



سنجش شناختی تشخیصی: مرور انتقادی مطالعات ایرانی Cognitive Diagnostic Assessment: Critical Review of Iranian Studies

Maryam Mohsenpour
Azam Moghadam

مریم محسن‌پور*
اعظم مقدم**

Abstract

Over the past decade, cognitive diagnostics has been favored by educational researchers as an approach to assessing student achievement, but the use of DCMs has been criticized. Due to the tendency of Iranian researchers towards this type of assessment, this study was conducted with the aim of critically reviewing Iranian research in this field in order to analyze its methodology. For this purpose, the main relevant Iranian databases including SID, Google Scholar, Scopus, Noormags, Magiran, and Elmnet were searched, and 32 studies were considered. Considering the criteria for inclusion and exclusion of studies, 20 studies in this field that were conducted between 2015 and 2020 were selected. In the next step, the selected studies were critically reviewed independently by the researchers using the indicators recommended by Sessoms & Henson (2018). The indicators included the thematic focus of the research, the number of competencies, the constructs measured, the structure of the test items, the Q matrix, the statistical model type, the model fit, the item fit, the question type, the sample size, the evidence of reliability and validity, the description of the competencies (latent categories, the correlation between competencies, the competency profile), and the application of the results. The results showed that a large number of the research conducted did not provide empirical support for general assumptions such as correlation between attributes, did not provide sufficient evidence for reliability and validity, did not report on the competency profile, and did not provide feedback to stakeholders. We provide some guidelines to improve the quality of such models.

Keywords: Cognitive Diagnostic Assessment, Competency, Q-Matrix, Diagnostic Classification Model, Critical Review.

چکیده

در چند دهه اخیر، به‌منظور سنجش عملکرد دانش‌آموزان، سنجش شناختی تشخیصی در میان پژوهشگران آموزشی محبوبیت فراوانی پیدا کرده، اما کاربرد آن با انتقادهایی مواجه شده است. با توجه به گرایش پژوهشگران ایرانی به این نوع سنجش، مطالعه حاضر با هدف مرور نقادانه پژوهش‌های ایرانی در این حوزه، به‌منظور تحلیل مباحث مربوط به بخش روش آن‌ها انجام شده است. برای این منظور، جست‌وجویی جامع در بانک‌های اطلاعاتی شامل ایرانداک، پایگاه داده‌های علمی جهاد دانشگاهی، گوگل اسکالر، اسکوپوس، نور، مگایران و موتور جست‌وجوی علم‌نت و با استفاده از کلمات کلیدی «سنجش شناختی تشخیصی»، «مدل شناختی تشخیصی»، «مدل طبقه‌بندی تشخیصی» و «تحلیل تشخیصی» صورت گرفت و ۳۲ پژوهش شناسایی شدند. با توجه به ملاک‌های ورود و خروج مطالعات، ۲۰ پژوهش در بازه زمانی ۱۳۹۹-۱۳۹۳ انتخاب شدند. در گام بعدی، این پژوهش‌ها را پژوهشگران مطالعه به‌طور مستقل و با استفاده از نشانگرهای دستورالعمل توصیه‌شده از سوی سسومز و هنسون (۲۰۱۸) ارزیابی کردند. نشانگرها شامل تمرکز موضوعی پژوهش‌ها، تعداد صلاحیت‌ها، سازه مورد اندازه‌گیری، ساختار سؤال‌های آزمون‌ها، ماتریس کیو، نوع مدل آماری، برازش مدل، برازش سؤال، نوع سؤال، حجم نمونه، شواهد پایایی و روایی، توصیف صلاحیت‌ها (طبقه‌های پنهان، همبستگی بین صلاحیت‌ها، نمرخ صلاحیت) و کاربست یافته‌ها بود. یافته‌ها حاکی از وارد بودن برخی انتقادهای مانند برقراری نبودن مفروضات و هم‌پوشانی بالای صلاحیت‌های شناختی، نبود شواهد پایایی و روایی کافی، ارائه نکردن نمرخ صلاحیت‌ها و بازخورد به ذی‌نفعان در بیشتر مطالعات بود. در مجموع رهنمودهایی برای بهبود کیفیت کاربرد این مدل‌ها ارائه شده است.

واژه‌های کلیدی: سنجش شناختی تشخیصی، صلاحیت، ماتریس کیو، مدل طبقه‌بندی تشخیصی، مرور انتقادی.

* نویسنده مسئول: استادیار گروه روان‌شناسی تربیتی، دانشکده علوم تربیتی و روان‌شناسی، دانشگاه الزهراء، تهران، ایران
** دکتری سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، دانشکده علوم تربیتی و روان‌شناسی تهران، ایران

Email: m.mohsenpour@alzahra.ac.ir

Received: 27 Sep 2020

Accepted: 20 Feb 2021

پذیرش: ۹۹/۱۲/۰۲

دریافت: ۹۹/۰۷/۰۶

مقدمه

در دهه‌های اخیر، استفاده از آزمون‌های شناختی تشخیصی^۱ به منظور سنجش توانایی‌های آزمون‌شوندگان مدنظر پژوهشگران حوزه اندازه‌گیری آموزشی و روان‌شناسی شناختی قرار گرفته است؛ زیرا اطلاعات به‌دست‌آمده از این نوع آزمون‌ها، به‌طور معناداری می‌تواند از سوی ذی‌نفعان^۲ آموزشی تفسیر شود (راپ، تمپلین و هنسون، ۲۰۱۰). برخلاف روش‌های مرسوم اندازه‌گیری آموزشی مانند نظریه کلاسیک آزمون و نظریه سؤال پاسخ تک‌بعدی که برآورد جایگاه فرد روی یک متغیر پنهان^۳، به کمک یک نمره کلی از تسلط فرد در حیطه‌ای محتوایی صورت می‌گیرد (چوی، لی و پارک، ۲۰۱۵). نتایج این نوع سنجش‌ها می‌تواند تصویری جامع از نیم‌رخ شناختی دانش‌آموزان درباره تکلیف‌های درسی فراهم کند (لیتون و گیرل، ۲۰۰۷).

طراحی ارزشیابی‌های آموزشی بر مبنای سنجش شناختی تشخیصی، مزایای فراوانی دارد که یکی از مهم‌ترین آن‌ها، پیوند دادن نظریه‌های شناختی و یادگیری به کمک آموزش است. از دیدگاه گیرل، ونگ و ژو (۲۰۰۸)، نتایج این سنجش‌ها با فراهم کردن تصویری جامع از نیم‌رخ شناختی دانش‌آموزان در زمینه تکلیف‌های درسی امکان شناسایی ضعف‌های دانش‌آموزان و طراحی‌های مناسب برای برطرف کردن آن‌ها را میسر می‌کند. تعریفی دقیق از سنجش شناختی تشخیصی در اندازه‌گیری آموزشی، به معنای طراحی آزمون برای سنجش و درک فرایندهای شناختی مستتر در پاسخ‌های دانش‌آموزان است (راپ، تمپلین و هنسون، ۲۰۱۰). همچنین از دیدگاه راپ، تمپلین و هنسون (۲۰۱۰) در این مدل‌ها، نه تنها درباره کیفیت پارامترهای سؤال بحث می‌شود، بلکه می‌توان به کمک احتمال موفقیت فرد در صلاحیت‌های گوناگون، کارنامه تشخیصی به آزمودنی‌ها ارائه کرد که در آن میزان تسلط هر فرد در هر یک از صلاحیت‌ها در قالب نیم‌رخ تسلط مشخص می‌شود. این نحوه ارائه نمرات، یعنی چندبعدی در نظر گرفتن آزمون به جای تک‌بعدی در نظر گرفتن آن، به تحقق هدف سنجش برای یادگیری به جای سنجش پیامدهای یادگیری کمک می‌کند. از دیدگاه جانگ (۲۰۰۸) رویکرد سنجش شناختی تشخیصی با هدف ارتقای سنجش برای یادگیری و به‌منزله فرایند یادگیری^۴، در مقابله با سنجش بازده‌های یادگیری^۵ به کمک فراهم کردن اطلاعات مورد نیاز به منظور اصلاح آموزش و یادگیری در کلاس درس از سوی معلم به‌وجود آمده است.

مراحل اجرای سنجش شناختی تشخیصی، عبارت است از: شناسایی صلاحیت‌های زیربنایی برای ارائه پاسخ صحیح به سؤال‌ها، طراحی سؤال‌های آزمون براساس مدل شناختی، تدوین ماتریس کیو^۶، استفاده از مدل

۱. Cognitive Diagnostic Tests

۲. stakeholders

۳. متغیر پنهان (Latent Variable) متغیری آماری است که ویژگی پنهان و مشاهده‌نشده‌ی پاسخ‌دهندگان (همان صلاحیت) را نشان می‌دهد.

۴. Assessment Used for Learning and as Learning Process

۵. Assessment of Learning Outcomes

۶. رابطه سؤال و صلاحیت در ماتریس کیو (دارای مرتبه $n \times k$) که نشان‌دهنده نیم‌رخ صلاحیت برای مجموعه سؤال‌هاست، نشان داده می‌شود. هر سطر این ماتریس، یک فرضیه درباره صلاحیت‌های مورد نیاز برای کسب پاسخ درست به یک سؤال است. تمام درایه‌های این ماتریس، اعداد صفر (سؤال صلاحیت را اندازه نمی‌گیرد) و یک (سؤال صلاحیت را اندازه می‌گیرد) هستند (راپ، تمپلین و هنسون، ۲۰۱۰).

های آماری روان‌سنجی برای تحلیل داده‌ها و ارائه گزارش نمره در هریک از صلاحیت‌ها و نیز طبقه‌بندی دانش‌آموزان. مدل‌های آماری روان‌سنجی که معمولاً آن‌ها را مدل‌های طبقه‌بندی تشخیصی می‌نامند، به طبقه‌بندی پاسخ‌های آزمون‌شوندگان براساس متغیرهای پنهان چندگانه می‌پردازد. در هریک از این مدل‌ها، عملکرد هر آزمون‌شونده، برحسب احتمال تسلط وی بر هریک از صلاحیت‌ها یا احتمال تعلق به یک طبقه پنهان تعیین می‌شود. باید توجه داشت، آنچه مدل‌های آماری سنجش شناختی تشخیصی را از دیگر مدل‌های متغیر پنهان متمایز می‌کند، ویژگی چندبعدی بودن، تأییدی بودن، پیچیده بودن و مقوله‌ای بودن متغیرهای پاسخ مشاهده شده است (راپ، تمپلین و هنسون، ۲۰۱۰). در دسته‌بندی‌های جدیدتر، مدل‌های سنجش شناختی تشخیصی با توجه به نوع ترکیب متغیرهای پیش‌بین پنهان به سه دسته جبرانی^۱، غیرجبرانی^۲ و کلی تقسیم می‌شوند (راوند و روبیتز، ۲۰۱۸).

در طراحی تحقیقات مبتنی بر سنجش شناختی دو رویکرد تشخیصی وجود دارد: ۱. استفاده از رویکردی *استقرایی*^۳ به منظور ساخت یک مجموعه سؤال تشخیصی برای استنباط فرایندهای صلاحیتی و ساختار دانش مورد نظر^۴، ۲. استفاده از رویکرد ریتروفیت^۵ با هدف استخراج فرایندهای شناختی و صلاحیت‌ها از آزمون‌های موجود به منظور کسب اطلاعات غنی‌تر در مقایسه با آنچه مقیاس‌پردازی تک‌بعدی به دست می‌دهد (جانگ، ۲۰۰۸). تفاوت عمده این دو رویکرد این است که در رویکرد استقرایی ابتدا صلاحیت‌هایی^۶ که ویژگی‌های مدل شناختی زیربنایی را برای ارائه پاسخ صحیح به سؤال‌ها مشخص می‌کند، تهیه می‌شود و سپس سؤال‌های آزمون براساس صلاحیت‌ها طراحی می‌شوند، اما در رویکرد ریتروفیت این فرایند به صورت معکوس انجام می‌شود. با توجه به مزیت‌های مدل‌های شناختی تشخیصی، مرور پیشینه این حوزه نشان می‌دهد که پژوهشگران بسیاری در خارج از کشور به استفاده از این مدل‌ها در حوزه آموزش ترغیب شده‌اند (عفت‌پناه، بقایی و بوری، ۲۰۱۹). در این باره، بررسی‌های دقیق‌تر راوند و بقایی (۲۰۲۰) نشان می‌دهد که در اوایل سال ۲۰۱۸، به دنبال جست‌وجوی گوگل با دو برچسب «مدل‌های شناختی تشخیصی»^۷ و «مدل‌های تشخیصی شناختی»^۸ بیش از ۲۴۰ مطالعه شناسایی شدند که بیش از ۹۵ درصد آن‌ها به مبحث روش‌شناسی، حدود ۴ درصد به رویکرد ریتروفیت و کمتر از ۱ درصد به رویکرد استقرایی مربوط بوده است. توجه بیش‌ازحد به مطالعه در حوزه روش‌شناسی به ایجاد دغدغه‌هایی مبنی بر فاصله گرفتن از اهداف اصلی سنجش شناختی تشخیصی

۱. compensatory

۲. nonc ompensatory

۳. inductive approach

۴. دو روش مهم در طراحی این نوع آزمون‌ها عبارت‌اند از طرح شاهد محور میسلوی (۱۹۹۴) و سیستم طراحی شناختی امبرسون (۱۹۹۴)

۵. Retrofit Approach

۶. در ادبیات سنجش شناختی تشخیصی، اسامی مختلفی مانند ویژگی‌های پنهان، صفات پنهان، مهارت‌ها و صلاحیت‌ها برای متغیرهای پنهان پیشنهاد شده است.

۷. Diagnostic Cognitive Models

۸. Cognitive Diagnostic Models

منجر شده است.

همچنین با وجود مزایای فراوان مدل‌های شناختی تشخیصی و گرایش به پژوهش در این حیطه، در عمل انتقاداتی نیز بر پژوهش‌های این حوزه وارد است؛ برای نمونه، برازش نامناسب مدل با داده‌ها به دلیل استفاده از رویکرد ریتروفیت برای آزمون‌های تک‌بعدی (بردشاو، ایزاک، تمپلین و جاکوبسون، ۲۰۱۳)، استفاده از مدل‌های شناختی تشخیصی در تحلیل سنجش‌های کلان^۱ و ارائه‌نکردن بازخوردهای عملی در تصمیم‌گیری‌های آموزشی (چن و دی‌لتوره، ۲۰۱۴)، کاربرد نداشتن شواهد سنتی پایایی برای مدل‌های شناختی تشخیصی (جانگ، ۲۰۰۸) و تأثیر تدوین نامناسب ماتریس کیو بر کیفیت طبقه‌بندی آزمودنی‌ها (راوند، ۲۰۱۶) از جمله انتقادهای وارد شده بر مطالعات سنجش شناختی تشخیصی است که در ادامه به برخی از این انتقادهای پرداخته شده است.

یکی از این نقدها ارائه‌نکردن شواهد پایایی متناسب با این مدل‌هاست؛ برای نمونه استفاده از ضرایب پایایی در رویکردهای سنتی که مبتنی بر سازه پیوسته تک‌بعدی بوده است، برای آزمون‌های شناختی تشخیصی مناسب نیستند؛ زیرا این آزمون‌ها بر پایه سازه چندبعدی طبقه‌ای طراحی می‌شوند (روسوس و همکاران، ۲۰۰۷). با توجه به اینکه مدل‌های شناختی تشخیصی، زیرمجموعه‌ای از مدل‌های طبقه‌بندی هستند، گزارش شواهد پایایی براساس شاخص‌های جدیدی مانند شاخص ثبات طبقه‌بندی^۲ افراد در این باره مهم است. براساس تعریف تمپلین و بردشاو (۲۰۱۳)، این شاخص به معنای آن است که با فرض سنجش مکرر آزمونی روی یک آزمودنی، طبقه‌بندی آزمودنی یکسان خواهد بود. همچنین از دیدگاه سوی، گیرل و جانگ (۲۰۱۲) براساس پایایی ثبات طبقه‌بندی آزمودنی‌ها مشخص می‌شود که آزمودنی تا چه حد به طبقه پنهان یکسانی تعلق دارد یا به عبارتی نشان‌دهنده تسلط داشتن یا نداشتن بر صلاحیت‌های یکسان در اجرای دوباره همان آزمون یا فرم‌های موازی آن آزمون است. از دیدگاه تمپلین و بردشاو (۲۰۱۳) این نقدها شاید به دلیل متفاوت بودن نوع شواهد پایایی در مدل‌های سنجش شناختی تشخیصی در مقایسه با مدل‌های اندازه‌گیری متداول است.

همچنین از نظر گورین (۲۰۰۹)، تدوین یا استخراج ماتریس کیو چالش‌برانگیزترین مرحله در مدل‌های شناختی تشخیصی است؛ بنابراین در طراحی این ماتریس و ارائه شواهد، روابودن آن باید با دقت لازم صورت بگیرد. در محیط‌های آموزشی، تدوین ماتریس کیو براساس استفاده از نظریه‌های یادگیری درباره محتوای مدنظر و بینش متخصصان، دبیران و محققان صورت می‌گیرد (مدیسون و بردشاو، ۲۰۱۴). همچنین براساس نظر یانگ و امبرتسون (۲۰۰۷) تدوین مناسب ماتریس کیو، بخشی مهم از شواهد روایی در مدل‌های شناختی تشخیصی است. با توجه به مطالعات پیشین، ساختار نامناسب ماتریس کیو می‌تواند بر کیفیت پارامترهای برآورد شده و در نتیجه دقت طبقه‌بندی آزمودنی‌ها اثرگذار باشد (لی و لی، ۲۰۱۶)؛ برای نمونه راپ و تمپلین (۲۰۰۸) دریافتند بدتوصیف شدن ماتریس^۳ کیو یعنی کم‌برآورد یا بیش‌برآورد^۴ ماتریس کیو (به ترتیب تبدیل

-
1. Large Scale Assessment
 2. Classification Consistency Index
 3. Q-matrix misspecification
 4. Underfitting and overfitting the Q-matrix

نادرست درایه‌های یک به صفر و تبدیل نادرست درایه‌های صفر به یک در ماتریس کیو) می‌تواند به بیش‌برآورد پارامترهای مدل دینا^۱ بینجامد. همچنین دی‌کارلو (۲۰۱۱) در تحلیل داده‌های تفریق کسر تاتسوکا (۱۹۹۰) با استفاده از مدل دینا، طبقه‌بندی غیردقیق افراد را احتمالاً به دلیل ساختار نامناسب ماتریس کیو می‌داند. چیو (۲۰۱۳) معتقد است، بیشتر تحقیقات دربارهٔ سنجش شناختی تشخیصی، با تدوین ماتریس کیو اولیه، دربارهٔ تناسب این ماتریس جست‌وجویی به عمل نمی‌آورند. بررسی پیشینه نشان می‌دهد که به‌منظور کاهش تأثیر تدوین نامناسب ماتریس کیو بر کیفیت پارامترهای مدل و طبقه‌بندی، روش‌های مختلفی در تأیید ساختار ماتریس کیو که با نظر متخصصان تدوین شده است، به کار برده می‌شود. درواقع این روش‌ها برای تعدیل کردن محتاطانهٔ ماتریس کیو طراحی شده‌اند که معروف‌ترین آن‌ها را روسوس و همکاران (۲۰۰۷)، دی‌لتوره (۲۰۰۸)، دی‌لتوره و چیو (۲۰۱۶) و لی (۲۰۱۶) معرفی کرده‌اند.

از دیگر انتقادات وارد بر سنجش شناختی تشخیصی، کافی نبودن شواهد مربوط به روایی طبقه‌بندی است. از نظر سینه‌ارای و هابرم (۲۰۰۹) کاربرد مدل‌های شناختی تشخیصی چنین شواهد روایی را فراهم نمی‌کنند؛ زیرا با استفاده از نتایج این مدل‌ها می‌توان تصمیم‌گیری‌های مهمی برای طبقه‌بندی دانش‌آموزان در نظر گرفت. همچنین باید شواهد مبتنی بر دقت این طبقه‌بندی‌ها گزارش شود؛ استفاده از شاخص‌هایی مانند دقت طبقه‌بندی^۲ آزمودنی‌ها به‌منزلهٔ شواهد روایی، اهمیت فراوانی دارد. از نظر سوی، گیرل و جانگ (۲۰۱۲) دقت طبقه‌بندی بدان معناست که تا چه اندازه طبقه‌بندی آزمودنی با طبقهٔ پنهان واقعی او منطبق است. به عبارت دیگر چقدر آزمودنی به‌درستی در یک صلاحیت به‌صورت مسلط یا غیرمسلط شناسایی شده است.

بخش دیگری از انتقادات وارد بر پژوهش‌ها در حوزهٔ شناختی تشخیصی، استفاده از مدل‌های طبقه‌بندی تشخیصی در قالب رویکرد ریتروفیت بر آزمون‌هایی است که لزوماً با هدف چندبعدی طراحی نشده‌اند (راوند، براتی و ویدیهارسو، ۲۰۱۳). براساس نظر جانگ (۲۰۰۹)، کاربرد رویکرد ریتروفیت برای آزمون‌های موجود که با اهداف تشخیصی طراحی نشده است، می‌تواند به مخدوش شدن روایی استنباط‌ها دربارهٔ صلاحیت‌های شناختی آزمودنی‌ها منجر شود. همچنین از نظر وی استفاده از چنین آزمون‌هایی، به از دست دادن ظرفیت تشخیصی این مدل‌ها می‌انجامد؛ چراکه سؤال‌ها در چنین آزمون‌هایی شامل دو سر طیف دشواری (خیلی آسان و خیلی سخت) هستند. از سوی دیگر، این نوع سؤال‌ها در ارزیابی صلاحیت‌های آزمودنی‌ها تأثیر فراوانی ندارند و برآزش نبودن مدل با داده‌های مشاهده‌شده در آزمودنی‌های با طیف گسترده نمرات، به دلیل وجود چنین سؤال‌هایی است. با این حال، با وجود محدودیت‌های رویکرد ریتروفیت، نتایج بررسی‌های عفت‌پناه، بقایی و بوری (۲۰۱۹) حاکی از تمایل پژوهشگران به استفاده از این رویکرد در مقایسه با رویکرد استقرایی است.

همچنین یکی دیگر از انتقادات وارد بر حوزهٔ سنجش شناختی تشخیصی این است که شواهدی مبنی بر برقراری برخی مفروضات در گزارش پژوهش‌ها وجود ندارد. از مفروضات مهم به‌کارگیری این مدل‌ها،

۱. دینا (DINA) یک مدل آماری غیرجبرانی شناختی تشخیصی است که در آن واژهٔ Deterministic، به جنبهٔ تعینی مدل، واژهٔ And-Gate، به فرایند ربطی مدل و واژهٔ Noisy به عنصر تطوری (احتمالی) مدل اشاره دارد.

۲. Classification Accuracy Index

چندبعدي بودن سازه استفاده شده است (سسومز و هنسون، ۲۰۱۸). يك روش بررسي چندبعدي بودن، وارسى ساختار ماتريس همبستگى صلاحيت هاست؛ به اين معنا كه همبستگى هاى كمتر از ۰/۹ ميان صلاحيت ها مى تواند حاكى از وجود ساختار چندبعدي در بين صلاحيت ها باشد (كانينا-هابنيخت، راپ و ويلهلم، ۲۰۰۹)؛ برائى نمونه در پژوهشى به منظور بررسي مفروضه چندبعدي بودن آزمون تشخيصى درك معلمان از ضرب و تقسيم اعداد گويا، از ماتريس همبستگى تتراكوريك^۱ استفاده شد. نتايج نشان مى دهد ضرايب درايه هاى اين ماتريس بين ۰/۷۸۱-۰/۶۲۶ بود كه اين همبستگى ها به دليل نزديك نبودن به مقدار يك، نشان از چندبعدي بودن آزمون داشتند (بردشاو و همكاران، ۲۰۱۳). به طور كلئى، برائى تحليل بعد^۲، چندين روش مبتنى بر كواريانس شرطى ناپارامترىك مانند رويكرد ديتكت^۳ (ژانگ و استوات^۴، ۱۹۹۹) وجود دارد كه پژوهشگران اين حوزه مى توانند از آن ها استفاده كنند (به نقل از جانگ، ۲۰۰۵). از آنجا كه بيشتر پژوهش هاى انجام شده در حوزه سنجش شناختى تشخيصى با استفاده از رويكرد ريتروفيت بر روى آزمون هاىي است كه اساساً ممكن است چندبعدي نباشند، بررسي مفروضه چندبعدي بودن صلاحيت هاى زيربنائى يك آزمون، موضوعى اساسى به شمار مى رود.

يكى ديگر از انتقادات وارد بر سنجش شناختى تشخيصى، محدود بودن كاربرد عملى پژوهش هاى مبتنى بر اين مدل ها در تصميم گيرى هاى آموزشى است. براساس چارچوب نظرى، اين مدل ها قابليت ارائه بازخورد جبرائى را به منظور برطرف كردن نقاط ضعف آزمودنى ها دارند؛ درحالى كه در عمل نتايج اين پژوهش ها، چنين انتظارى را برآورده نمى كنند؛ برائى نمونه سسومز و هنسون (۲۰۱۸) در ارزيايى پيشينه كاربرد اين مدل ها دريافتند كه تنها در ۸ درصد مطالعات از نتايج مدل هاى شناختى تشخيصى برائى ارائه بازخورد به دانش آموزان يا معلمان استفاده شده بود و هيچ مطالعه اى از نتايج مطالعات شناختى تشخيصى برائى تغيير در آموزش كلاسى استفاده نكرده بود. يكى از دلايل محتمل برائى محدوديت استفاده از نتايج اين مدل ها در عمل، نياز به داشتن نمونه هاى وسيع است؛ زيرا به زعم هنسون (۲۰۰۹)، كاربرد بيشتر مدل هاى آمارى شناختى تشخيصى مبتنى بر حجم نمونه بزرگ هستند. از نظر چيو و داگلاس (۲۰۱۳) استفاده از مدل هاى آمارى ناپارامترىك به دليل اتكا بر حجم نمونه كوچك، امكان به كارگيرى نتايج اين مدل ها در محيط هاى آموزشى را تسهيل مى كند. چالش اساسى ديگر در مدل هاى شناختى تشخيصى، برازش اين مدل ها با نرم افزارهاى مناسب است. دغدغه هاى اصلى در برآورد مدل هاى آمارى به دست آوردن هم گرايى مدل در نرم افزار و به عبارت ديگر توانايى نرم افزار در مواجهه با ساختار داده هاى پيچيده است. با پيچيده شدن ساختار آزمون و مدل آمارى سنجش شناختى تشخيصى، هم گرايى مدل به مسئله اى جدئى تبديل مى شود. برخى برآوردهاى مدل هاى آمارى مانند دينا يا دينو به نسبت ساده است؛ به گونه اى كه هم گرايى چنين مدل هاى بيشتر با يك نمونه چند صدتايى، با طول

۱. Tetrachoric Correlation

۲. Dimensionality Analysis

۳. DETECT procedure

۴. Zhang, J. & Stout, W.

آزمون متعادل (۲۰-۴۰ سؤال) و تعداد صلاحیت‌های متعادل (۶-۴ صلاحیت) به‌دست می‌آید، اما در برخی مدل‌ها مانند مدل غیرجبرانی یکپارچه با پارامترپردازی مجدد یا مدل شبکه‌ استنباط بیزی پیچیده، برآورد هم‌گرایی به انتخاب حجم نمونه بزرگ و تعداد سؤال‌های بیشتر نیاز دارد (راپ و تمپلین، ۲۰۰۸ الف). مرور پیشینه درباره‌ مدل‌های طبقه‌بندی چندگانه (برای مثال بولت و فو^۱، ۲۰۰۴؛ هارتز^۲، ۲۰۰۲؛ هنسون و تمپلین^۳، ۲۰۰۶؛ ماریس^۴، ۱۹۹۹) نشان می‌دهد که بیشتر نمونه‌های کاربردی از حدود ۸-۴ صلاحیت استفاده کرده‌اند (به نقل از راپ و تمپلین، ۲۰۰۸ ب).

در مجموع، با توجه به مزایای سنجش شناختی تشخیصی مانند ارائه نمره‌های تشخیصی خرد و در نتیجه مهیا شدن اطلاعاتی غنی‌تر درباره صلاحیت‌های اندازه‌گیری شده توسط یک آزمون، کاربرد این نوع سنجش در میان پژوهشگران داخلی رو به گسترش است. از سوی دیگر، با توجه به انتقادات مهم و اساسی وارد بر پژوهش‌های این حوزه، هدف مطالعه حاضر مرور نقادانه پژوهش‌های داخلی در این باره به منظور فراهم کردن تصویری از کیفیت مباحث مربوط به بخش روش و میزان عملی بودن نتایج آن‌هاست؛ چراکه شناسایی ضعف‌ها و قوت‌های مطالعات می‌تواند راهنمای عملی برای پژوهش‌های آتی در این حوزه باشد؛ بنابراین مطالعه حاضر به این پرسش پاسخ داده است که قوت‌ها و ضعف‌های مباحث مربوط به بخش روش مطالعات ایرانی در حوزه سنجش شناختی تشخیصی چیست.

روش

مطالعه حاضر از نوع مرور انتقادی است. براساس نظر گرنت و بوث (۲۰۰۹) بررسی ادبیات پژوهشی به‌طور گسترده و ارزیابی کیفیت آن به‌منزله مطالعه مروری نقادانه است. این نوع مطالعه فراتر از توصیف صرف مقالات شناسایی شده گام برداشته و شامل میزانی از تحلیل و نوآوری مفهومی است. یک مرور نقادانه اثربخش مشتمل بر ارائه، تحلیل و ترکیب محتویات منابع متنوع است. به‌منظور دستیابی به گزارش‌های پژوهشی در حوزه سنجش شناختی تشخیصی، با استفاده از کلمات کلیدی «سنجش شناختی تشخیصی»، «مدل شناختی تشخیصی»، «مدل طبقه‌بندی تشخیصی» و «تحلیل تشخیصی» جست‌وجویی گسترده و جامع در بانک‌های داده ایرانداک^۵ (پایگاه اطلاعات علمی ایران)، پایگاه داده علمی جهاد دانشگاهی^۶، گوگل اسکالر^۷، اسکوپوس^۸،

۱. Bolt, D., & Fu, J.

۲. Hartz, S. M.

۳. Henson, R. A. & Templin, J.

۴. Maris, E.

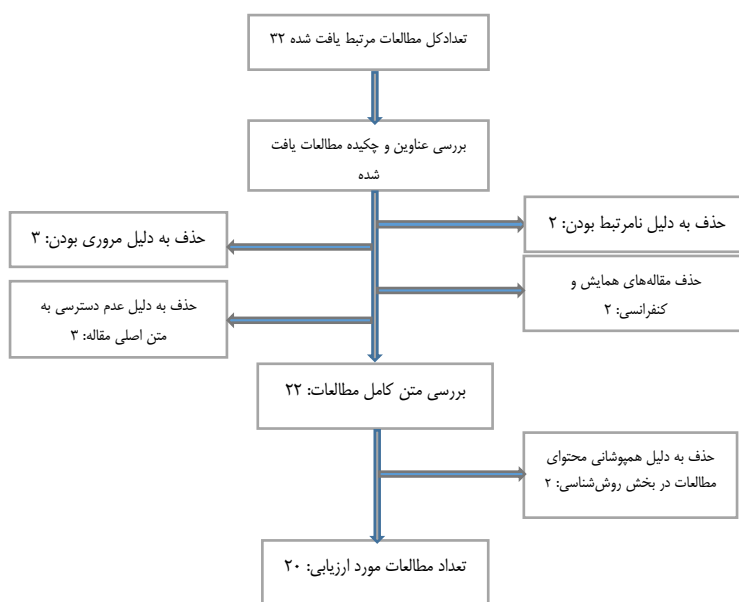
۵. Irandoc

۶. SID

۷. Google Scholar

۸. Scopus

نور^۱ (بانک اطلاعات کتاب‌های تخصصی علوم اسلامی و انسانی)، مگ‌ایران^۲ (بانک داده‌نشریات کشور) و موتور جستجوی علم‌نت^۳ انجام شد. نتیجه جست‌وجو در پایگاه‌های مذکور، استخراج ۲۱ پژوهش با استفاده از کلمات کلیدی بالا بود. سپس همین کلیدواژه‌ها به زبان انگلیسی همراه با عملگر بولی OR و ترکیب آن‌ها با عبارت «جامعه‌ایرانی» با استفاده از عملگر بولی AND جست‌وجو و ۱۲ پژوهش یافت شد. به‌منظور افزایش اطمینان از فهرست شناسایی‌شده مطالعات، فهرست منابع پژوهش‌های یافت‌شده نیز بررسی شدند؛ در نتیجه یک پژوهش یافت شد و ۳۲ مطالعه شناسایی شدند. اولین معیار انتخاب مطالعات برای ارزیابی، وجود یکی از کلیدواژه‌های فوق در عنوان یا واژگان کلیدی پژوهش بود. ملاک دیگر ورود مطالعات، علمی-پژوهشی بودن مقاله یا پایان‌نامه در دوره کارشناسی ارشد و دکتری منتشرشده به زبان فارسی و دسترسی داشتن به متن کامل پژوهش‌های یافت‌شده بود. همچنین مقاله ارائه‌شده در همایش‌ها، مقاله مروری، هم‌پوشانی محتوای مقالات در بخش مباحث مربوط به بخش روش (یکسانی داده‌ها، ماتریس کیو و مدل آماری سنجش شناختی تشخیصی) و نامرتب بودن (کنش افتراقی سؤال، تغییرناپذیری و مقایسه مدل‌ها) از ملاک‌های خروج مطالعه بود. در مجموع، ۲۰ مقاله علمی-پژوهشی در بازه زمانی ۱۳۹۳ تا بهار ۱۳۹۹ شناسایی و انتخاب شدند. در شکل ۱، نمودار راهبرد جست‌وجو آمده است.



شکل ۱. راهبرد جست‌وجو برای انتخاب پژوهش‌های قابل‌ارزیابی

۱. Noor
۲. Magiran
۳. Elmnet

پس از انتخاب پژوهش‌ها براساس ملاک‌های ورود و خروج برای ارزیابی، از برگه گزارش توصیفی مطالعات شامل تعدادی نشانگر استفاده شد. گفتنی است نشانگرهای شناسایی‌شده از دستورالعمل توصیه‌شده توسط سسومز و هنسون (۲۰۱۸) برای بررسی مدل‌های شناختی تشخیصی انتخاب شده است. در این مطالعه، ابتدا هر پژوهش به‌طور جداگانه از سوی یکی از دو پژوهشگر مطالعه شد و اطلاعات مربوط به هر کدام، در برگه گزارش توصیفی مطالعات وارد شد. برگه گزارش شامل نشانگرهای تمرکز موضوعی پژوهش‌ها بر مبنای تعداد صلاحیت‌ها، سازه مورد اندازه‌گیری، ساختار سؤال‌های آزمون‌ها، ماتریس کیو، نوع مدل آماری و نرم‌افزار، برازش مدل، برازش سؤال، نوع سؤال، حجم نمونه، شواهد پایایی و روایی، توصیف صلاحیت‌ها (طبقه‌های پنهان، همبستگی میان صلاحیت‌ها و نیم‌رخ صلاحیت) و کاربست یافته‌ها بود. این برگه با تکمیل برگه‌های گزارش و مقایسه یافته‌ها صورت گرفت.^۱ همچنین برای پژوهش‌هایی که برگه‌ها با هم مغایرت داشتند، ارزیابی مجدد انجام شد و در صورت لزوم، بین دو پژوهشگر گفت‌وگو و تبادل نظر ادامه یافت تا نتایج همسو شدند. در گام بعدی، داده‌های توصیفی این مطالعه‌ها استخراج شدند و ارزیابی نقادانه آن‌ها نیز صورت گرفت. جدول ۱ توزیع این مقالات در نشریات را نشان می‌دهد.

جدول ۱. توزیع مقالات علمی پژوهشی منتخب

ردیف	ناشر	پژوهشگر (سال)
۱	اندازه‌گیری تربیتی	مینایی، دلاور، فلسفی‌نژاد، کیامنش و مهاجر (۱۳۹۳)، مقدم، فلسفی‌نژاد، فرخی و استاجی (۱۳۹۵)، محمدی، دلاور، فرخی و مینایی (۱۳۹۶)، رحیمی، یونسی و مکرمی (۱۳۹۷)
۲	دستاورد‌های روان‌شناختی نوآوری‌های آموزشی	افضلی، دلاور، فلسفی‌نژاد، فرخی و برجعلی (۱۳۹۳) محسن‌پور، گویا، شکوهی‌یکتا، کیامنش و بازرگان (۱۳۹۴)، مفاخری، شاهورانی سمنانی، بهزادی و برهمند (۱۳۹۸)
۴	پژوهش‌های زبان‌شناختی در زبان دوم	رنجبران و علوی (۱۳۹۵)
۵	نقد زبان و ادبیات خارجی	جاویدان‌مهر و عنانی سراب (۱۳۹۶)
۶	آموزش زبان فارسی به غیر فارسی‌زبانان	تقیان، خدایی، بازرگان، مقدم زاده و کبیری (۱۳۹۷)
۷	مطالعات اندازه‌گیری و ارزشیابی	محسن‌پور (۱۳۹۸)
۸	پژوهش‌های زبان‌شناختی در زبان‌های خارجی	شهمیرزادی، سیری، مرعشی و گرامی‌پور (۱۳۹۹)
۹	Iranian Journal of Language Testing	راوند، براتی و ویدیهارسو (۲۰۱۳)
۱۰	Psychoeducational Assessment	راوند (۲۰۱۶)
۱۱	Practical Assessment, Research, and Evaluation	راوند و روبیتز (۲۰۱۶)
۱۲	Applied Measurement in Education	کبیری، قاضی طباطبایی، بازرگان، شکوهی‌یکتا و خرازی (۲۰۱۶)
۱۳	International Journal of Language Testing	جاویدان‌مهر و عنانی سراب (۲۰۱۷): عفت‌پناه (۲۰۱۹)
۱۴	Language Assessment Quarterly	جاویدان‌مهر عنانی سراب (۲۰۱۹)
۱۵	Language testing in Asia	عفت‌پناه، بقایی و بوری (۲۰۱۹)

۱. ابتدا به کمک نرم‌افزار SPSS یک فایل برای تک‌تک نشانگرهای مندرج در برگه توصیفی به‌عنوان یک متغیر براساس اجماع نظر پژوهشگران تعریف شد. گفتنی است برای هر نشانگر کدهایی در نرم‌افزار تعریف شد؛ برای مثال یک متغیر، شناسایی رویکرد تشخیصی مطالعه بود که برای رویکرد ریتروفیت کد ۱ و رویکرد استقرایی کد ۲ تعیین شد. سپس هر پژوهشگر به‌طور مستقل برای هر مطالعه وضعیت متغیرهای تعریف‌شده را بررسی و در فایل مربوط وارد کرد. سپس هم‌خوانی نتایج بررسی شد.

براساس جدول ۱، از این میان ۲۰ مقاله که در فاصله سال‌های ۱۳۹۳ تا بهار ۱۳۹۹ و در نشریات علمی-پژوهشی داخلی یا خارجی چاپ شده بودند، ۶۰ درصد از مقالات در نشریه‌های ایرانی و ۴۰ درصد در نشریه‌های خارجی که اعتبار پژوهشی داشتند، به چاپ رسیده‌اند. همچنین در میان نشریات علمی-پژوهشی داخلی، اندازه‌گیری تربیتی چاپ ۳۳ درصد از مقالات فارسی را به خود اختصاص داده بود. در میان مطالعات انجام‌شده، جامعه پژوهشی در هشت مطالعه، شامل دانش‌آموزان (۴۰ درصد) و ۹ مطالعه شامل داوطلبان آزمون ورودی به دانشگاه‌ها (۴۵ درصد) بود. همچنین یک مطالعه برای متقاضیان آیلتس (۵ درصد) و دو مطالعه برای دانشجویان (۱۰ درصد) بود.

یافته‌های مرور پژوهش‌های منتخب در حوزه سنجش شناختی تشخیصی

جدول ۲ نتایج ارزیابی مقالات منتخب حوزه سنجش شناختی تشخیصی را نشان می‌دهد.

جدول ۲. نتایج ارزیابی مقالات منتخب

ردیف	ویژگی	نتیجه
۱	سازه اندازه‌گیری شده	درک مطلب زبان انگلیسی (۵۰ درصد) و ریاضی (۲۵ درصد) دیگر سازه‌ها هرکدام تنها در یک مطالعه استفاده شده بودند.
۲	رویکرد تشخیصی	رویکرد ریتروفت (۷۰ درصد) و رویکرد استقرایی (۳۰ درصد)
۳	نام آزمون	درک مطلب زبان انگلیسی آزمون ورودی دانشگاه‌ها، آزمون پیشرفت ریاضی، آزمون وکسلر کودک ۴، آزمون آیلتس، آزمون زبان ورودی دانشگاه اصفهان، آزمون پیشرفت تحصیلی ریاضی (تیمز، ۲۰۰۷)، آزمون پیشرفت تحصیلی علوم (تیمز، ۲۰۱۱)، آزمون سواد ریاضی، آزمون خواندن فارسی، آزمون درک مطلب زبان انگلیسی و آزمون نوشتن انگلیسی
۴	نوع سؤال	آزمون چندگزینه‌ای (۶۰ درصد)، آزمون ترکیبی چندگزینه‌ای و تشریحی (۲۰ درصد)، نامعلوم (۱۰ درصد)، آزمون ترکیبی چندگزینه‌ای و کوتاه پاسخ (۵ درصد) و آزمون تشریحی (۵ درصد)
۵	تعداد سؤالات	۱۰ سؤال تا ۲۱۹ سؤال آزمون ۲۰ سؤالی با بیشترین فراوانی (۴۰ درصد)
۶	حجم نمونه	۱۰۰۰۰ - ۲۸۶ آزمودنی ۶۵ درصد مطالعات دارای حجم نمونه بالای ۱۰۰۰ آزمودنی
۷	تعداد صلاحیت‌ها	۲۰ - ۳ صلاحیت مطالعات ۵ صلاحیتی متداول‌ترین بود (۳۰ درصد)
۸	ماتریس کیو	۷۵ درصد مطالعات شامل بخشی از ماتریس کیو یا همه آن بودند.
۹	نوع مدل استفاده شده	متداول‌ترین مدل‌ها به ترتیب: GDINA (۳۵ درصد)، DINA (۲۰ درصد) و NC-RUM (۱۵ درصد)
۱۰	نرم‌افزار استفاده شده	نرم‌افزار R (۸۰ درصد)، Arpeggio (۱۵ درصد) و MDLTM (۵ درصد)
۱۱	ویژگی مدل	متداول‌ترین مدل‌های غیرجبرانی (۳۵ درصد)، کلی (۵۰ درصد) و جبرانی (۱۵ درصد)
۱۲	بررسی مفروضه‌ها	گزارش تحلیل بعد در ۱۰ درصد مطالعات گزارش داده‌های گمشده در ۵ درصد مطالعات
۱۳	همبستگی میان صلاحیت‌ها	گزارش‌ندادن در ۶۵ درصد مطالعات ضریب همبستگی ۰/۹ یا بیش از ۰/۹ در ۵۷ درصد مطالعات حاوی همبستگی‌ها
۱۴	برازش مطلق برازش نسبی	گزارش نکردن در ۵۰ درصد مطالعات گزارش نکردن در ۵۵ درصد مطالعات گزارش حداقل یک شاخص برازش سؤال در ۹۰ درصد مطالعات

ردیف	ویژگی	نتیجه
	برازش سؤال	
۱۵	توصیف صلاحیت‌ها	گزارش‌نکردن نسبت تسلط آزمودنی‌ها به هر صلاحیت در ۱۰ درصد مطالعات گزارش‌نکردن نسبت تعلق آزمودنی‌ها به هر طبقه پنهان در ۵۰ درصد مطالعات گزارش‌نکردن احتمال تسلط هر آزمودنی به هر صلاحیت در ۶۵ درصد مطالعات گزارش‌نکردن احتمال تعلق هر آزمودنی به هر طبقه پنهان در ۷۰ درصد مطالعات
۱۶	شواهد پایایی	گزارش‌نکردن هیچ نوع شاخص پایایی در ۵۵ درصد مطالعات گزارش پایایی ابزار (آلفای کرونباخ) در ۰/۱۵ مطالعات گزارش پایایی ثبات طبقه‌بندی توسط ۲۰ درصد مطالعات گزارش هم‌زمان پایایی ابزار (آلفای کرونباخ) و شاخص پایایی ثبات طبقه‌بندی در ۱۰ درصد مطالعات
۱۷	شواهد روایی	گزارش روایی کمی ماتریس کیو در ۶۰ درصد مطالعات گزارش روایی کیفی ماتریس کیو در ۹۰ درصد مطالعات گزارش روایی طبقه‌بندی در ۲۰ درصد مطالعات گزارش‌نکردن هیچ نوع شاخص روایی در ۱۰ درصد مطالعات
۱۸	کاربست نتایج	گزارش‌نکردن کاربرد یافته‌ها در تمام مطالعات

براساس جدول ۲، نتایج هریک از ویژگی‌های بررسی‌شده به شرح زیر است:

سازه‌اندازه‌گیری‌شده: یکی از شاخص‌های ارزیابی‌شده در این مطالعه، نوع سازه مورد اندازه‌گیری بود. متداول‌ترین این سازه‌ها در مطالعات بررسی‌شده مربوط به سازه درک مطلب زبان انگلیسی (۵۰ درصد) و ریاضی (۲۵ درصد) بودند. دیگر سازه‌های مورد اندازه‌گیری شامل علوم (کبیری و همکاران، ۲۰۱۶)، هوش (محمدی و همکاران، ۱۳۹۶)، آزمون خواندن فارسی (تقیان و همکاران، ۱۳۹۷)، بخش شنیداری آزمون آیلتس (عفت‌پناه، ۲۰۱۹) و آزمون نوشتن انگلیسی (عفت‌پناه، بقایی و بوری، ۲۰۱۹) بودند.

رویکرد تشخیصی: در ۷۰ درصد کل مطالعات بررسی‌شده از رویکرد ریتروفیت استفاده شده که در این میان آزمون درک مطلب زبان انگلیسی ورودی دانشگاه‌ها در ۴۵ درصد مطالعات استفاده شده بود. همچنین آزمون ریاضی تیمز^۱ ۲۰۰۷ (مینایی و همکاران، ۱۳۹۳)، آزمون علوم تیمز ۲۰۱۱ (کبیری و همکاران، ۲۰۱۶)، آزمون وکسلر کودک^۲ (محمدی و همکاران، ۱۳۹۶)، بخش شنیداری آزمون آیلتس (عفت‌پناه، ۲۰۱۹) و آزمون زبان ورودی دوره دکتری دانشگاه اصفهان (راوند، براتی و ویدهیارسو، ۲۰۱۳) هرکدام در یک پژوهش استفاده شدند. در رویکرد استقرایی، پژوهشگران به طراحی دو آزمون پیشرفت تحصیلی ریاضی در پایه نهم (افضلی و همکاران، ۱۳۹۳) و پایه دهم (مفاخری و همکاران، ۱۳۹۸)، سواد ریاضی در پایه نهم (محسن‌پور و همکاران، ۱۳۹۴) و محسن‌پور، ۱۳۹۸)، آزمون خواندن پایه ششم (تقیان و همکاران، ۱۳۹۷)، آزمون خواندن و درک مفاهیم زبان انگلیسی در دانشجویان (رنجبران و علوی، ۱۳۹۵) و آزمون نوشتن انگلیسی (عفت‌پناه، بقایی و بوری، ۲۰۱۹) پرداخته بودند.

۱. Trends in International Mathematics and Science Study (TIMSS)

۲. Wechsler Intelligence Scale for Children

تعداد و نوع سؤال‌ها: شاخص مهم در تدوین ماتریس کیو در مدل‌های شناختی تشخیصی، تعداد سؤال‌هاست. این تعداد در میان مطالعات از ۱۰ سؤال (شهمیرزادی و همکاران، ۱۳۹۹) تا ۲۱۹ سؤال (محمدی و همکاران، ۱۳۹۶) در تغییر بود. متوسط فراوانی سؤال‌ها ۵۴ (میانگین، میانه و انحراف ملاک به ترتیب برابر ۵۴/۱۵، ۲۳/۵ و ۷۰/۵۴) بود. ۴۰ درصد مطالعات آزمون ۲۰ سؤالی و ۸۰ درصد مطالعات ۴۰ یا کمتر از ۴۰ سؤال داشتند. در میان مطالعات ارزیابی‌شده، ساختار سؤال‌ها در ۶۰ درصد از مطالعات، چندگزینه‌ای و ۲۰ درصد سؤال‌های ترکیبی (چندگزینه‌ای و تشریحی) بود. ۱۰ درصد (دو مطالعه) به ساختار سؤال‌ها اشاره نداشتند. در دو مطالعه سؤال‌های ترکیبی از نوع چندگزینه‌ای، کوتاه‌پاسخ و تشریحی بودند.

حجم نمونه: تعداد آزمودنی‌ها در مطالعات ارزیابی‌شده، از ۱۰۰۰۰ - ۲۸۶ نفر در تغییر بودند. متوسط حجم نمونه در پژوهش‌های دارای میانگین، میانه و انحراف ملاک به ترتیب ۲۵۷۰/۸۵، ۲۲۴۳ و ۲۴۱۰/۰۹ بود. در ۶۵ درصد مطالعات، حجم نمونه بیشتر از ۱۰۰۰ آزمودنی بود. حجم نمونه ۳۵ درصد مطالعات دیگر نیز ۸۱۶ - ۲۸۶ نفر بودند.

تعداد صلاحیت‌ها: یکی از شاخص‌های مهم در مدل‌های شناختی تشخیصی، تعداد صلاحیت‌های زیربنایی سؤال‌هاست. در ارزیابی صورت‌گرفته، تعداد صلاحیت‌ها از سه صلاحیت (مفاخری و همکاران، ۱۳۹۸) تا ۲۰ صلاحیت (مینایی و همکاران، ۱۳۹۳) در تغییر بود. مطالعات پنج‌صلاحیتی متداول‌ترین نوع بودند که ۳۰ درصد مطالعات را شامل می‌شدند. متوسط تعداد صلاحیت‌ها ۶ (میانگین، میانه و انحراف ملاک به ترتیب ۶/۳۵، ۵ و ۳/۶۷) بود. همچنین ۳۵ درصد مطالعات بیش از شش صلاحیت داشتند. در میان مطالعات ارزیابی‌شده، ۷۵ درصد به‌طور کامل یا بخشی از ماتریس کیو را گزارش کرده بودند.

مدل برآوردشده: به‌طور کلی در میان ویژگی‌های مهم مدل‌های برآوردشده، سه رویکرد جبرانی، غیرجبرانی و کلی وجود دارد. در مطالعات ارزیابی‌شده، ۳۵ درصد مطالعات با رویکرد غیرجبرانی، ۵۰ درصد با رویکرد کلی و تنها ۱۵ درصد با رویکرد جبرانی همراه بودند. همچنین در میان مطالعات با ویژگی غیرجبرانی، مدل DINA (۲۰ درصد) و NC-RUM (۱۵ درصد) استفاده شده بود. در مطالعات دارای رویکرد کلی، مدل‌ها شامل GDINA (۳۵ درصد)، PGDINA (۵ درصد)، GDM (۵ درصد) و LLM (۵ درصد) بود. مطالعات دارای رویکرد جبرانی شامل مدل‌های ACDM (۱۰ درصد) و C-RUM (۵ درصد) بود. در برآورد مدل‌های آماری، ۸۰ درصد مطالعات از نرم‌افزار R، ۱۵ درصد از نرم‌افزار Arpeggio و ۵ درصد از نرم‌افزار MDLTM استفاده کرده بودند.

بررسی مفروضه‌ها: به‌منظور استفاده از مدل‌های آماری شناختی تشخیصی، در میان مطالعات بررسی‌شده تنها یک مطالعه به ارزیابی مفروضه چندبعدی پرداخته بود. همچنین درباره‌ی غربالگری داده‌ها تنها ۱۰ درصد مطالعات، یعنی دو مطالعه به بحث داده‌های گمشده پرداخته بودند. افزون بر این، ۶۵ درصد مطالعات گزارشی درباره‌ی همبستگی صلاحیت‌ها ارائه نکرده بودند. از ۳۵ درصد مطالعاتی که همبستگی‌ها را گزارش کرده بودند، ۵۷ درصد ضریب همبستگی ۰/۹ یا بیش از ۰/۹ داشتند.

برازش مدل: در مطالعات ارزیابی‌شده، ۵۰ درصد مطالعات شاخص‌های برازش مطلق را گزارش نکرده بودند، ۹۰ درصد حداقل یک شاخص برازش سؤال را گزارش کرده بودند. همچنین ۵۵ درصد مطالعات شاخص برازش نسبی مدل را گزارش نکرده بودند و ۲۰ درصد مطالعات هر سه نوع شاخص برازش را گزارش کرده بودند.^۱ در میان پژوهش‌هایی که به شاخص برازش نسبی اشاره کرده بودند، ۴۴ درصد هیچ‌گونه گزارشی درباره شاخص‌های برازش مطلق نداشتند.

توصیف صلاحیت‌ها: به‌طور کلی توصیف صلاحیت‌ها شامل همبستگی صلاحیت‌ها، نیم‌رخ طبقه پنهان، نیم‌رخ صلاحیت و نیم‌رخ آزمودنی است. در ارزیابی مطالعات، نسبت نیم‌رخ صلاحیت یا نیم‌رخ طبقه پنهان، درصد تعلق آزمودنی‌ها به هر طبقه پنهان است که ۵۰ درصد گزارشی از این شاخص نداشته‌اند. همچنین ۹۰ درصد مطالعات، نیم‌رخ صلاحیت (درصد تعلق آزمودنی‌ها به هر صلاحیت است) را گزارش کرده بودند. در بررسی مطالعات، نیم‌رخ آزمودنی (شامل احتمال تعلق هر آزمودنی به یک طبقه پنهان و احتمال تعلق هر آزمودنی به هر صلاحیت) به‌ترتیب در ۷۰ درصد (احتمال تعلق هر آزمودنی به یک طبقه پنهان) و ۶۵ درصد (احتمال تعلق هر آزمودنی به هر صلاحیت) اشاره نشده بود. همچنین ۶۰ درصد مطالعات تعداد طبقه‌های پنهان را گزارش کرده بودند.

شواهد پایایی و روایی: ۵۵ درصد مطالعات هیچ شاخصی از شواهد پایایی را در پژوهش خود گزارش نکرده بودند. ۱۵ درصد مطالعات، از رویکرد سنتی برای گزارش پایایی آزمون خود با استفاده از برآورد ضریب پایایی به روش آلفای کرونباخ استفاده کرده بودند. تنها ۲۰ درصد مطالعات ثبات طبقه‌بندی را به‌منزله شاخص پایایی گزارش کرده و ۱۰ درصد مطالعات هم‌زمان از ضریب آلفای کرونباخ و ثبات طبقه‌بندی استفاده کرده بودند. درباره شواهد روایی ماتریس کیو، ۹۰ درصد مطالعات از روش‌های کیفی برای تدوین ماتریس کیو براساس روش لیتون و گیرل (۲۰۰۷) استفاده کرده و ۶۰ درصد مطالعات شواهد روایی کمی برای ماتریس کیو را گزارش کرده بودند. در این میان، دو مطالعه از روش روسوس و همکاران (۲۰۰۷)، یک مطالعه از روش دی‌لتوره (۲۰۰۸) و هشت مطالعه از روش دی‌لتوره و چپو (۲۰۱۶) استفاده کرده بودند. در یک مطالعه نیز روش ارزیابی ماتریس کیو گزارش نشده بود. همچنین ۲۰ درصد مطالعات، گزارش روایی طبقه‌بندی را براساس روش سوی، گیرل و جانگ (۲۰۱۲) ارائه کرده بودند. در ۱۰ درصد مطالعات نیز هیچ گزارشی از شواهد روایی دیده نشد.

کاربست نتایج: در ارزیابی مطالعات، هیچ مطالعه‌ای به کاربست یافته‌ها مانند ارائه بازخورد به دانش‌آموزان و معلمان و طراحی آموزش مبتنی بر نتایج نپرداخته بودند.

^۱ برخی پژوهشگران از شاخص‌های برازش نسبی (مانند IAC و BIC) استفاده کردند که برازش مدلی را در مقایسه با مدل‌های دیگر نشان می‌دهد. عده دیگری از پژوهشگران از شاخص‌های برازش مطلق (مانند M_2 و شاخص HCI) استفاده کردند که نشان می‌دهد آیا مدل به‌طور مطلق با داده‌ها برازش خوبی دارد. بعضی دیگر، شاخص‌های برازش سؤال‌مانند (MAD شاخص در همبستگی‌های زوج سؤال‌ها، RMSE برای برآورد پارامترهای هر سؤال) را گزارش کردند.

بحث و نتیجه گیری

با توجه به یافته‌ها، نیمی از پژوهش‌های بررسی شده بر سنجش سازه درک مطلب زبان انگلیسی و یک‌چهارم آن‌ها بر سنجش سازه ریاضی تمرکز داشتند. این یافته با یافته‌های پژوهش سسومز و هنسون (۲۰۱۸) همسویی داشت؛ با این تفاوت که در پژوهش‌های ارزیابی شده از سوی آن‌ها، ابتدا ریاضی و بعد درک مطلب زبان انگلیسی بیشترین فراوانی پژوهش‌ها را شامل می‌شدند. همچنین در میان دیگر مطالعات ارزیابی شده، پنج سازه شامل هوش، علوم، بخش شنیداری آیلتس، آزمون خواندن فارسی و آزمون نوشتن انگلیسی وجود داشتند که هر کدام یک پژوهش را شامل می‌شدند. با توجه به نتایج بررسی مطالعات ایرانی، به نظر می‌رسد این مطالعات با پیروی از مطالعات خارجی صلاحیت‌های شناختی را انتخاب کرده‌اند. از نظر لیتون و گیرل (۲۰۰۷) کاربرد مدل‌های شناختی تشخیصی باید طیف متنوعی از حیطه‌های مورد اندازه‌گیری را پوشش دهد؛ بنابراین پیشنهاد می‌شود پژوهش‌های آتی در این زمینه، به سنجش طیف وسیع‌تری از سازه‌ها بپردازند.

همچنین در ارزیابی مطالعات منتخب، به‌طور متوسط شش صلاحیت در مطالعات در نظر گرفته شده بود و در بیش از یک‌سوم مطالعات، تعداد صلاحیت‌ها بیش از شش مورد بود؛ در حالی که براساس بررسی‌های سسومز و هنسون (۲۰۱۸) در بیشتر مطالعات در حوزه شبیه‌سازی از تعداد صلاحیت‌های کمتری استفاده شده است. این موضوع نشان‌دهنده وجود شکاف میان پژوهش‌های فنی و کاربردی است؛ بنابراین توصیه می‌شود پژوهشگرانی که به استفاده از تعداد صلاحیت‌های فراوان تمایل دارند، در بررسی اولیه از به‌کارگیری مدل‌های آماری که در مطالعات شبیه‌سازی مبتنی بر تعداد صلاحیت‌های فراوان هستند، مطمئن شوند؛ زیرا با توجه به ماهیت مدل‌های شناختی تشخیصی که با افزایش تعداد صلاحیت‌ها، پیچیدگی مدل‌ها (مانند افزایش طبقه‌های پنهان، فراوانی پارامترهای برآورد شده در مدل) بیشتر می‌شود و بحث همگرایی مدل به چالش ایجاد می‌شود، در این میان به‌کارگیری تعداد صلاحیت‌های کمتر می‌تواند محتاطانه‌تر باشد. از سوی دیگر، ارائه یافته‌های پژوهشی برای مریبان آموزشی، با زیادبودن تعداد صلاحیت‌ها می‌تواند به سردرگمی آن‌ها و دشواری در تحلیل یافته از سوی آن‌ها منجر شود؛ برای نمونه در مطالعه‌ای با شش صلاحیت، تعداد طبقه‌های پنهان شامل ۶۴ طبقه شده است که شناسایی این تعداد طبقه و ارائه تصویری از احتمال تعلق آزمودنی‌ها به هر یک از آن‌ها، کاربرد این مدل‌ها را محدود می‌کند؛ بنابراین در عمل توصیه می‌شود پژوهشگران در طراحی ماتریس کیو به محدودبودن تعداد صلاحیت‌ها توجه کافی داشته باشند.

یکی دیگر از شاخص‌های ارزیابی شده در مطالعات، حجم نمونه پژوهشی بود. به‌طور متوسط حجم نمونه استفاده شده در پژوهش‌های ارزیابی شده، حدود ۲۵۰۰ نفر بود. در این میان، نکته مهم اختصاص دوسوم مطالعات به حجم نمونه‌ای بالاتر از ۱۰۰۰ آزمودنی بود. از سوی دیگر، کارکرد نظری این مدل‌ها بر استفاده نتایج چنین سنجش‌هایی در محیط‌های آموزشی تأکید دارد که بیشتر دارای حجم نمونه‌های کوچکی هستند، اما به دلیل اینکه بیشتر مدل‌های آماری در چارچوب سنجش شناختی تشخیصی تأکید فراوانی به استفاده از حجم نمونه‌های بزرگ دارند و با توجه به ورود نسبتاً جدید مدل‌های آماری ناپارامتریک، بیشتر پژوهشگران ایرانی به کاربرد

مدل‌هایی که امکان استفاده نمونه‌ای با حجم کوچک را میسر می‌کند، بی‌توجه بوده‌اند؛ بنابراین جهت‌گیری نوین در پژوهش‌ها می‌تواند به دنبال شناسایی مدل‌های شناختی تشخیصی متناسب با حجم نمونه کوچک و به‌کارگیری آن‌ها در عمل باشد تا بتوان در راستای ورود این نوع سنجش‌ها به محیط‌های آموزشی واقعی گام برداشت.

از دیگر یافته‌ها، اختصاص دوسوم مطالعات ارزیابی‌شده به رویکرد ریتروفیت بود. این یافته با نتایج بررسی‌های عفت‌پناه، بقایی و بوری (۲۰۱۹) هم‌خوانی دارد؛ بنابراین پژوهشگران باید در راستای پرهیز از برآزش‌نداشتن مدل با داده‌ها و افزایش روایی استنباط‌های مبتنی بر صلاحیت‌های شناختی، تأکید بیشتری در به‌کارگیری مدل‌های شناختی تشخیصی از رویکرد استقرایی و طراحی آزمون به‌منظور ارائه تصویری مناسب از ساختار صلاحیت‌هاش شناختی داشته باشند. یکی دیگر از شاخص‌های ارزیابی‌شده در این مطالعه، ابعاد و ساختار ماتریس کیو بود. در مطالعه حاضر، فراوانی مطالعات ۲۰ سؤالی بیش از بقیه بود، اما در یک‌پنجم مطالعات، تعداد سؤال‌ها از ۴۰ بیشتر بود. همچنین تعداد صلاحیت‌ها از ۶ تا ۲۰ متغیر بود که یک‌سوم مطالعات، بیش از ۶ صلاحیت داشتند. گفتنی است طراحی مناسب ماتریس کیو می‌تواند به بحث روایی در مدل‌های شناختی تشخیصی کمک کند و بدتوصیف‌شدن ماتریس کیو بر کیفیت پارامترهای برآوردشده و دقت طبقه‌بندی آزمودنی‌ها اثرگذار باشد؛ بنابراین بهتر بود که پژوهش‌های بررسی‌شده به ارزیابی ساختار ماتریس کیو به‌منظور بررسی کیفیت آن پردازند؛ درحالی‌که نتایج این مطالعه نشان داد، در بیش از یک‌سوم مطالعات از روش‌های کمی ارزیابی مجدد ماتریس کیو استفاده نشده و تنها به روش کیفی اکتفا شده بود؛ بنابراین توصیه می‌شود پژوهشگران در مطالعات آتی به ارزیابی کمی ماتریس کیو با استفاده از روش‌های مرسوم توجه داشته باشند و به‌کمک این روش کیفیت ماتریس کیو را آزمایش کنند.

یکی دیگر از جنبه‌های ارزیابی‌شده در مطالعات حوزه سنجش شناختی تشخیصی، شاخص‌های برآزش مدل بود. مرور پژوهش‌های این حوزه نشان می‌دهد که پژوهشگران از روش‌های مختلف برآزش برای ارزیابی مدل استفاده کرده بودند. برخی پژوهشگران از شاخص‌های برآزش نسبی (مانند IAC و BIC) استفاده کردند که برآزش یک مدل را در مقایسه با مدل‌های دیگر نشان می‌دهد. عده دیگری از شاخص‌های برآزش مطلق (مانند M_2 و شاخص HCI) استفاده کردند که نشان‌دهنده برآزش مناسب مدل با داده‌ها به‌طور مطلق است. بعضی دیگر، شاخص‌های برآزش سؤال‌مانند (MAD شاخص در همبستگی‌های زوج سؤال‌ها، RMSE برای برآورد پارامترهای هر سؤال) را گزارش کردند. نکته مهم اینکه نیمی از پژوهش‌ها گزارشی از شاخص‌های برآزش مطلق مدل ارائه نکرده بودند؛ بنابراین وضعیت معناداری پارامترهای برآوردشده در آن‌ها مشخص نبود. از آنجا که براساس نظر راپ، هنسون و تمپلین (۲۰۱۰) معناداری پارامترهای برآوردشده می‌تواند تحت تأثیر برآزش مدل قرار گیرد، در ۵۰ درصد پژوهش‌های بررسی‌شده در این مطالعه، معناداری پارامترهای برآوردشده با تردید همراه است. نکته مهم این است که بررسی شاخص‌های برآزش نسبی به‌تنهایی کاربردی ندارد و تنها به‌منظور مقایسه مدل‌ها استفاده می‌شود؛ بنابراین پیشنهاد می‌شود پژوهشگران این حوزه که تنها به دنبال مقایسه مدل‌ها نیستند و برای برآورد پارامترها یک مدل را انتخاب می‌کنند، در بخش ارزیابی مدل، به گزارش شاخص‌های

برازش مطلق مدل، توجه لازم را داشته باشند.

یکی دیگر از شاخص‌های بررسی شده، ارائه نیمرخ وضعیت فردی و گروهی آزمودنی‌ها بود که در این مطالعه با عنوان میزان توصیف صلاحیت‌های شناختی در مطالعات ارزیابی شده بررسی شد. نتایج نشان می‌دهد که حدود دوسوم مطالعات، نیمرخ آزمودنی‌ها را گزارش نکرده بودند؛ درحالی که به منظور استفاده کاربردی از یافته‌ها قاعدتاً باید نیمرخ تسلط به صلاحیت‌ها و نیمرخ آزمودنی‌ها را به معلمان و متصدیان آموزشی ارائه داد و مفیدبودن این نتایج را از دیدگاه آن‌ها بررسی کرد. همچنین دیگر شاخص بررسی شده، مفروضات مدل‌های آماری بود. براساس نتایج، درصد بسیار اندکی از مطالعات تنها به گزارش دو ویژگی چندبعدی بودن و داده‌های گمشده اکتفا کرده بودند. همچنین در میان مطالعاتی که ساختار ماتریس همبستگی میان صلاحیت‌ها گزارش شده بود، یافته‌ها نشان داد که بیش از نیمی از آن‌ها، ضریب همبستگی $0/9$ یا بیشتر را داشتند. از سوی دیگر، وجود همبستگی‌های بالا در ساختار ماتریس صلاحیت‌ها، براساس دیدگاه کانینا-هابنخت، راپ و ویلهلم (۲۰۰۹) احتمالاً حاکی از تک‌بعدی بودن آزمون است؛ بنابراین در چنین مواردی تحلیل بعد ضرورت می‌یابد؛ زیرا کاربرد مدل‌های شناختی تشخیصی برای آزمون‌های چندبعدی است.

شاخص دیگری که در این مطالعه ارزیابی شد، شواهد پایایی بود. یافته‌ها نشان می‌دهد که بیش از نیمی از مطالعات هیچ‌گونه شواهد پایایی را در مطالعه خود گزارش نکرده بودند. ثبات برآورد طبقه‌بندی آزمودنی به‌منزله شاخصی مناسب در مدل‌های شناختی تشخیصی تنها در یک‌سوم مطالعات گزارش شده بود و در یک‌چهارم مطالعات شواهد پایایی ابزار اندازه‌گیری - آلفای کرونباخ - مشاهده شد. همان‌طور که پیش‌تر مطرح شد، مفروضه زیربنایی این روش برآورد پایایی برای سازه‌هایی که تک‌بعدی نیستند، تضاد دارد. براساس نظر گرین، لیسیتز و مولایک (۱۹۷۷) پیش‌فرض استفاده از ضریب آلفای کرونباخ به‌منزله یک شاخص همسانی درونی بر تک‌بعدی بودن سازه تأکید دارد؛ بنابراین استفاده از این شاخص در مدل‌های شناختی تشخیصی با توجه به پیش‌فرض چندبعدی بودن سازه مورد اندازه‌گیری تأثیری ندارد و ضروری است مطالعات آتی در ارزیابی شواهد پایایی، مفروضات این نوع سنجش را مدنظر قرار داده شود.

همچنین ارائه شواهد روایی از دیگر شاخص‌های ارزیابی شده بود که نتایج آن نشان داد بیشتر مطالعات، از شاخص‌های کیفی درباره روایی ماتریس کیو، مانند اظهارنظر متخصصان و تفکر بلند استفاده کرده بودند، اما تنها نیمی از پژوهش‌ها گزارشی از شاخص کمی روایی ماتریس کیو را مدنظر قرار داده بودند. براساس دیدگاه راپ و تمپلین (۲۰۰۸) ب) ماتریس کیو منعکس‌کننده ابزار سنجش صلاحیت‌های شناختی افراد و عنصر اصلی در تعیین کیفیت بازخورد تشخیصی ابزار است، همچنین براساس یافته‌های پژوهش لی (۲۰۱۶) کیفیت ماتریس کیو می‌تواند بر دقت پارامترهای برآوردشده اثرگذار باشد؛ بنابراین پژوهش‌های آتی در این حوزه باید درباره ارزیابی روایی ماتریس کیو مطالعات دقیق‌تری انجام دهند. همچنین تنها حدود یک‌پنجم مطالعات از شاخص روایی با عنوان دقت طبقه‌بندی استفاده کرده بودند؛ زیرا صحت طبقه‌بندی آزمودنی‌ها براساس تسلط داشتن و نداشتن مبنای تصمیم‌گیری‌ها در مداخلات آموزشی است؛ بنابراین توصیه می‌شود پژوهشگران به ارزیابی این شاخص به‌منظور ایجاد قابلیت اعتماد بیشتر به داده‌ها توجه داشته باشند. نکته مهم این است که در مطالعه

حاضر بررسی مطالعات شناسایی‌شده به صورت آنلاین بود و شاید اگر بررسی حضوری اسناد موجود در کتابخانه‌های دانشگاهی صورت می‌گرفت، آمار بیشتری از این مطالعات به دست می‌آمد. در مجموع براساس چارچوب نظری، نتایج حاصل از سنجش شناختی تشخیصی می‌تواند برای ارتقای یادگیری دانش‌آموزان استفاده شود، اما در عمل پژوهش‌های ارزیابی‌شده در این مطالعه به منظور ارائه بازخورد تشخیصی-تکوینی به کمک گزارش خرد نیم‌رخ تسلط آزمودنی در صلاحیت‌های شناختی به معلمان پیش‌نرفته است. درباره رعایت مفروضات و ملزومات مدل‌های شناختی تشخیصی مانند چندبعدی بودن سازه مورد اندازه‌گیری و ارائه شواهد پایایی و روایی با کاستی‌هایی همراه بود؛ بنابراین پیشنهاد می‌شود پژوهشگران برای وفاداری به پیش‌نیازها و کاربرد این مدل‌ها در عمل به منظور کمک به طراحی مداخله‌های آموزشی، براساس توصیه‌های زیر گام بردارند:

پس از گردآوری داده‌ها و پیش از برازش مدل به بررسی مفروضه‌های مدل‌های شناختی مانند ارزیابی داده‌های گم‌شده و چندبعدی بودن سازه اندازه‌گیری پردازند. همچنین به منظور اطمینان از معناداری پارامترهای مدل شاخص‌های برازش مطلق مدل در بخش یافته‌ها گزارش شوند. با توجه به نقش نیم‌رخ آزمودنی‌ها به صورت فردی و گروهی برای شناسایی ضعف‌ها و قوت‌های آن‌ها، پژوهشگران این حوزه باید به ارائه گزارش این بخش از تحلیل توجه ویژه داشته باشند. از آنجا که گزارش دقیق شواهد پایایی و روایی عامل مهمی در تعیین کیفیت بازخورد تشخیصی ابزار است، پژوهشگران باید به این مهم به طور دقیق پردازند. همچنین براساس رویکرد نظری سنجش شناختی تشخیصی، کاربست یافته‌ها می‌تواند مهم‌ترین قوت این نوع رویکرد اندازه‌گیری باشد؛ بنابراین پژوهشگران باید برای این مهم گام بردارند. طراحی مدل‌ها نیز باید به نحوی باشد که امکان طراحی مداخلات مبنی بر نتایج این سنجش‌ها میسر شود.

منابع

- افضلی، ا.، دلاور، ع.، فلسفی‌نژاد، م. ر.، فرخی، ن. ع.، و برجلی، ا. (۱۳۹۳). کاربرد مدل‌های تشخیصی شناختی در تعیین ماهیت تفاوت در عملکرد دانش‌آموزان دختر و پسر پایه اول دبیرستان در ریاضیات. *دست‌آوردهای روان‌شناختی*. ۸۹-۱۰۴، (۲)۳۱
- تقیان، ح.، خدایی، ا.، بازرگان، ع.، مقدم‌زاده، ع.، و کبیری، م. (۱۳۹۷). ساخت آزمون خواندن برای دانش‌آموزان پایه ششم ابتدایی با استفاده از انگاره سنجش شناختی تشخیصی. *پژوهش‌نامه آموزش زبان فارسی به غیرفارسی‌زبانان*. ۳-۳۰، (۱)۷
- جاویدان‌مهر، ز.، و عنانی‌سراب، م. ر. (۱۳۹۶). بررسی گستردگی و دشواری زیرمهارت‌های خواندن و درک مطلب با استفاده از مدل تشخیصی شناختی جی دینا. *نقد زبان و ادبیات خارجی*. ۹۹-۱۱۷، (۲)۱۴
- رحیمی، ر.، یونسی، ج.، و مکرمی، م. (۱۳۹۷). کاربرد سنجش تشخیصی شناختی به منظور تحلیل سؤال‌های درک مطلب کنکور کارشناسی ارشد زبان انگلیسی. *اندازه‌گیری تربیتی*. ۱۷-۴۰، (۳)۸

- رنجبران، ف.، و علوی، س. م. (۱۳۹۵). سنجش تشخیصی شناختی آزمون خواندن و درک مفاهیم برای بازخورد تشخیصی تکوینی. *پژوهش‌های زبان‌شناختی در زبان‌های خارجی*. ۳۴۲-۳۴۱، (۲)۶.
- شه‌میرزادی، ن.، سیری، م.، مرعشی، ح.، و گرامی‌پور، م. (۱۳۹۹). بررسی سوگیری در سؤالات درک مطلب آزمون مقطع دکترای رشته زبان انگلیسی تحت سنجش تشخیصی شناختی. *پژوهش‌های زبان‌شناختی در زبان‌های خارجی*. ۱۵۲-۱۶۵، (۱)۱۰.
- محسن‌پور، م.، گویا، ز.، شکوهی یکتا، م.، کیامنش، ع. ر.، و بازرگان، ع. (۱۳۹۴). سنجش تشخیصی صلاحیت‌های سواد ریاضی. *نوآوری‌های آموزشی*. (۱)۳۳-۷.
- محسن‌پور، م. (۱۳۹۸). سنجش صلاحیت‌های شناختی چندسطحی سواد ریاضی دانش‌آموزان پایه نهم: کاربردی از مدل پی جی دینا. *مطالعات اندازه‌گیری و ارزشیابی آموزشی*. ۱۳۴-۱۰۹، (۲)۹.
- محمدی، ن.، دلاور، ع.، فرخی، ن. ع.، و مینایی، ا. (۱۳۹۶). شناسایی صفات زیربنایی سؤال‌های آزمون هوش و کسلر چهار کودکان براساس توانایی باریک نظریه کتل-هورن-کارول با استفاده از مدل تشخیصی شناختی جی دینا. *اندازه‌گیری تربیتی*. (۲)۷، ۳۲-۱.
- مفاخری، ش.، شاهورانی سمنانی، ا.، بهزادی، م. ح.، و برهمند، ع. (۱۳۹۸). ارائه مدل ساختاری مؤلفه‌های مؤثر در یادگیری عبارت‌ها جبری دوره جدید متوسطه با استفاده از روش سلسله‌مراتبی ویژگی‌ها. *نوآوری‌های آموزشی*. ۱۲۷-۱۴۶، (۲)۱۸.
- مقدم، ا.، فلسفی‌نژاد، م. ر.، فرخی، ن. ع.، و استاجی، م. (۱۳۹۵). تحلیل تشخیصی سؤال‌های بخش درک مطلب زبان انگلیسی عمومی آزمون ورودی دوره‌های دکتری با استفاده از مدل غیرجبرانی فیوژن. *اندازه‌گیری تربیتی*. (۴)۶، ۴۱-۶۸.
- مینایی، ا.، دلاور، ع.، فلسفی‌نژاد، م. ر.، کیامنش، ع. ر.، و مهاجر، ی. (۱۳۹۳). مدل‌پردازی تشخیصی شناختی (CDM) سؤال‌های ریاضیات تیمز ۲۰۰۷ در دانش‌آموزان پایه هشتم ایران با استفاده از مدل یکپارچه با پارامترپردازی مجدد (RUM) و مقایسه مهارت‌های ریاضی دانش‌آموزان دختر و پسر. *اندازه‌گیری تربیتی*. (۲)۲، ۱۶۹-۱۳۷.

References

- Bradshaw, L., Izs'ak, A., Templin, J., & Jacobson, E. (2013). Diagnosing teachers' understandings of rational numbers: building a multidimensional test within the diagnostic classification framework. *Educational Measurement: Issues and Practice*. 33(1), 2-14.
- Chen, J., & de la Torre, J. (2014). A procedure for diagnostically modeling extant large-scale assessment data: The case of the programme for international student assessment in reading. *Psychology*. 5(18), 1967-1978.
- Choi, K. M., Lee, Y.-S., & Park, Y. S. (2015). What CDM can tell about what students have learned: An analysis of TIMSS eighth grade mathematics. *Eurasia Journal of Mathematics, Science, & Technology Education*. 11(6), 1563-1577.

- Chiu, C.-Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement*, 37(8), 598–618.
- Chiu, C., & Douglas, J. (2013). A nonparametric approach to cognitive diagnosis by proximity to ideal response patterns. *Journal of Classification*, 30(2), 225–250.
- Cui, Y., Gierl, M. J., & Chang, H.-H. (2012). Estimating classification consistency and accuracy for cognitive diagnostic assessment. *Journal of Educational Measurement*, 49(1), 19–38.
- Decarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-Matrix. *Applied Psychological Measurement*, 35(1), 8-26.
- De la Torre, J. (2008). An empirically based method of q-matrix validation for the Dina model: Development and applications. *Journal of Educational Measurement*, 45(4), 343-362.
- De la Torre, J., & Chiu, C. Y. (2016). General method of empirical Q-matrix validation. *Psychometrika*, 81(2), 253–273.
- Effatpanah, F. (2019). application of cognitive diagnostic models to the listening section of the international English language testing system (IELTS). *International Journal of Language Testing*, 9(1), 1-28.
- Effatpanah, F., Baghaei, P., & Boori, A. A. (2019). Diagnosing EFL learners' writing ability: A diagnostic classification modeling analysis. *Language Testing in Asia*, 9(12), 1-28.
- Gierl, M. J., Wang, C., & Zhou, J. (2008). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in algebra on the SAT. *Journal of Technology, Learning, and Assessment*, 6(6).
- Gorin, J. S. (2009). Diagnostic classification models: Are they necessary? *Measurement: Interdisciplinary Research and Perspectives*, 7(1), 30–33.
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, 26(2), 91–108
- Green, S. B., Lissitz, R. W., & Mulaik, S. (1977). Limitations of coefficient alpha as an index of text unidimensionality. *Educational and Psychological Measurement*, 37(4), 827–39.
- Henson, R. A. (2009). Diagnostic classification models: Thoughts and future directions. *Measurement: Interdisciplinary Research and Perspectives*, 7(1), 34–36.
- Jang, E. E. (2005). *A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL*. Ph. D. Dissertation. University of Illinois at Urbana-Champaign. Education Department.
- Jang, E. E. (2008). A framework for cognitive diagnostic assessment. In C. A. Chapelle,

- Y.-R. Chung, & J. Xu (Eds.), *Towards adaptive CALL: Natural language processing for diagnostic language assessment* (pp. 117-131). Ames, IA: Iowa State University.
- 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), 31–73.
- Javidanmehr, Z., & Anani Sarab, M. R. (2017). Cognitive diagnostic assessment: Issues and considerations. *International Journal of Language Testing*, 7(2), 73–98.
- Javidanmehr, Z., & Anani Sarab, M. R. (2019). Retrofitting non-diagnostic reading comprehension assessment: application of the G-DINA model to a high stakes reading comprehension test. *Language Assessment Quarterly*, 16(3), 294-311.
- Kabiri, M., Ghazi-Tabatabaei, M., Bazargan, A., Shokoohi-Yekta, M., & Kharrazi, K. (2016). Diagnosing competency mastery in science: An application of GDM to TIMSS 2011d. *Applied Measurement in Education*, 30(1), 27-38.
- 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. P., & Gierl, M. J. (2007). Why cognitive diagnostic assessment. In J. P. Leighton, & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 3-18). New York: Cambridge University Press.
- Li, H. (2016). *Estimation of Q-matrix for DINA model using the constrained generalized Dina framework*. Doctoral Ph. D. Dissertation. Graduate School of Arts and Sciences, Columbia University.
- Lei, P. W., & Li, H. (2016). Performance of fit indices in choosing correct cognitive diagnostic models and Q-matrices. *Applied Psychological Measurement*, 40(6), 405–417.
- Madison, M. J., & Bradshaw, L. P. (2014). Design on classification accuracy in the log-linear cognitive diagnosis model. *Educational and Psychological Measurement*, 75(3), 491–511.
- Ravand, H., Barati, H., & Widhiarso, W. (2013). Exploring diagnostic capacity of a high stakes reading comprehension test: A pedagogical demonstration. *Iranian Journal of Language Testing*, 3(1), 11-37.
- Ravand, H., & Robitzsch, A. (2016). Cognitive diagnostic modeling using R. *Practical Assessment, Research & Evaluation*, 20(11), 1-12.
- Ravand, H. (2016). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment*, 34(8), 782–799.
- Ravand, H., & Robitzsch, A. (2018). Cognitive diagnostic model of best choice: A study of reading comprehension. *Educational Psychology*, 38(10), 1255-1277.

- Ravand, H., & Baghaei, P. (2020). Diagnostic classification models: Recent developments, practical issues, and prospects. *International Journal of Testing*, 20(1), 24-56.
- Roussos, L., DiBello, L., Stout, W., Hartz, S., Henson, R., & Templin, J. (2007). The Fusion Model Skills Diagnosis System. In J. Leighton & M. Gierl (Eds.), *Cognitive Diagnostic Assessment for Education: Theory and Applications* (pp. 275-318). New York: Cambridge University Press. doi:10.1017/CBO9780511611186.010.
- Rupp, A. A., & Templin, J. (2008a). Unique characteristics of cognitive diagnosis models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, 6(4), 219-262
- Rupp, A., & Templin, J. (2008b). The effects of Q-matrix misspecification on parameter estimates and misclassification rates in the DINA model. *Educational and Psychological Measurement*, 68(1), 78-98.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement, theory, methods, and applications*. New York: The Guilford Press.
- Sessoms, J., & Henson, R. A. (2018). Applications of diagnostic classification models: A literature review and critical commentary. *Interdisciplinary Research and Perspectives*, 16(1), 1-17
- Sinharay, S., & Haberman, S. J. (2009). How much can we reliably know about what examinees know? *Measurement: Interdisciplinary Research and Perspectives*, 7(1), 46-49.
- Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453-488). Hillsdale, NJ: Erlbaum.
- Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, 30(2), 251-275.
- Yang, X., & Embretson, S. E. (2007). Construct validity and cognitive diagnostic assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 119-145). New York: Cambridge University Press.

