تشخیصی شناختی پیشنهاد می‌شوند؛ مدل‌های ریاضی را با مدل‌های روان‌سنجی معروف مثل نظریه سؤال پاسخ ترکیب می‌کنند تا از این طریق هم خواسته‌های شناختی سؤال را برآورده کنند و هم‌سطح توانایی آزمودنی‌ها را نشان دهند (لیتون و گریل[6] ۲۰۰۷).

مدل‌های تشخیصی شناختی مانند تحلیل عامل تأییدی[7] در ذات تأییدی هستند. در این مدل‌ها متغیر مکنون که همان مهارت‌ها هستند، به وسیله سؤال‌های موجود در

---

1. Cognitive Diagnostic Assessment
2. Skill
3. Embretson & Reise
4. Cognitive Diagnostic Models
5. Sinharay & Almond
6. Leighton & Gierl
7. Confirmatory Factor Analysis

ماتریس Q اندازه گرفته می‌شوند. ماتریس Q فرضیه‌هایی درباره مهارت‌های لازم برای پاسخ درست هستند (لی[1] ۲۰۱۱) و معمولاً عناصر این ماتریس بر اساس قضاوت متخصصان قبل از تجزیه‌وتحلیل مدل تشخیصی شناختی تعیین می‌شود و در تجزیه‌وتحلیل‌های آماری درستی این روابط بررسی می‌شود. درمجموع، مدل‌های تشخیصی شناختی به دنبال آن هستند که با بررسی فرایندهای ذهنی اطلاعات دقیقی را در زمینه درک و فهم آزمودنی‌ها از حیطه‌ای که مورد سنجش قرارگرفته 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 به مهارت‌های مختلف و استفاده از فرایندهای عالی ذهنی نیاز دارد. از همین روی، بیشتر پژوهش‌های صورت گرفته در زمینه کاربرد مدل‌های تشخیصی شناختی به آزمون‌های ریاضیات و زبان (به‌ویژه مهارت درک مطلب) که نیازمند مهارت‌های شناختی هستند، اختصاص یافته است.

تصمیم مهمی که باید پژوهشگران حوزه سنجش تشخیصی شناختی در طی فرایند پژوهش خود بگیرند، انتخاب یک مدل تشخیصی شناختی بر اساس معیار جبرانی یا غیر جبرانی بودن روابط بین مهارت‌ها است (روسو، تمپلین و هنسو[2] ۲۰۰۷). در یک مدل جبرانی[3]داشتن سطح بالا در یک شایستگی می‌تواند سطح پایین در شایستگی دیگر را در انجام کار جبران کند. در مقابل مدل‌های جبرانی، مدل‌های غیر جبرانی[4] قرار دارند؛ با این ویژگی که در یک مدل غیر جبرانی داشتن سطح بالایی از شایستگی در یک مهارت نمی‌تواند سطح پایین شایستگی در مهارت دیگر را جبران کند. در سال‌های اخیر، دسته دیگری به نام مدل‌های کلی یا مدل‌های اشباع‌شده[5]به مدل‌های تشخیصی شناختی اضافه شده است. تمپلین و هنسون (۲۰۱۰) در کتاب خود مدل‌های اشباع‌شده تشخیصی شناختی را به مدل‌های معادلات ساختاری نامحدود که بین همه متغیرهای‌شان رابطه وجود دارد، تشبیه کرده‌اند؛ همچنین در این کتاب

---

1. Li
2. Roussos, Templin & Henson
3. .Compensatory Model
4. Non-Compensatory Model
5. Saturated model

مدل‌های اشباع‌شده در زمره مدل‌های جبرانی طبقه‌بندی شده است. ویژگی منحصربه‌فرد مدل‌های اشباع‌شده این است که در این مدل‌ها علاوه بر اثرات اصلی، اثرات تعاملی بین مهارت‌ها نیز مورد توجه قرارگرفته است. درواقع مدل‌های اشباع‌شده انواع مختلفی از روابط بین مهارت‌ها را در نظر می‌گیرند. مدل تشخیصی شناختی لگاریتم خطی[۱] (هنسون، تمپلین و ویلز[۲] ۲۰۰۹)، مدل تشخیصی کلی[۳] (داویر[۴] ۲۰۰۵) و مدل DINAG-[۵] (دی لاتوره[۶] ۲۰۱۱) ازجمله این مدل‌ها هستند.

کاربردهای اولیه مدل‌های تشخیصی شناختی بیشتر در ریاضیات بوده است که در آن راه‌حل به مراحل کوچک‌تری تقسیم می‌شود و اگر تمام مراحل موفقیت‌آمیز پشت سر گذاشته شود مسئله ریاضی مورد نظر حل می‌شود؛ بنابراین تا حدودی توافق شده است که مدل‌های غیر جبرانی برای آزمون‌های ریاضیات به کار رود (روسو، ۲۰۰۷). بااین‌حال در پژوهش‌های مربوط به درک مطلب خواندن وضعیت به این‌گونه نیست؛ زیرا در مورد جبرانی یا غیر جبرانی بودن مهارت‌های درک مطلب اختلاف‌نظر وجود دارد. به‌عنوان‌مثال استانویچ[۷](۱۹۸۰) یک مدل تعاملی- جبرانی[۸]را پیشنهاد داد. او در مدل خود چنین استدلال کرد که صرف‌نظر از سطح مهارت‌ها در سلسله‌مراتب پردازش، نبود یک فرایند خاص با تکیه بر دیگر منابع دانش قابل جبران است. گلداسمیت-فیلیپس[۹](۱۹۸۹) مدل پیشنهاد شده استانویچ را با استفاده از داده‌هایی که از دانش آموزان پایه‌های دوم، چهارم و ششم به‌دست‌آمده بود، آزمود. او در پژوهش خود به یک سیر تکاملی دست یافت؛ به این شرح که بچه‌ها با افزایش سن، برای شناسایی لغات استفاده از نشانه‌های آوایی[۱۰]را افزایش و استفاده از اطلاعات معنایی[۱۱]

---

1. the log-linear CDM
2. Willse
3. General Diagnostic Model
4. von Davier
5. Generalized DINA Model
6. De La Torre
7. Stanovich
8. compensatory-interactive model
9. Goldstmith-Philips
10. phonetic cues
11. sematic information

را کاهش می‌دهند. گلداسمیت–فیلیپس این یافته‌ها را به‌عنوان شاهدی برای تأیید مدل استانویچ و ماهیت جبرانی بودن مهارت‌های درک مطلب قلمداد کرد. برنارد[1](۲۰۰۵) نیز یک مدل جبرانی از خواندن زبان دوم را معرفی کرد؛ به این صورت که وجود دانش زبان دوم یادگیرنده و توانایی خواندن زبان اول، دو توانایی هستند که در طول فرایند خواندن ضعف‌های یکدیگر را جبران می‌کنند. یوسوآ- جوان[2](۲۰۰۶) نیز به این نتیجه رسید که دانش مرتبط با یک رشته و چیرگی بر زبان انگلیسی می‌توانند در خواندن زبان انگلیسی برای تحصیلات، مکمل و جبران‌کننده کمبودهای یکدیگر باشند. از سوی دیگر، مدل نمایش ساده‌ای از خواندن[3](گاف و تومر[4]، ۱۹۸۶) ارائه شد که بر طبق آن درک مطلب خواندن RC محصول درک زبانی C و رمزگشایی D است؛ یعنی RC=C×D. منظور از رمزگشایی استخراج سریع جمله و تفسیر گفته‌ها از اطلاعات واژگانی است. هوور[5]و گاف (۱۹۹۰) معتقد بودند که این دو مهارت اهمیت یکسانی داشته و هیچ‌کدام به‌تنهایی کافی نیست؛ درواقع آنها به یک رابطه غیر جبرانی اعتقاد داشتند. درمجموع شواهد نظری آمیخته‌ای درباره جبرانی یا غیر جبرانی بودن فرایند درک مطلب خواندن وجود دارد. زمانی که روابط بین مهارت‌های شناختی به‌طور کامل شناخته‌شده نیست، راه سومی هم پیشنهاد می‌شود؛ یعنی استفاده از یک مدل اشباع‌شده که به‌اندازه کافی انعطاف‌پذیر است. البته مدل‌های اشباع‌شده نیز مشکلات خاص خود را دارند. این مدل‌ها در مقایسه با مدل‌های قطعی پیچیده‌تر هستند و به حجم نمونه بزرگ‌تری برای برآورد دقیق عملکرد نیاز دارند و همچنین در تفسیر نتایج از مدل‌های قطعی دشوارتر هستند.

با توجه به اهمیت ویژه مدل‌های تشخیصی شناختی CDMs در سنجش آموزشی انتخاب یک مدل مناسب از بین انواع مدل‌های موجود به دغدغه‌ای مهم تبدیل شده است. تا جایی که در سال‌های اخیر موضوع بسیاری از پژوهش‌های این حوزه به مقایسه برازش بین مدل‌ها اختصاص یافته است. به‌عنوان مثال، در پژوهش باسوکو و

---

1. Bernhardt
2. Uso-Juan
3. Simple View of Reading
4. Gough & Tummer
5. Hoover

همکارانش (۲۰۱۳) مدل DINA و G-DINA با استفاده از شاخص‌های برازش AIC، BIC و 2LL- مقایسه شدند. تحلیل نتایج این پژوهش نشان داد که مدل G-DINA برازش بهتری نسبت به مدل DINA با داده‌های پژوهش دارد. البته در این پژوهش فقط به مقایسه دو مدل DINA و G-DINA بسنده شده و هیچ مدل جبرانی برای مقایسه در نظر گرفته نشده است؛ همچنین شاخص‌های برازش مطلق و میزان آنها برای این دو مدل گزارش نشده است. لی، هانتر و لی[1](۲۰۱۶) در پژوهشی روی آزمون خواندن MELAB پنج مدل شامل یک مدل اشباع شده "G-DINA" دو مدل جبرانی "ACDM,DINO" و دو مدل غیر جبرانی "RRUM,DINA"را با همدیگر مقایسه کردند. در مقایسه بین مدل G-DINA و مدل ACDM این دو مدل شاخص‌های برازش همانند و نیز طبقه‌بندی مشابهی از مهارت‌ها را نشان دادند. مدل RRUM در برازش مدل و نتایج طبقه‌بندی از دو مدل DINA و DINO بهتر اما نسبت به مدل ACDM و G-DINA ضعیف‌تر بود؛ بنابراین طبق این پژوهش دو مدل G-DINA و ACDM بهترین برازش را با مدل داشتند. این پژوهش در مقایسه با پژوهش باسوکو و همکارانش (۲۰۱۳) مدل‌های بیشتری را مورد توجه قرار داده و شاخص‌های برازش مطلق را نیز برای هر مدل محاسبه کرده است؛ اما به شاخص SRMSR که یکی از شاخص‌های برازش مطلق مهم است و برای رفع نارسایی‌های و مشکلات شاخص خی‌دو پیشنهاد می‌شود، اشاره‌ای نشده است. در پژوهش راوند و روبیچ[2] (۲۰۱۵) روی آزمون خواندن درک مطلب ورودی دانشگاه‌های ایران، شاخص‌های برازش نسبی مدل G-DINA با دو مدل DINA و ACDM مقایسه شد. نتیجه این مقایسه برتری مدل G-DINA بر دو مدل دیگر را تأیید کرد. با اینکه پژوهش راوند و روبیچ در مقایسه با دو پژوهش قبلی به شاخص‌های برازش بیشتری توجه کرده است؛ ولی فقط سه مدل مورد مقایسه قرار گرفته‌اند. بنابراین، در پژوهش حاضر تلاش شده است که مدل‌های بیشتری از طریق شاخص‌های برازش متنوع‌تر مقایسه شوند. با توجه به توضیحات ارائه شده و مرور پیشینه تجربی، هدف از اجرای پژوهش حاضر پاسخ به این پرسش است که آیا بین

---

[1]. Li, Hunter & Lei
[2]. Ravavd & Robitzsch

مدل‌های مختلف تشخیصی شناختی تفاوت وجود دارد و بهترین مدل ازنظر برازش با داده‌های آزمون خواندن و درک مطلب کدام مدل است؟

روش پژوهش

پژوهش حاضر را می‌توان پژوهشی توصیفی-مقایسه‌ای دانست؛ زیرا در این پژوهش هدف مقایسه مدل‌های تشخیصی شناختی با یکدیگر و تعیین مدل با برازش بهتر و معتبرتر است. به‌طور کلی، هر پژوهشی که توصیف یک پدیده یا یک موقعیت با هدف شناخت بهتر شرایط و کمک به تصمیم‌گیری‌های مناسب انجام گیرد در زمره پژوهش‌های توصیفی قرار می‌گیرد (سرمد، بازرگان و حجازی؛ ۱۳۹۱).

جامعه آماری این پژوهش شامل همه داوطلبانی می‌شود که در آزمون سراسری کارشناسی ارشد سال ۱۳۹٤ رشته زبان انگلیسی شرکت کرده‌اند. طبق آمار دریافتی از سازمان سنجش آموزش کشور، تعداد شرکت‌کنندگان رشته زبان در سال ۱۳۹٤، ١٦٠٤٤ نفر بودند که از این تعداد ١٢٠٧٦ نفر زن و ۳۹٦۸ نفر مرد بودند. تعداد کل افراد گروه نمونه که پاسخنامه آنها بررسی شد، ۳۰۰۰ نفر بودند. از این ۳۰۰۰ نفر افراد گروه نمونه، ۲۰۵۸ نفر زن و ۹٤۲ نفرشان مرد هستند. با توجه به اینکه نمونه مورد نظر بنا به درخواست پژوهشگر از بانک داده‌های سازمان سنجش آموزش کشور تهیه شد و پژوهشگر هیچ دخالتی در امر انتخاب نداشت و سازمان سنجش به‌صورت تصادفی داده‌های درخواستی پژوهشگر را از بین همه داده‌های 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 ۱۳۹٤ است که توسط سازمان سنجش آموزش کشور طراحی شده است؛ از بین ۲٤۰ 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 ٤۱ تا ٦۰ دفترچه عمومی). روش نمره‌گذاری سؤال‌های آزمون به‌صورت صفر و یک (دو ارزشی) است. درصورتی‌که فرد در پاسخ به یک سؤال گزینه درست را انتخاب کرده باشد، نمره یک و اگر گزینه‌های اشتباه را انتخاب کرده یا اصلاً پاسخ نداده نمره صفر می‌گیرد؛ بنابراین نمره کل داوطلبی که موفق شود به همه سؤال‌ها پاسخ درست بدهد برابر با نمره بیست و نمره کل داوطلبی که نتواند به هیچ‌یک از این بیست سؤال، پاسخ صحیح بدهد برابر با نمره صفر است.

یافته‌ها

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 قرارگرفته، به معنی نبود رابطه بین مهارت و سؤال است؛ درحالی‌که عدد یک نشان‌دهنده الزام وجود آن مهارت شناختی برای پاسخ به سؤال مورد نظر است. به منظور 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 درس‌های مربوط به خواندن و درک مطلب سابقه داشتند؛ ۹ مهارت پیشنهادی جانگ (۲۰۰۹) به‌عنوان 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 هریک از آنها در جدول (۱) ارائه شده است.

جدول (۱) مهارت‌های پایه و تعریف آنها

ردیف	مهارت	تعریف
۱	CDV	استنباط معنی یک کلمه یا عبارت با جستجو و تجزیه‌وتحلیل متن با کمک سرنخ‌هایی 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.
٤	TEI	خواندن سریع جمله‌های سراسر یک پاراگراف برای معنی کردن تحت‌الفظی جمله‌های یک متن.
٥	TIM	خواندن گزینشی یک پاراگراف یا همه پاراگراف‌ها برای تشخیص عقیده اصلی

ردیف	مهارت	تعریف
		که بیان شده بر اساس اطلاعات ضمنی در متن.
٦	INF	خواندن سطحی یک پاراگراف و استنتاج‌های گزاره‌ای درباره استدلال‌ها یا هدف نویسنده با مراجعه به اطلاعات بیان شده ضمنی یا دانش استقرایی.
۷	NEG	خواندن دقیق یا مؤثر به‌منظور تعیین مکان اطلاعات مناسب در متن و تعیین اطلاعات درست یا نادرست.
۸	SUM	تجزیه‌وتحلیل و ارزیابی اهمیت نسبی اطلاعات موجود در متن با کمک تفاوت قائل شدن بین ایده اصلی و جزئیات.
۹	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.

این ۹ مهارت همراه با سؤال‌های آزمون در اختیار سه متخصص سنجش زبان با سابقه تدریس درک مطلب و خواندن قرار گرفت و از آنها 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 ادغام‌شده 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 جمله‌های سراسر یک پاراگراف برای معنی کردن تحت‌الفظی جمله‌های یک متن) حذف شدند. دلیل اصلی حذف این مهارت‌ها این بود که برای قابل اعتماد 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 به بی‌ثباتی و عدم اطمینان طبقه‌بندی آزمودنی‌ها منجر 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 داده‌شده 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
٤	۱	۱	۰	۰	۰	۰	۰
٥	۰	۱	۱	۱	۰	۰	۱
٦	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
۱٤	۰	۱	۱	۰	۰	۰	۰
۱٥	۰	۱	۱	۰	۰	۰	۰
۱٦	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

برای نیل به هدف پژوهش (انتخاب مدل تشخیصی شناختی مناسب برای آزمون‌های خواندن و درک مطلب)، ٦ مدل شامل دو مدل جبرانی (ACDMو

DINO)، دو مدل غیر جبرانی (RRUMوDINA) و دو مدل کلی (G-DINA و LCDM) انتخاب شدند؛ تا برازش آنها با داده‌های موجود بررسی شود. برای تجزیه‌وتحلیل‌های کمی و محاسبه شاخص‌های برازش از نرم‌افزار R و پکیج CDM استفاده شد.

شاخص‌های برازش در مدل‌های تشخیصی شناختی به دو دسته کلی شاخص برازش مطلق[1] و شاخص برازش نسبی[2] تقسیم می‌شوند. شاخص‌های برازش مطلق، برازش مدل با داده‌ها را بررسی می‌کند، درحالی‌که شاخص‌های برازش نسبی برای مقایسه مدل‌ها با همدیگر مناسب است. درواقع شاخص‌های برازش نسبی به‌تنهایی ملاک قضاوت قرار نمی‌گیرد و در مقایسه با شاخص‌های مدل دیگر معنا پیدا می‌کنند.

یکی از شاخص‌های برازش مطلق آماره ماکزیمم χ^2 هست (چن، دی لاتوره و ژانگ[3] ۲۰۱۳). $\chi^2_{jj'}$ تفاوت بین احتمال پاسخ پیش‌بینی‌شده و مشاهده‌شده را برای دو سؤال j و j' محاسبه می‌کند. درحالی‌که احتمال پاسخ‌های پیش‌بینی‌شده فرض استقلال موضعی را نیز در نظر می‌گیرند، امکان دارد که احتمال پاسخ‌های مشاهده شده از این فرض منحرف شوند. آماره $\chi^2_{jj'}$ به وسیله تابع زیر قابل‌محاسبه است.

$$\chi^2_{jj'} = \sum_{n=0}^{1}\sum_{m=0}^{1} \frac{[N(X_j = n, X_{j'} = m) - \hat{N}(X_j = n, X_{j'} = m)]^2}{\hat{N}(X_j = n, X_{j'} = m)}$$

ماکزیمم χ^2 یک شاخص برازش مطلق است که با $\chi^2_{1-\alpha}$ با درجه آزادی یک مقایسه می‌شود. بنابراین اگر CDM تحلیل‌شده داده‌ها را به‌خوبی تقریب بزند ارزش χ^2 پایین می‌شود.

---

[1]. Absolute Fit
[2] . Relative Fit
[3]. Chen, de la Torre & Zhang

یک شاخص برازش مطلق دیگر ریشه دوم استاندارد میانگین مجذورات باقی‌مانده[1](SRMSR) است. شاخص SRMSR متوسط اندازه باقی‌مانده‌های بین کوواریانس مشاهده‌شده و کوواریانس پیش‌بینی‌شده است که با تابع زیر محاسبه می‌شود:

$$SRMSR = \sqrt{\sum_{i<j} \frac{(r_{ij} - \hat{\rho}_{ij})^2}{n(n-1)/2}}$$

SRMSR با ارزش کوچک‌تر از ۰/۰۵ نشان‌دهنده برازش خوب مدل با داده‌هاست (مای دیا- اولیوارس[2] ۲۰۱۳). در جدول (۳) شاخص‌های برازش مطلق گزارش‌شده است.

جدول (۳) شاخص‌های برازش مطلق مدل‌ها

مدل	MADcor	SRMSR	maxx^2	p
DINA	۰/۰۳۱	۰/۰٤۰	۳۲/۵۱	۰/۰۰۰۱
G-DINA	۰/۰۲۰	۰/۰۲٦	۳۸/۱۲	۰/۰۰۰۱
DINO	۰/۰۳۲	۰/۰٤۰	۳۹/۱۰	۰/۰۰۰۱
RRUM	۰/۰۲۷	۰/۰۳٦	۵۰/۷۵	۰/۰۰۰۱
ACDM	۰/۰۲۳	۰/۰۲۹	۳۹/۸۳	۰/۰۰۰۱
LCDM	۰/۰۲٤	۰/۰۳۰	۳۲/۸۸	۰/۰۰۰۱

در جدول (۳) سه شاخص ماکزیمم $\mathbf{x^2}$ ، SRMSR و MADcor به‌عنوان شاخص‌های برازش مطلق و نیز سطح معنی‌داری $\mathbf{x^2}$ نشان داده شده است. برای هر شش مدل سطح معنی‌داری ماکزیمم $\mathbf{x^2}$ کمتر از ۰/۰۱ شده است و این نتیجه به

[1]. The Standardized Root Mean Square Residual
[2]. Maydeu-Olivares

معنای آن است که مدل با محیط (داده‌ها) متجانس نیست. اما ازآنجایی‌که $\mathbf{x^2}$ به‌شدت تحت تأثیر حجم نمونه قرار دارد دو شاخص دیگر نیز گزارش شده است. شاخص SRMSR برای هر شش مدل کمتر از ۰/۰۵ است که نشان‌دهنده برازش مطلوب هر شش مدل با داده‌ها است. برای شاخص MADcor که نشان‌دهنده میانگین قدر مطلق انحراف همبستگی مشاهده شده بین دو سؤال با همبستگی مورد انتظار است، نقطه برش مشخصی تعیین نشده است؛ ولی هرچه میزان این شاخص به صفر نزدیک‌تر باشد برازش مدل مطلوب‌تر ارزیابی می‌شود. در مقایسه بین شش مدل مورد بررسی میزان شاخص MADcor مربوط به مدل G-DINA از همه مدل‌ها کوچک‌تر است.

شاخص‌های برازش نسبی نمی‌توانند برازش بین مدل منفرد و داده‌ها را نشان دهند. این شاخص‌ها به‌جای فراهم کردن اطلاعات درباره برازش داده‌ها، از دو مدل برای مقایسه و تعیین مدل بهتر استفاده می‌کنند. ازجمله مهم‌ترین شاخص‌های برازش نسبی می‌توان به لگاریتم درست‌نمایی، شاخص اطلاعات آکائیک[1](AIC) و شاخص اطلاعات بیزی[2](BIC) اشاره کرد. AIC یک معیار انتخاب مدل، قوی هست که در مقایسه بین دو مدل با اندازه‌های متفاوت به کار می‌رود (باندولوس[3] ۱۹۹۳). BIC نیز یک معیاری است که برای رفع مشکلات مدل انتخابی در رگرسیون در نظر گرفته شده است (باسوکو، اجرتمن، کلسیوگلو[4]۲۰۱۳).

$$AIC = -2logL(X) + 2.P$$

و

$$BIC = -2logL(X) + log\ I\ .\ P$$

---

[1] . Akaike Information Criterion
[2]. Bayesian Information Criterion
[3]. Bandolos
[4]. Basokcu,Ogretmen & Kelecioglu

در این فرمول‌ها منظور از $logL(X)$ همان بیشینه لگاریتم درست‌نمایی است و P نیز نشان‌دهنده تعداد کل پارامترهاست. نحوه قضاوت درباره برتری مدل با توجه به شاخص‌های AIC و BIC به این صورت است که مدل با شاخص‌های AIC و BIC کوچک‌تر برازش مطلوب‌تری نسبت به مدل با AIC و BIC بزرگ‌تر خواهد داشت (کاوانا[1] ۲۰۰۹). در جدول (٤) شاخص‌های AIC، BIC و -2log likelihood برای مقایسه بین مدل‌ها گزارش شده است.

جدول (٤) شاخص‌های برازش نسبی در مقایسه مدل‌ها

مدل	-2Log like	Nobs	Npars	AIC	BIC
DINA	٥٣٤١٦/٨٠	٣٠٠٠	١٦٧	٥٣٧٥٠/٧٩	٥٤٧٥٣/٨٦
G-DINA	٥٢٥٤٣/٥٢	٣٠٠٠	١٧٥	٥٢٨٩٣/٥١	٥٣٩٤٤/٦٢
DINO	٥٣٤٤٢/٥٢	٣٠٠٠	١٦٧	٥٣٧٧٦/٥٣	٥٤٧٧٩/٥٩
RRUM	٥٣٠٧٩/٧٨	٣٠٠٠	١٠٠	٥٣٢٧٩/٧٨	٥٣٨٨٠/٤١
ACDM	٥٢٧٧١/٨٤	٣٠٠٠	١٠٠	٥٢٩٧١/٨٤	٥٣٥٧٢/٤٨
LCDM	٥٢٩١٤/١٠	٣٠٠٠	١٩٨	٥٣٣١٠/١٠	٥٤٤٩٩/٣٦

همان‌طور که در جدول (٤) دیده می‌شود کمترین میزان شاخص AIC به ترتیب متعلق به مدل G-DINA و ACDM است. کمترین مقدار شاخص BIC نیز به ترتیب به مدل‌های ACDM، RRUM و G-DINA تعلق دارد. در مورد شاخص 2LL- نیز به ترتیب مدل G-DINA، ACDM و LCDM دارای کمترین مقدار هستند. در هر سه شاخص بیشترین مقدار به ترتیب به مدل‌های DINO و DINA تعلق دارد.

آزمون LR یک آزمون آماری است که برای مقایسه شاخص‌های برازش بین دو مدل که یکی از آنها در دیگری آشیانه[2]می‌کند، کاربرد دارد. آزمون LR که با تابع زیر قابل محاسبه است؛ توزیعی شبیه به توزیع خی‌دو با درجه آزادی برابر با تفاوت بین پارامترهای مدل کلی و مدل کاهش‌یافته دارد (گارسیا، اولا و دی لاتوره،[3] ۲۰۱٤).

---

[1]. Cavanaugh
[2]. nested
[3]. García, Olea & de la Torre

$$LR = [-2LL^{reduced\ model}] - [-2LL^{general\ model}]$$

در جدول (۵) چهار مدل DINO، ACDM،DINA و RRUM با مدل G-DINA از طریق آزمونLR مقایسه شده‌اند.

جدول (۵) نتایج آزمون LR

مدل ۱	مدل ۲	Chi2	df	p
DINA	G-DINA	۸۷۳/۲۸	۸	۰/۰۰۰۱
DINO	G-DINA	۰۲/۸۹۹	۸	۰/۰۰۰۱
RRUM	G-DINA	۵۳۶/۲۷	۷۵	۰/۰۰۰۱
ACDM	G-DINA	۲۲۸/۳۳	۷۵	۰/۰۰۰۱

بر اساس نتایج آزمون LR بین مدل G-DINA با هریک از چهار مدل DINO، ACDM،DINA و RRUM تفاوت معنی‌دار وجود دارد. به این معنی که مدل G-DINA برازش بهتری با داده‌ها دارد و این برازش بهتر حاصل شانس و تصادف نیست.

در مدل‌های تشخیصی شناختی به منظور بررسی برازش سؤال با مدل از شاخص RMSEA به معنای ریشه میانگین مجذور خطای تقریب[1] استفاده می‌شود. شاخص برازش RMSEA از سؤال j احتمال پاسخ سؤال پیش‌بینی‌شده توسط مدل را $P(X_j = 1|\alpha_l)$ با پاسخ‌های درست پیش‌بینی‌شده $\widehat{N}(X_j|\alpha_l)$ در هر کلاس مهارت α_l مقایسه می‌کند.

[1]. Root Mean Square Error Approximation

$$RMSEA_j = \sqrt{\sum_{l=1}^{L} p(\alpha_l)\left[P(X_j = 1|\alpha_l) - \frac{\widehat{\mathrm{N}}(X_j = 1|\alpha_l)}{\widehat{\mathrm{N}}(X_j|\alpha_l)}\right]^2}$$

$p(\alpha_l)$ نشان‌دهنده فراوانی آزمودنی‌های طبقه‌بندی‌شده در کلاس مهارت α_l و $\widehat{\mathrm{N}}(X_j|\alpha_l)$ تعداد کل پاسخ‌های پیش‌بینی برای سؤال j است که توسط آزمودنی‌ها در کلاس مهارت α_l ارائه‌شده است. سؤال‌هایی که میزان شاخص RMSEA آنها بیشتر از ۰/۱باشد برازش ضعیفی با مدل دارند، ولی سؤال‌هایی که میزان شاخص RMSEA آنها بین ۰/۱ و ۰/۰۵ باشد برازش متوسطی با مدل خواهند داشت؛ مطلوب‌ترین میزان برای شاخص RMSEA مقادیر کمتر از۰/۰۵ است که نشان‌دهنده برازش خوب سؤال با مدل هستند (کانینا–هبینت، راپ و ویلهلم[1] ۲۰۰۹). در جدول (٦) شاخص‌های برازش مربوط به هر سؤال برای هریک از شش مدل مورد نظر گزارش شده است.

جدول (٦) شاخص برازش سؤال

سؤال	DINA	G-DINA	DINO	RRUM	ACDM	LCDM
۱	۰/۰۲۵	۰/۰۱۸	۰/۰۱۳	۰/۰٤۷	۰/۰۱۷	۰/۰۳۸
۲	۰/۰۲٤	۰/۰۲٦	۰/۰۲٦	۰/۰۳۲	۰/۰۰۷	۰/۰۳۱
۳	۰/۰۲۳	۰/۰۲٦	۰/۰۳۰	۰/۰۲۲	۰/۰۱۹	۰/۰٤۱
٤	۰/۰۲٤	۰/۰۲٦	۰/۰۱۹	۰/۰۳۱	۰/۰۳۲	۰/۰۲٦
٥	۰/۰۲٦	۰/۰۱۱	۰/۰۳٤	۰/۰۲۷	۰/۰۱۷	۰/۰۲۸
٦	۰/۰۲٦	۰/۰۲۲	۰/۰۳۳	۰/۰۲٥	۰/۰۱۸	۰/۰۲۷
۷	۰/۰٥٤	۰/۰۲٥	۰/۰۱۹	۰/۱۰۸	۰/۰۲٥	۰/۰٤۱
۸	۰/۰۳۷	۰/۰۰۹	۰/۰۲۰	۰/۰۱۹	۰/۰۱۳	۰/۰۳۳
۹	۰/۰۲۲	۰/۰۲٤	۰/۰۱٤	۰/۰۱۲	۰/۰۱۱	۰/۰۳۱
۱۰	۰/۰۲۳	۰/۰۳۰	۰/۰۳٤	۰/۰۲۳	۰/۰۱۷	۰/۰۰۹

[1]. Kunina-Habenicht, Rupp & Wilhelm

سؤال	DINA	G-DINA	DINO	RRUM	ACDM	LCDM
۱۱	۰/۰۳۹	۰/۰۳۰	۰/۰٤۰	۰/۰۳۱	۰/۰۳۳	۰/۰٤۹
۱۲	۰/۰۱۳	۰/۰۲۳	۰/۰۳۳	۰/۰۱۷	۰/۰۱۷	۰/۰۳٤
۱۳	۰/۰۲۰	۰/۰۲۷	۰/۰۱۱	۰/۰۲۲	۰/۰۱۳	۰/۰۳۰
۱٤	۰/۰۲۳	۰/۰۲۳	۰/۰۲۲	۰/۰۳۳	۰/۰٥٤	۰/۰۳٤
۱٥	۰/۰۲۲	۰/۰۲۹	۰/۰۱۷	۰/۰۲۸	۰/۰۲۹	۰/۰۳۸
۱٦	۰/۰۲۰	۰/۰۲٥	۰/۰۱٦	۰/۰۲۲	۰/۰۱۷	۰/۰۳۹
۱۷	۰/۰۲٤	۰/۰۱٥	۰/۰۲۳	۰/۰۰٥	۰/۰۱٤	۰/۰۱٤
۱۸	۰/۰۲٤	۰/۰۳۷	۰/۰۲۷	۰/۰۳۷	۰/۰۱۸	۰/۰۳۳
۱۹	۰/۰۱۹	۰/۰۱۳	۰/۰٥۱	۰/۰۳٥	۰/۰۲۸	۰/۰۳۱
۲۰	۰/۰۲٤	۰/۰۱٦	۰/۰٤۸	۰/۰۲۷	۰/۰۳۰	۰/۰۲۱

بر اساس اطلاعات درج شده در جدول (٦)، پرسش ۷ با مدل DINA و RRUM، پرسش ۱٤ با مدل ACDM و پرسش ۱۹ با مدل DINO برازش مطلوبی ندارند (RMSEA>۰/۰٥).

بحث و نتیجه‌گیری

با توجه به ادبیات پژوهشی موجود در حوزه مدل‌پردازی تشخیصی شناختی آزمون‌های خواندن و درک مطلب، راجع به جبرانی یا غیر جبرانی بودن مهارت‌های زیربنایی خواندن در بین صاحب‌نظران اتفاق‌نظر وجود ندارد؛ تا جایی که موضوع بسیاری از پژوهش‌ها به مقایسه برازش مدل‌های مختلف اختصاص یافته است. در این پژوهش نیز بعد از تدوین مدل شناختی در قالب ماتریس Q که دربردارنده رابطه بین ۲۰ سؤال خواندن و درک مطلب با ۷ مهارت لازم برای پاسخ به این سؤال‌ها بود؛ ٦ مدل تشخیصی شناختی پرکاربرد در این حوزه به منظور مقایسه شاخص‌های برازش‌شان با یکدیگر انتخاب شدند. این مدل‌ها عبارت‌اند از: مدل G-DINA و LCDM (مدل‌های کلی)، مدل DINO و ACDM (مدل‌های جبرانی) و مدل DINA و RRUM (مدل‌های غیر جبرانی).

در ابتدا شاخص‌های برازش مطلق شامل ماکزیمم خی‌دو MADcor (میانگین قدر مطلق انحراف همبستگی‌های مشاهده‌شده با همبستگی‌های مورد انتظار) و

SRMSR (ریشه دوم استاندارد میانگین مجذورات باقی‌مانده) برآورد شدند. شاخص ماکزیمم خی‌دو برای هر شش مدل معنی‌دار شد؛ یعنی همه شش مدل برازش مطلوبی با داده‌ها ندارند. در مقابل میزان شاخص SRMSR برای هر شش مدل کمتر از ۰/۰۵ شد؛ که این میزان گویای برازش مطلوب مدل با داده‌ها است (مای دیا- اولیوارس، ۲۰۱۳). در پژوهش راوند و روبیچ (۲۰۱۵) نیز برازش داده‌ها با سه مدل G-DINA، DINA و ACDM بررسی شد؛ نتیجه این پژوهش نشان داد که ازنظر شاخص خی‌دو هیچ‌کدام از مدل‌ها با داده‌های پژوهش برازش ندارند؛ ولی شاخص SRMSR برازش مدل با داده‌ها را تأیید کرد. همچنین در پژوهش جورج و روبیچ[1] (۲۰۱۵) سه مدل شناختی مختلف که هرکدام از حذف یک مهارت شناختی به وجود آمده بودند در قالب مدل DINA مقایسه شدند؛ در این پژوهش نیز هر سه این مدل‌ها که فقط از نظر نوع مهارت‌ها با یکدیگر تفاوت داشتند، بر اساس آماره خی‌دو با داده‌های پژوهش برازش نشان ندادند؛ ولی بر طبق شاخص SRMSR برازش هر سه مدل با داده‌ها مطلوب بود. این وضعیت عدم برازش بر اساس آزمون خی‌دو شبیه به عملکرد آزمون خی‌دو در مدل‌های معادلات ساختاری است. بنتلر[2] (۱۹۹۰) در مورد عدم برازش مدل معادلات ساختاری بر اساس خی‌دو این‌چنین استدلال کرد که خی‌دو به حجم نمونه حساس است؛ درنتیجه با افزایش حجم نمونه آزمون خی‌دو تفاوت‌های جزئی بین کوواریانس مشاهده‌شده و پیش‌بینی‌شده را تشخیص می‌دهد؛ بنابراین همان‌گونه که در مدل معادلات ساختاری شاخص‌های برازش دیگری به‌عنوان جایگزین برای خی دو معرفی شده است، در مدل‌های تشخیصی شناختی نیز شاخص SRMSR، به‌عنوان یک شاخص برازش مطلق دیگر ارائه شده است. ازآنجایی‌که شاخص SRMSR توسط تعداد سؤال‌ها و هیچ‌یک از عوامل دیگر که ثابت نگه داشته شده‌اند، تحت تأثیر قرار نمی‌گیرد، شاخص مناسبی برای قضاوت در مورد برازش مدل با داده‌ها است.

چنان‌که هر شش مدل در شاخص‌های برازش مطلق وضعیت تقریباً یکسانی داشتند؛ از شاخص‌های برازش نسبی استفاده شد که توانایی مقایسه مدل‌ها با یکدیگر

---

1. George & Robitzsch
2. Bentler

را دارند. شاخص لگاریتم درست‌نمایی، شاخص اطلاعات آکائیک (AIC) و شاخص اطلاعات بیزی (BIC) در این پژوهش، به‌عنوان شاخص برازش نسبی مورد بررسی قرار گرفتند. با توجه به یافته‌های پژوهش و در نظر گرفتن این معیار که هرچه میزان AIC، BIC و 2LL– یک مدل کوچک‌تر باشد، آن مدل برازش بهتری با داده‌ها دارد (کاوانا، ۲۰۰۹)، از بین این شش مدل، مدل‌های G-DINA و ACDM برازش بهتر و مدل‌های DINO و DINA برازش ضعیف‌تری در مقایسه با مدل‌های دیگر نشان دادند؛ بنابراین مدل‌های G-DINA و ACDM نسبت به دو مدل دیگر، مناسب‌ترین مدل برای تحلیل داده‌های این پژوهش هستند. ازجمله پژوهش‌های همسو با این پژوهش می‌توان به پژوهش باسوکو و همکارانش (۲۰۱۳) اشاره کرد. در این پژوهش برازش مدل‌های DINA و G-DINA بر اساس شاخص‌های AIC، BIC و -2LL مقایسه شدند که برازش مدل G-DINA با داده‌های پژوهش بهتر از برازش مدل DINA گزارش شد؛ اما این پژوهش روی داده‌های یک آزمون ریاضیات پایه ششم اجرا شده و از نظر نوع داده‌ها با داده‌های پژوهش حاضر متفاوت است. در پژوهشی دیگر (لی، هانتر و لی، ۲۰۱۶) که روی آزمون خواندن MELAB انجام گرفت، پنج مدل با یکدیگر مقایسه شدند. نتیجه مقایسه این پنج مدل به این صورت بود که مدل G-DINA و ACDM بهترین شاخص‌های برازش را نشان دادند و بعد از این دو، مدل RRUM در رتبه بعدی قرار گرفت و مدل DINO و DINA نیز بدترین شاخص‌های برازش را به‌دست آوردند. همچنین در پژوهش راوند و روبیچ (۲۰۱۵) روی یک آزمون خواندن، برازش مدل G-DINA نسبت به دو مدل DINA و ACDM بهتر بود. با وجود برتری مدل G-DINA در بیشتر شاخص‌های برازش، یک نکته قابل تأمل است؛ در بیشتر موارد به هنگام مقایسه شاخص 2LL-، مدل‌های کلی همچون G-DINA نسبت به مدل‌هایی مانند DINA و DINO برازش بهتری نشان می‌دهند؛ بنابراین شایسته است معنی‌داری آماری این تفاوت بررسی شود (سورل و همکاران[1] ۲۰۱۶). درواقع یک مدل کلی با تولید پارامترهای بیشتر شانس داشتن برازش بهتر نسبت به سایر مدل‌ها را دارد؛ اما ازآنجایی‌که ساده‌تر بودن یک مدل مزیت محسوب می‌شود؛ پیش از هر تصمیم‌گیری

[1] . Sorrel et al

باید مشخص شود که آیا تفاوت مشاهده‌شده، معنی‌دار هست یا نه و در صورت معنی‌دار بودن این تفاوت برتری با مدل کلی (با پارامترهای بیشتر) خواهد بود. از همین روی، در این پژوهش پس از مشخص شدن برتری مدل G-DINA بر سایر مدل‌ها از آزمون LR استفاده شد و نتیجه آن گویای این مطلب بود که برتری مدل G-DINA بر سایر مدل‌ها حاصل شانس و تصادف نبوده و واقعاً بین این مدل با سایر مدل‌ها به لحاظ آماری تفاوت معنی‌دار وجود دارد. در آخر نیز شاخص RMSEA که برازش هریک از سؤال‌های آزمون با مدل را نشان می‌دهد برای هر یک از این ٦ مدل محاسبه شد. در بررسی شاخص RMSEA، مدل‌های G-DINA و LCDM با تمام سؤال‌های آزمون برازش مطلوبی داشتند؛ زیرا مقدار RMSEA همه سؤال‌ها در این دو مدل کمتر از ۰/۰۵ بود. از بین تمام سؤال‌ها، تنها سؤال ۷ با مدل DINA و RRUM، سؤال ١٤ با مدل ACDM و سؤال ۱۹ با مدل DINO برازش مطلوبی نداشتند.

بنا بر یافته‌های این پژوهش و نتایج پژوهش‌های دیگری که به آنها اشاره شد؛ مدل G-DINA در برازش با داده‌ها نسبت به سایر مدل‌ها موفق‌تر عمل می‌کند. این نتیجه با توجه به اینکه مدل G-DINA یک مدل اشباع‌شده است و همه روابط موجود را در نظر می‌گیرد، قابل توجیه است. با این وجود، برای تأیید این موضوع که آیا تحت هر شرایط و هر نوع آزمونی این نتیجه صادق است یا نه، به پژوهش‌های بیشتر و دقیق‌تری در این حیطه نیاز است.

منابع

امبرتسون، سوزان ای؛ و رایس، استیون پی. (۲۰۰۰). *نظریه‌های جدید روان‌سنجی برای روان شناسان*؛ ترجمه حسن پاشا شریفی، ولی‌اله فرزاد، مجتبی حبیبی عسگرآباد و بلال ایزانلو (۱۳۸۸). تهران: انتشارات رشد.

سرمد، زهره؛ بازرگان، عباس؛ و حجازی، الهه.(۱۳۹۱). *روش‌های تحقیق در علوم رفتاری*. تهران: نشر آگاه.

Bandalos, D. L. (1993). Factors influencing cross-validation of confirmatory factor analysis models. *Multivariate Behavioral Research*, *28*(3), 351-374.

Basokcu, T. O., Ogretmen, T., & Kelecioglu, H. (2013). Model data fit comparison between DINA and G-DINA in cognitive diagnostic models. *Education Journal*, *2*(6), 256-262.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238.

Bernhardt, E. (2005). Progress and procrastination in second language reading. *Annual Review of Applied Linguistics*, *25*, 133-150.

Cavanaugh, J. E. (2009). *171: 290 model selection, lecture VI: The Bayesian information criterion*. Department of Biostatistics, Department of Statistics and Actuarial Science, University of Iowa.

Chen, J., Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, *50*(2), 123-140.

Davier, M. (2005). A general diagnostic model applied to language testing data. *ETS Research Report Series*, *2005*(2), i-35.

De La Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*(2), 179-199.

García, P. E., Olea, J., & De la Torre, J. (2014). Application of cognitive diagnosis models to competency-based situational judgment tests. *Psicothema*, *26*(3), 372-377.

George, A. C., & Robitzsch, A. (2015). Cognitive Diagnosis Models in R: A Didactic. *The Quantitative Methods for Psychology*, *11*(3), 189-205.

Goldsmith-Phillips, J. (1989). Word and context in reading development: A test of the interactive-compensatory hypothesis. *Journal of Educational Psychology*, *81*(3), 299.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial & Special Education*, *7*(1), 6-10.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191-210.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading & Writing*, *2*(2), 127-160.

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.

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 & Perspectives*, *11*(3), 71-101.

Ravand, H., & Robitzsch, A. (2015). Cognitive Diagnostic Modeling Using R. *Practical Assessment, Research & Evaluation*, *20*.

Roussos, L. A., Templin, J. L., & Henson, R. A. (2007). Skills Diagnosis Using IRT-Based Latent Class Models. *Journal of Educational Measurement*, *44*(4), 293-311

Sinharay, S., & Almond, R. G. (2007). Assessing fit of cognitive diagnostic models, a case study. *Educational & Psychological Measurement*, *67*(2), 239-257.

Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016). Validity and reliability of situational judgement test scores: A new approach based on cognitive diagnosis models. *Organizational Research Methods*, *19*(3), 506-532.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*(1), 32-71.

Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.

Usó-Juan, E. (2006). The compensatory nature of discipline-related knowledge and English-language proficiency in reading English for academic purposes. *Modern Language Journal*, *90*(2), 210-227.