

بررسی گسترده‌گی و دشواری زیرمهارت‌های خواندن و درک مطلب با
استفاده از مدل تشخیصی شناختی جی‌دینا

زهرا جاویدان‌مهر^۱

محمدرضا عنانی سراب^۲

چکیده

مهارت خواندن و درک مطلب یکی از مهمترین مهارت‌های زبان انگلیسی، به‌ویژه در سطح آکادمیک می‌باشد. این مهارت به تنهایی و در کنار مهارت‌های دیگر و از زوایای مختلف در حوزه آموزش و همچنین سنجش مورد بررسی قرار گرفته است. یکی از مولفه‌های مورد تحقیق، در زمینه سنجش آموزشی، زیرمهارت‌های ساختاری مهارت خواندن است. آنچه در پژوهش حاضر بررسی می‌شود نوع، گسترده‌گی و درجه دشواری این زیرمهارت‌ها و همچنین دشواری سوالات با استفاده از مدل تشخیصی شناختی جی‌دینا^۳ است. سنجش تشخیصی شناختی^۴، چارچوب نظری جدیدی در حوزه روانسنجی است که بر خلاف مدل‌های سنتی سنجش، به جای رتبه‌بندی آزمودنی‌ها بر اساس عملکردشان در آزمون، اطلاعاتی در مورد آزمودنی و تسلطش بر زیرمهارت‌های یک حوزه و همچنین خود سوالات آزمون ارائه می‌دهد. به منظور انجام این پژوهش، سوالات بخش خواندن و درک مطلب آزمون ورودی دانشگاه در مقطع کارشناسی ارشد انتخاب شد و زیرمهارت‌های آزمون از طریق سه منبع، مطالعات انجام شده قبلی حوزه، تکنیک فکری و همچنین داوری متخصصین حیطه، استخراج گردید. این زیرمهارت‌ها به همراه پاسخنامه‌های آزمون مورد نظر در بسته جی‌دینا (Ma & de la Torre, 2017) در نرم افزار برنامه نویسی R تجزیه و تحلیل شدند. چهار زیرمهارت از این آزمون استخراج و در ادامه، دشواری و گسترده‌گی این زیرمهارت‌ها و پس از آن دشواری برخی سوالات^۵ به روش سنجش تشخیصی شناختی گزارش شد. در پایان، دلایل احتمالی نتایج به‌دست آمده در مقایسه با دیگر مطالعات این حوزه مورد تحلیل و بررسی قرار گرفت.

واژگان کلیدی: مدل تشخیصی شناختی- زیرمهارت-مدل جی‌دینا-مهارت خواندن و درک مطلب

دوره چهاردهم شماره ۱۹، پاییز و زمستان ۱۳۹۶

۱. دانش آموخته دکتری زبان و ادبیات انگلیسی - دانشگاه شهید بهشتی

bzarrinjooee@yahoo.com

۲. هیات علمی دانشگاه شهید بهشتی

reza_ananisarab@yahoo.co.uk

3. G-DINA model

4. Cognitive Diagnostic Assessment

5. Item Difficulty

مقدمه

به دنبال گسترش مطالعات در حوزه تاثیر سنجش بر روی جامعه، سیستم آموزشی و افراد (Messick, 1989; Shohamy, 1992)، توجه جامعه سنجش تحصیلی به کمبودهای نتایج ارائه شده توسط سنجش به روشهای سنتی، که بر اساس آنها آزمودنی ها بر اساس عملکردشان در آزمون رتبه بندی می شدند، معطوف شد. گزارش نمره به آزمودنی ها به این شیوه در اواسط دهه ۱۹۸۰ میلادی به چالش کشیده شد (Shohamy, 1992). در همین راستا، سنجش تشخیصی شناختی، به عنوان یک شیوه جدید روانسنجی که قادر است بازخورد تشخیصی خرد در خصوص میزان تسلط آزمودنی در زیرمهارت‌های^۱ زیرساختی یک حوزه مشخص بود، معرفی گردید (Lee & Swaki, 2009). این رویکرد، نقطه‌ی تلاقی روانشناسی شناختی و روانسنجی محسوب می‌شود. از این رو، هدف آن ارزیابی دانش آزمودنی‌ها و مهارت‌های پردازش شناختی می‌باشد (DiBello, Roussos, & Stout, 2006; Gierl & Leighton, 2007).

همان‌طور که گفته شد نیاز به سنجش تشخیصی شناختی به دلیل نواقص مدل‌های ارزیابی قبلی در ارائه گزارش پیشرفت تحصیلی به کاربران بوجود آمد. نظریه کلاسیک آزمون^۲ و نظریه سوال-پاسخ^۳ با ارائه یک نمره کلی سطح دانش آزمودنی و رتبه او را بر اساس میزان سنجش تعیین می‌کنند (Choi, Rupp, & Pan, 2012). براساس نظر راپ، تمپلین و هنسون (Rupp, Templin, & Henson, 2010)، گزارش نتیجه سنجش به این شکل، بازخورد تکوینی^۴ برای اهداف تشخیصی کیفی ارائه نمی‌دهد. روش تشخیصی شناختی، در مقابل، می‌تواند سطح مهارت پنهان^۵ آزمودنی را، که نمایانگر دانش زیرساختی وی می‌باشد، نشان دهد. بدین ترتیب نقاط ضعف و قوت یک فرد و به دنبال آن نقایص برنامه آموزشی تبیین می‌شود (صادقی و خضری، ۲۰۱۱، ۲۰۱۱). مدل‌های تشخیصی شناختی^۶ قادر به ارائه اطلاعات دقیق در خصوص ویژگی‌های آزمون و همچنین آزمودنی می‌باشند. هدف عمده این مقاله، گزارش گسترده‌ی مهارت‌های زیرساختی سنجش خواندن و درک مطلب زبان انگلیسی آزمون ورودی کنکور سراسری دانشگاه‌های ایران و همچنین گزارش دشواری این مهارت‌ها و

1. Attributes/Sub-skills

2. Classical Test Theory (CTT)

3. Item Response Theory (IRT)

4. Formative

5. Latent

6. Cognitive Diagnostic Models (CDMs)

دشواری سوالات آزمون با استفاده از مدل تشخیصی شناختی جی دینامی باشد. با استناد به اهداف تحقیق، سه پرسش زیر مطرح گردید:

۱- چه زیرمهارت‌هایی توسط سوالات آزمون خواندن و درک مطلب زبان انگلیسی ارزیابی می‌شوند؟

۲- درجه دشواری این زیرمهارت‌ها چگونه است؟

۳- درجه دشواری سوالات آزمون چگونه است؟

پیشینه تحقیق

مدل‌های تشخیصی شناختی

به دلیل ماهیت تشخیصی این نوع سنجش، به منظور ارائه نمایه زیرمهارت‌های^۱ آزمودنی‌ها، که شامل یک سلسله از زیرمهارت‌ها و میزان احتمال تسلط آزمودنی‌ها بر آن‌ها می‌باشد، مدل‌های تشخیصی شناختی مورد استفاده قرار می‌گیرند. هدف این مدل‌ها، دسته‌بندی آزمودنی‌ها به دو گروه مسلط و غیرمسلط در زیرمهارت‌های موجود در آزمون است. (von Davier, 2005) از زمان معرفی سنجش تشخیصی شناختی، مدل‌های آماری زیادی معرفی شده‌اند که اگرچه روش استفاده از آن‌ها متفاوت است، اما هدف مشترکی را دنبال می‌کنند. مهم‌ترین مدل‌ها به ترتیب توالی زمانی عبارتند از: روش‌شناسی فضا قاعده^۲ (Tatsuoka, 1983)، مدل مولفه کلی و ویژگی نهفته^۳ (Embretson, 1998)، مدل دینا (Junker & Sijtsma, 2001)، مدل فیوژن (Hartz, 2002)، روش مشخصه سلسله مراتبی^۴ (Leighton, Gierl, & Hunka, 2004) و مدل جی دینا (Torre, 2011) می‌باشند.

سنجش تشخیصی شناختی در دیگر مطالعات

تعداد زیادی از مطالعات انجام شده در این حوزه بر بخش نظری و مدل‌های ارزیابی این روش تمرکز داشته‌اند و تعدادی دیگر با استفاده از داده‌های حقیقی، به مطالعه بخش عملی پرداخته‌اند. در بخش نظری، مطالعاتی در خصوص اصول سنجش تشخیصی شناختی و چالش‌های پیش رو صورت گرفته‌اند. وان داوییر (von Davier, 2009)،

1. Sub-skill profiles

2. Rule space methodology

3. General component latent trait model

4. Attribute hierarchical method

راپ و تمپلین (Rupp & Templin, 2008)، و لی و ساواکی (Lee & Sawaki, 2009) نمونه هایی از این دست هستند. دسته ای دیگر از مطالعات، ساخت ماتریس کیو، نحوه اعتبارسنجی آن و قوت آن را مورد بررسی قرار داده اند (e.g., Alderson, 2010; Chiu, 2013; De Carlo, 2012; Li & Suen, 2013; Liu, Xu, & Ying, 2012; Sawaki, Kim, & Gentile, 2009).

در این میان، پاره ای دیگر از محققان از مدل‌های مختلف تحلیلی، جهت ارزیابی آزمون های موجود استفاده کرده اند. به عنوان مثال در زمینه آموزش زبان انگلیسی، مدل فیوژن (Hartz, 2002) در تحلیل بخش خواندن و درک مطلب (e.g., Jang, 2005, 2008) و همچنین در تحلیل آزمون نوشتن (e.g., Kim, 2011; Xie, 2016) به کار گرفته شده است. مدل جی دینا نیز در ارزیابی تشخیصی شناختی مهارت‌های مختلف زبان انگلیسی استفاده شده است (e.g., Basoksou, 2014; Chen & Chen, 2015; Ravand, 2015). این مطالعات، گستردگی زیرمهارت‌های موجود در یک آزمون، درجه تسلط آزمودنی‌ها بر این زیر مهارت‌ها و اعتبار مدل های مورد استفاده در ارزیابی آزمودنی‌ها را گزارش می‌دهند.

با معرفی این مدل در حوزه سنجش، مطالعاتی نیز از بازخوردهای به دست آمده در ارتباط با دیگر حوزه ها سود جستند. آزمون تطبیقی کامپیوتری^۱ یکی از این حوزه هاست که از بازخورد مدل‌های تشخیصی شناختی استفاده کرده است (Cheng, 2008; McGlohen & Chang, 2010). عملکرد افتراقی سوالات آزمون^۲ دسته دیگری از مطالعات مرتبط با این حوزه می باشد. مهمترین مطالعات به ترتیب توسط لی (Li, 2011) و لی و سوئن (Li & Suen, 2013) انجام گرفته اند.

زیرمهارت‌های خواندن و درک مطلب در دیگر مطالعات

در راستای هدف پژوهش حاضر که تشخیص زیرمهارت‌های آزمون خواندن و درک مطلب و گستردگی و دشواری این زیرمهارت‌ها است، زیرمهارت‌های به دست آمده که در تحقیقات قبلی این حوزه بسیار حائز اهمیت می‌باشند به ترتیب بیان می‌شوند. یانگ (Jang, 2009)، که با استفاده از مدل فیوژن اعتبار آزمون تافل آی‌بی‌تی را سنجیده، نه زیرمهارت از این آزمون استخراج کرده است که عبارتند از: واژه وابسته به متن،

1. Computerized Adaptive Testing

2. Differential Item Functioning

واژه مستقل از متن، ارتباط معنایی و نحوی، اطلاعات صریح درون متن، اطلاعات ضمنی درون متن، استنتاج، منفی‌سازی، خلاصه کردن، ترسیم نظرات متضاد در نقشه ذهنی. این زیرمهارت‌ها در ماتریس کیو و سپس در نرم‌افزار آرچیو^۱ مورد استفاده قرار گرفته‌اند.

سواکی و همکاران (Sawaki et al., 2009) نیز آزمون تافل آی‌بی‌تی را مورد تحقیق قرار داده‌اند، اما در مقایسه با زیرمهارت‌های استخراجی یانگ، لیست کوتاه تری شامل چهار زیرمهارت تعریف کرده‌اند که عبارتند از: ۱. درک معنی واژه، ۲. درک اطلاعات مشخص، ۳. ارتباط اطلاعات با یکدیگر، ۴. ترکیب اطلاعات. آن‌ها در این تحقیق علاوه بر مهارت خواندن، مهارت شنیداری را نیز با همان مدل فیوژن مورد تجزیه و تحلیل قرار داده‌اند.

در زمینه تافل، افراد دیگری مهارت خواندن را تحلیل کرده‌اند (Gao, 2006; Kim, 2011). در زمینه آزمون‌های ورودی دانشگاه در ایران، تحقیقات زیادی صورت نگرفته است. یکی از مطالعاتی که در این زمینه انجام شده است، آزمون ورودی دانشگاه در رشته زبان انگلیسی توسط راوند (Ravand, 2015) است. پنج زیرمهارت برای سنجش خواندن و درک مطلب این آزمون تعریف شده است که عبارتند از: ۱. خواندن برای تشخیص جزئیات، ۲. خواندن برای استنتاج، ۳. خواندن برای ایده اصلی، ۴. نحو، ۵. واژگان. راوند در این پژوهش، برای ارائه اطلاعات تشخیصی در خصوص سوالات آزمون و آزمودنی‌ها از مدل تحلیلی جی‌دینا استفاده کرده است.

ماتریس کیو^۲

بخش عمده سنجش تشخیصی شناختی، مجموعه زیرمهارت‌هایی است که به وسیله هر سوال ارزیابی می‌شود. رابطه سوال-زیرمهارت در ماتریسی به نام ماتریس کیو ارائه می‌شود (Li & Suen, 2013). به بیان دیگر، تخصیص زیرمهارت‌ها به هر سوال یک آزمون، با استفاده از اعداد صفر و یک در ماتریس کیو انجام می‌شود (Tatsuo-ka, 1983). در این ماتریس، عدد یک نشان‌دهنده حضور زیرمهارت و صفر نشانه عدم حضور زیرمهارت است. پاسخ درست به یک سوال ممکن است به بیش از یک زیرمهارت بستگی داشته باشد.

استخراج زیرمهارت‌ها

زیرمهارت‌های هر حوزه با شیوه‌های مختلف استخراج می‌شوند. در هر مطالعه از یک یا چند روش متفاوت استفاده می‌شود تا دقت و اعتبار ماتریس کیو را، که براساس آن زیرمهارت‌ها مشخص می‌شوند، افزایش دهد (Choi et al., 2012). مهم‌ترین رویکردهایی که به کرات مورد استفاده قرار گرفته‌اند (Rupp & Templin, 2008) عبارتند از:

۱. جدول مشخصات آزمون^۱:

یکی از منابع برای تعریف زیرمهارت‌های یک آزمون، الگوهایی هستند که در مشخصات آزمون ذکر شده‌اند. نکته قابل توجه این است که این الگوها در مقیاس دانه درشت^۲ تعریف شده‌اند و شیوه‌های سنجش تشخیصی به الگوها در مقیاس دانه ریز^۳ نیاز دارند (Leighton & Gierl, 2007).

۲. تکنیک فکرگویی^۴:

این روش به منظور استخراج اطلاعات مربوط به فرآیند ذهنی احتمالی آزمودنی در زمان پاسخ دادن به سوالات استفاده می‌شود. این روش در مقایسه با روش‌های دیگر، اعتبار و دقت بالاتری دارد، چرا که نشان‌دهنده فرآیند های ذهنی در حین عمل است (Ericson & Simon, 1998).

۳. داوری متخصصین حیطه:

این رویکرد بر کسب نظرات متخصصین حیطه مورد نظر در مورد حضور یا عدم حضور زیرمهارت‌ها در آزمون براساس تجارب و مطالعات قبلی آن‌ها استوار است. برای بهینه کردن این روش، راپ و همکاران (Rupp et al., 2010) پیشنهاد می‌دهند که بهتر است متخصصین در حوزه مربوطه و همچنین در حوزه سنجش صاحب‌نظر باشند (Rupp et al., 2010; Sawaki, Kim, & Gentile, 2009).

1. Test specification

2. Coarse-grain

3. Fine-grain

4. Think-aloud

۴. پیشینه تحقیق:

یکی دیگر از منابع مفید برای تشخیص زیرمهارت‌های یک آزمون، مطالعات انجام‌شده قبلی در آن حوزه است. برای مثال در حوزه خواندن و درک مطلب، مطالعاتی که در این زمینه انجام‌شده است می‌تواند قالب اولیه تشخیص زیرمهارت‌های مورد نیاز باشد.

روش تحقیق

شرکت‌کنندگان:

پاسخ‌نامه آزمون ۲۵۰۰ آزمودنی کنکور ورودی کارشناسی ارشد رشته علوم انسانی سال ۱۳۹۲، که هم شامل شرکت‌کنندگان زن و هم شرکت‌کنندگان مرد می‌شود، به روش تصادفی انتخاب شدند.

آزمون:

بخش خواندن و درک مطلب زبان عمومی آزمون کارشناسی ارشد (۱۳۹۲)، شامل سه متن در موضوعات مختلف، به همراه بیست سوال چهار گزینه‌ای مربوط به آن‌ها مورد بررسی و تحلیل قرار گرفت.

روند اجرای تحقیق:

استفاده از مدل‌های تشخیصی شناختی در سه مرحله انجام می‌گیرد و پس از آن نتایج اعلام می‌شود:

مرحله اول: استخراج زیرمهارت‌های آزمون

برای استخراج مهارت‌های زیرساختی آزمون، از سه منبع اصلی شامل مطالعات انجام‌شده قبلی، نظر متخصصین حیطه و تکنیک فکرخوانی، استفاده گردید. مطالعات انجام‌شده قبلی شامل مطالعه یانگ (Jang, 2009)، جائو (Gao, 2006) ساواکی و همکاران (Sawaki et al., 2009) و راوند (Ravand, 2015) می‌باشند. سه دانشجوی دکتری تخصصی آموزش زبان که در حوزه‌های مختلف آزمون‌سازی و روانسنجی مشغول تحقیق بودند، به عنوان متخصصین حیطه انتخاب شدند. ده دانشجوی کارشناسی ارشد رشته‌های مختلف علوم انسانی نیز در بخش فکرگویی شرکت کردند.

با استفاده از ترکیب این سه روش، چهار زیرمهارت زیرساختی برای آزمون مورد نظر، تعریف گردید (Javidanmehr & Anani Sarab, in press). این مهارت‌ها عبارتند از درک معنی واژه، استنتاج، درک اطلاعات بیان شده در متن و ترکیب اجزای متن.

مرحله دوم: تولید و سنجش روایی ماتریس کیو

در این مرحله، ماتریس کیو، که رابطه سوال-مهارت را ترسیم می‌کند (Tatsuoka, 1983)، توسط متخصصین حیطه ساخته شد و روایی آن در نرم‌افزار برنامه‌نویسی R مورد ارزیابی قرار گرفت. نسخه نهایی ماتریس کیو با تایید متخصصین حیطه برای تحلیل مورد استفاده قرار گرفت (جدول ۱).

جدول شماره ۱

ماتریس کیو اولیه، پیشنهادی توسط نرم افزار و نهایی

ترکیب متنی				ترکیب متنی				ترکیب متنی						
لستاج لغت				لستاج لغت				لستاج لغت						
41	0	1	1	0	41	0	1	0*	0	41	0	1	0	0
42	0	0	1	1	42	0	0	1	1	42	0	0	1	1
43	1	1	0	1	43	1	1	0	1	43	1	1	0	1
44	1	1	0	1	44	1	1	0	1	44	1	1	0	1
45	1	0	1	1	45	1	0	1	1	45	1	0	1	1
46	0	0	1	1	46	0	0	1	1	46	0	0	1	1
47	0	0	1	0	47	1*	1*	1	0	47	1	1	1	0
48	1	1	0	1	48	1	1	0	1	48	1	1	0	1
49	0	0	1	0	49	1*	1*	1	1*	49	0	1	1	1
50	1	0	1	0	50	1	1*	1	1*	50	1	0	1	1
51	0	1	0	1	51	0	1	0	1	51	0	1	0	1
52	1	1	0	1	52	1	1	0	1	52	1	1	0	1
53	0	0	1	1	53	0	0	1	1	53	0	0	1	1
54	1	0	1	0	54	1	0	1	0	54	1	0	1	0
55	0	0	1	0	55	1*	0	1	1*	55	0	0	1	0
56	0	0	1	0	56	0	0	1	0	56	1	0	1	0
57	1	0	0	1	57	1	0	0	1	57	1	0	0	1
58	0	0	1	0	58	0	0	1	0	58	0	0	1	0
59	0	0	1	0	59	1*	0	1	1*	59	0	0	1	0
60	1	0	1	0	60	1	0	1	0	60	1	0	1	0

ماتریس نهایی ماتریس پیشنهادی توسط نرم افزار ماتریس اولیه
آرایه‌های تغییر یافته مشخص شده اند

مرحله سوم: تحلیل داده

داده‌های تحقیق در بسته جی‌دینا در نرم‌افزار برنامه‌نویسی R (Ma & de la Torre,

(2017) تجزیه و در راستای اهداف تحقیق، خروجی‌های مربوط به درجه دشواری زیرمهارت‌ها و سوالات و همچنین گستردگی زیرمهارت‌ها در میان آزمودنی‌ها گزارش گردیدند، که در مرحله بعد مورد بحث قرار می‌گیرند.

نتایج، بحث و بررسی

نتیجه محاسبه گستردگی مهارت‌ها در جدول شماره ۲ نشان داده شده است. نمایه‌ها نشانگر درجه تسلط نمونه آماری در هر زیرمهارت و همچنین درجه دشواری این زیرمهارت‌ها می‌باشند.

جدول شماره ۲

گستردگی زیرمهارت‌های آزمون

مهارت‌های شناختی درک مطلب	گستردگی مهارت‌ها
درک معنی واژه	۰,۸۲
استنتاج	۰,۶۹
درک اطلاعات بیان شده در متن	۰,۶۶
ترکیب اجزای متن	۰,۸۰

گستردگی و دشواری زیرمهارت‌ها:

چهار زیرمهارت برای خواندن و درک مطلب استخراج و توسط مدل آماری تایید شدند. نتایج تحلیل آماری نشان می‌دهد که دانش واژگانی، آسان‌ترین زیرمهارت برای آزمودنی‌ها می‌باشد، زیرا ۸۲ درصد از آزمودنی‌ها بر این مهارت مسلط می‌باشند. درک اطلاعات بیان شده در متن، مهارت نسبتاً دشواری تخمین زده می‌شود، زیرا تعداد کمتری از آزمودنی‌ها بر آن تسلط دارند. گستردگی و درجه دشواری هر یک از زیرمهارت‌های آزمون به ترتیب گستردگی بررسی می‌شوند.

۱. **درک معنی واژه:** دانش واژگانی در تمام متون این آزمون مشاهده شد و تخمین زده می‌شود ۸۲ درصد از آزمودنی‌ها بر این مهارت مسلط می‌باشند. نتیجه به دست آمده در راستای نظریه سلسله مراتبی خواندن و درک مطلب^۱ می‌باشد (Alderson & Lukmani, 1989). نتایج مطالعات این محققان حاکی از این است که میان مولفه‌های

1. Hierarchical theories of reading

خواندن و درک مطلب، رابطه‌پلکانی وجود دارد. براساس این نظریه، سطوح بالای درک مطلب، نیازمند تسلط بر مهارت‌های سطوح پایین تر خواندن و درک مطلب است. در این مطالعات، درک تحت الفظی^۱ متن که با تسلط بر واژگان و دستورممكن می شود، شامل فرآیند شناختی سطوح پایین خواندن و درک مطلب است. درک استنتاجی^۲ که شامل تجزیه و ارزیابی متن می‌شود، نیاز به فرآیند‌های شناختی سطوح بالای خواندن دارد، بنابراین، تسلط بر این مهارت‌ها برای آزمودنی‌ها دشوارتر خواهد بود. از نظر محققان حوزه خواندن و درک مطلب، واژگان جایگاه بسیار ویژه‌ای دارد. حتی افرادی که با تقسیم‌پذیری مولفه‌های خواندن و درک مطلب مخالف هستند و آن را یک توانایی کلی و یکپارچه معرفی می‌کنند، اظهار می‌دارند که اهمیت واژگان به حدی است که می‌توان آن را یک مولفه‌ی مستقل دانست. این افراد طرفدار رویکرد دو وجهی^۳ مهارت خواندن می‌باشند که شامل واژگان و درک مطلب استنتاجی است (Davey, 1988; Rost, 1993; Weir, 1993). این میزان توجه به واژگان، قطعاً تأثیرات چشم‌گیری بر سنجش خواندن و درک مطلب از دیدگاه کاربران و طراحان آزمون داشته است.

۲. ترکیب اجزای متن: در مقایسه با زیرمهارت واژگان، ترکیب اجزای متن در مرحله دوم گستردگی قرار دارد. تخمین زده می‌شود در حدود ۸۰ درصد آزمودنی‌ها بر این مهارت مسلط می‌باشند. این‌گونه سوالات که بالاتر از سطح یک جمله مجزا تعریف می‌شوند، رابطه آشکار و پنهان میان اجزای متن، انسجام و ساختار لفظی آن را می‌سنجند. براساس نظریه‌ی سلسله مراتبی خواندن و دشواری این زیرمهارت، انتظار می‌رفت تعداد کمتری از افراد بر این زیرمهارت تسلط داشته باشند. دلیل عدم مشاهده این دشواری ممکن است مربوط به نحوه آموزش مهارت خواندن در کلاس‌های زبان دانشگاهی باشد (Javidanmehr & Anani Sarab, in press).

۳. استنتاج کردن: استنتاج کردن از عنوان متن، دانش زمینه‌ای آزمودنی، مقصود اصلی نویسنده و به بیان گری (Gray, 1960)، خواندن میان خطوط^۴، از زیرمهارت‌های موجود در این آزمون می‌باشد که در درجه سوم گستردگی قرار دارد و حدود ۶۹ درصد آزمودنی‌ها بر آن مسلط هستند. اگرچه این زیرمهارت دشوار نیست، اما در گروه زیرمهارت‌های بسیار آسان نیز جای نمی‌گیرد. بخش عمده درک مطلب به قرار دادن

1. Literal understanding

2. Inferential understanding

3. Bi-component approach

4. Reading between lines

متن در دنیای واقعی و توجه به مولفه‌های برون‌متنی بستگی دارد. پیشینه موضوعی، آزمودنی را در موقعیت جهان واقعی و جایی فراتر از دانش واژگانی قرار می‌دهد. موضوع‌های مرتبط با بعضی از رشته‌های خاص به نظر می‌رسد که به آزمودنی‌های آن رشته، کمک بیشتری می‌کند.

موضوعات مربوط به استنتاج در درک مطلب به وفور در نوشته‌های این حوزه به چشم می‌خورند و دسته‌بندی‌های متفاوتی برای آن ارائه شده است. زینگر، گریسرو تراباسو (Singer, Graesser, & Trabasso, 1994) استنتاج درون‌متنی^۱ و استنتاج برون‌متنی^۲ را معرفی می‌کنند. در طبقه‌بندی نسبت‌امشابه، آزدو و کراملی (Azevedo & Cromley, 2004) از استنتاج متن به متن^۲ و دانش قبلی به متن^۳ نام می‌برند. از سوی دیگر پرسلی و افلرباخ (Pressley & Afflerbach, 1995) استنتاج را به خودآگاه و ناخودآگاه تقسیم می‌کنند. به نظر می‌رسد که استنتاج در این آزمون بیشتر مربوط به استفاده از دانش پیشینه تحصیلی باشد. برای بررسی انواع استنتاج نیاز به مطالعات کیفی است تا درجه‌اهمیت و ترتیب استفاده آن‌ها مشخص شود.

۴. درک اطلاعات بیان شده در متن :

۶۶ درصد از آزمودنی‌ها بر این زیرمهارت مسلط هستند. گسترده‌گی آن در مقایسه با دیگر زیرمهارت‌ها در مرحله آخر قرار دارد. در سنجش این زیرمهارت، آزمودنی می‌باید اطلاعات بیان شده در متن را با اطلاعات همسان در متن مطابقت دهد و از این رو، طبق رویکرد سلسله‌مراتبی خواندن، از فرآیند پردازش سطوح پایین استفاده کند. گری (Gray, 1960) این نوع خواندن را، خواندن خطوط (در مقابل خواندن بین خطوط) می‌نامد.

دشواری سوالات

دشواری سوالات در مدل‌های تشخیصی شناختی از جمله مدل جی‌دینا، براساس احتمال موفقیت در پاسخ دادن به سوال^۴ تخمین زده می‌شود. اگر احتمال موفقیت آزمودنی‌هایی

1. Text-connecting inference

2. Extra-textual inference

3. Text-to-text

4. Background-to-text

5. Item's probability of success

که بر تمام زیرمهارت‌های موجود در یک سوال مسلط بوده اند کمتر از ۰/۴۰ باشد، آن سوال دشوار خوانده می‌شود. در این آزمون، سوالات ۶۰-۵۲، یعنی هشت سوال آخر، در گروه سوالات دشوار طبقه بندی شدند. این بدین معنی است که احتمال پاسخگویی صحیح افراد مسلط بر تمام زیرمهارت‌های هر سوال، کمتر از ۰/۴۰ می‌باشد. الدرسون (Alderson, 2005)، دشواری آزمون خواندن و درک مطلب را به دو موضوع نسبت می‌دهد: دشواری متن و دشواری سوالات. او برای هر یک، متغیرهایی را در نظر می‌گیرد که در مطالعه حاضر، سه متغیر در تحلیل دشواری سوالات مورد بررسی قرار گرفت که عبارتند از: محدودیت زمان^۱، دانش پیشینه و دشواری زیرمهارت‌های سوال.

۱. محدودیت زمان:

تحلیل آماری این سوالات نشان می‌دهد که هشت سوال آخر، سوالات دشواری برای آزمودنی‌ها هستند. یک دلیل منطقی برای دشواری این دسته از سوالات محدودیت زمان است. این احتمال وجود دارد که آزمودنی‌ها قادر به خواندن و تحلیل متن نبوده‌اند. آزمون‌های چهار گزینه‌ای را نه می‌توان آزمون صرفاً سرعتی و نه صرفاً قدرتی به حساب آورد. لی و چن (Lee & Chen, 2011) این آزمون‌ها را آزمون‌های با محدودیت زمان می‌نامند و معتقدند که توانایی این نوع آزمون‌ها در تشخیص درجه دشواری سوالات در ابتدا و انتهای آزمون متفاوت است.

۲. دانش پیشینه:

دانش پیشینه و آشنایی با موضوع، دومین دلیل دشواری برخی سوالات آزمون است. گفته می‌شود که دانش پیشینه آزمودنی در یک موضوع خاص، عامل موفقیت وی در درک بهتری از مطلب می‌باشد (Butcher & Kintsch, 2003). بنابراین ارتباط میان متن و رشته تحصیلی آزمودنی اهمیت قابل توجهی دارد. دو متن آخر آزمون، درباره موضوعات زمین‌شناسی و شیمی هستند که برای آزمودنی‌های رشته علوم انسانی چندان آشنا به نظر نمی‌رسند. این نتیجه در راستای نظریه طرحواره ذهنی^۲ است که براساس آن، افراد دانش زمینه‌ای خود را به این وسیله دسته‌بندی می‌کنند و در نتیجه به خاطر سپاری آن‌ها راحت‌تر می‌شود (Anderson & Pearson, 1984; Nar-

1. Time intensity

2. Schema theory

۲۰۰۱). هرگاه خواننده با یک مفهوم آشنا روبه‌رو شود، قادر است به این سیستم اطلاعات دسته‌بندی شده به راحتی دسترسی یابد و به واسطه آن می‌تواند مفهوم و معنی متن را درک کند. هیل (Hale, 1988) نیز در بررسی آزمون تافل به همین نتیجه دست یافته است به این صورت که آزمودنی‌های رشته‌های علوم انسانی/علوم اجتماعی و علوم پایه هرکدام در پاسخ دادن به سوالات متن‌های مربوط به رشته خود بهتر عمل کرده‌اند.

۳. دشواری زیرمهارت‌های سوال:

دشواری سوالات همچنین می‌تواند با توجه به دشواری زیرمهارت‌های مورد نیاز در هر سوال توجیه شود. با بررسی ماتریس کیو (جدول ۱)، چندین الگو قابل دریافت است. سوالات ۵۸ و ۵۹، تنها به زیرمهارت "درک اطلاعات بیان‌شده در متن" نیاز دارند. شایان ذکر است که زیرمهارت "استنتاج" که یک زیرمهارت سطح بالا محسوب می‌شود، در این سوالات وجود ندارد. از این رو دشواری سوالات را نمی‌توان منسوب به زیرمهارت "استنتاج" دانست. این نتیجه در تقابل با نتایج مطالعات دیگر است. برای مثال دیوی (Davey, 1988) گزارش کرده است که سوالاتی با زیرمهارت "استنتاج" نسبتاً دشوارتر از سوالات با زیرمهارت "درک اطلاعات بیان‌شده در متن" هستند.

نتیجه‌گیری

در این پژوهش، زیرمهارت‌های استخراجی از آزمون خواندن و درک مطلب با استفاده از مدل تشخیصی شناختی جی‌دینا مورد تجزیه و تحلیل قرار گرفت. بر خلاف روش‌های سنتی تحلیل درجه دشواری سوال، که عبارت بود از نسبت تعداد سوالات صحیح پاسخ‌دهندگان به تعداد کل سوالات، در مدل‌های شناختی تشخیصی احتمال موفقیت در پاسخ دادن به سوال، ملاک این ارزیابی است. علاوه بر دشواری سوال، دشواری زیرمهارت‌های یک آزمون نیز در این نوع تحلیل گزارش می‌شود. بر اساس نتایج بدست آمده، زیرمهارت درک معنی لغت، برای آزمودنی‌ها ساده‌ترین و درک اطلاعات بیان‌شده در متن، دشوارترین زیرمهارت‌ها بودند. دلایل احتمالی دشواری سوالات بر اساس سه عامل زمان، دانش قبلی آزمودنی و دشواری زیرمهارت‌های هر سوال بررسی شد. نتایج به‌دست آمده از این پژوهش پیام‌های عملی برای طراحان سوالات آزمون و همچنین حوزه تدریس این مهارت دارد. با توجه به اهداف آموزشی،

طراحان سوال می‌توانند زیرمهارت‌های مربوط به هر مهارت را تعریف کرده و بر اساس آن، به انتخاب متن و همچنین طراحی سوال بپردازند. نتایج آزمون‌ها را می‌توان با مدل‌های تشخیصی شناختی بررسی کرد و اهمیت و درجه دشواری این زیرمهارت‌ها را در اختیار حوزه آموزش و تدریس قرار داد. مدرسان زبان، به عنوان استفاده‌کنندگان این نتایج می‌توانند با آموزش صریح، به افزایش آگاهی دانش‌آموزان از این زیرمهارت‌ها بپردازند و بر اساس درجه اهمیت آنها از مطالب و محتوای مورد نیاز استفاده کنند.

در حوزه سنجش تشخیصی شناختی و همچنین در خصوص زیرمهارت‌های خواندن و درک مطلب، نیاز به انجام مطالعات بیشتری می‌باشد. در حال حاضر مدل‌های آماری تشخیصی شناختی نیاز به جامعه آماری گسترده‌ای دارند، از این رو آزمون‌های مقیاس بزرگ در سطح منطقه‌ای و کشوری می‌توانند مورد تحلیل و تجزیه قرار بگیرند. بنابراین اولین قدم بررسی این آزمون‌ها و استفاده از نتایج به دست آمده در حوزه تدریس و سنجش می‌باشد. قدم بعدی، تولید مدل‌های آماری تحلیلی جدید است که قادر به تجزیه و تحلیل شناختی جامعه‌های آماری کوچکتر، به عنوان مثال کلاس‌های آموزشی، می‌باشد.

منابع

- صادقی، ک.، و خضزلو، س. (۲۰۱۱). بررسی نقش آموزش راهبردهای فراشناختی در خواندن متون مختلف: مقایسه دانشجویان رشته‌های مختلف تحصیلی. نقد زبان و ادبیات خارجی، ۴(۲)، ۹۱-۱۰۹.

- Alderson, J. C. (2000). *Assessing reading*. Cambridge: CUP.
- Alderson, J. C., & Lukmani, Y. (1989). Cognition and reading: Cognitive levels as embodied in test questions. *Reading in a Foreign Language*, 5(2), 253-270.
- Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. *Handbook of Reading Research*, 1, 255-291.
- Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology*, 96(3), 523.
- Butcher, K. R., & Kintsch, W. (2003). Text comprehension and discourse processing. *Handbook of psychology*.
- Chen, H., & Chen, J. (2015). Exploring reading comprehension skill relationships through the G-DINA model. *Educational Psychology*, 36(6), 1049-1064.
- Cheng, Y. (2010). Improving cognitive diagnostic computerized adaptive testing by balancing attribute coverage: The modified maximum global discrimination index method. *Educational and Psychological Measurement*, 70(6), 902-913.
- Chiu, C. Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement*, 37(8), 598-618.
- Choi, H-J, Rupp, A. A., & Pan, M. (2012). Standardized diagnostic assessment design and analysis: Key Ideas from modern measurement theory. 61-85.
- Davey, B. (1988). Factors affecting the difficulty of reading comprehension items for successful and unsuccessful readers. *The Journal of Experimental Education*, 56(2), 67-76.
- DeCarlo, L. T. (2012). Recognizing uncertainty in the Q-matrix via a bayesian

- extension of the DINAMIC model. *Applied Psychological Measurement*, 36(6), 447-468.
- de la Torre, J. (2011). The generalized DINAMIC model framework. *Psychometrika*, 76(2), 179-199.
- DiBello, L. V., Roussos, L. A., & Stout, W. (2006). Review of cognitively diagnostic assessment and a summary of psychometric models. *Handbook of statistics*, 26, 979-1030.
 - Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods*, 3(3), 380.
 - Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, 5(3), 178-186.
 - Gao, L. (2006). Toward a cognitive processing model of MELAB reading test item performance. *SPAAN FELLOW*, 1001, 1.
 - Garcia, P. E., Olea, J., & de la Torre, J. (2014). Application of cognitive diagnosis models to competency-based situational judgment tests. *Psicothema*, 26(3), 372-377.
 - Gierl, M. J., Leighton, J. P., & Hunka, S. M. (2000). An NCME instructional module on exploring the logic ofatsuoka's Rule-Space Model for test development and analysis. *Educational Measurement: Issues and Practice*, 19(3), 34-44.
 - Gray, W. S. (1960). The major aspects of reading. In H. Robinson (ed.), *Sequential development of reading abilities* (vol. 90, pp. 8-24). Chicago: Chicago University Press.
 - Hale, G. A. (1988). Student major field and text content: interactive effects on reading comprehension in the Test of English as a Foreign Language. *Language Testing*, 5(1), 49-61.
 - Hartz, S. M. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality*. University of Illinois at Urbana-Champaign.
 - Jang, E. E. (2005). *A validity narrative: Effects of reading skills diagnosis on teach-*

ing and learning in the context of NG TOEFL. University of Illinois at Urbana-Champaign

- Jang, E. E. (2008). A Review of cognitive diagnostic assessment for education: Theory and application. *International Journal of Testing*, 8(3), 290-295.
- Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion Model application to LanguEdge assessment. *Language Testing*, 26(1), 031-073.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric Item Response Theory. *Applied Psychological Measurement*, 25(3), 258-272.
- Kim, Y-H. (2011). Diagnosing EAP writing ability using the reduced reparameterized unified model. *Language Testing*, 28(4), 509-541.
- Lee, Y. H., & Chen, H. (2011). A review of recent response-time analyses in educational testing. *Psychological Test and Assessment Modeling*, 53(3), 359-379.
- Lee, Y-W., & Sawaki, Y. (2009). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly*, 6(3), 239-263.
- Leighton, J., & Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications*: Cambridge University Press.
- Li, H., & Suen, HK. (2013). Constructing and validating a Q-matrix for cognitive diagnostic analyses of a reading test. *Educational Assessment*, 18(1), 1-25.
- Li, H. (2011). A cognitive diagnostic analysis of the MELAB reading test. *Spaan Fellow Working Papers in Second or Foreign Language Assessment*, 9, 17-46.
- Liu, J., Xu, G., & Ying, Z. (2013). Theory of the self-learning -matrix. *Bernoulli (Andover)*, 19(5A), 1790-1817.
- Ma, W., & de la Torre, J. (2017). The generalized DINA model framework (Version 1.2.1) [Computer Software]. *Rutgers University, New Brunswick, NJ*.
- McGlohen, M., & Chang, H-H. (2008). Combining computer adaptive testing tech-

nology with cognitively diagnostic assessment. *Behavior Research Methods*, 40(3), 808-821.

- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11.
- Narvaez, D. (2001). Moral text comprehension: Implications for education and research. *Journal of Moral Education*, 30(1), 43-54.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Routledge.
- Ravand, H. (2015). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment*.
- Rost, D. (1993). Assessing different components of reading comprehension: Fact or fiction? *Language Testing*, 10(1), 79-92.
- Rupp, A. A., & Templin, J. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, 68(1), 78-96.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic assessment: Theory, methods, and applications*. New York: Guilford.
- Sawaki, Y., Kim, H-J, & Gentile, C. (2009). Q-matrix construction: Defining the link between constructs and test items in large-scale reading and listening comprehension assessments. *Language Assessment Quarterly*, 6(3), 190-209.
- Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal*, 76(4), 513-521.
- Singer, M., Graesser, A. C., & Trabasso, T. (1994). Minimal or global inference during reading. *Journal of Memory and Language*, 33(4), 421-441.
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20(4), 345-354.
- von Davier, M. (2005). *A general diagnostic model applied to language testing data*.

ETS Research Report RR-05-16. ETS, Princeton, NJ: ETS.

- Von Davier, M. (2009). Some notes on the reinvention of latent structure models as diagnostic classification models. *Measurement: Interdisciplinary Research and Perspectives*, 7(1), 67-74.
- Weir, C. J. (1993). *Understanding and developing language tests*: Hemel Hempstead: Prentice-Hall International (UK) Ltd.
- Xie, Q. (2016). Diagnosing university students' academic writing in English: Is cognitive diagnostic modelling the way forward? *Educational Psychology*, 1-22.