DOI: https://doi.org/10.22067/jls.v0i0.71976

بررسی بخش خواندن در آزمون بسندگیِ زبان فارسی دانشگاه فردوسی مشهد

محسن رودمعجنی[۱]
دکتر بهزاد قنسولی [۲]
دکتر احسان قبول [۳]
دکتر محمّد جعفر یاحقی[٤]

جستارهای نوین ادبی مجلهٔ علمی - پژوهشی، شمارهٔ ۲۰۱، تابستان ۱۳۹۷

چکیده

مهارتِ خواندن و درک مطلب و محیط دانشگاه پیوندی ناگسستنی با یکدیگر دارند و ضعف و ناتوانی در این مهارت می‌تواند به زندگی دانشگاهی دانشجویان لطمه‌ای جدّی بزند. از همین رو سنجش این توانمندی بخش بسیار مهمّی از آزمون‌های بسندگی زبان فارسی برای غیر فارسی‌زبانان را به خود اختصاص می‌دهد؛ ازجمله آزمون زبان فارسی دانشگاه فردوسی مشهد که بیش از چهار سال است به‌عنوان معیاری برای سنجش دانشجویان غیرفارسی‌زبان برای ورود به دانشگاه‌های ایران، در برخی شهرهای ایران و عراق برگزار می‌شود. در این پژوهش با کمک مدل آماری رش، به بررسی داده‌های گردآوری‌شدهٔ بخشِ خواندن یکی از آزمون‌های برگزارشده در مرکز بین‌المللی آموزش زبان فارسی دانشگاه فردوسی مشهد پرداختیم. بررسی‌ها حاکی از این بود که پایایی بخشِ خواندن آزمون ۰/۷٦ است که برای یک آزمون سرنوشت‌ساز رقم بالایی به‌شمار نمی‌آید. همچنین نقشهٔ آزمودنی- پرسش، بیانگر این نکته بود که سؤالات آزمون عمدتاً در سطح داوطلبانِ متوسّط هستند. اما نقطهٔ قوت این آزمون را می‌توان تعداد نسبتاً زیاد پرسش‌ها و نیز تنوع آنها از نظر نوع اطلاعات خواسته شده دانست. مقایسهٔ آزمون با مبانی نظری ارائه‌شده توسّط طرّاحان تافل این امکان را فراهم کرد تا راهکارهایی برای بهبود پایایی آزمون ارائه بدهیم. سؤالات آزمون بسندگی فارسی دانشگاه فردوسی مشهد از نظر نوع

۱. دانشجوی دکتری زبان و ادبیات فارسی دانشگاه فردوسی مشهد mo.roudmajani@gmail.com
۲. استاد زبان و ادبیات انگلیسی دانشگاه فردوسی مشهد (نویسنده مسئول) ghonsooly@ferdowsi.um.ac.ir
۳. استادیار زبان و ادبیات فارسی دانشگاه فردوسی مشهد ghabool@ ferdowsi um.ac.ir
٤. استاد زبان و ادبیات فارسی دانشگاه فردوسی مشهد ferdows@ferdowsi um.ac.ir

تاریخ دریافت: ۱۳۹۷/۱/۲۱ تاریخ پذیرش: ۱۳۹۷/۹/۵

مطابقت، عمدتاً از نوع محل‌یابی و یا سلسله‌ای هستند و پرسش‌هایی از نوع ترکیبی و تولیدی می‌تواند به سخت‌تر شدن سؤالات بینجامد. اگر از منظر اهداف خواندن به آزمون نگاهی بیندازیم، مشاهده می‌کنیم که سؤالات عمدتاً متناسب با اهداف خواندن برای پیداکردن اطّلاعات و نیز خواندن برای رسیدن به درک مقدّماتی هستند. به نظر می‌رسد که این آزمون به سؤالاتی متناسب با اهداف خواندن برای یادگیری و نیز هدف خواندن برای ادغام اطّلاعات در بین چند متن مختلف نیاز دارد که از سطح دشواری بالاتری برخوردار هستند.

کلیدواژه‌ها: آزمون بسندگی زبان فارسی، سنجش تکلیف‌محور، سنجش مهارت محور، مهارت خواندن، پایایی آزمون.

۱- مقدّمه

انسان‌ها با توانمندی خواندن[1] زاده نمی‌شوند و این پدیده در تاریخ عمر بشر امری نوظهور به‌شمار می‌آید. هنوز هم می‌توان افرادی را یافت که از این توانایی محروم‌اند. باوجوداین، امروزه شرایط زندگی در جوامع انسانی به‌گونه‌ای است که ضعف در مهارتِ خواندن و درک متون، مشکلات بسیاری را برای افراد به‌وجود می‌آورد.شهروندان جوامع امروزی برای موفّق‌بودن باید متن‌خوان‌های خوبی باشند. هرچند توانمندی خواندن، موفّقیت فرد را تضمین نمی‌کند، امّا بی‌بهره‌بودن از این مهارت رسیدن به موفّقیت را برای او دشوار می‌گرداند. اختراع رایانه و اینترنت نیز نتوانسته است از اهمّیت این پدیده بکاهد (Grabe, 2009 :5).

مهارتِ خواندن و محیط دانشگاهی نیز پیوندی ناگسستنی دارند و بسیاری از فعّالیت‌های دانشجویان در دانشگاه با خواندن متون درهم تنیده است. خواندن کتاب‌ها، مقاله‌ها، اطّلاعیه‌ها، یادداشت‌های کلاسی، گزارش‌های آزمایشگاه و انتخاب واحد، همه نیازمندِ داشتن توانمندی خواندن است. ضعف در این مهارت می‌تواند به فعّالیت‌های دانشگاهی فرد آسیبی جدّی بزند.

به‌دلیل اهمّیت توانمندی خواندن در محیط دانشگاه، داوطلبانی که قصد ادامهٔ تحصیل در دانشگاه‌های ایران را دارند، می‌بایست از این مهارت به شکل کارآمدی بهره‌مند باشند. بدین‌جهت است که سنجش این مهارت، بخش بسیار مهمّی از آزمون‌های بسندگیِ زبان فارسی را به خود

1. Reading Skill

اختصاص می‌دهد. آزمون جامع زبان فارسی دانشگاه فردوسی مشهد نیز به‌عنوان یکی از آزمون‌های شناخته‌شدهٔ زبان فارسی با هدف سنجش توانمندی زبانی متقاضیانِ غیرفارسی‌زبانِ ورود به دانشگاه‌های ایران، سالانه در دو نوبت در شهرهای مشهد و اهواز و چند مرکز در کشور عراق برگزار می‌شود. هدف از این پژوهش، بررسی پایایی بخشِ خواندن این آزمون و ارائهٔ پیشنهادهایی برای بهبود آن است. این پژوهش تلاش می‌کند تا به دو پرسش پاسخ دهد: ۱ــــ بخشِ خواندن آزمون بسندگی زبان فارسی دانشگاه فردوسی مشهد از چه میزان پایایی برخوردار است؟ ۲ــــ در این آزمون چه رابطه‌ای بین سطح دشواری سؤالات و سطح توانمندی آزمون‌دهندگان برقرار است؟

مدل آماری‌ای که برای بررسی داده‌های این پژوهش استفاده می‌شود، رش[1] است. این مدل که ساختاری پیچیده‌تر از مدل‌های کلاسیک دارد، از یک مزیت اساسی برخوردار است و آن وابسته‌نبودن به نمونهٔ آماری[2] پژوهش است (Bachman, 2004 :139; Mousavi, 2012). مدل‌های آماری کلاسیک مانند کودر ریچاردسون[3]، روش دونیمه[4]، آلفا کرونباخ[5] و آزمون بازآزمون[6] بر این فرض اساسی استوار هستند که نمره بدست آمده از یک آزمون حاصل جمع نمرهٔ حقیقی فرد و میزانی از خطا در محاسبه است. مبنی قرار دادن چنین فرضی باعث شده است که این مدل‌ها دارای دو ضعف اساسی باشند. نخست اینکه توانای تمییز خطاهای نظام‌مند از خطاهای تصادفی را نداشته باشند (Bachman,1990:186) دوم اینکه با تغییر نمونهٔ پژوهش نتایج بدست آمده نیز دچار تغییر شود. به دلیلی محدودیت‌های موجود در مدل‌های آماری کلاسیک روان‌سنجان مدل‌های آماری پیچیده‌تری را طراحی کرده‌اند. این مدل‌ها که به طور کلی آی آر تی[7] نامیده می‌شوند این قابلیت را دارند که عملکرد آزمون‌دهندگان در آزمون را با توجه به توانمندی زبانی آنها مورد بررسی قرار دهند. همین امر باعث می‌شود که دیگر به نمونه آماری پژوهش وابسته نباشند. مدل آماری رش نیز یک نمونه از مدل‌های آی آر تی به حساب می‌آید. به همین خاطر ما این مدل را برای تحلیل داده‌های خود بکار بردیم.

1. Rasch Model
2. Sample Free
3. Kuder-Richardson
4. Split half Experiment
5. Cronbach’s Alpha
6. Test-Retest experiment
7. IRT: Item Response Theory

۲- پیشینهٔ پژوهش

پژوهش‌های موجود در زمینهٔ مهارت خواندن را می‌توان به سه دسته تقسیم کرد. دستهٔ نخست تحلیل نیازها نام دارند که به بررسی نیازهای موجود در زمینهٔ مهارت خواندن می‌پردازند. نتایج این پژوهش‌ها می‌تواند مبنای طراحی آزمون‌های زبانی قرار بگیرد. تحلیل نیازها این امکان را فراهم می‌آورند که سنجش، متناسب با نیازهای زبانی زبان‌آموزان باشد و در نتیجه تصویری حقیقی از توانایی زبانی آنها ارائه دهد. در جدول شمارهٔ ۱ به برخی تحلیل نیازها مرتبط با مهارت خواندن اشاره شده است.

جدول ۱- تحلیل نیازهای مرتبط با سنجش مهارت خواندن

نویسنده	تاریخ	مراکز مورد بررسی	مخاطب	موضوع	روش پژوهش	تمرکز
اوستلر[1]	۱۹۸۰	دانشگاه کالیفرنیای شمالی	دانشجویان بین المللی	رشته‌های مختلف	پرسش‌نامه‌ای	چهارمهارت
جانز[2]	۱۹۸۱	دانشگاه ایالتی سن دیگو	-----	رشته‌های مختلف	پرسش‌نامه‌ای	چهارمهارت
کریستیسن و کرنک[3]	۱۹۸٦	۵ دانشگاه در امریکا	دانشجویان بین المللی	رشته‌های مختلف	پرسش‌نامه‌ای	چهارمهارت
بنسون[4]	۱۹۹۱	دانشگاه‌های امریکا	دانشجویان بین المللی	گروه انسان-شناسی	------	مهارت خواندن
گینتر و گرنت[5]	۱۹۹٦	دانشگاه‌های امریکا	دانشجویان بومی	رشته‌های مختلف	--------	چهار مهارت
واترز[6]	۱۹۹٦	بررسی انتقادی تحلیل نیازهای پیشین				

دستهٔ دیگر پژوهش‌ها در زمینهٔ سنجش مهارت خواندن مربوط به تعریف ماهیت این مهارت می‌شود. این پژوهش‌ها سعی می‌کنند به این سؤال اساسی پاسخ دهند که مهارت خواندن از چه

1. Ostler
2. Jonhs
3. Christison and Krahnke
4. Benson
5. Ginther and Grant
6. Waters

سازه‌هایی تشکیل شده است. بدیهی است که بدون داشتن درک روشنی از سازه‌های مهارت خواندن سنجش آن از اعتبار[1] چندانی برخوردار نخواهد بود. از جمله تلاش‌های صورت گرفته در این حوزه می‌توان به گریپ[2] و همکاران (۲۰۱۳)، گریپ (۲۰۰۹)، الدرسون[3] (۲۰۰۰)، انرایت[4] و همکاران (۲۰۰۰) و هودسون[5] (۱۹۹٦) اشاره کرد.

در دسته سوم پژوهشگران به بررسی و یا طراحی یک آزمون خاص می‌پردازند. در این پژوهش‌ها مفاهیمی مانند اعتبار، پایایی و بازشویی[6] آزمون و یا ارتباط سنجش خواندن با سایر متغیرها مورد بررسی قرار می‌گیرند. بدیهی است که به دلیل گستردگی آزمون‌های زبانی، پژوهش‌های بسیاری در این دسته قرار می‌گیرند. به عنوان نمونه می‌توان به برخی پژوهش‌های صورت گرفته بر روی بخش خواندن آزمون تافل اشاره کرد: چپل[7] و همکاران (۲۰۱۱)، چن[8] و شیهان[9] (۲۰۱۵)، لیو[10] و همکاران (۲۰۰۹)، انرایت[11] و همکاران (۲۰۰۰).

در زبان فارسی هیچ پژوهشی مربوط به تحلیل نیازها یافت نشد اما پژوهشگرانی مانند موسوی، ۱۳۷۹؛ قنسولی، ۱۳۸۹؛ جلیلی، ۱۳۹۰؛ گلپور، ۱۳۹٤ تلاش کرده‌اند تا یک آزمون بسندگی زبان فارسی طراحی کنند و بخشی از آزمون خود را به مهارتِ خواندن اختصاص داده‌اند. مشخصات هر یک از این پژوهش‌ها در جدول شمارهٔ ۲ آمده است.

1. Validity
2. Grabe
3. Alderson
4. Enright
5. Hudson
6. Washback
7. Chappelle
8. Chen
9. Sheehan
10. Liu
11. Enright

جدول ۲- پژوهش‌های صورت گرفته در زبان فارسی مرتبط با سنجش مهارت خواندن

نویسنده	تاریخ	هدف از انجام پژوهش	روش آماری بررسی پایایی پژوهش
موسوی	۱۳۷۹	طراحی آزمون بسندگی زبان فارسی برای سنجش دانش زبانی غیر فارسی زبانان	روش‌های آماری کلاسیک
قنسولی	۱۳۸۹	طراحی آزمون بسندگی زبان فارسی	آلفا کرونباخ
جلیلی	۱۳۹۰	طراحی آزمون بسندگی زبان فارسی برای سنجش دانش زبانی غیر فارسی زبانان	بدون تحلیل آماری
گلپور	۱۳۹٤	طراحی آزمون بسندگی زبان فارسی	روش‌های کلاسیک

در پژوهش حاضر تلاش می‌شود تا بخش خواندن آزمون بسندگی فارسی دانشگاه فردوسی مورد بررسی قرار بگیرد. بنابراین این تحقیق در دستهٔ سوم پژوهش‌های یادشده قرار می‌گیرد. بر خلاف پژوهش‌های پیشین در زبان فارسی، تلاش شده است که از مدل آماری آی آر تی برای تحلیل داده‌ها استفاده شود. همان‌طور که گفته شد این مدل به نمونهٔ پژوهش وابسته نبوده و آمار دقیق‌تری را در اختیار می‌گذارد. همچنین به منظور بحث و بررسی بهتر داده‌های بدست آمده سعی شده است تا بر اساس پژوهش‌های صورت گرفته در زبان انگلیسی مدلی برای سنجش مهارت خواندن در محیط دانشگاه معرفی شود. این مدل امکان تحلیل پایایی آزمون و ارائه پیشنهادهایی برای بهبود کیفیت آن را فراهم می‌سازد.

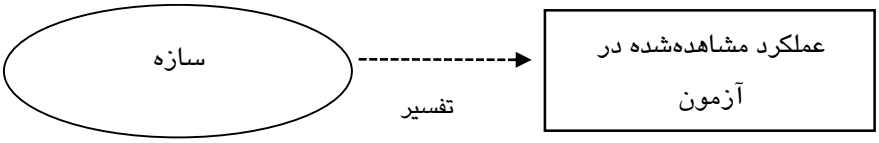
۲- مبانی نظری

سنجش دانش زبان را می‌توان از دو منظر متفاوت بررسی کرد. مسیک از این دو با نام‌های توانایی‌محور[1] و تکلیف‌محور[2] یاد می‌کند (Messick, 1994: 14). در سنجش توانایی‌محور، هدف اندازه‌گیری سازه‌ها[3] ی تشکیل‌دهندهٔ دانش زبانی است که درحقیقت مفاهیمی انتزاعی در ذهن آزمون‌دهندگان هستند. رابطهٔ سازه‌ها با یکدیگر و همچنین رابطهٔ سازه‌ها با نمودهای عینی‌شان

1. Competency- centered
2. Task-centered
3. Constructs

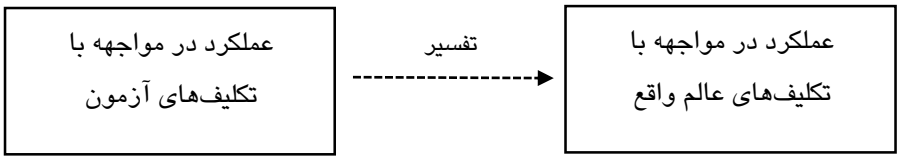
(عملکرد آزمون‌دهندگان در آزمون) از طریق یک نظریهٔ زبانی مشـخّص می‌شـود. بنابراین در سـنجش توانایی‌محور مبنا یک نظریهٔ زبانی است.

Figure 1. Interpretation in Competency Assessment



شکل ۱- تفسیر در سنجش مهارت‌محور (Chapelle, Enright et al., 2008: 3)

در مقابل، مبنای تفسـیر نمرات در سـنجش تکلیف‌محور بافت[1] است. قضاوت صحیح دربارهٔ اینکه آیا دانشـجویان توانایی اسـتفاده از دانش خود را در محیط موردنظر دارند، زمانی می‌تواند محقّق شـود که عملکرد آن‌ها در مواجهه با تکلیف‌های مشـابه آن محیط مشـاهده گردد. آنچه در یک سـنجش تکلیف‌محور اهمّیت دارد، شـناسـایی تکلیف‌های مهم در محیط عالم واقعی[2] و شـبیه‌سـازی آن‌ها در آزمون است.



شکل ۲- تفسیر در سنجش تکلیف‌محور (Chapelle, Enright et al., 2008: 4)

بنابراین رسـیدن به یک سـنجش دقیق از مهارتِ خواندن، نیازمند پاسـخ‌گویی به دو پرسش اساسی اسـت: نخسـت اینکه سـازه‌های تشکیل‌دهندهٔ مهارتِ خواندن چیست؛ دوم اینکه دانشجویان در محیط دانشگاهی با چه تکلیف‌هایی روبه‌رو می‌شوند.

1. Context
2. Real world

۳-۱- فرآیندهای تشکیل‌دهندهٔ مهارتِ خواندن

مهارتِ خواندن ساختار بسیار پیچیده‌ای دارد و از فرایندهای مختلفی تشکیل شده است. هرچند برخی تلاش کرده‌اند تا تعریفی یک‌خطّی از آن ارائه دهند، امّا به‌دلیل ابعاد گستردهٔ این مهارت، انجام چنین امری ناممکن است (Grabe: 2009:14). بهترین راه شناخت مهارتِ خواندن، بررسی فرایندهای تشکیل‌دهندهٔ آن است.

فرایندهای خواندن را می‌توان به دو دستهٔ فراپایه[1] و فروپایه[2] تقسیم کرد (قنسولی، ۱۳۹۲: ۷۲؛ Hudson,1996: 3; Grabe, 2009: 21). عنوان فروپایه به معنای کم‌اهمّیت‌بودن نیست، بلکه این نام‌گذاری ازاین‌رو صورت گرفته است که این‌دست فرایندها می‌توانند به‌شکل اتوماتیک و خودکار و بدون آگاهی فرد رخ دهند و نظر و خواست متن‌خوان تأثیری بر آن‌ها ندارد. فرایندهای فروپایه عبارت‌اند از تشخیص واژه[3]، تجزیهٔ نحوی[4] و کدشکنی واژی/نویسه‌ای (قنسولی، ۱۳۹۲: ۷۲). البتّه گریب رمزگذاری گذاره‌های معنایی[5] را نیز جزء فرایندهای فروپایه می‌داند (Grabe, 2009: 22).

فرایندهای فروپایه به‌دلیل خودکاربودن سرعت بسیار بالایی دارند. یک متن‌خوان خوب می‌تواند یک واژه را در مدّت‌زمان ۱۰۰ میلی‌ثانیه تشخیص دهد و یک متن ۲۵۰ تا ۳۰۰ کلمه‌ای را به‌راحتی در مدّت یک دقیقه بخواند. بنابراین می‌توان مهم‌ترین ویژگی فرایندهای فروپایه را سرعت بسیار بالای پردازش آن‌ها دانست (Grabe, 2009: 23).

اطّلاعاتی که از طریق فرایندهای فروپایه گردآوری می‌شوند، در کنار هم قرار می‌گیرند تا به متن‌خوان در ساختن یک تصویر کلّی از متن کمک کنند. درحقیقت، متن‌خوان برمبنای کلمات و جملات خوانده‌شده شبکه‌ای از عناصر تشکیل‌دهندهٔ متن را می‌سازد. همچنان که فرد به خواندن متن ادامه می‌دهد، اطّلاعات جدید در درون این شبکه قرار گرفته و شبکه‌ای بزرگ‌تر و کامل‌تر شکل می‌گیرد. شبکهٔ به‌وجودآمده مدل متنی[6] نام دارد که در درک متن‌خوان از متن نقش بسیار برجسته‌ای ایفا می‌کند و ازجمله فرایندهای فراپایه به‌شمار می‌آید.

---

1. Higher Level
2. Lower Level
3. Word Recognition
4. Syntax Parsing
5. Meaning Proposition Encoding
6. Text Model

امّا درک افراد از یک متن تنها به اطّلاعات موجود در آن محدود نمی‌شود، بلکه اطّلاعات شخصی افراد نیز در پردازش آن تأثیرگذار است. تجربهٔ متن‌خوان در مواجهه با متون مشابه، هدف متن‌خوان از خواندن متن، اطّلاعات او دربارهٔ موضوع متن، میزان لذّتی که او از خواندن می‌برد، دیدگاه او نسبت به نویسنده و یا ژانر موردنظر، نکاتی هستند که هرچند در متن موجود نیستند، امّا در درک افراد از آن تأثیرگذارند. گریب از این نکات با عنوان مدل موقعیتی[1] نام می‌برد. بنابراین می‌توان از دو جنبه به بررسی یک متن پرداخت: نخست از زاویهٔ دید نویسنده و آنچه او در نظر دارد که ما از متن برداشت کنیم و دیگر درک ما از متن و احساسمان نسبت به آن.

علاوه بر مدل متنی و موقعیتی، فرایندهای فراپایهٔ دیگری نیز در درک ما از متن تأثیرگذارند که به مکانیزم‌های کنترل اجرایی[2] موسوم‌اند. مکانیزم‌های کنترل اجرایی نقشِ جهت‌دادن به فرایند توجّه را ایفا می‌کنند. حذف اطّلاعات غیرضروری و غیرمرتبط، تغییردادن مسیر توجّه به شکل آگاهانه و بروزرسانی حافظهٔ کاری ازجمله کارکردهای دیگر مکانیزم‌های کنترل اجرایی هستند (Grabe, 2009: 43).

پس از شناسایی فرایندهای فروپایه و فراپایه با پرسش اساسی دیگری روبه‌رو می‌شویم و آن این است که این فرایندها چگونه با یکدیگر در ارتباط هستند. نخستین پژوهش‌ها تحت تأثیر الگوی خواندن گو[3] خواندن را فرایندی پایین به بالا[4] می‌دانست (قنسولی، ۱۳۹۲: ۵۹) که در آن شکل‌گیری معنی در ذهن متن‌خوان به ترتیب از رهگذر بررسی نویسه‌ها، واژه‌ها و جملات شکل می‌گرفت. در مقابل، در رویکرد بالا به پایین[5] متن‌خوان‌ها کمترین استفاده را از درک حرف‌به‌حرف و واژه‌به‌واژه می‌کنند و به پیش‌بینی و حدس‌زدن معنی متن از طریق به‌کارگیری دانش خود دربارهٔ جهان و زبان می‌پردازند (Aldrson, 2000: 17).

امّا رویکرد سوم در رابطه با فرایندهای فروپایه و فراپایه، رویکرد هم‌کنشی[6] نام دارد که در آن زبان‌آموز برای خواندن مؤثّر از هر دو روش بالا به پایین و پایین به بالا به‌صورت موازی استفاده

1. Situational Model
2. Executive control Mechanism
3. Gough
4. Bottom-up
5. Top-Down
6. Interactive

می‌کند. هریک از اجزای تشکیل‌دهندهٔ پردازش متنی (چه بالا به پایین و پایین به بالا) می‌توانند با اجزای دیگر، کنش متقابل داشته باشند (Aldrson, 2000: 17).

به‌طورکلّی چهار عامل را در موفّقیت و یا ناکامی افراد در خواندن و درک متن می‌توان تأثیرگذار دانست: ۱ـ خودکاربودن فرایندها ۲ـ طرح محتوایی و طرح صوری[1] ۳ـ راهبردها و مهارت‌های فراشناختی ٤ـ شناخت و آگاهی از هدف و بافت متن (Hudson, 1996: 4).

۳-۲- سنجش تکلیف‌محور

همان‌طور که پیش‌تر گفته شد، یکی از شیوه‌های طرّاحی آزمون، بررسی تکلیف‌های زبانی است که افراد در محیط موردنظر با آن‌ها روبه‌رو می‌شوند. به این منظور، جمیسون[2] و همکاران (۲۰۰۰) چارچوبی برای بررسی تکلیف‌های زبانی با نام چارچوب مشخّصات تکلیف طرّاحی کرده‌اند. چارچوب مشخّصات تکلیف شیوه‌ای کارآمد برای بررسی مشخّصات فعّالیت‌های زبانی در محیط مقصد ارائه می‌دهد. استفاده از این چارچوب می‌تواند به اصالت آزمون کمک کند و محصول آن در واقع ارائهٔ تصویری واقعی‌تر از توانمندی آزمون‌دهندگان است. همان‌طور که در جدول ۱ آمده است، این چارچوب شامل سه بخش اصلی می‌شود: موقعیت، محتوای متن و مشخّصات آزمون (Jamieson et al, 2000).

۳-۲-۱- موقعیت

محیط فرازبان‌شناسی‌ای که تولید زبان در آن اتّفاق می‌افتد، موقعیت نام دارد. برای مشخّص‌کردن موقعیت تکلیف، باید به پنج پرسش پاسخ داد؟ ۱ـ چه کسانی در انجام تکلیف شرکت دارند؟ ۲ـ محتوای تکلیف چیست؟ ۳ـ محیطی که تکلیف در آن اتّفاق می‌افتد کجاست؟ ٤ـ هدف از انجام تکلیف چه می‌باشد؟ ۵ـ گونهٔ کاربردی تکلیف چیست؟

1. Context Schema and Formal Schema
2. Jamieson

جدول ۳- چارچوب مشخّصات تکلیف برگرفته از (Jamieson et al,2008: 64)

مشخّصات تکلیف	متغیّرهای تکلیف	تعریف متغیّرها
موقعیت	شرکت-کنندگان	افرادی که در فعّالیت زبانی شرکت دارند (مانند دانشجویان و استادان). می‌توان آن‌ها را از نظر جنسیت، نژاد، سن و همچنین نقش اجتماعی دسته‌بندی کرد.
	محتوا	محتوا همان موضوع تکلیف‌های زبانی است. ـــ تخصّصی ـــ دانشگاهی که می‌بایست در حوزه‌های علوم اجتماعی و طبیعی باشد؛ زیرا گسترده بوده و موضوعات عمومی‌تری را دربر می‌گیرند. ـ کلاس درس: مربوط به فعّالیت‌ها و تمرین‌های کلاسی است. ـــ غیردانشگاهی: موضوعات پرکاربردی که لزوماً به فضای کلاسی مربوط نمی‌شوند (بیماری، آب‌وهوا و...).
	محیط	فضایی که تکلیف در آن اتّفاق می‌افتد: ـ آموزشی، مانند کلاس درس آزمایشگاه و... ـ دانشگاهی ـ غیرآموزشی، مانند سالن مطالعه، خوابگاه و... ـ غیردانشگاهی، مانند درمانگاه، سلف سرویس و...
	هدف	هدف شرکت در فعّالیت زبانی یا همان کاربرد تکلیف به‌عنوان مثال آموزشی، بیان احساسات فردی، برقراری تعامل
	گونه[1]	میزان رسمی‌بودن متن که شامل سه سطح است: ـ رسمی، مانند: سخنرانی، ارائهٔ کلاسی، مقاله ـ نیمه‌رسمی، مانند گفت‌وگو در اتاق استاد، گفت‌وگوی تلفنی با یک فرد ناشناس ـ غیررسمی
محتوای متن	ویژگی-های دستوری	ساختارهای نحوی و واژگانی متون که می‌توان آن‌ها را با استفاده از فرمول‌های خوانایی، بررسی بسامد واژگان، نوع افعال و بندهای وابسته بررسی کرد.
	ویژگی-های کاربردی[2]	ازآنجایی‌که عمدهٔ متون دانشگاهی تشریحی هستند، کاربرد اصلی این متون انتقال و کاویدن اطّلاعات است و کارکردهای دیگر می‌تواند دفاع‌کردن از یک ایده و یا ارائهٔ یک تحلیل (بحث، جدل) باشد. تعداد محدودی از متون تشریحی، هدف اقناعی و یا تعاملی دارند.

1. Register
2. Pragmatic Features

ادامه جدول ۳

مشخّصات تکلیف	متغیّرهای تکلیف	تعریف متغیّرها
	ویژگی‌های گفتمان	هدف نویسنده و یا گوینده است که می‌تواند معنی‌کردن، توصیف‌کردن، بیان علّت و معلول، بیان مسئله و راه‌حلّ آن، آوردن مثال و نمونه، مقایسه‌کردن، ارائهٔ یک شرح ساده مانند لیست اقلام و یا بیان قوانین و مقرّرات باشد.
مشخّصات آزمون	**سؤالات**	نوع اطّلاعات خواسته‌شده: هرچه اطّلاعات خواسته‌شده انتزاعی‌تر باشد، سؤال سخت‌تر خواهد بود (به پیوست ۱ مراجعه کنید).
		نوع مطابقت: برای پاسخ‌گویی به سؤال باید با چه روشی متن را بررسی کرد (به پیوست ۲ مراجعه کنید). ـ محل‌یابی: پیداکردن محلّ جواب در متن ـ سلسله‌ای: پیداکردن چند محل جواب در متن ـ ترکیبی: از دو مرحله تشکیل می‌شود؛ نخست چند محلّ جواب در متن شناسایی شود، سپس رابطهٔ بین آن‌ها مشخّص گردد. ـ تولیدی: سؤالاتی که در آن داوطلبان باید برای پیداکردن محلّ پاسخ به تفسیر بپردازند.
		حضور گزینه‌های انحرافی: ساده‌ترین سؤالات سؤالاتی هستند که گزینهٔ انحرافی ندارند. هرچه تعداد گزینه‌های انحرافی بیشتر باشد و گزینه‌های انحرافی اشتراکات بیشتری با پاسخ داشته باشند، سؤال سخت‌تر می‌شود.
	نوع پاسخ‌دهی	سؤالات چندگزینه‌ای، سؤالات تشریحی، کامل‌کردن یک نمودار
	قواعد نمره‌دهی	مراحل نمره‌دهی به پاسخ‌های آزمون‌دهندگان

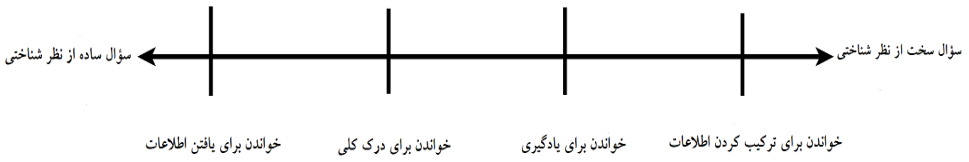
انرایت[1] و همکارانش (۲۰۰۰) تلاش کردند تا چارچوب مشخّصات تکلیف طرّاحی‌شده را با مهارتِ خواندن تطبیق دهند. آن‌ها معتقدند که موقعیت در مهارتِ خواندن به برجستگی مهارت‌هایی مانند صحبت‌کردن و یا شنیدن نیست (Enright et al, 2000: 14). در واقع می‌توان چنین گفت که زبانِ خواندن نسبتاً از بافت خارج می‌شود[2]. شرکت‌کنندگان مهارتِ خواندن نیز در وهلهٔ نخست،

1. Enright
2. Decontextualize

شـامل خودِ متن‌خوان‌ها می‌شوند که به ساختن مدل متنی و موقعیتی می‌پردازند. نقش نویسنده به‌عنوان شـرکت‌کنندهٔ دیگر در برخی متون مانند متون تشـریحی کم‌رنگ‌تر اسـت، امّا در متون استدلالی[1] نقش برجسته‌ای ایفا می‌کند.

انرایت معتقد اسـت که محیط[2] تأثیر چندانی در سـختی مهارتِ خواندن ندارد و عاملی نیسـت که بخواهیم در طرّاحی تکلیف‌هـای مهـارتِ خوانـدن در نظر بگیریم. تنهـا هنگـامی که مهارتِ خواندن با سـایر مهارت‌ها ترکیب می‌شـود، می‌توان محیط را عاملی تأثیرگذار قلمداد کرد. به‌عنوان مثال، در حین شنیدن یک سخنرانی، مطالبی بر روی تخته و یا صفحهٔ نمایش وجود دارد که فرد باید آن‌ها را بخواند.

انرایت محتوای متونِ خواندن را هر موضـوع دانشـگاهی در رشـته‌های علوم‌انسانی، علوم‌اجتماعی، هنر و علوم‌تجربی می‌داند، البتّه درصـورتی‌که فهم متن نیازمند داشـتن دانش تخصـّصـی نباشـد. او همچنین چهار هدف برای خواندن قائل است: ۱ـــ خواندن به‌منظور پیداکردن اطّلاعات ۲ـــ خواندن برای رسیدن به درک ابتدایی از متن ۳ـــ خواندن با هدف یادگیری ٤ـــ خواندن برای ادغام اطّلاعات در بین چند متن مختلف. این اهداف دارای سـطوح سـختی متفاوتی هستند که در شکل ۳ نشان داده شده است.



شکل ۳- مقیاس مفهومی فرضی برای اهداف خواندن (Jamieson et al, 2008: 55)

۳-۲-۲- محتوای متن

در بخش محتوای متن، ویژگی‌های متنِ تکلیف‌ها[3] بررسـی می‌شود. این بخش شامل سه پرسش است: متن تکلیف‌های زبانی در محیط مقصـد دارای چه ویژگی‌های دسـتوری و واژگانی هسـتند؟ این متون از چه ساخت گفتمانی‌ای برخوردار می‌باشند؟ و از نظر کاربردشناختی چه تفاوتی با یکدیگر دارند؟

1. Argumentation
2. Setting
3. Task's textual features

پاسخ‌گویی به دو پرسش نخست به‌دلیل پژوهش‌های معدودی که در این زمینه در زبان فارسی صورت گرفته، بسیار دشوار است. ما اطّلاعات چندانی دربارهٔ ویژگی‌های نحوی و واژگانی متون دانشگاهی نداریم. مختصر تحقیقاتی که در این حوزه نیز صورت گرفته، در کار طرّاحی سؤال چندان راه‌گشا نیست. غفارثمر و همکاران (۱۳۹۲) پیش‌تر به این ضعف در حوزهٔ واژگان اشاره کرده‌اند.

ازآنجایی‌که عمدهٔ متون دانشگاهی تشریحی[1] هستند، هدف این متون از نظر کاربردشناختی انتقال و کاویدن اطّلاعات است. بااین‌حال کارکردهای دیگری نیز می‌توانند داشته باشند، مانند دفاع‌کردن از یک ایده و یا ارائهٔ یک تحلیل (بحث، جدل). تعداد محدودی از متون تشریحی نیز هدف اقناعی و یا تعاملی دارند.

۳-۲-۳- مشخّصات آزمون

تاکنون ما به بررسی ویژگی‌های فعّالیت‌های زبانی در محیط دانشگاه پرداختیم که می‌توانند در سنجش مهارت خواندن تأثیرگذار باشند. امّا اگر بخواهیم این تکلیف‌ها را به سؤالات امتحانی تبدیل کنیم، باید چند نکته را نیز روشن نماییم که بخش پایانی چارچوب انرایت با نام مشخّصات آزمون را تشکیل می‌دهد. پرسش‌های آزمون دارای چه مشخّصاتی هستند؟ نوع سؤالات آزمون چیست؟ چه قواعدی برای نمره‌دهی به‌کار رفته‌است؟

پرسش‌ها تأثیر بسیار مهمّی در سختی آزمون دارند و ما باید میزان سختی آن‌ها را به‌دقّت محاسبه نماییم. به همین منظور، جمیسون و همکاران (۲۰۰۰) به بررسی ویژگی‌های پرسش‌ها و تأثیر آن‌ها در سختی آزمون پرداخته‌اند که آگاهی از این نکات می‌تواند در طرّاحی آزمون بسیار تأثیرگذار باشد. پرسش‌ها را می‌توان از دو منظر نوع اطّلاعات خواسته‌شده[2] و نوع رابطهٔ سؤال و پاسخ[3] (نوع مطابقت) تحلیل و بررسی کرد. هرچه اطّلاعات خواسته‌شده انتزاعی‌تر باشند، سؤال نیز از میزان سختی بیشتری برخوردار خواهد بود. رابطهٔ بین نوع اطّلاعات و سختی سؤال در جدول ٤ به‌طور کامل آمده است.

1. Expository
2. Type of question requested
3. Type of match

جدول ٤- نوع اطّلاعات خواسته‌شده (Jamieson et al,2000: 63-65)

نوع اطّلاعات خواسته‌شده	سطح سختی
تشخیص یک شخص، گروه، مکان یا شیء	۱
پرسش‌هایی دربارهٔ زمان، مقدار، صفت، نوع، فعّالیت، محل، موقعیت، گروه و یا مراحل انجام یک کار	۲
پرسش‌هایی دربارهٔ حالت، برنامهٔ یک شخص، هدف و یا کارکرد، گزینهٔ انتخابی (مثال: نویسنده چه گزینه‌هایی را برای کاهش آلودگی ارائه می‌دهد؟)؛ اقدام (او برای رئیس‌جمهور شدن چه کارهایی انجام داد؟)؛ وضعیت، مراحل، ارجاع ضمیر، تأیید (آیا درست است که مارکوپلو دو بار به چین سفر کرد؟)؛ بیان یک مسئله، بیان راه‌حل	۳
سؤالاتی که برای پاسخ‌دادن به آن‌ها می‌بایست رابطهٔ علّت و معلولی و شباهت چند چیز را بررسی کرد و یا سؤالاتی دربارهٔ نظر و عقیده، بیان الگو و یا تبیین یک موضوع	٤
سؤالاتی مربوط به موضوع یک متن یا بند، بررسی تفاوت‌ها و یا معادل‌یابی برای واژگان کم‌بسامد	۵

مطلب دیگری که در طراحی سؤال باید بررسی کرد، رابطهٔ بین سؤال و پاسخ است. آیا داوطلب تنها باید محلّ یک پاسخ را در متن پیدا کند (محل‌یابی) و یا پاسخ در چند نقطه از متن قرار گرفته است و داوطلب می‌بایست محل‌های مختلفی را شناسایی کند (سلسله‌ای). در برخی از سؤالات، داوطلب علاوه بر پیداکردن چند محلّ پاسخ در متن، باید رابطهٔ آن‌ها را نیز با یکدیگر بررسی کند (ترکیبی) و یا در مواردی برای یافتن پاسخ، به تفسیر بپردازد (تولیدی). میزان سختی هرکدام از این نوع سؤالات در جدول ۵ آمده است.

جدول ۵- نوع مطابقت (Jamieson et al,2000:66)

نوع مطابقت	سطح سختی
محل‌یابی: باید محلّ جواب را در متن پیدا کرد.	۱
سلسله‌ای: باید چندین عمل محلی‌یابی صورت گیرد.	۲
ترکیبی: اوّل باید چند محلّ جواب در متن پیدا شود، سپس ارتباط آن‌ها با هم شناخته شود.	۳
تولیدی: سؤالاتی که داوطلب برای پاسخ دادن به آن‌ها نیازمند تفسیر متن است.	٤

۳-۳- مدلی برای سنجش مهارت خواندن در محیط دانشگاه

براساس مطالبی که دربارهٔ سازه‌های آزمون و تکلیف‌های زبانی در محیط دانشگاه گفته شد، می‌توان مدلی نهایی برای سنجش مهارت خواندن ارائه داد. این مدل در جدول ٦ معرّفی شده و مبنای طرّاحی آزمون جدید تافل است که ما آن را برای سنجش مهارتِ خواندن زبان فارسی در سطح پیشرفته در محیط دانشگاه پیشنهاد می‌دهیم.

جدول ٦- مشخّصات چارچوب خواندن (Jamieson, 2008: 70)

مدلی برای سنجش مهارت خواندن در محیط دانشگاه			
هدف خواندن	**توانایی‌ها**	**نوع تکلیف**	**نوع متن**
خواندن برای پیداکردن اطّلاعات	تشخیص سریع و کارآمد واژه	تندخوانی برای پیداکردن واژه و یا نکته‌ای در متن	
خواندن برای رسیدن به درک مقدّماتی	تشخیص معنای اصلی گزاره‌های مهم	ساده‌سازی جملات	
	درک گزاره‌های اصلی به‌شکل سریع و کارآمد	خواندن سریع و کارآمد	
خواندن برای یادگیری	ترکیب‌کردن و یا به‌خاطرسپردن نکات اصلی و اطّلاعات فرعی متن، دسته‌بندی‌کردن اطّلاعات مهم متن	خلاصه‌کردن اطّلاعات متن به‌شکل منثور و یا در قالب یک جدول نمودار گراف (سؤالات پاسخ‌گزیدهٔ پیچیده)	تاریخی روایی، تشریحی، جدلی و ارزیابی
	به‌خاطرسپردن جزئیات مهم متن	تشخیص جمله	
	بیان‌کردن نکات مهم و اطّلاعات فرعی متن	خلاصه کردن: پاسخ دادن به سؤالاتی دربارهٔ نکات مهم و فرعی متن از طریق نوشتن و یا بیان شفاهی	
خواندن برای ادغام اطّلاعات در بین چند متن مختلف	توانایی ساختن یک چارچوب سازمان‌دهی‌شده و برقراری ارتباط بین ایده‌ها و اطّلاعات از چند منبع مختلف	خلاصه‌کردن اطّلاعات دو متن به‌شکل منثور و یا در قالب یک جدول نمودار گراف (سؤالات پاسخ‌گزیدهٔ پیچیده)	
		خلاصه‌کردن: سازمان‌دهی‌کردن و ربط دادن ایده‌ها و اطّلاعات فرعی از چند منبع از طریق نوشتن و یا بیان شفاهی	

٤- روش تحقیق

٤-١- ابزار تحقیق

آزمون‌ها متداول‌ترین ابزارِ گردآوری اطّلاعات هستند (Ary, Donald et al, 2014: 215). اطّلاعات در این پژوهش نیز از طریق برگزاری آزمون بسندگی زبان فارسی در مرکز بین‌المللی آموزش زبان فارسیِ دانشگاه فردوسی مشهد گردآوری گردیده است. این آزمون برای متقاضیان غیرفارسی‌زبان ورود به دانشگاه‌های ایران طرّاحی شده است و داوطلبان باید نمرهٔ حدّ نصاب ٦٠ را در آن کسب کنند. آزمون بسندگی زبان فارسی دانشگاه فردوسی مشهد، شامل چهار مهارتِ خواندن، شنیدن، شفاهی و نگارش و همچنین دستور و واژه می‌شود که بارم هر بخش و مدّت‌زمان موردنیاز برای پاسخ‌گویی به آن در جدول ۵ آمده است.

جدول ۷- بخش‌های مختلف آزمون بسندگی فارسی دانشگاه فردوسی مشهد

نام مهارت	تعداد پرسش‌ها	زمان (دقیقه)	بارم
شنیدن	۱۱	۳۰	۱۰ نمره
خواندن	۲۵	٦٠	۲۵ نمره
دستور و واژگان	۲۰	۳۰	۱۰ نمره
نگارش	شامل یک قسمت	٦٠	۱۵ نمره
شفاهی	------	۱۰	۲۰ نمره

ما در ادامه، تنها به بررسی داده‌های بخشِ خواندن آن خواهیم پرداخت. این بخش از آزمون از سه متن حدوداً ۵۰۰ کلمه‌ای تشکیل شده است و درمجموع ۲۵ سؤال را دربر می‌گیرد. ویژگی سؤالات این آزمون به‌تفصیل در جدول ۸ آمده است.

جدول ۸ـ سؤالات بخشِ خواندن آزمون بسندگی زبان فارسی دانشگاه فردوسی مشهد

شمارهٔ متن	نوع متن	شمارهٔ سؤال	نوع اطّلاعات خواسته‌شده	نوع مطابقت
۱	تشریحی	۱	موضوع	تولیدی
		۲	صفت	سلسله‌ای
		۳	تبیین	محل‌یابی

ادامه جدول ۸

شمارهٔ متن	نوع متن	شمارهٔ سؤال	نوع اطّلاعات خواسته‌شده	نوع مطابقت
۱	تشریحی	٤	علّت	تولیدی
		٥	علّت	تولیدی
		٦	شباهت	ترکیبی
۲	تشریحی	۷	موضوع	تولیدی
		۸	ارجاع ضمیر	محل‌یابی
		۹	ارجاع ضمیر	محل‌یابی
		۱۰	مکان	محل‌یابی
		۱۱	کارکرد	سلسله‌ای
		۱۲	ارجاع ضمیر	محل‌یابی
		۱۳	مقدار	سلسله‌ای
		١٤	محل	محل‌یابی
		۱٥	موضوع	تولیدی
۳	تشریحی	١٦	مکان	محل‌یابی
		۱۷	شیء	ترکیبی
		۱۸	ارجاع ضمیر	محل‌یابی
		۱۹	مراحل	سلسله‌ای
		۲۰	شیء	محل‌یابی
		۲۱	تفاوت	ترکیبی
		۲۲	تفاوت	ترکیبی
		۲۳	موضوع	تولیدی
		٢٤	موضوع	تولیدی
		۲٥	محل	سلسله‌ای

٤-۲- شرکت‌کنندگان

تعداد ۹۰ نفر در این آزمون شرکت کردند که ۳۳ نفر آن‌ها زن و ۵۷ نفر مرد بودند. عراق با ٦۳ نفر بیشترین میزان شرکت‌کننده در این آزمون را داشته‌است و سایر شرکت‌کنندگان از کشورهای اندونزی،

لبنان، سوریه، استرالیا، کرهٔجنوبی، پاکستان و یمن بوده‌اند. از نظر رشتهٔ تحصیلی، گروه مهندسی با ۳۸ نفر، پُرمخاطب‌ترین حوزهٔ دانشگاهی را تشکیل می‌دهد. علوم‌انسانی با ۱۷ نفر، علوم‌پایه ۱۳ نفر و پزشکی ۷ نفر در رتبه‌های بعدی از نظر تعداد شرکت‌کننده بوده‌اند. ۱۵ نفر از شرکت‌کنندگان نیز رشتهٔ تحصیلی‌شان نامشخّص بوده است. سطح تحصیلی برخی از شزکت‌کنندگان کارشناسی و برخی کارشناسی ارشد و دکتری بوده است.

٤-۳- شیوهٔ جمع‌آوری و تحلیل داده‌ها

آزمون بسندگی زبان فارسی دانشگاه فردوسی مشهد در ساعت هشت روز ۲۵ مرداد ۱۳۹٦ در مرکز بین‌المللی زبان فارسی این دانشگاه در مدّت سه‌ساعت‌ونیم برگزار گردید. ازآنجایی‌که بخشِ خواندن قسمت دوم این آزمون را به خود اختصاص می‌دهد و نیز بعد از بخش شنیدن قرار می‌گیرد، پاسخ‌گویی به سؤالات این بخش در ساعت ۸:۳۰ آغاز شد و داوطلبان در مدّت‌زمان حدود ٦۰ دقیقه به ۲۵ سؤال آزمون پاسخ دادند. در این آزمون، برگهٔ خواندن و بخش دستور و واژهٔ آزمون هم‌زمان به داوطلبان داده می‌شود و اگر داوطلبی بخواهد، می‌تواند از زمان بخش سوم آزمون، یعنی دستور و واژه، برای پاسخ‌گویی به سؤالات خواندن استفاده کند. تمامی برگه‌ها در ساعت ۱۰ جمع‌آوری گردید.

داوطلبان در ده کلاس مختلف بودند و در هر کلاس، حدود ۹ آزمون‌دهنده و یک مراقب حضور داشت. پس از آزمون، مدرّسان مجرب مرکز به کمک کلید آزمون که پیش‌تر طرّاحی شده بود، برگه‌ها را تصحیح کردند.

۵- نتایج

نرم‌افزار وینستپس[1] (Linacre,2009)۲۵ پرسش بخش خواندن را توسّط رش تحلیل کرد. آماره‌های مربوط به پرسش‌ها در جدول ۱ آمده است. ستون {measure} دشواری پرسش‌ها را نشان می‌دهد. هرچه مقدار آن بیشتر باشد، پرسش دشوارتر است. جدول نشان می‌دهد که پرسش ۲۵ دشوارترین و پرسش ۱۰ ساده‌ترین پرسش‌ها هستند.

1. WINSTEPS

ستون‌های درون‌برازش[1] و برون‌برازش[2]، برازش پرسش‌ها به مدل راش را نشان می‌دهند. مقادیر MNSQ باید بین ۰/۷۰ تا ۱/۳۰ باشند (Bond and Fox, 2008). برازش پرسش بدین‌معناست که پرسش با دیگر پرسش‌های آزمون هم‌ردیف است و به تعریف یک مقیاس برای سنجش سازهٔ موردنظر ـــ که در اینجا توانایی خواندن است ـــ کمک می‌کند. به عبارت دیگر، پرسش‌ها همخوان با سازهٔ موردنظر هستند. آماره‌های برازش پرسش‌ها همچنین نشان می‌دهند که آیا آزمون تک‌بعدی است[3] یا خیر. لازم به ذکر ات که یکی از شروط استفاده از مدل آماری رش تک‌بعدی بودن آزمون است. جدول ۷ بیانگر این است که همهٔ پرسش‌ها به‌جز پرسش ۲۱ آماره‌های درون‌برازش و برون‌برازش مناسب دارند؛ بنابراین آزمون تک‌بعدی است. مقدار برون‌برازش پرسش اوّل، ۱/۵۳ است که از بیشینهٔ مقدار پیشنهادی بیشتر است. این پرسش چندگزینه‌ای و ساده است (دشواری آن ۰۵۵/- است) و با توجّه به اینکه برون‌برازش آن زیاد است، احتمالاً دوپهلو است و باعث سردرگمی برخی از آزمون‌دهندگان‌می‌شود. همچنین نتیجه نشان داد که پایایی رش آزمون ۰/۷٦ است که این عدد برای یک آزمون سرنوشت‌ساز[4] چندان رقم بالایی نیست. لازم به ذکر است که پایای آلفای کرونباخ آزمون ۰/۷۵ است. این در حالیست که برای سنجش مهارت خواندن پایایی آزمون بهتر است بین ۰/۹۰ تا ۰/۹۹ باشد (Lado, 1961 cited in Hughes,1989 :32).

۵-۱- نقشهٔ آزمودنی ـ پرسش

شکل ٤ نقشهٔ رایت[5] یا نقشهٔ آزمودنی ـــ پرسش است. در این نقشه سطح توانایی زبانی آزمون‌دهندگان و سطح دشواری سؤالات به طور همزمان نمایش داده می‌شود و در نتیجه مستقیماً قابل‌مقایسه هستند و همانگونه که پیشتر گفته شد این امکان مقایسه همزمان، از جمله مزیت‌های مدل رش است که در مدل‌های کلاسیک وجود ندارد. در سمت راست نقشه، سطح سختی پرسش‌ها و در سمت چپ سطح توانایی آزمون‌دهندگان قرار دارند. انتظار می‌رود که سؤالات آزمون متناسب با سطوحِ زبانیِ مختلفِ آزمون‌دهندگان طراحی شده باشند تا توانایی تمییز داوطلبان ضعیف، متوسط و قوی را داشته

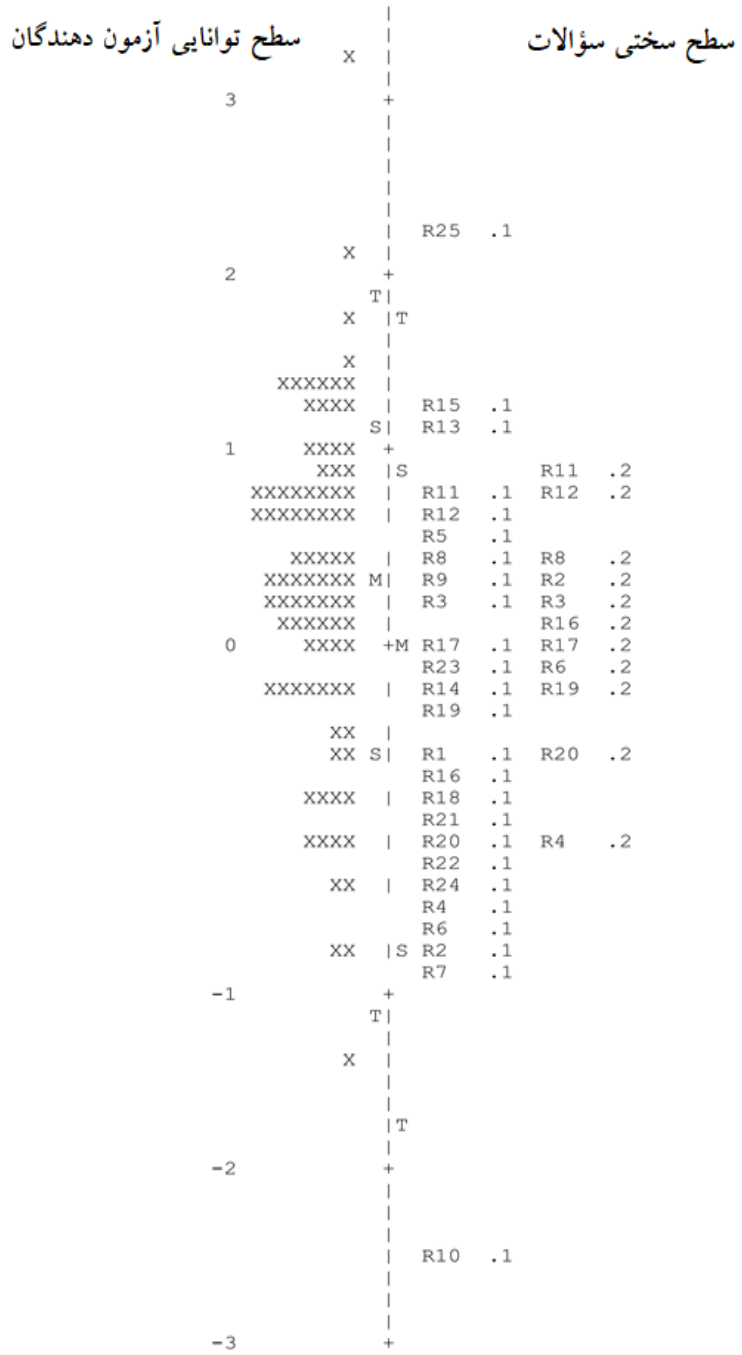
1. Infit
2. Outfit
3. Unidimensional
4. High Stake
5. Wright Map

باشـند، به عبارتی تودهٔ سـؤالات در سـمت راست باید مقابل تودهٔ آزمون‌دهندگان در سمت چپ قرار بگیرند. اگر در نقشـه‌ای در مقابل پرسـش و مجموعه‌ای از پرسـش‌ها، هیچ داوطلبی قرار نگیرد و یا تعداد کمی داوطلب دیده شـود؛ این نکته نشـان می‌دهد که آن پرسش یا آن مجموعهٔ پرسش‌ها کارکرد چندانی در تمییز توانایی داوطلبان از یکدیگر ندارند و باید مورد بازبینی قرار بگیرند.

بررسـی نقشـه نشـان می‌دهد که هرچند تا حدودی تودهٔ پرسش‌ها مقابل تودهٔ آزمون‌دهندگان قرار گرفته اسـت، امّا بااین‌حال آزمون به تعدادی پرسـش با سـطح دشواری بالاتر نیاز دارد. تنها پرسش‌های ۲۵ و ۱۵ آزمون از نظر سـختی، متناسـب با ١٤ داوطلب نخست آزمون هستند. همچنین نقشهٔ آزمودنی برای ٣٦ داوطلب اوّل فقط ٤ سـؤال را نسبتاً هم‌سطح نشان می‌دهد؛ این بدان معناست که آزمون عمدتاً در سطح آزمون‌دهندگان با توانمندی متوسّط و ضعیف است.

جدول ۹- اندازهٔ دشواری پرسش‌ها و برازش آن‌ها به مدل راش

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM	G
1	59	90	-.33	.23	1.06	.7	1.08	.7	.21	.30	64.4	68.2	R1	D
2	117	90	-.21	.15	.81	-1.7	.77	-1.7	.57	.42	45.6	47.9	R2	0
3	98	90	.25	.13	1.17	1.5	1.20	.9	.39	.49	34.4	36.5	R3	0
4	146	90	-.64	.16	1.15	.9	1.19	.6	.30	.40	61.1	64.8	R4	0
5	39	90	.68	.22	.97	-.3	.96	-.4	.35	.32	66.7	64.0	R5	D
6	127	90	-.38	.15	.84	-1.2	.74	-1.5	.55	.42	56.7	51.0	R6	0
7	68	90	-.86	.26	.98	-.1	1.02	.2	.28	.27	74.4	75.9	R7	D
8	80	90	.53	.12	.95	-.4	.99	.0	.52	.50	40.0	39.0	R8	0
9	44	90	.43	.22	1.07	1.0	1.05	.7	.23	.32	57.8	62.9	R9	D
10	84	90	-2.46	.43	1.00	.1	1.14	.4	.14	.16	93.3	93.3	R10	D
11	60	90	.87	.13	.97	-.2	.89	-.4	.51	.49	44.4	43.7	R11	0
12	71	90	.67	.13	1.09	.9	1.13	.6	.44	.50	34.4	37.3	R12	0
13	31	90	1.10	.23	.96	-.4	.96	-.3	.36	.31	72.2	68.6	R13	D
14	54	90	-.07	.23	1.11	1.4	1.12	1.1	.16	.31	62.2	65.2	R14	D
15	29	90	1.21	.24	1.07	.7	1.11	.8	.21	.31	70.0	70.1	R15	D
16	116	90	-.10	.14	.88	-1.0	.83	-.9	.53	.45	37.8	41.0	R16	0
17	113	90	.01	.13	.91	-.8	.79	-.7	.54	.48	44.4	44.5	R17	0
18	61	90	-.44	.24	.97	-.3	.97	-.2	.33	.29	71.1	69.7	R18	D
19	124	90	-.16	.13	.87	-1.1	.91	-.1	.53	.47	54.4	51.7	R19	0
20	136	90	-.45	.15	.98	-.1	.81	-.6	.46	.42	61.1	57.5	R20	0
21	63	90	-.55	.24	1.07	.7	1.53	3.1	.13	.28	71.1	71.2	R21	D
22	64	90	-.61	.24	1.00	.1	1.00	.0	.27	.28	72.2	72.1	R22	D
23	53	90	-.02	.23	1.05	.7	1.08	.8	.23	.31	62.2	64.8	R23	D
24	66	90	-.73	.25	.96	-.3	1.00	.1	.31	.27	76.7	73.9	R24	D
25	14	90	2.25	.30	1.03	.2	1.08	.4	.21	.27	85.6	85.0	R25	D
MEAN	76.7	90.0	.00	.20	1.00	.0	1.01	.1			60.6	60.8		
S.D.	35.4	.0	.87	.07	.09	.8	.17	.9			15.4	14.9		



شکل ٤- نقشهٔ آزمونی-پرسش

٦- بحث و نتیجه‌گیری

از رهگذر این پژوهش مشخص شد که آزمون بسندگی زبان فارسی دانشگاه فردوسی مشهد از پایایی ۰/۷٦ برخوردار است؛ یعنی اگر شخصی در دفعات متفاوت این آزمون را بدهد (با این فرض که دانش زبانی‌اش تغییری نکند) احتمال دارد نمرات متفاوتی کسب کند و این یک ضعف برای آزمون محسوب می‌شود. پایایی آزمون خواندن بهتر است حدود ۰/۹۰ تا ۰/۹۹ باشد. به عنوان مثال در آزمون تافل پایایی بخش خواندن ۰/۸۷ است (ETS, 2011: 7). بنابراین طراحان آزمون بسندگی زبان فارسی دانشگاه فردوسی مشهد بهتر است پایایی آزمون خود را ارتقاء دهند.

از سوی دیگر بررسی نقشهٔ آزمودنی-پرسش نیز حاکی از این بود که سؤالات آزمون عمدتاً در سطح آزمون‌دهندگان متوسّط و ضعیف هستند؛ یعنی آزمون نمی‌تواند آزمون‌دهندگان سطح پیشرفته را به طور دقیق بسنجد. به منظور پی بردن به علت سادگی پرسش‌ها و راه‌های دشوارتر کردن می‌توان آنها را از منظر نوع اطلاعات خواسته شده و نوع مطابقت بررسی کرد.

همانطور که گفته شد آزمون دارای سه متن و بیست و پنج پرسش است که در مقایسه با آزمون‌های بسندگی در زبان انگلیسی مانند تافل[1] و آیلتس[2] که هر کدام حدود ٤۰ پرسش دارند (Cullen et al., 2014:52; ETS, 2009:37) از تعداد سؤال کمتری برخوردار است که خود می‌تواند یکی از دلایل پایین بودن پایایی آن باشد. اما سؤالات آزمون از نظر نوع اطّلاعات خواسته‌شده[3] دارای تنوع خوبی بوده و عمدتاً از سطح سختی متوسّط و یا بالایی برخوردار هستند (هفت پرسش در سطح سختی ۵، چهار پرسش در سطح سختی ٤، شش پرسش در سطح سختی ۳ و هشت پرسش در سطح سختی دو و یک). همچنین به‌کارگیری متون بلند حدود ۵۰۰ کلمه‌ای که نمونهٔ نسبتاً مناسبی از متون دانشگاهی هستند، ازجمله ویژگی‌های مثبت دیگر این آزمون به‌شمار می‌آید. هرچند پیشنهاد می‌شود از متون بلندتر حدود ٦۵۰ کلمه‌ای نیز در آزمون استفاده شود؛ زیرا این‌گونه متون مدل‌های متنی[4] بزرگ‌تری را می‌سنجند و تصویر واقعی‌تری از خواندن در محیط

1. TOEFL
2. IELTS
3. Type of Questions Requested
4. Text Model

دانشـــگـاه ارائـه می‌دهنـد. لازم بـه ذکر متون بکـار رفتـه در آزمون تـافـل حـدود ۷۰۰ واژه دارنـد (ETS,2009: 7).

با این حال ریشـهٔ اصـلی سـادگی آزمون را باید در نوع مطابقت‌ها دانست. ١٤ سؤال آزمون از نوع محل‌یابی و یا سـلسله‌ای هستند که از سطح سختی پایینی برخوردارند (پرسش‌های شمارهٔ ۲-۳-۸-۹-۱۰-۱۱-۱۲-۱۳-١٤-١٦-۱۸-۱۹-۲۰-۲۵) بکارگیری پرســش‌های بیشــتری از نوع ترکیبی و تولیدی می‌تواند به ســخت‌ترشــدن ســؤالات بینجامد. همچنین تمامی متون این آزمون از جنس تشــریحی و روایی هسـتند و می‌بایست متون جدلی برای طرّاحی سؤال به‌کار گرفته شوند، زیرا این نوع متون سطح سختی بالاتری دارند.

از منظر اهداف خواندن سـؤالات عمدتاً دو هدف را دنبال می‌کنند: پیداکردن اطّلاعات و خواندن برای رسـیدن به درک مقدّماتی. بدونِ‌تردید آزمون به سؤالاتی متناسب با اهداف خواندن برای یادگیری و نیز خواندن برای ادغام اطّلاعات در بین چند متن مختلف، نیاز دارد که از ســطح دشــواری بالاتری برخوردار هستند. به عنوان مثال به یک نمونه پرسش برگرفته از آزمون تافل اشاره می‌کنیم:

– با توجه به متن از میان گزینه‌های زیر مشــخص کنید کدام مورد مربوط به عکاســی در دورهٔ پهلوی و کدام یک مربوط به دورهٔ قاجار می‌شود (۲ گزینه اضافی است و این سؤال ۲ نمره دارد.)

دورهٔ قاجار:	دورهٔ پهلوی:

– اعزام دانشجو برای آموختن فن عکاسی به اروپا
– ترساندن مردم از عکاسی
– متداول شدن عکس در میان عموم مردم
– علاقه‌مندی شخص پادشاه به عکاسی
– اعمال محدودیت برای گرفتن عکس از زنان

در بخش اجرا نیز مهارتِ خواندن می‌بایست از سایر بخش‌ها جدا باشد تا آزمون‌دهندگان نتوانند از زمان سایر بخش‌ها برای پاسخ‌گویی سؤالات خواندن استفاده کنند. بدیهی است که این امر خود می‌تواند در کاهش پایایی آزمون تأثیرگذار باشد.

کتابنامه

جلیلی، سیّد اکبر. (۱۳۹۰). «آزمون مهارتی فارسی (آمفا) برپایهٔ چهار مهارت اصلی زبانی». پایان‌نامهٔ کارشناسی ارشد. دانشکدهٔ ادبیات فارسی و زبان‌های خارجی دانشگاه علاّمه طباطبائی.

غفّارثمر، رضا؛ شیرازی‌زاده، محسن؛ کیانی، غلامرضا. (۱۳۹۲). «بسترها، چشم‌اندازها، کاربردها و چالش‌های مطالعهٔ واژگان در متون دانشگاهی: ضرورت توجّه بیشتر به زبان فارسی و زبان‌آموزان فارسی زبان». *دوماهنامهٔ جستارهای زبانی*، ٦ (٤)، (پیاپی ۲۵). صص ۱۵۳ـ۱۸۱.

قنسولی، بهزاد. (۱۳۸۹). «طرّاحی و روایی‌سنجی آزمون بسندگی زبان فارسی». *پژوهش زبان‌های خارجی*، (پیاپی ۵۷)، بهار. صص ۱۱۵ـ ۱۲۹.

قنسولی، بهزاد. (۱۳۹۲). *برون‌فکنی اندیشه در مهارت خواندن در زبان دوم*. مشهد: تابران

گلپور، لیلا. (۱۳۹٤). «طرّاحی و اعتباربخشی آزمون بسندگی زبان فارسی برپایهٔ چهار مهارت زبانی». پایان‌نامهٔ دکتری. دانشکدهٔ ادبیات فارسی و زبان‌های خارجی دانشگاه پیام‌نور مرکز.

موسوی، ثریّا. (۱۳۷۹). «طرّاحی و اعتباربخشی آزمون استاندارد زبان فارسی برپایهٔ چهار مهارت زبانی». پایان‌نامهٔ کارشناسی ارشد. دانشکدهٔ ادبیات و علوم انسانی دانشگاه شیراز.

Ary, D., Jacobs, L. C., Sorensen, C., Walker, D. A. (2014). *Introduction to Research in Education*. Belmont: Wadsworth.

Bachman, Lyle. (1990). *Fundamental Considerations in Language Testing*. New York: Oxford University Press.

Bachman, Lyle. (2004). *Statistical Analysis for Language Assessment*. New York: Cambridge University Press.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Chapelle, C. A., Enright M. K., Jamieson, J. M. (2008). *Test Score Interpretation and Use*. In *Building a Validity Argument for the Test of English as a Foreign Language* Edited by Chapelle, C. A., Enright M. K., Jamieson, J. M. New York: Routledge. pp. 1-25

Chen, Jing; Sheehan, Kathleen M. (2015). "Analyzing and Comparing Reading Stimulus Materials Across the TOEFL Family of Assessments". (TOEFLiBT-26). Princeton, NJ: Educational Testing system.

Cullen, P., French, A., Jakeman, V. (2014).*The Official Cambridge Guide to IELTS*. China: Cambridge University Press.

Enright, M. K., Grabe, W., Koda, K., Mosenthal, P., Mulkahy-Ernt, P., & Shedl, M. (2000). *TOEFL 2000 Reading Framework: A Working Paper*. (TOEFL monograph No. 17). Princeton, NJ: Educational Testing system.

ETS. (2009). *The Official guide to the TOEFL Test*. Ed. 3. New York: McGraw-Hill

ETS. (2011). Reliability and Comparability of TOEFL IBT scores. *TOEFL Research Insight Series*. Vol.3.

Grabe, W. Stoller F. L. (2013). *Teaching and Researching Reading*. Ed. 2. New York: Routledge.

Grabe, William (2009*). Reading in a Second Language: Moving from Theory to Practice*. New York: Cambridge University Press.

Hudson, Thom. (1996). *Assessing Second Language Academic Reading from a Communicative Perspective: Relevance for TOEFL 2000.* (TOEFL monograph No. 17). Princeton, NJ: Educational Testing system.

Hughes, Artur. (1989). *Testing for Language Teaching*. New York: Cambridge University Press.

Jamieson, J. M., Eigner, D., Grabe W., Kunnan, A. J. (2008). "*Frameworks for a new TOEFL*". In Building a Validity Argument for the Test of English as a Foreign Language Edited by Chapelle, C. A., Enright M. K., Jamieson, J. M. New York: Routledge. pp. 1-25.

Jamieson, J.Jones, S., Kirsch, I., Mosenthal, P. & Taylor, C. (2000). *TOEFL 2000 Framework: A Working Paper* (TOEFL monograph No. 16). Princeton, NJ: Educational Testing system.

Lado, R. (1961). Language Testing. London: Longman.

Linacre, J. M. (2009). *WINSTEPS Rasch Measurement* [Computer program]. Chicago, IL: Winsteps.

Liu, O. L., M. Schedl, J. Malloy, & N. Kong. (2009). "Does Content Knowledge Affect TOEFL iBT Reading Performance? A Confirmatory Approach to Differential Item Functioning". (ETS Research Rep. No. TOEFLiBT-09). Princeton, NJ: Educational Testing system.

Messick, Samuel. (1994). "The interplay of evidence and Consequences in the Validation of Performance Assessment". *Educational Researcher*, 23(2), 13-23.

Messick, Samuel. (1994). "The interplay of evidence and consequences in the validation of performance assessments". *Educational Researcher*, 23(2), 13-23.

Mousavi, Seyyed Abbas. (2012). "Item Response Theory". In *An Encyclopedic Dictionary of Language Testing*. 5th ed. Tehran. Rahnama.

Patricia, C, Gordon, Ann, Schedl Mary A., Tang k. Linda. (1996) "An Analysis of the Dimensionality of TOEFL Reading Comprehension Items" (ETS Research Rep. TOEFL-RR-53). Princeton, NJ: Educational Testing system.