

بررسی بخش پذیری سازه درک مطلب در زبان خارجی (مطالعه‌ای با استفاده از مدل پرسش - پاسخ چند بعدی)

پوریا بقایی مقدم (دانشگاه آزاد اسلامی، واحد مشهد، گروه زبان انگلیسی، مشهد)

pbaghaei@mshdiau.ac.ir

محسن محمود روشن ضمیر (دانشگاه آزاد اسلامی، واحد مشهد، گروه زبان انگلیسی، مشهد)

چکیده

چگونگی بخش پذیری توانایی عمومی زبان خارجه و مهارت‌های آن، پرسشی است که سال‌هاست توجه زبان شناسان کاربردی را به خود معطوف داشته است. در گذشته تصور می‌شد توانایی عمومی زبان خارجه و مهارت‌های آن، واحد و غیر قابل تفکیک هستند؛ اما نظریات جدید نه تنها توانایی عمومی زبان خارجه را بخش پذیر می‌دانند، بلکه معتقدند مهارت‌های زبانی هم قابل تقسیم به زیرمهارت‌ها هستند. با این حال، محققین در مورد تعداد و ماهیت زیرمهارت‌هایی که مهارت‌های اصلی شنیدن، صحبت کردن، خواندن و نوشتن را تشکیل می‌دهند، اتفاق نظر ندارند. هدف این مطالعه، بررسی بخش پذیری سازه درک مطلب (خواندن و شنیدن) در زبان خارجی است. برای این منظور یک آزمون درک مطلب زبان انگلیسی، که از دو بخش خواندن و شنیدن تشکیل شده بود، به وسیله مدل پرسش - پاسخ چند بعدی تحلیل شد. سه مدل رقیب با یک بعد، دو بعد و چهار بعد برآزش شدند. یافته‌ها نشان داد مدل چهار بعدی برآزش بهتری نسبت به مدل دو بعدی دارد و مدل دو بعدی برآزش بهتری نسبت به مدل تک بعدی دارد. این بدان معناست که سازه درک در زبان دوم قابل تفکیک به زیرمهارت‌هاست. کاربرد یافته‌ها در تدریس و آزمون زبان‌های خارجه بحث شده است.

کلیدواژه‌ها: درک در زبان خارجی، بخش پذیری، تئوری پرسش - پاسخ چند بعدی، برآزش.

۱- مقدمه

مدت‌هاست این سؤال که: «آیا توانایی زبان در انسان، واحد و یا بخش پذیر است؟» ذهن محققین را به خود مشغول کرده است. به‌طورکلی بر اساس نتایج تحقیقات متعدد، دانش زبان به احتمال زیاد از یک عامل مرتبه بالا و چندین عامل توانایی مرتبه اول مجزا از هم تشکیل شده است (باکمان و پالمر، ۱۹۸۱؛ باکمان، دیویدسون و فولکز، ۱۹۹۰؛ باکمان، دیویدسون، راین و چویی، ۱۹۹۵).

هدف این مطالعه، بررسی بخش‌پذیری ساختار درک مطلب در زبان دوم است. از نظر تئوری، مهارت‌های شنیدن و خواندن، فرایندهای چند بعدی و پیچیده‌ایی هستند و تلاش‌های زیادی برای توصیف آنها از لحاظ دسته بندی زیرمهارت‌هایی که زیر بنای این مهارت‌ها محسوب می‌شوند تا کنون صورت گرفته است. آن‌گونه که الدرسون و لوکمانی (۱۹۸۹) اعلام کرده اند، محققین زیادی تعداد و انواع متفاوتی از زیرمهارت‌ها را براساس تحقیقات تجربی و غیر تجربی برای اثبات ادعایشان ذکر کرده اند (مامبی، ۱۹۷۸؛ ریچاردز، ۱۹۸۳؛ ویبر، ۱۹۹۳). اما هنوز اطلاعات دقیقی درباره اینکه آیا مهارت‌های درک مطلب بخش پذیرند یا نه و اینکه زیرمهارت‌های آن از چه اجزایی تشکیل شده اند و دسته بندی آنها در دست نیست (الدرسون، ۲۰۰۰). مساله دیگری که هنوز لاینحل باقی مانده این است که آیا زیرمهارت‌های تشکیل دهنده دو مهارت شنیدن و خواندن در زبان دوم یکسان هستند یا خیر؟ (سانگ، ۲۰۰۸).

اگرچه مفهوم زیرمهارت‌ها بحث انگیز است، اما در عمل، تعیین زیرمهارت‌ها یا سطوح مختلف درک توسط معلمین زبان و تهیه کنندگان آزمون و مواد آموزشی بسیار رایج است. فهرست زیرمهارت‌ها معمولاً به عنوان پایه‌ای برای طرح درس، توصیف معلومات ذهنی دانش آموزان، و تهیه سؤالات آزمون به زبان مادری و زبان خارجی به کار می‌رود (الدرسون و لوکمانی، ۱۹۸۹؛ الدرسون، ۲۰۰۰). به عنوان مثال، معلمی ممکن است در مورد توانایی زبان آموزانش در درک مفاهیم مستقیم و مفاهیم تلویحی صحبت کند. یا معلمی ممکن است تشخیص دهد زبان آموزانش در یافتن اطلاعات ریز در متن یا حدس زدن واژگان دارای مشکل هستند. نویسندگان کتاب‌های آموزشی هم در تهیه مواد، زیرمهارت‌ها را در نظر می‌گیرند و فعالیت‌های هدفمندی تهیه می‌کنند که زیرمهارت‌های خاصی را در زبان آموزان تقویت کنند. به همین ترتیب، تهیه کنندگان آزمون‌های زبان، زیرمهارت‌هایی را که باید اندازه گیری شوند مشخص کرده سپس پرسش‌هایی می‌نویسند که هریک از زیرمهارت‌های تعیین شده را بسنجند؛ یعنی آزمون سازان آزمون‌هایی

را که دربرگیرنده چندین زیرمهارت هستند تعریف می‌کنند و در آزمون‌هایشان سؤالاتی را جهت سنجش آن زیرمهارت‌ها می‌گنجانند. این‌گونه رویکردها دلیلی است بر اینکه مدرسین و آزمون‌سازان زبان، خودآگاه و یا ناخودآگاه به وجود زیرمهارت‌ها و اهمیت تدریس و سنجش آنها به صورت مجزا معتقدند. این مطالعه با استفاده از مدل پرسش- پاسخ چند بعدی، ساختار عاملی^۱ سازه درک در زبان دوم را از نظر تعداد و ماهیت زیرمهارت‌های تشکیل دهنده آن بررسی می‌کند. گرچه مطالعات بسیاری در زمینه ساختار دانش عمومی زبان دوم و قابلیت بخش پذیری مهارت‌های زبان دوم انجام شده است (باکمان و پالمر ۱۹۸۱، ۱۹۸۲؛ باک، ۱۹۹۱، ۱۹۹۲) اما هیچ مطالعه‌ای بر روی بخش پذیری مهارت‌های درک مطلب با استفاده از مدل پرسش- پاسخ چند بعدی صورت نگرفته است. مساله مورد مطالعه در این تحقیق این است که تا چه حد سوالات چند گزینه‌ای مهارت‌های شنیدن و خواندن و انواع متفاوتی را از زیرمهارت‌های درک مطلب در زبان دوم مورد ارزیابی قرار می‌دهند.

پیشینه تحقیق

یکی از مهم‌ترین مسائلی که به ارزشیابی زیرمهارت‌های درک مطلب کمک می‌کند دانستن تعداد عوامل زیر بنایی، بخش پذیری و ماهیت این مهارت‌هاست. بیشتر پژوهش‌های تجربی در زمینه زیرمهارت‌های درک مطلب از طریق استفاده از تحلیل عاملی اکتشافی (باک و تاتسوکا، ۱۹۹۸) و نیز تحلیل رگرسیون چند گانه و عمدتاً در زمینه خواندن و درک مطلب (درام، کالفی و کوک، ۱۹۸۱؛ دی وی، ۱۹۸۸) انجام شده است. در ارتباط با بخش پذیری مهارت‌ها، محققین هستند مهارت خواندن را می‌توان حداقل به سه گروه تقسیم کرد. گروه اول معتقدند مدارک و شواهد دال بر وجود زیرمهارت‌های جداگانه بندرت یافت می‌شوند، بنابراین خواندن یک مهارت واحد و به هم پیوسته است (روست، ۱۹۹۳).

گروه دوم بر این اعتقادند که خواندن به چندین زیرمهارت قابل بخش است. اگرچه که مستندات تعداد زیرمهارت‌ها را به شکل تجربی و قابل تشخیص نشان دهند، وجود ندارد. بسیاری از محققین به تعداد قابل توجهی از زیرمهارت‌ها از ۳ تا ۳۶ عدد اشاره کرده اند (الدرسون و لوکمانی، ۱۹۸۹). اما لنون

^۱ factor structure

(۱۹۶۲) اشاره می‌کند که تنها ممکن است بطور معتبر فقط حد اکثر چهار زیرمهارت خواندن را شناسایی و ارزیابی کرد. این چهار زیرمهارت عبارتند از: دانش لغت، درک مفاهیمی که به صورت مستقیم بیان شده‌اند، درک مفاهیم تلویحی، و لذت ادبی. از طرف دیگر، کارول (۱۹۹۳) بیش از ۳۰ مطالعه تحلیل عاملی را مورد بررسی مجدد قرار داد و چهار عامل مشترک را در خواندن شناسایی نمود که عبارتند از: خواندن و درک مطلب (عمومی)، خواندن و درک مطلب تخصصی، رمز گشایی خواندن و سرعت در خواندن. به همین روش، تحقیقات متفاوت در زمینه مهارت شنیدن نشان دادند که این مهارت نیز در بر گیرنده مجموعه‌ای از زیرمهارت‌هاست، اما درباره اینکه این زیرمهارت‌ها دقیقا چه ماهیتی دارند توافقی وجود ندارد.

گروه سوم معتقدند خواندن به دو فرایند رمز گشایی (تشخیص کلمه) و درک مطلب بخش پذیر است. رمز گشایی شامل فهم و تحلیل محرک شنیداری و یا چاپی به شکل واحدهای زبانی است و درک مطلب شامل ساخت معنا با استفاده از زبان رمز گشایی شده و ترکیب آن با دانش پیشین فرد درک کننده است (لاند، ۱۹۹۱). این فرایند درک در خواندن و شنیدن مشترک است؛ در حالی که فرایند رمز گشایی کاملا متفاوت می‌باشد. بنابراین درک مطلب اساسا فرایندی است مرکزی که مستقل از کانال دریافت پیام (شنیداری و دیداری) است. این بدان معناست که درک در شنیدن و خواندن فرایند یکسانی است. به همین ترتیب مهارت شنیدن را نیز می‌توان به یک فرایند دو مرحله‌ای تقسیم بندی نمود. مرحله اول، که در آن اطلاعات اساسی زبان شناختی استخراج می‌شوند، و سپس مرحله دوم که در آن اطلاعات برای اهداف ارتباطی مناسب مورد استفاده قرار می‌گیرند (کلارک و کلارک، ۱۹۷۷).

در درک مطلب در زبان دوم، تفاوت بین شنیدن و خواندن مرز بندی دقیقی ندارد. لوین و ریورز (۱۹۸۸) با استفاده از تحلیل رگرسیون چند گانه وهم بستگی به این نتیجه رسیدند که دو مهارت خواندن و شنیدن می‌توانند از یک عامل واحد و چندین زیرمهارت تشکیل شده باشند. آنها همچنین مطرح کردند که شنیدن و خواندن ممکن است از یکدیگر مجزا باشند، بدین ترتیب که شنیدن یک مهارت بهم پیوسته است و درک کلی تری از متن را نسبت به خواندن در بر می‌گیرد. باک (۱۹۹۲) و بائه و باکمان (۱۹۹۸) با استفاده از تحلیل عاملی تاییدی، تاکید کردند که همپوشانی قابل توجهی بین توانایی شنیدن و خواندن وجود دارد. اما نباید فراموش کرد که هر کدام از این مهارت‌ها دارای جنبه‌های منحصر به خود نیز هستند.

علاوه بر این، فریدل و کاستین (۱۹۹۴، ۱۹۹۹) زیرمهارت‌های شنیدن را با زیرمهارت‌های خواندن مقایسه نمودند و به نتایج مشابهی دست یافتند. براساس شواهد تجربی و ملاحظات نظری، باک (۲۰۰۱) بیان داشت که بسیاری از ویژگی‌های درک مطلب در شنیدن و خواندن، در واقع ویژگی‌های همه انواع درک هستند، اما هر مهارت در برخی از جنبه‌ها منحصر به فرد است.

نظریه پرسش - پاسخ چند بعدی

یکی از بنیادی‌ترین فرضیه‌های مدل‌های نظریه پرسش-پاسخ^۱ فرضیه تک بعدی^۲ بودن آنهاست. تک بعدی بودن به این معناست که تمام سؤالات در یک آزمون باید یک مهارت یا بعد را مورد ارزشیابی قرار دهند. بطور سنتی، در علوم فیزیکی و انسانی ارزشیابی تک بعدی بوده است. به عبارت دیگر، ما سعی داریم فقط یک متغیر واحد را در هر زمان اندازه‌گیری کنیم و معمولاً اندازه‌ها را با یکدیگر ادغام نمی‌کنیم. به واقیت رساندن مفروضه تک بعدی بودن در عمل بسیار دشوار و یا حتی غیر ممکن است. چند بعدی بودن واقعیت غیر قابل اجتنابی در سنجش و اندازه‌گیری است و مفروضه تک بعدی بودن به آسانی نقض می‌شود. علاوه بر این در بسیاری از آزمون‌های سنجش آموزشی، تعدد آزمون‌های چند بعدی ساخته می‌شوند تا توانایی‌های دانش آموزان در چندین بعد یا زیرمهارت سنجیده شوند. مدل‌های پرسش-پاسخ در مقابل نقض این گونه مفروضه تک بعدی بودن مقاوم هستند، مخصوصاً اگر این ابعاد تا حد زیادی با یکدیگر همبستگی داشته باشند. اما اظهاراتی وجود دارد مبنی بر اینکه زمانی که یک آزمون از چندین زیر آزمون که ابعاد زیر بنایشان به هم مرتبط نیستند تشکیل شده باشد، مدل‌های تک بعدی در برآورد عوامل مدل دقت لازم را ندارند (فالک و گرین، ۱۹۸۹). مدل‌های چند بعدی پرسش-پاسخ، چند بعد را پردازش کرده و اندازه توانایی‌های دانش آموزان را در چندین بعد به صورت همزمان برآورد می‌کنند. برآزش کردن یک مدل تک بعدی به یک آزمون چند بعدی باعث از دست رفتن اطلاعات و ناپدید شدن زیر مقیاس‌ها می‌شود (ادمز، ویلسون و ونگ، ۱۹۹۷).

^۱ Item Response Theory

^۲ unidimensionality principle

مدل‌های چند بعدی برای تحلیل‌های تاییدی

مدل‌های نظریه پرسش- پاسخ چند بعدی^۱ می‌توانند در حالت تاییدی^۲ برای تحقیق در مورد روایی سازه‌ای آزمون‌هایی که از چند بخش و یا زیر مقیاس تشکیل شده اند مورد استفاده قرار گیرند (امبرتسون، ۱۹۸۰/۱۹۸۳؛ بقایی و گروتیان، [زیر چاپ]؛ سنت لیس و کاسپاری، ۲۰۰۹؛ یانسن و دی بک، ۱۹۹۹). اجزا یا زیر مقیاس‌ها در حقیقت فرایندهای ذهنی یا توانایی‌های متفاوتی هستند که در حل مسائل دخیل‌اند. به طور سنتی، روش‌های «تحلیل عاملی»^۳ برای مطالعه ساختار آزمون‌ها مورد استفاده قرار می‌گرفته‌اند. روش‌های تحلیل عاملی بر اساس همبستگی بین زیر مقیاس‌ها بنا شده‌اند. در این تکنیک‌ها اجزای تشکیل دهنده سازه مورد نظر با همبستگی بین زیر مقیاس‌ها شناخته می‌شوند. اما روش‌های پرسش- پاسخ چند بعدی تعامل بین پاسخ دهندگان و پرسش‌ها را مدل می‌کنند تا اجزای تشکیل دهنده آزمون را مشخص کنند.

در سنجش و آزمون سازی نوین، ما به چیزی بیش از خلاصه کردن توانایی‌های آزمون دهندگان در یک نمره واحد علاقه مندیم. ما می‌خواهیم بدانیم آزمون دهندگان هنگام پاسخ به سؤالات از چه راهبردها و مهارت‌هایی استفاده می‌کنند (امبرتسون، ۱۹۸۳؛ ادمز، ویلسون، و ونگ، ۱۹۹۷). مدل‌های پرسش- پاسخ چند بعدی می‌توانند راهبردها و فرایندهای ذهنی زیر بنایی را که آزمون دهندگان در هنگام دست و پنجه نرم کردن با پرسش‌ها از آنها استفاده می‌کنند آشکار سازند. از این رو مدل‌ها نقش مهمی را در درک فرایندهای ذهنی پاسخگویان آزمون می‌توانند داشته باشند. بنابراین از مدل‌های پرسش- پاسخ چند بعدی می‌توان به عنوان تکنیک‌های قدرتمند آماری برای بررسی روایی سازه‌ای آزمون‌ها استفاده کرد.

تحقیقات درباره اجزا ساختار آزمون با مفهوم "بازنمایی سازه"^۴ امبرتسون ارتباط نزدیک دارد. بنا به گفته امبرتسون (۲۰۰۷/۱۹۸۳) بازنمایی سازه که یک جنبه از روایی درونی می‌باشد قابلیت آزمون در بازنمایی فرایندها و راهبردهایی است که آزمون دهندگان هنگام پاسخ به آزمون استفاده می‌کند. بازنمایی سازه، معنای نمرات آزمون در قالب مهارت‌ها، فرایندها، و دانشی است که آزمون دهنده در فرایند آزمون به

^۱ Multidimensional Item Response Theory

^۲ confirmatory

^۳ factor analysis

^۴ construct representation

کار می‌گیرد تا به پرسش‌ها پاسخ دهد. او همچنین مفهوم "گستره فراگیر"^۱ را به عنوان یک جنبه خارجی از روایی سازه‌ای مطرح کرد که همان رابطه بین نمرات یک آزمون و دیگر آزمون‌ها و معیارهای خارجی است که به طور سنتی در تحقیقات روایی مورد استفاده قرار می‌گرفته است.

بررسی بازنمایی سازه بوسیله نظریه پرسش - پاسخ چند بعدی نیاز به یک آزمون با طراحی دقیق مبتنی بر نظریه دارد. در چنین آزمونی فرایندهای ذهنی آزمون دهندگان باید بر اساس یک تئوری مشخص به شکل زیرمهارت‌های آزمون ساخته شوند (امبرتسون، ۱۹۸۳؛ یانسن و دی بک، ۱۹۹۹). سپس مدل‌های نظریه پرسش - پاسخ چند بعدی می‌توانند برای بررسی اعتبار مدل فرض شده، از نظر برازش داده‌ها به مدل و بررسی معیار اطلاعات^۲ به کار روند. اگر مدل ساختاری خاصی از آزمون در مدل پرسش - پاسخ برازش شود، یا اگر یک مدل ساختاری فرضی خاص نسبت به یک مدل رقیب بهتر برازش شود، مدارک و شواهد کافی دال بر تناسب و اعتبار ساختار فرض شده و فرایندهای ذهنی که توسط هر زیر مقیاس فعال می‌شوند به دست آمده است.

مدل‌های رقیب با مقایسه راستی‌نمایی^۳ راه حل‌هایشان با هم مقایسه می‌شوند. هر چه میزان راستی‌نمایی آنها بالاتر می‌رود، مدل برازش شده به مدل حقیقی نزدیک تر می‌شود. لگاریتم راستی‌نمایی^۴ منفی یا انحراف^۵ شاخص تفاوت بین مدل تخمینی و مدل واقعی است. بنابراین انتظار داریم که انحراف کوچک باشد و مدل‌هایی با انحراف‌های کوچک تر انتخاب می‌گردند (یانسن و دی بک، ۱۹۹۹). معیارهای اطلاعات نیز برای مقایسه مدل‌های مختلف مورد استفاده قرار می‌گیرد. این معیارها بر اساس میزان راستی‌نمایی و تعداد پارامتر برآورد شده محاسبه می‌شوند. هر چه معیار اطلاعات به دست آمده کوچک تر باشد، مدل محاسبه شده به مدل واقعی نزدیک تر است و مدل برازش بهتری دارد.

۱ nomothetic span

۲ information criteria

۳ likelihood

۴ loglikelihood

۵ deviance

روش تحقیق

۱- نمونه و ابزار

نمونه مورد استفاده در این مطالعه از ۹۱۲ دانش آموز دبیرستانی ایرانی در مقطع اول و دوم (۵۳۷ دختر و ۳۷۵ پسر) از شهر مشهد در استان خراسان رضوی تشکیل شده است. از یک آزمون درک زبان انگلیسی که ۴۰ پرسش از دو مهارت شنیدن و خواندن درک مطلب را پوشش می‌داد، به عنوان ابزار استفاده شد. آزمون درک شنیدن از دو زیر آزمون مهارت شنیدن اطلاعاتی^۱ و مهارت شنیدن تعاملی^۲ تشکیل شده بود و همچنین آزمون خواندن خود شامل دو زیر آزمون مهارت دقیق خوانی^۳ و مهارت سریع خوانی^۴ بود (هیوز، ۲۰۰۳). هر زیر آزمون از ده پرسش چهار گزینه‌ای تشکیل شده بود. بنابراین ابزار از چهار زیر آزمون مهارت‌های شنیدن اطلاعاتی، مهارت‌های شنیدن تعاملی، مهارت‌های سریع خوانی و دقیق خوانی، تشکیل شده بود. در این نوشتار از کلمات بعد و زیرمهارت به عنوان جایگزین استفاده شده است.

۲- مدل سازی و تحلیل داده‌ها

مدل چند بعدی که در این تحقیق مورد استفاده قرار گرفت، مدل لوجیت چند جمله‌ای ضرایب تصادفی چند بعدی^۵ نام دارد (ادمز، ویلسون، و ونگ، ۱۹۹۷). این مدل گونه‌ای چند بعدی و ترمیمی^۶ از مدل تک بعدی راش^۷ است که همه پارامترهای مدل را به صورت همزمان بر روی یک یک مقیاس مشترک لوجیت برآورد می‌کند. از آنجایی که مدل لوجیت چند جمله‌ای ضرایب

۱ Informational listening

۲ Interactional listening

۳ careful reading

۴ expeditious reading

۵ Multidimensional Random coefficient Multinomial Logit Model

۶ Compensatory

۷ Rasch model unidimensional

تصادفی چند بعدی^۱ عضوی از خانواده مدل‌های راش است، دارای تمام ویژگی‌های این مدل‌ها از جمله وجود آماره کافی^۲ برای محاسبه پارامترها است. اما برخلاف مدل‌های راش، نمرات خام زیرآزمون‌ها آماره کافی برای تخمین پارامتر توانایی افراد در زیر آزمون‌ها به حساب نمی‌آیند. بردارهای نمرات خام که در بر گیرنده نمره خام همه زیر آزمون‌ها هستند آماره‌های کافی برای تخمین توانایی افراد در زیر آزمون‌ها بحساب می‌آیند (چنک، ونگ و هو، ۲۰۰۹). مدل لوجیت چند جمله‌ای ضرایب تصادفی چند بعدی، مدلی بسیار انعطاف پذیر است و می‌تواند گستره‌ای از مدل‌های راش از جمله مدل دو گانه^۳ (راش، ۱۹۶۰/۱۹۸۰)، مدل مقیاس مقیاس ارزیابی^۴ (اندریچ، ۱۹۷۸)، مدل امتیاز نسبی^۵ (مسترز، ۱۹۸۲) و مدل چند جنبه‌ای^۶ (لیناکر، ۱۹۸۹) را تحت پوشش قرار دهد. مدل لوجیت چند جمله‌ای ضرایب تصادفی چند بعدی از طریق نرم افزار کان کوئست (وو، ادمز، و هالدین، ۲۰۰۷) قابل اجرا است. این مدل از همبستگی میان بعدها برای ترمیم برآورد پارامترها و افزایش دقت اندازه گیری استفاده می‌کند. سه مدل به داده‌ها برازش شدند و برازش کلی آنها با هم مقایسه شد. ابتدا مدل تک بعدی دو گانه راش (۱۹۶۰/۱۹۸۰) برازش شد؛ یعنی آزمون به گونه‌ای مدل شد که همه ۴۰ پرسش بر روی یک بعد قرار گیرند. در مرحله بعد، یک مدل دو گانه دو بعدی برازش شد. در این مدل، پرسش‌های شنیدن روی یک بعد و پرسش‌های خواندن روی بعد دوم بار گذاری شدند. در مرحله سوم، یک مدل دو گانه چهار بعدی برازش شد. در این مدل زیرآزمون‌های شنیدن اطلاعاتی، شنیدن تعاملی، خواندن سریع و خواندن دقیق هر یک بر روی یک بعد جداگانه

۱ Mrcml

۲ sufficient statistic

۳ dichotomous model

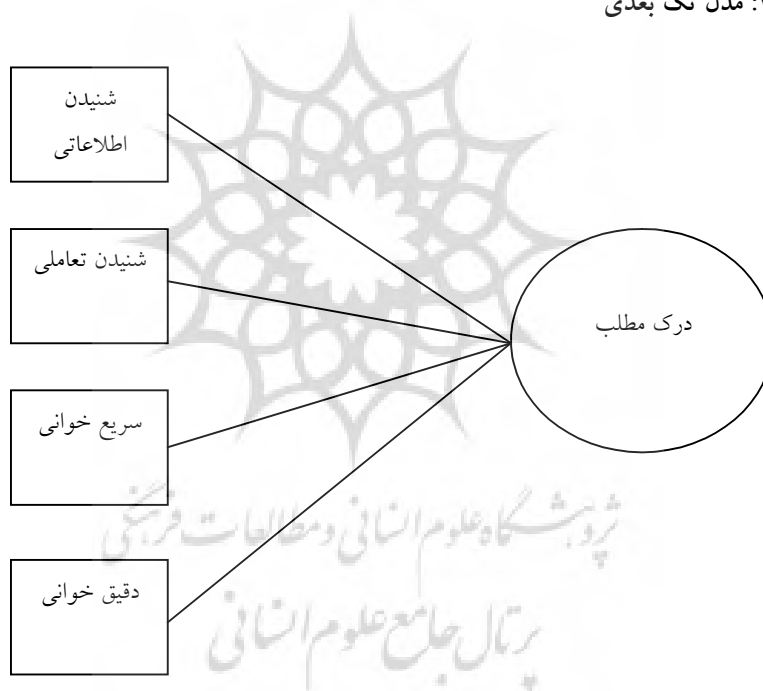
۴ rating scale model

۵ rating scale model

۶ multifaceted model

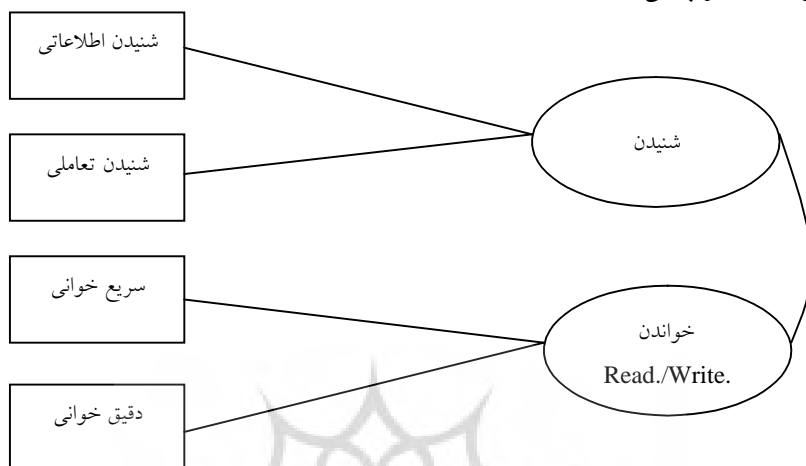
بارگذاری شدند. نرم افزار کان کوئست (وو و همکاران، ۲۰۰۷) برای تحلیل و برازش مدل‌ها به داده‌ها استفاده شد. نمودارهای ۱، ۲ و ۳ این سه مدل را نشان می‌دهند. در این نمودارها مستطیل‌ها نماینده پرسش‌ها و بیضی‌ها نماینده صفات مکنون^۱ یا همان ابعاد اندازه‌گیری مورد نظر هستند. خطوط مستقیم، پرسش‌های آزمون را به ابعادی که فرض می‌شود آن پرسش‌ها اندازه‌گیری می‌کنند وصل می‌نمایند. خطوط منحنی نشان‌دهنده همبستگی است که فرض می‌شود بین ابعاد وجود دارد.

نمودار ۱: مدل تک بعدی

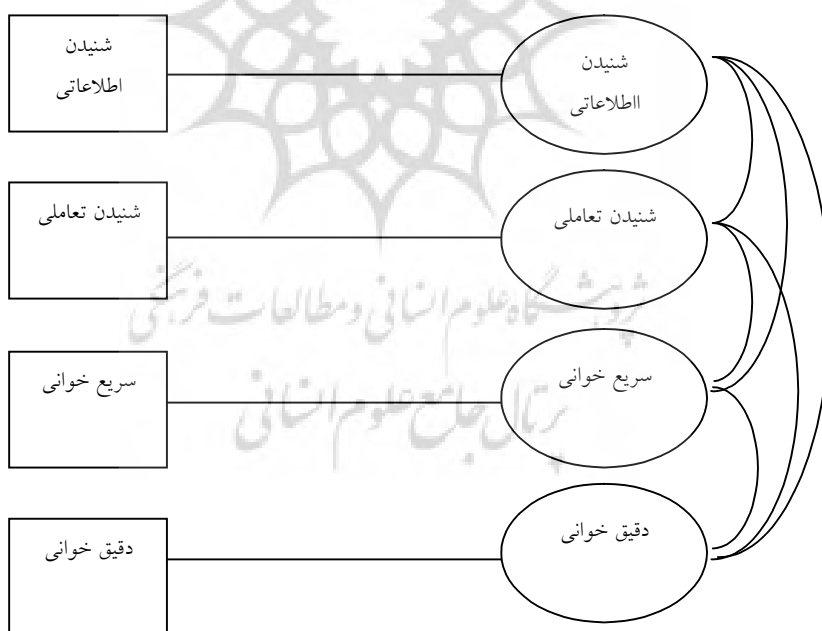


^۱ latent traits

نمودار ۲: مدل دو بعدی



نمودار ۳: مدل چهار بعدی



۳- نتایج

برازش مدل - داده

مدل‌های رقیب از طریق راستی‌نمایی راه حل‌هایشان با هم مقایسه شدند. از آنجایی که مدل‌های چند بعدی به صورت سلسله‌مراتبی با مدل تک بعدی در ارتباط هستند، یعنی این مدل‌ها در یکدیگر قابل‌جا گذاری هستند، برازش مدل‌ها می‌تواند با تغییر در انحراف^۱ یا G^2 با هم مقایسه شوند. هر چه راستی‌نمایی بزرگ‌تر باشد، مدل برازش شده به مدل واقعی نزدیک‌تر است. لگاریتم راستی‌نمایی منفی یا شاخص انحراف، تفاوت بین مدل محاسبه شده و مدل واقعی را نشان می‌دهد. بنابراین انتظار داریم که انحراف کوچک باشد و مدل‌هایی با انحراف کوچک‌تر انتخاب می‌شوند (یانسن و دی بک، ۱۹۹۹). تفاوت انحراف بین دو مدل تقریباً به شکل χ^2 توزیع شده است و تفاوت بین تعداد پارامترها درجه آزادی آن است (بریگز و ویلسون، ۲۰۰۳).

جدول ۱: آماره‌های برازش کلی و معیار اطلاعات برای مدل‌ها

مدل	G^2	تغییر در G^2	# پارامترها	AIC	BIC
۱- بعدی	۴۸۳۹۵.۲۰	-	۴۱	۴۸۴۷۷	۴۸۶۷۹
۲- بعدی	۴۸۲۲۵.۰۱	۱۷۰.۱۹	۴۳	۴۸۳۱۱	۴۸۵۲۳
۴- بعدی	۴۷۹۰۰.۸۴	۳۲۴.۱۷	۵۰	۴۸۰۰۰	۴۸۲۴۶

جدول (۱) نشان می‌دهد مدل دو بعدی انحراف کوچک‌تری از مدل تک بعدی دارد و نیز مدل چهار بعدی انحراف کمتری نسبت به مدل دو بعدی دارد. تغییر در انحراف از مدل تک بعدی به دو بعدی از نظر آماری معنا دار است، $p < .01$ ، $\chi^2(2) = 170.19$. همچنین تغییر در انحراف از مدل دو بعدی به مدل چهار بعدی از نظر آماری معنا دار است، $p < .01$ ، $\chi^2(7) = 324.17$. بنابراین، مدل چهار بعدی که دارای کوچک‌ترین انحراف است بهترین برازش را دارد. معیار اطلاعات آکایک (AIC) (آکایک، ۱۹۷۴) و

۱ deviance

۲ chi square

معیار اطلاعات بایز (BIC) (شوارتز، ۱۹۷۸) که با تلفیق انحراف و تعداد پارامترهای محاسبه شده و اندازه نمونه محاسبه می‌شوند نشان می‌دهند مدل چهار بعدی بهترین برازش را به دلیل داشتن کوچک‌ترین AIC و BIC داراست.

جدول ۲: همبستگی بین چهار بعد

بعد ۴	بعد ۳	بعد ۲	بعد ۱	
۰.۴۹	۰.۵۲	۰.۵۸		بعد ۱
۰.۴۹	۰.۵۲			بعد ۲
۰.۵۴				بعد ۳

بحث و نتیجه‌گیری

این مطالعه از طریق مقایسه برازش‌های چندین مدل موجود و به کمک آماره‌های انحراف و معیار اطلاعات نشان داد که مدل‌های چند بعدی، مانند تحلیل عامل تاییدی، از نظر ماهیت تاییدی هستند و می‌توانند برای تحقیق درباره ساختار تشکیل دهنده آزمون‌ها مورد استفاده قرار گیرند (بقایی و گروتیان، [زیر چاپ]؛ یانسن و دی باک، ۱۹۹۹). این تحقیق نشان داد مدل چهار بعدی آزمون درک مطلب نسبت به مدل دو بعدی و مدل تک بعدی بطور معنا دار، برازش بهتری دارد. این یک دلیل واضح بر تایید بخش پذیری توانایی درک مطلب در زبان دوم و نشان دهنده این است که یک مدل چهار بعدی می‌تواند به نحو بهتری نشان دهنده ساختار سازه درک در زبان دوم باشد. این مفهوم بطور مستقیم به جنبه ساختاری روایی سازه ای^۱ مرتبط است. جنبه ساختاری روایی سازه‌ای اساساً با فرایند نمره دادن مرتبط است. بسیار مهم است که به ساختار یک آزمون هنگام نمره دادن توجه شود. جمع بستن نمرات قسمت‌های یک آزمون با هم، زمانی که هر قسمت یک بعد متفاوت را اندازه‌گیری می‌کند، صحیح نیست. در حالی که تنها یک بعد می‌تواند عملکرد یک فرد را در یک آزمون تک بعدی نشان دهد، نمرات بعدهای متفاوت باید

^۱ structural aspect of construct validity

بطور مجزا گزارش شوند. به عبارت دیگر، مدل‌های نمره دادن باید با توجه به ساختار هر آزمون تعیین شوند (بقایی و امراهی، ۲۰۱۱؛ مسیک، ۱۹۸۹).

این تحقیق سعی بر آن داشت تا ساختار عاملی پرسش‌های شنیدن و خواندن را با توجه به چهار زیرمهارتی که آزمون سازان قصد ارزشیابی آن را داشتند بررسی کند. پرسش‌های خواندن و شنیدن بنظر می‌رسند که دو زیرمهارت را مورد ارزشیابی قرار می‌دهند. شاخص‌های برآزش بهتر (جدول ۱) برای مدل‌های دو و چهار بعدی، این ادعا را که پرسش‌های درک مطلب در زبان دوم به دو یا چهار نوع زیرمهارت تقسیم می‌شوند تایید می‌کند. تا جایی که به تحقیق حاضر مرتبط است، می‌توان نتیجه‌گیری کرد که سازه درک در زبان دوم به زیرمهارت‌های مختلفی قابل تقسیم است.

این مطالعه همچنین نشان داد دو زیرمهارت خواندن و درک مطلب و شنیدن و درک مطلب به آسانی در زبان دوم به آسانی قابل تفکیک هستند. نتایج نشان داد این دو سازه مجزا اما مرتبط به هم هستند. این موضوع با یافته‌های لوین و ریورز (۱۹۸۸) همخوانی دارد. اما باک (۱۹۹۲) و بائه و باکمان (۱۹۹۸) بر وجود میزان بالایی از همپوشانی بین توانایی شنیدن و خواندن تاکید کرده اند. با این حال، محققین فوق همچنین اشاره کرده اند که هر مهارت به تنهایی دارای جنبه‌های منحصر به فرد خود نیز هست.

با توجه به برآزش مدل‌ها و همبستگی متوسطی که در این پژوهش دیده شد، نتیجه می‌گیریم که شنیدن و خواندن در زبان دوم دو سازه مجزا هستند. همچنین نشان داده شد دو مهارت شنیدن و خواندن می‌تواند باز هم به زیرمهارت‌های کوچکتر مهارت شنیدن تعاملی و اطلاعاتی و سریع و دقیق خوانی تقسیم شوند.

دلیل این که مهارت‌ها در این پژوهش به وضوح و راحتی تفکیک شدند می‌تواند با سطح دانش عمومی زبان شرکت کننده گان ارتباط داشته باشد. شرکت کنندگان در مطالعه حاضر را زبان آموزان سطح متوسط پایین تشکیل می‌دادند. الدرسون (۲۰۰۰) بیان می‌کند زیرمهارت‌ها در آزمون‌های خواندنی که برای مبتدیان طراحی شده اند آسان‌تر قابل تشخیص هستند. بنابراین روش تفکیک مهارت‌های جزئی در تدریس و آزمون می‌تواند برای خوانندگان مبتدی زبان دوم معتبرتر و موجه تر باشد، در حالی که این امر برای خوانندگان سطح پیشرفته صدق نمی‌کند. ممکن است بتوانیم این مفهوم را به مهارت درک شنیدن

نیز تعمیم بدهیم. براساس این بیانیه، می‌توان نتیجه‌گیری کرد که تشخیص زیرمهارت‌ها در آزمون‌های طراحی شده برای زبان آموزان سطوح ابتدائی و متوسط آسان‌تر است.

سانگ (۲۰۰۸) معتقد است دلیل اینکه تحقیقات تجربی قبلی نتوانستند به یک پاسخ یکسان در مورد تعداد زیرمهارت‌ها قابل تشخیص دست یابند، می‌تواند این باشد که در پژوهش‌های مختلف از آزمون‌های مختلف و گروه‌های متفاوت آزمون دهندگان استفاده شده است. بنابراین نتیجه چنین مطالعاتی ممکن است بسته به سطح دانش عمومی زبان آزمون دهندگان، ویژگی‌های آزمون اجرا شده، و تعامل بین دو عامل بستگی داشته باشد.

توجه به این نکته ضروری است که تمام پرسش‌ها در این مطالعه چند گزینه‌ای بودند، بنابراین تفکیک پذیری به دست آمده نمی‌تواند به تفاوت‌های شکل پرسش‌ها ارتباطی داشته باشد. بنابراین احتمال اینکه تفاوت‌های مشاهده شده مربوط به پردازش‌های ذهنی متفاوت باشد بسیار قوی‌تر است. توجه به این مساله تجزیه پذیری سازه درک را که در این پژوهش به آن رسیدیم تأیید می‌کند.

این مطالعه می‌تواند کاربردهای عملی در تهیه و نمره گذاری آزمون‌ها داشته باشد. همچنین کمک می‌کند تا درباره زیرمهارت‌هایی که باید در آزمون‌های زبان گنجانده شوند تصمیم‌گیری کنیم. بنابراین آزمون‌سازان ممکن است بخواهند آزمون‌هایی تهیه کنند که تعداد بیشتری زیرمهارت را برای برخی موقعیت‌ها و شرایط ارزشیابی - در خود جای دهند. به عنوان مثال برای یک آزمون بسندگی آکادمیک مانند تافل یا آیلتس آکادمیک گنجاندن زیرمهارت دقیق خوانی ضروری است. همچنین در یک آزمون بسندگی عمومی که به منظور مهاجرت و زندگی مورد استفاده قرار می‌گیرد گنجاندن سریع خوانی و شنیدن تعاملی، ضروری به نظر می‌رسد. بدیهی است انتخاب زیرمهارت‌ها برای یک آزمون بر اساس هدف آزمون و نوع تصمیم‌گیری‌هایی که در آینده بر پایه نمرات آزمون صورت می‌گیرد انجام می‌شود. بر اساس نتایج این تحقیق منطقی است که آزمون‌های شنیدن و خواندن از لحاظ مهارت‌های شنیدن تعاملی و اطلاعاتی و مهارت‌های خواندن دقیق و سریع به خصوص در آزمون‌های تشخیصی^۱ تفکیک شوند.

^۱ diagnostic tests

کاربرد عملی دیگر این مطالعه، به فرایند نمره گذاری در آزمون‌های شنیدن و خواندن مربوط است. بیشتر آزمون‌ها برای هر یک از مهارت‌های خواندن و شنیدن یک نمره جدا در نظر می‌گیرند. با توجه به نتایج این تحقیق، کاملاً منطقی بنظر می‌رسد که نمرات به تفکیک زیرمهارت‌ها گزارش شوند. نمرات زیرمهارت‌ها به مدرسین انگلیسی به عنوان زبان خارجی کمک می‌کند که دروسشان را با در دست داشتن اطلاعات دقیق‌تر و جزئی درباره توانایی‌های مهارت‌های خواندن و شنیدن دانش آموزان طراحی کنند.

کتابنامه

- Adams, R. J., Wilson, M. R., & Wang, W.-C. (۱۹۹۷). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, ۲۱, ۱-۲۳.
- Akaike, H. (۱۹۷۴). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, ۱۹: ۶, ۷۱۶-۷۲۳.
- Alderson, J. C. (۲۰۰۰). *Assessing reading*. Cambridge, UK: Cambridge University Press.
- Alderson, J. C., & Lukmani, Y. (۱۹۸۹). Cognition and reading: Cognitive levels as embodied in test questions. *Reading in a Foreign Language*, ۵, ۲۵۳-۲۷۰.
- Andrich, D. (۱۹۷۸). A rating formulation for ordered response categories. *Psychometrika*, ۴۳, ۵۶۱-۵۷۳.
- Bachman, L. F., & Palmer, A. (۱۹۸۱). The construct validation of the FSI oral interview. *Language Learning*, ۳۱, ۶۷-۸۶.
- Bachman, L. F., & Palmer, A. (۱۹۸۲). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, ۱۶, ۴۴۹-۴۶۵.

- Bachman, L. F., Davidson, F. G., & Foulkes, J. (۱۹۹۰). A comparison of the abilities measured by the Cambridge and Educational Testing Service EFL Test Batteries. *Issues in Applied Linguistics*, ۱, ۳۰-۵۵.
- Bachman, L. F., Davidson, F. G., Ryan, K., & Choi, I-C. (۱۹۹۵). *An investigation into the comparability of two tests of English as a foreign language: The Cambridge-TOEFL comparability study*. Cambridge: UCLES.
- Bae, J., & Bachman, L. F. (۱۹۹۸). A latent variable approach to listening and reading: Testing factorial invariance across two groups of children in the Korean/English Two-way Immersion program. *Language Testing*, ۱۵ (۳), ۳۸۰-۴۱۴.
- Baghaei, P. & Grotjahn, R. (in press). Establishing the construct validity of conversational C-Tests using a multidimensional Item Response Model. In R. Grotjahn (Ed.). *Der C-Test: Aktuelle Tendenzen/The C-Test: Current trends*. Frankfurt/M.: Lang.
- Baghaei, P. & Amrahi, N. (۲۰۱۱). Validation of a multiple choice English vocabulary test with the Rasch model. *Journal of Language Teaching and Research*, ۲(۵), ۱۰۵۲-۱۰۶۰.
- Bock, R.D., & Aitkin, M. (۱۹۸۱). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika*, ۴۶, ۴۴۳-۴۵۹.
- Briggs, D.C. & Wilson, M. (۲۰۰۳). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement*, ۴, ۸۷-۱۰۰.
- Buck, G. (۲۰۰۱). *Assessing listening*. Cambridge, UK: Cambridge University Press.
- Buck, G. (۱۹۹۱). The testing of listening comprehension: An introspective study. *Language Testing*, ۸(۱), ۶۷-۹۱.

- Buck, G. (۱۹۹۲). Listening comprehension: Construct validity and trait characteristics. *Language Learning*, ۴۲(۳), ۳۱۳-۳۵۷.
- Buck, G., & Tatsuoka, K. (۱۹۹۸). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, ۱۵(۲), ۱۱۹-۱۵۷.
- Carroll, J. B. (۱۹۹۳). *Human cognitive abilities*. Cambridge: Cambridge University Press.
- Cheng, Y.-Y., Wang, W.-C., & Ho, Y.-H. (۲۰۰۹). Multidimensional Rasch analysis of a psychological test with multiple subtests. *Educational and Psychological Measurement*, ۶۹, ۳۶۹-۳۸۸.
- Clark, H. H., & Clark, E. V. (۱۹۷۷). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.
- Davey, B. (۱۹۸۸). Factors affecting the difficulty of reading comprehension items for successful and unsuccessful readers. *Experimental Education*, ۵۶, ۶۷-۷۶.
- Drum, P. A., Calfee, R. C., & Cook, L. K. (۱۹۸۱). The effects of surface structure variables on performance in reading comprehension tests. *Reading Research Quarterly*, ۱۶, ۴۸۶-۵۱۴.
- Embretson, S. E. (۲۰۰۷). Mixed Rasch Models for Measurement in Cognitive Psychology. In Matthias von Davier & Claus H. Carstensen (eds.), *Multivariate and mixture distribution Rasch models: extensions and applications*. New York: Springer.
- Embretson (Whitely), S. E. (۱۹۸۳). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, ۹۳, ۱۷۹-۱۹۷.
- Embretson (Whitely), S. E. (۱۹۸۰). Multicomponent latent trait models for ability tests. *Psychometrika*, ۴۵, ۴۷۹-۴۹۴.

- Folk, V. G. & Green, B. F. (۱۹۸۹). Adaptive estimation when the unidimensionality assumption of IRT is violated. *Applied Psychological Measurement*, ۱۳, ۳۷۳-۳۸۹.
- Freedle, R., & Kostin, I. (۱۹۹۴). Can multiple-choice reading tests be construct valid? *Psychological Science*, ۵, ۱۰۷-۱۱۰.
- Freedle, R., & Kostin, I. (۱۹۹۹). Does text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing*, ۱۶(۱), ۲-۳۲.
- Hughes, A. (۲۰۰۳). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Janssen, R., & De Boeck, P. (۱۹۹۹). Confirmatory analyses of componential test structure using multidimensional item response theory. *Multivariate Behavioral Research*, ۳۴, ۲۴۵-۲۶۸.
- Lennon, R. T. (۱۹۶۲). What can be measured? *Reading Teacher*, ۱۵, ۳۲۶-۳۳۷.
- Levine, A., & Revers, T. (۱۹۸۸). The FL receptive skills: Same or different? *System*, ۱۶, ۳۲۶-۳۳۶.
- Linacre, J. M. (۱۹۸۹). *Many-faceted Rasch measurement*. Chicago: MESA Press.
- Lund, R. J. (۱۹۹۱). A comparison of second language listening and reading comprehension. *Modern Language Journal*, ۷۵, ۱۹۶-۲۰۴.
- Masters, G. N. (۱۹۸۲). A Rasch model for partial credit scoring. *Psychometrika*, ۴۷, ۱۴۹-۱۷۴.
- Messick, S. (۱۹۸۹) Validity. In R.L. Linn (ed.) *Educational measurement* (pp. ۱۳-۱۰۳). New York: Macmillan.
- Munby, J. (۱۹۷۸). *Communicative syllabus design*. Cambridge: Cambridge University Press.

- Rasch, G. (۱۹۶۰/۱۹۸۰) *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research, ۱۹۶۰. (Expanded edition, Chicago: The university of Chicago Press, ۱۹۸۰).
- Richards, J. C. (۱۹۸۳). Listening comprehension: Approach, design, procedure. *TESOL Quarterly*, ۱۷(۲), ۲۱۹-۲۴۰.
- Rost, D. (۱۹۹۳). Assessing the difference components of reading comprehension: fact or fiction? *Language Testing*, ۱۰(۱), ۷۹-۹۲.
- Santelices, M. V. & Caspary, K. (۲۰۰۹). Development of a Multidimensional Measure of Academic Engagement. *Journal of Applied Measurement*, ۱۰: ۴, ۳۷۱-۳۹۳.
- Song, M.-Y. (۲۰۰۸) Do divisible subskills exist in second language (L۲) comprehension? A structural equation modeling approach. *Language Testing*, ۲۵ (۴), ۴۳۵-۴۶۴.
- Weir, C. (۱۹۹۳). *Understanding and developing language tests*. New York: Prentice Hall.
- Wu, M. L., Adams, R. J., & Haldane, S. A. (۲۰۰۷). *ACER ConQuest*. Australian Council for Educational Research.