

سنجش تشخیصی شناختی آزمون خواندن و درک مفاهیم برای بازخورد تشخیصی تکوینی

فاطمه رنجبران*

دکتری آموزش زبان انگلیسی، دانشکده زبان‌ها و ادبیات خارجی، دانشگاه تهران، تهران، ایران

سید محمد علوی**

پروفسور زبان شناسی کاربردی، دانشکده زبان‌ها و ادبیات خارجی، دانشگاه تهران، تهران، ایران

(تاریخ دریافت: ۹۵/۱۲/۰۶، تاریخ تصویب: ۹۶/۰۶/۱۴، تاریخ چاپ: مهر ۱۳۹۶)

چکیده

یکی از موضوع‌های مهم در سنجش تشخیصی شناختی، نبود پژوهشی در زمینه ساخت آزمون‌های تشخیصی^۱ به منظور پیشبرد اهداف تشخیصی شناختی است. طی سالهای اخیر، پژوهشگران تلاش کرده‌اند تا نتایج آزمون را به صورت کاربردی‌تر برای دانشجو و معلمان فراهم کنند و نتیجه‌ها را فراتر از نمره میانگین ارائه دهند تا دانشجو بتواند به وضعیت درسی خود بهبود بخشد.

به همین منظور این پژوهش به تدوین یک آزمون خواندن و درک مفاهیم بر اساس چارچوب شناختی پرداخته است. ابتدا مهارت‌های لازم برای پاسخدهی به هر سوال شناسایی و سپس یک ماتریس کیو^۲ برای شناسایی رابطه بین هر سوال و مهارت‌های مورد نیاز ساخته شده است. سرانجام این آزمون برای تشخیص نقاط قوت و ضعف ۱۹۸۶ دانشجوی دانشگاه تهران برگزار شد و از مدل فیوژن^۳ برای به دست آوردن گزارش خرد نیمرخ تسلط آزمودنی در مهارت^۴ استفاده شد. نتیجه‌های به دست آمده می‌تواند برای بازخورد تشخیصی تکوینی^۵ دانشجویها مفید واقع شوند.

واژه‌های کلیدی: بازخورد تشخیصی تکوینی؛ سنجش تشخیصی شناختی؛ مدل تشخیصی شناختی؛ مدل فیوژن؛ ماتریس کیو؛ مهارت‌ها.

* E-mail: franjbaran@ut.ac.ir

** E-mail: smalavi@ut.ac.ir

1. Cognitive diagnostic assessment
2. Diagnostic tests
3. Q-Matrix
4. Fusion Model
5. Learner's skill mastery profile
6. cognitive diagnostic feedback

۱. مقدمه

طی سال‌های اخیر، سنجش و آزمون به‌طور کلی به‌عنوان ابزاری برای به‌دست آوردن نمره فردی، میانگین نمره‌ها یا نمره کلی یادگیری دانشجویان بوده است. البته سنجش تکوینی گامی فراتر از نهاد و آموزش را با هدف آگاهی از میزان و روش یادگیری دانشجویان پیش می‌برد. به این ترتیب، نقاط قوت و ضعف یادگیری آن‌ها را و نیز تشخیص دشواری‌های روش آموزشی معلم را فراهم می‌کند که می‌تواند به توسعه یادگیری و تشخیص نقاط قوت و ضعف و پاسخ به نیازهای آینده منجر شود. از طرف دیگر، آزمون‌های تدوین شده توسط معلم، اخیراً به دلیل عملکرد آن‌ها برای سنجش تکوینی و یادگیری دانشجویان جایگاه ویژه‌ای یافته‌اند (هوانگ و وو، ۲۰۱۳). برای تسهیل و بهبود فرایند یادگیری، از معلم‌ها انتظار می‌رود که در ساخت آزمون و یادگیری تشخیصی در کلاس مهارت لازم را داشته باشند. از آنجایی که انتظار می‌رود این آزمون‌ها خطای دانشجویان را در فرایند یادگیری شناسایی کنند، کاربرد آزمون‌های تشخیصی برای بهبود درک مفهومی دانشجویان اهمیت بسیاری دارد و شناسا تلقی می‌شوند (هارتمن، ۲۰۰۱). بنابراین، محققان به‌طور گسترده‌تری برای بهبود آموزش و یارگیری بر اتحاد روانشناسی شناختی و سنجش آموزشی متمرکز شده‌اند (سنو و لومن، ۱۹۸۹؛ لیتون، گیرل و هانوکا، ۲۰۰۴؛ تاتسواکا، ۱۹۹۵؛ میسلوی، ۲۰۰۶).

خواندن و درک مفاهیم به‌عنوان مهارتی برای به‌دست آوردن دانش در تمام رشته‌ها شناخته شده است. بنابراین، این نیاز وجود دارد که اجزای مختلف مهارت خواندن برای درک بهتر آن و همچنین دستیابی به دشواری‌های زبان‌آموزان در این گستره سنجیده شود. اگر بخش‌های مشکل ساز این مهارت، در طول ترم تشخیص داده شود، بازخورد کافی به دانشجویان داده می‌شود تا بتوانند فرایند یادگیری را بهبود بخشند و ضعف‌ها را جبران کنند.

انتقاد بر سر هدف اصلی آزمون‌های آموزشی است که سنجش کمی از توانایی‌ها و مهارت‌های عمومی دانشجویان را در مقایسه با دانشجویان دیگر در گروه معیار فراهم می‌سازد. این گونه سنجش معیار محور به‌طور گسترده‌ای برای رتبه‌بندی و گزینش دانشجویان به منظور اهداف آموزشی گوناگون استفاده شده است. با وجود اینکه این آزمون‌ها اطلاعات کلی درباره مهارت دانشجویان و عملکرد آن‌ها در آزمون فراهم می‌کند، به ندرت می‌تواند اطلاعات جزئی درباره نقاط قوت و ضعف دانشجویان را فراهم کند که می‌تواند در بهبود یادگیری مهارت‌ها به آن‌ها کمک کند. این امر می‌تواند حتی به معلم در برنامه‌ریزی آموزشی و یا به عنوان راهنمای تدریس کمک کند. اینک، محققان بر این باورند که سنجش تشخیصی شناختی می‌تواند نقشی

کلیدی در بهبود جایگاه اطلاعاتی سنجش داشته باشد (لیتون و گیرل، ۲۰۰۷؛ آلدرسون، ۲۰۱۰؛ روپ و همکاران، ۲۰۱۲؛ دو لا تور، ۲۰۰۹؛ ینگ، ۲۰۰۵). در تفسیر خود از "تشخیص شناختی و کیو ماتریس در سنجش زبان"، آلدرسون (۲۰۱۰) با ناراحتی به این امر اشاره می‌کند که هیچ بررسی درباره مشکل نبود سنجش‌های تشخیصی واقعی وجود ندارد. در واقع، تقریباً تمام مطالعات انجام شده تاکنون بر آزمون‌های موجود در مقیاس بزرگ بوده است، و هیچ آزمونی برای انجام تجزیه و تحلیل تشخیصی شناختی طراحی نشده است. بنابراین، این مطالعه سعی دارد با ساختن یک آزمون خواندن و درک مفاهیم بر اساس معیارهای تشخیصی شناختی به این ندا پاسخ دهد، آزمونی که بتواند اطلاعات جزئی از نقاط قوت و ضعف دانشجو را در مهارت خواندن فراهم کند و در نهایت به بازخورد تشخیصی تکوینی منجر شود.

۲. مبانی نظری

۲-۱. سنجش تشخیصی شناختی

سنجش تشخیصی شناختی، چارچوب نظری جدیدی در زمینه اندازه‌گیری آموزشی و روانشناختی است که به جای رتبه‌بندی آزمودنی‌ها بر اساس نمره‌های به‌دست آمده از آزمون برای ارائه بازخورد تشخیصی تکوینی، گزارش خرد نیمرخ تسلط آزمودنی در مهارت را ارائه می‌دهد (ینگ، ۲۰۰۹). اگرچه آزمون‌های مرسوم برای درجه‌بندی و رتبه‌بندی عملکرد آزمودنی در آزمون تهیه شده‌اند، ولی آن‌ها اطلاعات تشخیصی مفیدی درباره عملکرد آزمودنی‌ها ارائه نمی‌دهند. سنجش تشخیصی شناختی، اطلاعات مفصلی در نقاط قوت و ضعف آزمودنی‌ها در ساختار خاص دانش و مهارت‌های پردازش تدارک می‌بیند، به گونه‌ای که آزمودنی‌ها می‌توانند به دلایل موفقیت یا شکست‌شان در سوالی خاص پی برده و عملکرد آینده خود را بهبود بخشند (لیتون و گیرل، ۲۰۰۷).

وقتی هدف سنجش ارزشیابی یادگیرندگان و بررسی جنبه‌های خاصی از مهارت‌هاست، ارائه خرد شایستگی‌ها ضروری به‌نظر می‌رسد. هدف سنجش تشخیصی شناختی هدایت آزمون‌سازان به‌منظور ساخت آزمون‌هایی است که قادر به‌چنین ارائه‌ای باشند. در این روش به‌جای توجه به سنجش پیامدهای یادگیری بر سنجش برای یادگیری و سنجش به‌عنوان فرایند یادگیری تاکید می‌شود (ستینگنز، آرتر و چاپپوس، ۲۰۰۴). در رویکرد سنجش پیامدهای یادگیری، سنجش به‌عنوان ابزاری برای ارزشیابی زبان‌آموزان به‌منظور اطلاع از تحقق اهداف تعیین شده در برنامه درسی و میزان آمادگی فراگیران برای رفتن به مقطع بعدی (یعنی ارتقا،

دانش آموختگی، و دریافت گواهینامه) در نظر گرفته می‌شود. در حالی که در رویکرد سنجش برای یادگیری، اطلاعاتی در اختیار معلمان قرار می‌گیرد تا به کمک آن، آموزش و یادگیری را در کلاس درس اصلاح کنند (ستیگنز و همکاران، ۲۰۰۴). معلمان می‌توانند از اطلاعات تشخیصی تکوینی برای طراحی دوباره رویکردهای آموزشی خودشان، بازنگری منابع آموزشی و رفع نقاط ضعف فراگیران استفاده کنند. افزون بر این، در این رویکرد زبان‌آموزان در فعالیت‌های مختلف یادگیری و سنجش درگیر می‌شوند (ینگ، ۲۰۰۹).

به‌طور کلی، آزمون تشخیصی به‌منظور شناسایی نقاط قوت و ضعف آزمودنی‌ها طراحی می‌شود و بیش‌تر بر نقاط ضعف آزمودنی‌ها تاکید دارد تا نقاط قوت آن‌ها. چنین آزمونی باید به‌اصلاح آموزش‌های بعدی منجر شود، پس خوراند^۱ مفصلی درباره عملکرد آزمودنی‌ها ارائه دهد که می‌توان بر اساس آن اقدامات اصلاحی را صورت داد، نتایج فوری به آزمودنی‌ها دهد که مبتنی بر نظریه‌های شناختی باشد (آلدرسون، ۲۰۰۵). جنبه دیگر بررسی در سنجش تشخیصی شناختی، مدل‌های شناختی در زمینه محتوایی مورد آزمون است که برای تبیین عملکرد آزمودنی‌ها لازم است. مدل‌های شناختی باید تبیین‌گر عملکرد آزمودنی‌ها از طریق ارائه بینش‌هایی در این زمینه باشد که آیا درک آزمودنی، عامل اصلی عملکرد او است یا عوامل دیگر؟ (پالگرینو و چوداوسکی، ۲۰۰۳)

۲-۲. ماتریس کیو

بخش اصلی طراحی سنجش تشخیصی، مجموعه مهارت‌ها یا صفت‌های^۲ پنهانی است که با هر سوال اندازه‌گیری می‌شوند. رابطه سوال-مهارت در ماتریسی به‌نام ماتریس کیو بیان می‌شود (که تاتسوکا آن را در ۱۹۸۳ ارائه داده است) که مجموعه فرضیه‌های آزمون‌پذیر ارائه شده توسط متخصص درباره مهارت‌ها زیربنایی یک آزمون، با توجه به ساختار آزمون، آزمودنی‌ها و داده‌ها است (دی‌پلو و استوت، ۲۰۰۸). درایه‌های صفر و یک در این ماتریس یعنی اگر سوال i نیاز به صفت a داشته باشد، به درایه ia عدد یک و در غیر این صورت عدد صفر اختصاص داده می‌شود. در واقع ترسیم سوالات یک آزمون در قالب مهارت‌های زیربنایی آن، ماتریس کیو را تشکیل می‌دهد. در جدول زیر نمونه‌ای از ماتریس کیو برای سه سوال و

1. feedback

2. attributes

چهار مهارت ارائه شده است. در این ماتریس، سوال یک، مهارت دوم و چهارم، سوال دو، مهارت اول و چهارم، و سوال سه، مهارت اول و سوم را اندازه گیری می کند.

جدول ۱. ماتریس کیو مهارت‌ها ۳×۴

سوال	مهارت ۱	مهارت ۲	مهارت ۳	مهارت ۴
۱i	۰	۱	۰	۱
۲i	۱	۰	۰	۱
۳i	۱	۰	۱	۰

تدوین درست ماتریس کیو فرایند مهمی است که می تواند بر دقت طبقه بندی که هدف مدل های تشخیصی شناختی است اثر بگذارد (مدیسن و برادشاو، ۲۰۱۴). ماتریس کیو نشان دهنده فرضیه ای درباره مهارت های لازم برای پاسخدهی به هر سوال در آزمون است (لی و سوان، ۲۰۱۳). بنابراین هر سوال نیازمند داشتن بیش از یک مهارت برای پاسخ دهی صحیح است. بنابر اظهارات باک و همکاران (۱۹۹۷)، ساخت ماتریس کیو یک فرایند را دنبال می کند. ابتدا باید فهرستی از مهارت ها تدوین شود و سپس هر سوال بنا بر مهارت های لازم برای پاسخدهی رمزگذاری شود. اولین گام در این فرایند با مراجعه به مطالعات قبلی، نظرات پروتکل کلامی تفکر با صدای بلند و داوری درجه بندی محتوا برای مهارت های هر سوال بدست می آید. گام بعدی، تجزیه و تحلیل داده ها با استفاده از مدل تشخیصی شناختی^۳ منتخب، مدل فیوژن در این مطالعه، با ماتریس کیو تدوین شده است. سرانجام، ماتریس کیو می تواند بر اساس آمار بدست آمده برای هر مهارت اصلاح شود.

۲-۳. رویکرد ساخت آزمون تشخیصی
در رویکرد ساخت آزمون تلاش می شود تا در راستای هدف آزمون، سوال های تشخیصی تدوین شود. یک آزمون تشخیصی خوب آزمونی است که اطلاعاتی فراتر از برآورد سطوح

1. think-aloud verbal protocols
2. content raters judgment
3. cognitive diagnostic models (CDMs)

توانایی کلی آزمودنی ارائه می‌دهد. اطلاعات ویژه‌ای باید در پاسخ‌های آزمودنی وجود داشته باشد و نقاط ضعف یا کج‌فهمی‌ها در دانشجو را مشخص کرده و تصویر کامل‌تری از توانایی‌هایش را ارائه دهد. به عبارت دیگر، سوالات باید به گونه‌ای طراحی شود تا نشان دهد که چرا دانشجو به سوالات این گونه پاسخ داده است. سوالات طراحی شده در آزمون‌های مرسوم، اطلاعات کافی برای تشخیص ارائه نمی‌دهد در حالی که یک آزمون تشخیصی مناسب، می‌تواند بین مهارت‌هایی که یک آزمودنی در آن به تسلط رسیده و مهارت‌هایی که هنوز یاد نگرفته است، تمایز قایل شود. سوال‌های تشخیصی خوب، فرصتی برای بررسی فرایند پاسخ‌های فراگیران تدارک می‌بیند و میزان اطلاعات به دست آمده از پاسخ‌های آزمودنی‌ها را افزایش می‌دهد. دو روش مهم در طراحی این نوع آزمون‌ها عبارتند از طرح شاهد محور میسلوی^۱ (۱۹۹۴) و نظام طراحی شناختی امبرتسون^۲ (۱۹۹۰) که با وجود تفاوت‌هایی که با هم دارند، شباهت‌های بیش‌تری در آن دیده می‌شود. مهم‌ترین شباهت آن‌ها عبارت است از نقش نظریه شناختی و شواهدی که از طریق مدل تجربی زیربنایی تدوین آزمون به دست می‌آید (گورین، ۲۰۰۶). طرح سنجش مبتنی بر شاهد، رویکردی مبتنی بر سازه است که دارای لایه‌های بررسی زمینه، مدل‌سازی زمینه، چارچوب سنجش مفهومی، اجرای سنجش و ارائه سنجش بوده و هدف آن طراحی سنجش مبتنی بر بحث‌های دارای شاهد^۳ است تا بتواند استنباط‌هایی درباره نقاط قوت و ضعف آزمودنی‌ها داشته باشد. پنج مولفه اصلی این نوع سنجش عبارت است از مدل‌های دانش آموز، شاهد، تکلیف، ارائه و ترکیب (میسلوی، استاینبرگو آلموند، ۲۰۰۳).

نظام طراحی شناختی امبرتسون نیز به منظور تاکید بر نقش نظریه شناختی در طراحی آزمون به منظور بهبود تفسیر و استفاده از نمره‌های آزمون‌های توانایی و پیشرفت تحصیلی ارائه شده است. این روش شامل مجموعه مراحل برای مهارت‌های مورد نظر در تدوین، نوشتن و تحلیل آزمون است. هفت گام طراحی آزمون در این رویکرد عبارت است از تعیین هدف اندازه‌گیری، شناسایی ویژگی‌های زمینه تکلیف، تدوین مدل شناختی، ساخت سوال، بررسی مدل سوالات ساخته شده، تولید بانک سوال با توجه به پیچیدگی شناختی و روایی (گورین، ۲۰۰۶).

-
1. Mislevy's Evidence-Centered Design
 2. Embretson's Cognitive Design System
 3. Evidentiary reasoning

۲-۴. مدل‌های تشخیصی شناختی

مدل‌هایی احتمالاتی چند بعدی است که با شناسایی مهارت‌های مورد نیاز برای ارایه پاسخ درست به سوالات یک آزمون، تقاطق قوت و ضعف آزمودنی‌ها را تعیین و آن‌ها را به صورت تشخیصی طبقه‌بندی ارایه نیمرخ مسلط در مهارت می‌کند (لیتون و گیرل، ۲۰۰۷). اغلب مطالعات انجام شده تا کنون محدود به تجزیه و بررسی آزمون‌های موجود بوده‌اند، و نه تدوین آزمون‌های جدید، در حالی که این مطالعه به تدوین آزمون تشخیصی شناختی جدید بر اساس چارچوبی تشخیصی شناختی می‌پردازد. مدل‌های تشخیصی شناختی مدل‌هایی آماری با هدف پیوند دادن نظریه شناختی با ویژگی‌های روانسنجی هر سوال آزمون می‌باشند (گورین، ۲۰۰۶). نمونه‌هایی از مدل‌های تشخیصی شناختی شامل مدل روش شناسی فضا قاعده^۱ (تاتسوکا، ۱۹۹۵)، روش سلسله مراتبی صفت^۲ (لیتون، گیرل، و هونکا، ۲۰۰۴)، و مدل فیوژن (هارتز، ۲۰۰۲) است. بیشتر مدل‌های تشخیصی شناختی مدل‌های مولفه پنهانی با محوریت نظریه سوال پاسخ‌اند که در آن‌ها ویژگی چند بعدی از اهمیت بیشتری برخوردار است. در مدل‌های نظریه سوال پاسخ قبلی، توانایی زبان‌آموز توسط یک بهین‌گزینه‌توانایی عمومی مدل سازی می‌شود. این در حالی است که ویژگی چند بعدی مدل‌های تشخیصی شناختی این امکان را فراهم می‌کنند تا فرایندهای ذهنی نهفته در پاسخ دانشجو را بررسی کنیم و توانایی کلی را به چندین بخش تقسیم کنیم. تعداد ابعاد بستگی به تعداد مولفه‌های مهارت موجود در آزمون دارد. مولفه‌های پنهان^۳ این مدل‌ها دو بخشی، مانند تسلط یا عدم تسلط، یا سطوح چند بخشی مانند متغیر رتبه‌بندی شامل عالی، خوب، متوسط، ضعیف و غیره‌اند.

۲-۴-۱. مدل فیوژن

مدل فیوژن، مدل سوال پاسخ تشخیصی شناختی مبتنی بر مهارت است که پاسخ آزمودنی‌ها به پرسش‌ها را به عنوان تابعی از ویژگی‌های آزمودنی (بردار نیمرخ سطح تسلط آزمودنی در مهارت‌ها) و بهین‌گزینه‌های سوال را مدل‌سازی می‌کند. این مدل توسط هارتز (۲۰۰۲) که مدل یکپارچه دیبلو، استوت و روسو^۴ (۱۹۹۵) را دوباره پارامتربندی کرد، پیشنهاد شده و مبتنی

1. Tatsuoka's Rule-Space Methodology
2. Attribute Hierarchy Method
3. Latent traits
4. DiBello, Stuart and Roussos's Unified Model

بر الگوریتم زنجیره‌های مارکف مونته کارلو^۱ است. در این مدل احتمال ارائه پاسخ درست به دو صورت کاهش یافته (مبتنی بر دو بهین گزینه) و کامل (مبتنی بر سه بهین گزینه) است. احتمال ارائه پاسخ درست در مدل کاهش یافته عبارت است از:

$$p(X_{ni} = 1 | \alpha_n, \eta_n; \pi^*, r_{ik}^*) = \pi^* \prod_{k=1}^k r_{ik}^{*(1-\alpha_{nk})}$$

احتمال پاسخ درست در مدل کامل عبارت است از:

$$p(X_{ni} = 1 | \alpha_n, \eta_n; \pi^*, r_{ik}^*, c_i) = \pi^* \prod_{k=1}^k r_{ik}^{*(1-\alpha_{nk})} P_{ci}(\eta_n)$$

ویژگی‌های آزمودنی شامل تسلط یا عدم تسلط آزمودنی n در مهارت‌های ارائه شده توسط بردار نیمرخ مهارت^۲ ($\alpha = (\alpha_n, \dots, \alpha_{nk})$) و پارامتر آزمودنی (η_n) است. افزون بر این، بهین گزینه‌های سوال در این مدل به شرح زیر تعریف می‌شوند (دی‌بلو و استوت، ۲۰۰۸؛ ینگ، ۲۰۰۵ و ۲۰۰۹؛ لی، ۲۰۱۱، تمپلین ۲۰۰۴؛ روسو و همکاران، ۲۰۰۷).

بهین گزینه سوال π_i^* یا خط پایه، یا دشواری عبارت است از احتمال ارائه پاسخ درست به یک سوال، با این فرض که آزمودنی در همه مهارت‌های مورد نیاز به تسلط رسیده و به فرض اینکه مقدار η زیاد است، مقادیر آن از صفر تا یک متغیر است. مقادیر کم‌تر از ۰/۶ نشان دهنده این است که سوال برای مهارت (های) اختصاص داده شده به آن، مشکل است. بنابراین مهارت‌های بیش‌تر یا متفاوت‌تری (به احتمال سخت‌تری) باید به‌چنین سوالی اختصاص یابد. افزون بر این، مقادیر پایین این بهین گزینه را می‌توان اینگونه تفسیر کرد که آزمودنی‌ها شانس کمی برای ارائه پاسخ صحیح به سوال، حتی پس از تسلط در مهارت (های) زیربنایی یک سوال دارند.

بهین گزینه سوال r_{ik}^* یا جریمه یا تشخیص به‌رابطه سوال-مهارت (ارایه شده در ماتریس کیو) مربوط می‌شود و مشخص می‌کند که تا چه حد تسلط در یک مهارت خاص، بر شانس ارائه پاسخ درست به یک سوال اثر دارد. بهین گزینه جریمه، یک عامل ضربی است که در اثر آن اگر فرد در مهارت k در سوال i به تسلط نرسیده باشد احتمال پاسخ درست کاهش می‌یابد. بهین گزینه جریمه برای هر درایه ماتریس کیو برآورد می‌شود زیرا آزمودنی‌ها در طبقه‌های پنهان مختلف در ترکیب‌های متفاوتی از مهارت‌ها به تسلط رسیده‌اند. مقادیر آن به طور معمول

1. Markov Chain Monte Carlo
2. Skills mastery profile vector

بین صفر تا یک است. مقادیر 0.5 تا 1.0 برای پارامتر r_{ik}^* نشان‌دهنده پرسش‌هایی با ضریب تشخیص خیلی بالا است. وقتی اغلب مقادیر این بهین‌گزینه برای یک مهارت به‌صفر نزدیک باشد (میانگین آن‌ها کم‌تر از 0.5 باشد)، آزمون برای آن مهارت دارای ساختار شناختی بالایی است که نشان می‌دهد آزمون برای تشخیص تسلط در آن مهارت به خوبی طراحی شده است. اگر r_{ik}^* بزرگ‌تر از 0.9 باشد، سوال برای مهارت مربوطه، دارای قدرت تشخیص بالایی نیست و درایه ماتریس کیو برای آن سوال در مهارت اختصاص داده شده به آن باید از یک به صفر تغییر یابد. بهین‌گزینه C_i یا بهین‌گزینه توانایی باقیمانده اطلاعاتی در مورد این که تا چه حد مهارت‌های مورد نیاز برای ارایه پاسخ صحیح در ماتریس کیو در نظر گرفته نشده است (ماتریس کیو غیر کامل) ارایه می‌دهد.

هدف از این مطالعه تدوین یک آزمون خواندن و درک مفاهیم بر اساس چارچوب تشخیصی شناختی به منظور فراهم نمودن بازخورد تشخیصی تکوینی برای دانشجویان است. در این راستا به ساخت ماتریس کیو برای اعتباربخشی به آزمون می‌پردازیم تا مهارت‌های لازم برای پاسخدهی صحیح به هر سوال شناسایی شود. به‌همین منظور به دو پرسش تحقیق پاسخ خواهیم داد:

۱. آزمون خواندن و درک مفاهیم تدوین شده جدید چه مهارت‌هایی را می‌سنجد و تکرارپذیری آن‌ها برای هر سوال آزمون چگونه است؟
۲. عملکرد دانشجویان نسبت به مهارت‌های بکار رفته در آزمون خواندن به‌طور کلی و به‌طور انفرادی چگونه است؟

۳. روش پژوهش

این مطالعه از نظر هدف پژوهشی دارای کاربردی با چارچوب مدل سازی تشخیصی شناختی است. در این چارچوب مبتنی بر نظر باک، تاتسوکا و کاستین^۱ (۱۹۹۷) ابتدا فهرست اولیه‌ای از مهارت‌های زیربنای آزمون تهیه شد که از آنها و نتایج بدست آمده از آزمون برای ساخت ماتریس کیو استفاده شد. با استفاده از مدل آماری فیوژن، بهین‌گزینه‌های توانایی هر سوال برای مشخص کردن نقاط قوت و ضعف آزمودنی‌ها و بررسی توان تشخیص آزمون بررسی شد.

۱۹۸۶ دانشجو از دانشگاه تهران در این آزمون شرکت کردند که همگی دانشجوی مقطع کارشناسی درس زبان عمومی از رشته‌های مختلف علوم انسانی، فنی مهندسی و هنر بودند. پاسخ ۱۹۸۶ دانشجو به ۲۰ پرسش این آزمون تدوین شده برای اعتبارسنجی تجربی بکارگرفته شد.

از ۱۳ دانشجو برای انجام پروتکل کلامی تفکر با صدای بلند دعوت شد تا کاربرد آن‌ها از مهارت‌های خواندن بهتر درک شود. هر یک از شرکت کنندگان پیشتر آزمون را داده بودند و از آن‌ها خواسته شد تا فرایند تفکر خود هنگام پاسخدهی به سوالات را شفاهی بازگو کنند. این روند، شامل مهارت‌ها و رویکردهایی بود که در آغاز امر پس از خواندن سوال به ذهن آن‌ها خطور می‌کرد. آزمون بار دیگر به عنوان مرجع در اختیار آنان قرار گرفت و صدای آنان ضبط شد.

از گروه شش نفری، برای داوری درجه‌بندی محتوا دعوت شد. آنان دانشجویهای دکتری دانشگاه تهران در رشته آموزش زبان انگلیسی با سابقه پژوهش در زبانشناسی کاربردی و آموزش خواندن و درک مفاهیم بودند، تا درباره مهارت‌های لازم نگرش‌هایشان را برای دادن پاسخ صحیح به هر پرسش بازتاب دهند. آنها هر سوال را بررسی کردند و از فهرست مهارت‌ها آن‌هایی را انتخاب کردند که لازمه صحیح پاسخ دادن به هر پرسش بود. از پیشنهادهای آنان، در کنار پروتکل کلامی تفکر و نتیجه‌های بدست آمده از مطالعات قبلی برای تدوین ماتریس کیو استفاده شد. از داده‌های کارکرد دانشجویان در آزمون برای درک ویژگی‌های هر سوال و تصحیح ماتریس کیو از مدل‌سازی آماری مدل فیوژن استفاده شد.

۳-۱. مهارت‌های خواندن زبان دوم

فهرست مهارت‌های خواندن (نشان داده شده در جدول ۲) بر اساس مطالعات قبلی انجام شده (ینگ، ۲۰۰۹؛ کوهن و آپتون، ۲۰۰۶؛ فرنسیس و همکاران، ۲۰۰۶؛ برچ، ۲۰۰۲؛ فلتچر، ۲۰۰۶؛ روپ و همکاران، ۲۰۰۶)، نظریه‌های داوری درباره درجه‌بندی محتوا و پروتکل کلامی تفکر با صدای بلند دانشجویها به دست آمده است. فهرست شامل آن مهارت‌های استنباطی است که در فرایند خواندن بکار رفته است.

1. Jang

2. Cohen and Upton

جدول ۲. مهارت‌های خواندن زبان دوم

مهارت	
۱A	تعیین معنابراساس محتوا
۲A	تعیین معنابدون توجه به محتوا
۳A	درک اطلاعات آشکاردرمتن
۴A	درک اطلاعات ضمنی درمتن
۵A	خواندن سطحی
۶A	خلاصه کردن
۷A	ارجاع دادن
۸A	بکارگیری دانش قبلی
۹A	استنباط ایده اصلی یاهدف نویسنده

۲-۳. روش کار

این مطالعه در سه مرحله انجام شد: (۱) تدوین آزمون خواندن و درک مفاهیم بر اساس چارچوب سنجش شناختی؛ (۲) ساخت و اعتباربخشی به ماتریس کیو مهارت‌های خواندن؛ (۳) تجزیه و تحلیل آماری داده‌ها. در اولین گام، مطالعه گسترده‌ای از پژوهش‌های قبلی در زمینه سنجش تشخیصی شناختی و تدوین آزمون با چارچوب تشخیصی شناختی انجام شد. از بررسی ادبیات در این زمینه، استنباط اولیه از ویژگی‌های آزمون بر اساس طراحی مبتنی بر شواهد ارائه شده توسط میسلوی (۱۹۹۶) برای تدوین ۲۰ سوال آزمون بکارگرفته شد. آزمون تدوین شده برای این منظور با بکارگیری چارچوب تشخیصی شناختی انجام پذیرفت، شامل سه متن خواندن با موضوعاتی از قبیل علوم طبیعی، مهندسی، و علوم انسانی.

این آزمون در کلاس‌های زبان عمومی دانشگاه تهران با تعداد کل ۱۹۸۶ شرکت کننده به صورت مقطعی برگزار شد. داده‌های این آزمون برای بررسی آماری بکار گرفته شد. مرحله بعدی ساخت ماتریس کیو براساس مهارت‌های زبان دوم بود. برای ساخت ماتریس کیو، نخست فهرست مهارت‌های زبان دوم، با بهره‌گیری از پژوهش‌های پیشین، پروتکل کلامی تفکر با صدای بلند، و داوری درجه‌بندی محتوا مشخص شد. مرحله نهایی شامل اعتبارسنجی

تجربی ماتریس کیو با تجزیه و بررسی مدل فیوژن است. داده‌های آزمون خواندن در کنارسازه ماتریس کیو با بکارگیری نرم افزار آرپیجیو^۱ و اجرای مدل فیوژن انجام پذیرفت.

۳-۳. واکاوی داده ها

در فرایند تدوین آزمون و ساخت ماتریس کیو، واکاوی کیفی و کمی انجام شد. ابتدا، بررسی کیفی برای شناسایی مهارت‌های خواندن در پرسش‌های آزمون انجام شد. به‌همین منظور طبقه‌بندی گوناگون از مهارت‌ها و راهبردهای خواندن و درک مفاهیم در مطالعات قبلی بررسی شد. سپس پروتکل‌های تفکر کلامی با صدای بلند به‌طور کیفی بررسی شد تا به درک بهتری از فرایند شناختی و مهارت‌های بکارگرفته، توسط دانشجوها دست یابیم و بتوانیم مهارت‌های اولیه را شناسایی کنیم. از داوری درجه‌بندی محتوا نیز استفاده شد تا میزان بکارگیری این مهارت‌ها برای پاسخ صحیح به پرسش‌های آزمون بررسی شود.

اولین گام در تجزیه و بررسی با مدل فیوژن، بررسی همگرایی زنجیره مارکوف مانتی کارلو برای تایید مقدار پایدار بهین‌گزینه‌های مدل است (روسو و همکاران، ۲۰۰۷). نرم افزار آرپیجیو از رویکرد بیزی با الگوریتم زنجیره مارکوف مانتی کارلو استفاده می‌کند. همگرایی زنجیره مارکوف مانتی کارلو اصولاً با بررسی چشمی نمودارهای زنجیره‌ای و نمودارهای چگالی به دست می‌آید. درباره مدل فیوژن، حاصل زنجیره مارکوف مانتی کارلو از مقادیر به دست آمده برای ارزیابی تمام بهین‌گزینه‌ها بکارگرفته می‌شود (لی و سوان، ۲۰۱۳). در بین بهین‌گزینه‌های مختلف، ابتدا همگرایی احتمال تسلط هر مهارت (p_k) آزمون شوندگان به‌طور کلی بررسی شد. سپس، سه بهین‌گزینه هر سوال که نشان دهنده دشواری سوال^۲ (π_i^*)، جریمه یا تشخیص^۳ (r_{ik}^*) و عبارت کامل^۴ (c_i) برآورد شده است. در این مرحله از یک الگوریتم کاهش گام به گام در چارچوب رهنمودهای ارائه شده توسط طراحان مدل (هارتز، ۲۰۰۲؛ روسو و همکاران، ۲۰۰۷) برای اصلاح مدل به‌منظور به‌بیشینه رساندن ظرفیت تشخیصی پرسش‌ها استفاده شده است.

پس از اطمینان از همگرایی مدل و بهینه‌گزینه‌های مناسب سوال به‌بررسی برآزش مدل^۵ با

1. Arpeggio
2. Item difficulty
3. Item discrimination
4. Completeness index
5. model fit

داده‌ها پرداخته می‌شود. برای ارزیابی برازش مدل با داده‌ها روایی درون^۱ یا همسانی درونی (اندازه‌گیری تفاوت در رفتار مشاهده شده آزمودنی‌هایی که در طبقه‌های مختلف قرار دارند) شامل آمار تسلط در سوال^۲ و محاسبه قدر مطلق تفاوت بین درجه دشواری مشاهده شده و درجه دشواری برآورد شده سوال‌ها و مقایسه احتمال تراکمی مشاهده شده و برآورد شده آزمودنی‌ها بررسی می‌شود (هارتز و روسو، ۲۰۰۵).

۳-۴. نقاط قوت و ضعف آزمودنی‌ها

به منظور شناسایی نقاط قوت و ضعف داوطلبان به ملاک طبقه‌بندی نیاز است. ملاک طبقه‌بندی افراد مسلط یا غیرمسلط را به روش‌های مختلفی می‌توان تعیین کرد. در نظام آرپیجیو، آزمودنی‌ها به روش‌های مختلفی طبقه‌بندی می‌شوند. یکی از این روش‌ها تقسیم آزمودنی‌ها به دو دسته است: افراد مسلط که احتمال پسین تسلط آن‌ها، بیش‌تر از ۰/۵ است یا افراد غیر مسلط که احتمال پسین تسلط آن‌ها کم‌تر از ۰/۵ است. این ملاک، همیشه مناسب نیست، زیرا به طور معمول احتمال‌های پسین تسلط نزدیک به ۰/۵، آگاهی‌بخش نیستند (هارتز و روسو، ۲۰۰۸). ملاک‌های طبقه‌بندی دیگری نیز با توجه به هدف سنجش می‌توان در نظر گرفت که یکی از آن‌ها، طبقه‌بندی آزمودنی‌ها با احتمال پسین تسلط کم‌تر از ۰/۴ به عنوان غیرمسلط، بیش‌تر از ۰/۶ به عنوان مسلط و بین ۰/۴ و ۰/۶ در ناحیه خنثی است. قرار گرفتن در ناحیه خنثی یعنی این آزمودنی‌ها را نمی‌توان در طبقه مسلط‌ها یا غیر مسلط‌ها قرار داد، زیرا اطلاعات کافی برای طبقه‌بندی آن‌ها در داده‌ها وجود ندارد. این شیوه اگر چه باعث می‌شود که بعضی از آزمودنی‌ها در طبقه مشخصی قرار نگیرند ولی طبقه‌بندی‌ها با این نقطه برش پایاتر است (روسو و همکاران، ۲۰۰۷). مقادیر ۰/۴ و ۰/۶ به صورت قراردادی تعیین شده است و می‌توان مقادیر دیگر برای مثال ۰/۳ و ۰/۸ نیز تعیین کرد. انتخاب این مقادیر برای طبقه‌بندی آزمودنی‌ها به جای نقطه ۰/۵ منجر به دقت بیش‌تر طبقه‌بندی می‌شود که مهم‌تر از طبقه‌بندی همه افراد است.

همچنین، اعتبار مدل فیوژن با به دست آوردن نرخ طبقه‌بندی درست^۳ (CCR) بررسی شد، که در واقع ثبات طبقه‌بندی آزمون شوندگان، به گروه مسلط و غیر مسلط مهارت هاست. در

1. Internal validity

2. ImStats

3. Correct Classification Rate

گام آخر، نقاط قوت و ضعف آزمون شوندگان در توانایی خواندن زبان دوم در سطح مهارت از طریق احتمال تسلط برای هر مهارت (P_k) بررسی شد.

۳-۵. ساخت و ماتریس کیو

نتیجه‌های به‌دست آمده از پروتکل کلامی تفکر با صدای بلند و داوری طبقه بندی شده محتوا، برای تدوین فهرست مهارت‌های خواندن بررسی شد. سپس این فهرست مهارت‌ها برای تدوین ماتریس کیو اولیه بکار گرفته شد. همان طور که در جدول زیر مشاهده می‌شود، ردیف‌ها نشان دهنده ۲۰ سوال آزمون و ستون‌ها نشان دهنده ۹ مهارت خواندنند. به دلیل طبیعت پیچیده خواندن و درک مفاهیم، همانطور که در این مطالعه نشان داده شد، پاسخ دادن به سوالات، نیازمند تعداد زیادی مهارت خواندن است (اورکهارت و ویر، ۱۹۹۸؛ آلدرسون، ۲۰۰۰).

جدول ۳. ماتریس کیو مهارت‌ها

I/A	A1	A2	A3	A4	A5	A6	A7	A8	A9
1	1	0	1	0	1	0	1	0	0
2	1	0	0	0	1	0	1	1	0
3	0	0	1	0	1	0	1	0	1
4	1	0	0	1	1	0	0	0	1
5	1	0	1	0	1	1	0	0	0
6	0	0	1	0	1	1	0	0	1
7	1	0	1	0	1	0	1	0	0
8	1	0	1	0	1	0	0	1	0
9	1	0	1	0	1	0	1	0	0
10	1	0	1	0	1	0	0	1	0
11	1	1	1	0	1	0	1	0	0
12	1	0	1	0	1	0	1	1	0
13	1	1	1	0	1	0	0	0	0
14	1	0	1	1	1	0	1	0	1
15	1	0	1	0	1	0	1	0	1
16	1	0	1	0	1	0	0	1	0
17	1	0	1	1	1	0	1	1	0
18	1	0	1	1	1	1	1	0	1
19	0	0	1	0	1	0	1	0	0
20	1	0	0	1	1	1	1	0	1

۴. تجزیه و بررسی بامدل فیوژن

در این پژوهش میانگین π به دست آمده برابر با ۰,۸۳۳ و نشان دهنده آن است که مهارت‌های شناسایی شده در پرسش‌های آزمون مناسب و معقولند. نتایج نشان می‌دهد که سه پرسش (پرسش‌های ۵، ۲، و ۱۳) دارای مقادیر π^* پایین‌تر از ۰,۶ اند و به معنای آن است که آنها بسیار دشوار بوده و نیاز به اصلاح دارند. از طرف دیگر، پرسش‌های ۲ و ۱۳ با مقادیر π^* ۰,۵۸ و ۰,۵۶ کمتر دچار مشکل بودند تا سوال ۵ با مقدار ۰,۳۸۸ π که نیاز به جایگزینی دارد. تخمین بهین‌گزینه پرسش به دست آمده از مطالعه اخیر در جدول ۴ نشان داده شده است.

جدول ۴. تخمین پارامتر سوال

Item	π^*	$r^* 1$	$r^* 2$	$r^* 3$	$r^* 4$	$r^* 5$	$r^* 6$	$r^* 7$	$r^* 8$	$r^* 9$	c_i
1	0.879	0.895	0	0.827	0	0.568	0	0	0	0	1.97
2	0.583	0.870	0.847	0	0	0	0	0	0	0.678	2.12
3	0.963	0.923	0	0.899	0	0.726	0	0	0	0	2.73
4	0.786	0	0	0.835	0	0.631	0	0	0.877	0	2.43
5	0.376	0	0	0	0.843	0.926	0	0.722	0	0	2.76
6	0.804	0.694	0	0.709	0	0.490	0	0	0	0	2.24
7	0.993	0.860	0	0	0	0.809	0	0.944	0	0	2.61
8	0.950	0.423	0.443	0	0	0	0	0	0	0.811	2.68
9	0.962	0	0.694	0.536	0	0.853	0	0	0	0	2.70
10	0.769	0.531	0.580	0	0	0	0	0	0	0	2.02
11	0.974	0	0	0.594	0	0.759	0	0	0	0	1.60
12	0.789	0.828	0.521	0	0	0	0	0	0.757	0	2.47
13	0.557	0	0	0	0.704	0.883	0	0.730	0	0	2.04
14	0.894	0	0	0.544	0.863	0	0.650	0.822	0	0.786	2.54
15	0.979	0	0	0.812	0	0.713	0	0	0	0	0.87
16	0.940	0.939	0.941	0	0	0	0	0	0.878	0	1.05
17	0.661	0	0	0.887	0	0.821	0	0	0	0	0.65
18	0.934	0	0	0	0.909	0.860	0	0.884	0	0	0.74
19	0.997	0.981	0	0	0	0.938	0.980	0	0	0	1.72
20	0.863	0	0	0	0.738	0	0.804	0.826	0	0.719	0.80

با نگاهی به جدول می‌بینیم که همه r^*s کمتر از ۰,۹ اند، که نشان دهنده پرسش‌هایی است که آزمون‌شوندگان مسلط را از غیر مسلط متمایز می‌کند. مشاهده می‌شود که پرسش‌های ۷، ۱۶، ۱۸ و ۱۹ همه دارای میانگین بالای ۰,۸ هستند که نشان می‌دهد بیشتر آزمون‌شوندگان به

این پرسش‌ها به خوبی پاسخ دادند. بنابراین، می‌توان گفت که این پرسش‌ها به خوبی آزمون شوندگان مسلط را از غیرمسلط نسبت به مهارت‌های سنجیده شده با پرسش‌ها متمایز نکرده است. زمانی که چنین نتایجی به دست می‌آید، می‌توان نتیجه‌گیری کرد که اصلاح این پرسش‌های آزمون بهترین گزینه برای ایجاد آزمون تشخیصی بهتر و سوال‌هایی با قدرت تمایز بالاتر مسلط‌ها از غیرمسلط‌ها است. در نتیجه، پارامتر I^* می‌تواند برای ارزیابی کیفیت سوال از لحاظ قدرت تشخیص مورد استفاده قرار گیرد (لیتون و گیرل، ۲۰۰۷).

همانطور که پیشتر گفته شد، شاخص عبارت کامل (C_i) اطلاعاتی در این باره که تا چه حد مهارت‌های مورد نیاز برای ارایه پاسخ صحیح در ماتریس کیو در نظر گرفته نشده است (ماتریس کیو غیر کامل)، ارایه می‌دهد. اگر مقادیر C_i بزرگ باشد (بیش‌تر از ۲) ماتریس کیو برای آن سوال تقریباً کامل است و پارامتر C_i آن سوال باید حذف شود (یعنی برابر با ۱۰ در نظر گرفته شود). مقادیر ۰ تا ۱،۵ نشان دهنده این است که یا مهارت‌های لازم برای پاسخ به یک سوال در ماتریس کیو ارائه نشده است یا (اگر برای بسیاری از سوالات اتفاق بیفتد) یک مهارت در کل ماتریس کیو ارائه نشده است (وان دویر، دی بلو و یاماموتو، ۲۰۰۶). با نگاهی به جدول بالا می‌بینیم که پرسش‌های ۱ تا ۱۴ دارای شاخص هم‌کنشی ۲ و بالاتر از ۲ اند، در حالی که سوالات ۱۵ تا ۲۰ دارای شاخص پایین ۰،۶۵ تا ۱،۷۲ اند. این نشان دهنده آن است که مهارت‌های انتخاب شده برای پرسش‌های ۱۵ تا ۲۰ به خوبی نتوانسته پاسخ آزمون شوندگان به این پرسش‌ها را توضیح دهد و مهارت‌های دیگری برای این منظور باید مشخص شود. از شاخص هم‌کنشی پنهان در این مطالعه، می‌توان نتیجه گرفت که ماتریس کیو تدوین شده در انتخاب مهارت برای سوال‌های ۱۵ تا ۲۰ با فقدان‌هایی روبرو است و باید اصلاح شود. به طور کلی، شاخص هم‌کنشی پنهان اطلاعات معنادار قابل ملاحظه‌ای در رابطه با کیفیت سوال‌ها فراهم می‌کند.

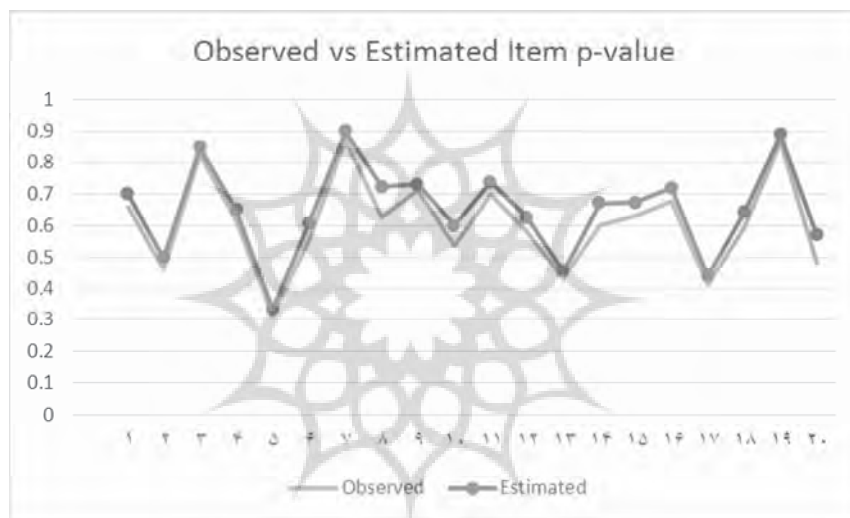
برای بررسی سازگاری طرح فیوژن با داده‌ها، دو نوع سازگاری سنجیده شد: ۱. آمار تسلط سوال^۲ و ۲. آمار فیوژن^۳ که تفاوت بین میزان پرسش‌های صحیح مشاهده شده و میزان پرسش‌های صحیح تخمین‌زده را مقایسه می‌کند. تفاوت اندک این دو مقدار سازگاری خوب مدل را نشان می‌دهد. همان‌طور که در روسو و همکاران (۲۰۰۷) اشاره شد، تفاوت مطلق بین

1. Item completeness index

2. FusionStats

3. ImStats

هر مقدار p مشاهده شده و تخمین زده شده، باید کمتر از مقدار پیشنهادی $0,05$ برای تمام سوالها باشد. البته، در این مورد مشخص این مقدار برای چهار پرسش بالاتر از $0,05$ بود (سوالات ۸، ۱۰، ۱۴ و ۲۰). افزون بر این، قدر مطلق تفاوت بین مقادیر p پایین بود و برابر با $0,04$. شکل ۱ نتیجه‌های به دست آمده از مقادیر p مشاهده شده و تخمین زده شده را نمایش می‌دهد. آمار تسلط پرسش برای مقایسه عملکرد مشاهده شده آزمون شوندگان مسلط و غیر مسلط در سطح سوال بکار گرفته شد.



شکل ۱. مقادیر p مشاهده شده و تخمین زده شده

اعتبار بخشی مدل فیوژن نیز با شاخص نرخ طبقه‌بندی درست بررسی شد. فایل خروجی از Tabulator به نام `csv.classfile` نرخ طبقه‌بندی درست را برای هر مهارت تخمین می‌زند. نرخ طبقه‌بندی درست به ثبات طبقه‌بندی آزمون شوندگان به تسلط یا عدم تسلط نسبت به مهارت‌های همان آزمون که از یک گروه چندین بار گرفته شده است اشاره دارد (روسو و همکاران، ۲۰۰۷). این مقدار بین صفر و یک است. در این سری داده، نرخ طبقه‌بندی درست برابر با $0,826$ ، نشان دهنده اعتبار بخشی بالای مدل فیوژن بود. احتمال تسلط مهارت‌های خواندن درباره عملکرد خواندن زبان دوم آزمون شوندگان نیز

بررسی شد. برای گروه به طور کلی، احتمال کل تسلط برای هر مهارت (pk) بررسی شد. جدول ۵ نشان دهنده مهارت‌های خواندن زبان دوم و مقادیر pk است. با مقادیری حدود ۰,۷۶۹ (تعیین معنا بر اساس محتوا) تا ۰,۶۲۷ (درک اطلاعات ضمنی در متن). این نشان می‌دهد که ۷۶,۹٪ آزمون شوندگان، دارای تسلط در تعیین معنا بر اساس محتوا، و این مهارت آسان‌ترین در نظر گرفته می‌شود. این در حالی است که آزمون شوندگان ۶۲,۷٪ تسلط بر درک اطلاعات ضمنی در متن را به دست آوردند، که آن دشوارترین مهارت در نظر گرفته شد. سطح میانگین کلی احتمال تسلط برای ۹ مهارت خواندن زبان دوم برابر با ۰,۷۱ به دست آمد.

جدول ۵. احتمال تسلط مهارت برای گروه کلی

مهارت خواندن زبان دوم	احتمال تسلط (pk)
تعیین معنا بر اساس محتوا	۰,۷۶۹
تعیین معنا بدون توجه به محتوا	۰,۶۷۶
درک اطلاعات آشکار در متن	۰,۶۶۲
درک اطلاعات ضمنی در متن	۰,۶۲۷
خواندن سطحی	۰,۷۲۲
خلاصه کردن	۰,۶۹۸
ارجاع دادن	۰,۷۵۴
بکارگیری دانش قبلی	۰,۶۶۶
استنباط ایده اصلی یا هدف	۰,۷۳۷

۵. بحث و نتیجه گیری

بنابر اظهارات آلدerson^۱ (۲۰۱۰)، از آنجایی که بیشتر سنجش‌های تشخیصی شناختی بر روی آزمون‌های غیرتشخیصی انجام گرفته است، نیاز برای ساخت آزمون‌های تشخیصی شناختی رو به افزایش است. در همین راستا، هدف اصلی این مطالعه، تلاش برای تدوین

1. Alderson

آزمونی بر اساس چارچوب تشخیصی شناختی به منظور تشخیص نقاط قوت و ضعف زبان آموزان در مهارت خواندن زبان دوم بود، با هدف نهایی فراهم کردن اطلاعات جزئی برای بازخورد تشخیصی تکوینی. بازخورد تشخیصی تکوینی در اهداف آموزشی به معلم و زبان آموز به طور همزمان کمک می‌کند، بطوریکه زبان آموز با استفاده از بازخورد تکوینی، از نقاط ضعف خود آگاه می‌شود و خواهد توانست عملکرد خود را بهبود بخشد. داشتن گزارش‌های جزئی از عملکرد آزمون شوندگان می‌تواند در تسهیل یادگیری از طرف دانشجو مفید واقع شود و همچنین به آمادگی معلم و طرح برنامه درسی کمک کند. با داشتن اطلاعات جزئی درباره نتیجه‌های آزمون، معلم از نقاط شکست آزمون شونده آگاه می‌شود و از این رو می‌تواند در برنامه‌ریزی درسی بیشتر بر آنها تمرکز کند و محتوای درسی مناسب‌تری را فراهم کند. از آنجایی که آزمون خواندن تدوین شده در این پژوهش بر اساس چارچوب تشخیصی شناختی به همراه تجزیه و تحلیل با مدل فیوژن انجام گرفت، پرسش‌های دشواری‌زا شناسایی شدند که باید اصلاح و جایگزین شوند تا آزمون شناختی بهتر و کارآمدتری برای آینده فراهم شود. بنابراین، می‌توان بانک پرسشی برای سوال‌هایی در زمینه پیشرفت تشخیصی شناختی در آزمون فعلی و در پژوهش‌های آینده فراهم کرد.

از میان مهارت‌های چهارگانه زبان انگلیسی، می‌توان از خواندن و درک مفاهیم به عنوان مهارت‌های ضروری برای موفقیت در دنیای علمی نام برد. بنابراین، سنجش صحیح توانایی خواندن زبان آموزان برای اندازه‌گیری یادگیری آنها، بسیار حیاتی است، تا تسهیل‌گر روند بهبود مهارت‌های خواندن آنها باشیم. در همین راستا، دو مولفه اصلی در این پژوهش در نظر گرفته شد. اولی بررسی مهارت‌های زبان دوم در به دست آوردن پاسخ صحیح برای هر سوال بود. فهرست نه مهارت به دست آمده در این پژوهش برای ساخت ماتریس کیو بکار گرفته شد. دوم کارکرد دانشجوها در آزمون به منظور اهداف تشخیصی با استفاده از مدل فیوژن بکار گرفته شد. نتیجه‌های به دست آمده از بررسی مدل فیوژن نشان داد که در آزمون، کدام پرسش‌ها برای تشخیص نقاط قوت و ضعف دانشجو کارآمدند و می‌توانند در روند دادن بازخورد به دانشجو مفید واقع شوند. دانشجو می‌تواند این بازخورد را معیاری برای پژوهش‌های بعدی و یادگیری درس بکار گیرد.

این پژوهش دارای چندین پیامد برای معلم‌ها و کارشناسان است، هم به لحاظ آموزشی و هم به لحاظ نظری. ابتدا و از همه مهمتر پیامد آموزشی برای اهداف سنجش در تلاش برای تدوین آزمون تشخیصی شناختی برای اندازه‌گیری مهارت خواندن زبان دوم در نظر گرفته

می‌شود. ضرورت برای تدوین آزمون‌هایی بر اساس چارچوب‌های تشخیصی شناختی اینک در گستره سنجش و آزمون زبان دوم بسیار شناخته‌تر شده است. این نوع بازخورد تشخیصی که برای معلم‌ها فراهم می‌شود شامل احتمال تسلط مهارت برای کل گروه‌ها، سطح توانمندی‌های متفاوت و تک تک آزمون‌شوندگان است. معلم‌ها می‌توانند برای بهبود و ارتقای مطالب درسی خواندن زبان دوم برای پاسخدهی به نیازهای هر سطح مهارت به داده‌های به‌دست آمده از سنجش تشخیصی شناختی مراجعه کنند.

منابع

- Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. A&C Black.
- Alderson, J. C. (2010). "Cognitive Diagnosis and Q-Matrices in Language Assessment": A Commentary, 96-103.
- Birch, B. M. (2002). *English L2 reading: Getting to the bottom*. Routledge.
- Buck, G., Tatsuoka, 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), 423-266.
- Cohen, A. D., & Upton, T. A. (2006). *Strategies in responding to the new TOEFL reading tasks* (TOEFL Monograph No. MS-33). Princeton, NJ: ETS.
- de la Torre, J. (2009). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement*, 33(3), 163-183.
- DiBello, L. V., Stout, W. F., & Roussos, L. (1995). Unified cognitive psychometric assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-390). Hillsdale, NJ: Erlbaum.
- DiBello, L., & Stout, W. (2008). Arpeggio documentation and analyst manual. *Chicago: Applied informative assessment research enterprises (AIARE)—LLC*.
- Embretson, S. E. (1990). Diagnostic testing by measuring learning processes: Psychometric considerations for dynamic testing. In N. Frederiksen, R. Glaser, A. Lesgold, & M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 407-432). Hillsdale, NJ: Erlbaum.
- Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading*, 10(3), 323-330.

- Francis, D.J., Snow, C.E., August, D., Carlson, C.D., Miller, J., & Iglesias, A. (2006). Measures of Reading Comprehension: A Latent Variable Analysis of the Diagnostic Assessment of Reading Comprehension, *Scientific Studies of Reading*, 10(3), 301-322.
- Gorin, J. S. (2006). Test design with cognition in mind. *Educational measurement: Issues and practice*, 25(4), 21-35.
- Hartman, H. J. (2001). Developing students' metacognitive knowledge and skills. In H. J. Hartman (Ed.). *Metacognition in learning and instruction: Theory, Research and Practice* (pp. 33-68). Dordrecht, the Netherlands: Kluwer.
- Hartz, S. M. (2002). A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 63(2-B), 864.
- Hartz, S. M., & Roussos, L. A. (2005). *The Fusion Model for skills diagnosis: Blending theory with practice*. ETS Research Report. Princeton, NJ: Educational Testing Service.
- Huang, T.W., & Wu, P.C. (2013). Classroom-based Cognitive Diagnostic Model for a Teacher-made Fraction- Decimal Test. *Educational Technology & Society*, 16 (3), 347-361.
- Jang, E. E. (2005). A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG-TOEFL (Unpublished doctoral dissertation). University of Illinois at Urbana Champaign. Available from ProQuest Dissertations and Theses database. (AAT 3182288)
- Jang, E. E. (2009). Demystifying a Q-matrix for making diagnostic inferences about L2 reading skills. *Language Assessment Quarterly*, 6(3), 210-238.
- Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The Attribute Hierarchy Method for Cognitive Assessment: A Variation on Tatsuoka's Rule - Space Approach. *Journal of Educational Measurement*, 41(3), 205-237.
- Li, H. (2011). A cognitive diagnostic analysis of the MELAB reading test. *Spaan Fellow*, 17.
- Li, H., & Suen, H. K. (2013). Constructing and Validating a Q-Matrix for Cognitive Diagnostic Analyses of a Reading Test. *Educational Assessment*, 18(1), 1-25.
- Madison, M. J., & Bradshaw, L. P. (2015). The Effects of Q-Matrix Design on Classification Accuracy in the Log-Linear Cognitive Diagnosis Model. *Educational and Psychological Measurement*, 75(3), 491-511.
- Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika*, 59, 439-483.
- Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement*, 33(4), 379-416.

- Mislevy, R. J. (2006). Cognitive psychology and educational assessment. *Educational measurement*, 4, 257-305.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspective*, 1, 3-67.
- Pellegrino, J. W., & Chudowsky, N. (2003). The Foundations of Assessment. *Measurement: Interdisciplinary Research and Perspectives*, 1(2), 103-148.
- Roussos, L. A., DiBello, L. V., Stout, W. F., Hartz, S. M., Henson, R. A., & Templin, J. H. (2007). The fusion model skills diagnostic system. In J. Leighton & M. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 275-318). New York, NY: Cambridge University Press.
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23(4), 441-474.
- Rupp, A. A., Templin, J., & Henson, R. A. (2012). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.
- Snow, R. E., & Lohman, D. F. (1989). *Implications of cognitive psychology for educational measurement*. American Council on Education.
- Stiggins, R., Arter, J., & Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right—using it well*. Dover, NH: Assessment Training Institute.
- Tatsuoka, K. K. (1995). Architecture of knowledge structure and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman, and R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment* (pp. 327-361). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Templin, J. L. (2004). *Generalized linear mixed proficiency models for cognitive diagnosis*. Available from ProQuest Dissertations and Theses database. (AAT3160960)
- Urquhart, S., & Weir, C. J. (1998). *Reading in a second language: Process, product and practice*. New York: Longman.
- Von Davier, M., DiBello, L., & Yamamoto, K. Y. (2006). Reporting test outcomes with models for cognitive diagnosis (ETS Research Rep. No. RR-06-28). *Princeton, NJ: ETS*.