

بررسی میزان روایی آزمون زبان انگلیسی کنکور دکتری، با استفاده از چارچوب بررسی سازه استدلال محور

مهربان هماوندی^۱، غلامرضا کیانی^{۲*}، رضا غفارنمر^۳، رامین اکبری^۴

۱. دانشجوی دکتری آموزش زبان انگلیسی، دانشگاه تربیت مدرس، تهران، ایران
۲. دانشیار آموزش زبان انگلیسی، دانشگاه تربیت مدرس، تهران، ایران
۳. دانشیار آموزش زبان انگلیسی، دانشگاه تربیت مدرس، تهران، ایران
۴. دانشیار آموزش زبان انگلیسی، دانشگاه تربیت مدرس، تهران، ایران

پذیرش: ۹۳/۳/۲۶

دریافت: ۹۳/۲/۱۳

چکیده

می‌توان مدعی بود که سنجش توانایی زبان انگلیسی داوطلبان ورود به مقاطع تحصیلات تکمیلی، به‌ویژه دوره دکتری، اهمیت چشمگیری دارد. آزمونی که با عنوان آزمون زبان انگلیسی کنکور دکتری در حال حاضر از سوی سازمان سنجش و آموزش کشور برگزار می‌شود، موافقان و منتقدان بسیاری دارد. هدف اصلی این پژوهش، بررسی میزان روایی این آزمون با استفاده از چارچوب ارزشیابی استدلال‌محور است. برای رسیدن به این هدف، از گروه‌های مختلف ذی‌نفع در آزمون، از جمله داوطلبان آزمون دکتری، مدرسان دوره‌های آمادگی آزمون زبان دکتری، استادان رشته‌ی تعلیم و تربیت و آزمون‌سازان زبان با استفاده از روش کیفی و ابزارهای مصاحبه (گروهی و فردی) داده‌ها جمع‌آوری شد. برپایه روش کیفی تحلیل محتوا، الگوهای یافت‌شده در مجموعه داده‌ها براساس قواعد ازپیش‌تعیین‌شده بررسی محتوا و با رعایت ویژگی‌های پژوهش ارائه می‌شوند. یافته‌های پژوهش نشان دادند که آزمون فعلی از نظر شرکت‌کنندگان به دگرگونی‌های گاه اساسی در ابعاد محتوا، شیوه برگزاری و ضریب اختصاصی آزمون نیازمند است. علاوه بر ارائه نتایج، در پایان، پیشنهادهای ارائه و دلالت‌ها، چالش‌ها و ملزومات اجرایی‌سازی آن‌ها بیان شده‌اند.

واژگان کلیدی: آزمون زبان دکتری، بررسی روایی آزمون، ارزشیابی استدلال‌محور، ابعاد محتوا، ضریب اختصاصی آزمون.

۱. مقدمه

روایی آزمون^۱ از دیرباز یکی از کلیدی‌ترین ویژگی‌های آزمون به‌شمار می‌رفت. روایی در گذشته و به‌صورت معمول، از چهار منظر روایی محتوا^۲، معیار^۳، سازه^۴ و صورتی^۵ می‌شد. پس از ارائه مفهوم وحدت روایی^۶ ساموئل مسیک (۱۹۸۹)، به‌منظور بررسی روایی آزمون و مقبولیت این نظریه از منظر عالمان تعلیم و تربیت، چارچوب‌های متفاوتی برای عملیاتی کردن و شیوه‌مهندسازی آن ارائه شد. از سوی دیگر، پیشرفت فناوری، ارتباطات و انواع پایگاه‌های داده مرتبط با رشته‌های مختلف، توانش زبان انگلیسی^۷ را به امری حیاتی در زندگی فردی و دانشگاهی افراد تبدیل کرده است. در این راستا، متقاضیان دوره‌های تحصیلات تکمیلی، به‌ویژه دوره دکتری، به توانش زبان انگلیسی کافی و مناسبی نیاز دارند تا بتوانند خود را در زمینه‌های تخصصی به‌روز نگه دارند و دستاوردهای علمی، پژوهشی و صنعتی خود را به جهان خارج عرضه کنند؛ از این‌رو، ارزیابی توانش زبان انگلیسی متقاضیان ورود به دوره‌های دکتری اهمیتی دوچندان می‌یابد.

از آنجا که هرچه درجه اهمیت آزمون بیشتر باشد، تأثیر و تأثر آن در ابعاد فردی، سازمانی، دانشگاهی و حتی اجتماعی بیشتر می‌شود، مسئله استانداردسازی آزمون‌های پراهمیت و تلاش به‌منظور بررسی نظرات مختلف و ارائه راهکاری برای بهینه‌سازی آن‌ها، به فرآیندی مشخص و بدیهی تبدیل شده است. در زمینه آزمون‌های زبان انگلیسی پراهمیت در کشور، همچنان وضعیت روبه‌رشدی دیده نمی‌شود و «تقریباً همه آزمون‌های ملی کشور با آزمون‌های استاندارد فاصله بسیار فراوان دارند» (کیانی، ۱۳۸۸: ۲). با تغییر شیوه برگزاری آزمون ورودی دوره‌های دکتری داخل، متمرکز شدن برگزاری آن و عهده‌دار شدن برگزاری آزمون ازسوی سازمان سنجش و آموزش کشور به‌عنوان تنها متولی، شیوه آزمون‌گیری از سطح توانش زبان انگلیسی متقاضیان نیز دستخوش تغییراتی شد؛ تغییراتی چون شدن آزمون‌های متنوع ورودی زبان انگلیسی که در سال‌های پیش از آن برگزار می‌شد، تغییرات در محتوای آزمون‌ها گاه به‌صورت اساسی و شیوه متفاوت بازم‌بندی و تعداد سؤالات بخش‌های مختلف محتوای آن. همه این تغییرات درحالی انجام شده است که تاکنون هیچ تحقیق مدون و جامعی به‌منظور آشکارسازی درجه کیفی و میزان استاندارد بودن بخش زبان این آزمون صورت نگرفته است. ناگفته پیداست که آزمون ورودی دکتری بنابر ماهیت آن و به‌عنوان آزمون ورودی به عالی‌ترین مقطع دانشگاهی، تأثیرات شگرفی در کوتاه‌مدت، میان‌مدت و بلندمدت بر ابعاد آموزشی، فرهنگی،

سیاسی و اقتصادی کشور خواهد داشت.

با توجه به اهمیت این موضوع و کمبود کار پژوهشی که به طور جدی آزمون دکتری در آن بررسی شده باشد، در پژوهش فرارو، آزمون یادشده را براساس مدل ارزشیابی استدلال محور انجام دادیم که شیوه‌ای نوین در زمینه بررسی روایی آزمون‌های زبان است. برای این کار، ذی‌نفع‌های آزمون را در مصاحبه‌هایی انفرادی و گروهی شرکت دادیم که سؤالات آن‌ها را براساس چارچوب ارزشیابی استدلال محور تهیه کرده بودیم. سپس، نتایج پژوهش را براساس شیوه کیفی تحلیل محتوا بررسی کردیم تا به پرسش اساسی زیر پاسخ دهیم:

- آیا آزمون زبان انگلیسی عمومی که در حال حاضر به عنوان بخشی از کنکور ورودی دکتری داخل ازسوی سازمان سنجش و آموزش کشور برگزار می‌شود ویژگی‌های یک آزمون روا، از منظر ارزشیابی استدلال محور را دارد؟

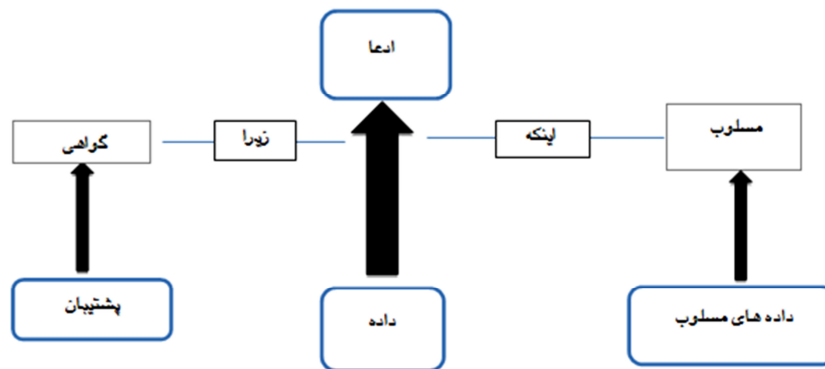
۲. اهمیت و شیوه‌های ارزیابی روایی آزمون‌های زبان

یکی از ویژگی‌هایی که هر آزمون باید داشته باشد تا استاندارد به‌شمار آید، پایایی^۱ و روایی است. روایی در آزمون‌سازی به صورت معمول عبارت است از «میزان هم‌سنگی بین نمره به‌دست آمده از آزمون با ویژگی که آزمون مورد نظر برای اندازه‌گیری آن ساخته شده است» (Cronbach & Meehl, 1955: 23). مهم‌ترین عاملی که در ارزیابی یک آزمون باید به آن توجه شود، روایی آزمون است. مقصود از روایی آزمون مناسب بودن، بامعنا بودن و مفید بودن استنباط‌های خاصی است که از روی نمرات آزمون انجام می‌شود (کین، ۲۰۰۷). برای تأیید این‌گونه استنباط‌ها، باید شواهدی جمع‌آوری شود که در فرهنگ روان‌سنجی^۲ به آن رواسازی آزمون^۱ گفته می‌شود.

۲-۱. چارچوب استدلال محور بک من

جدیدترین شیوه بررسی روایی آزمون‌های زبان، براساس نظریه ارزشیابی استدلال محور بک من است (بک من، ۲۰۰۳ و ۲۰۰۵، بک من و پالمر، ۲۰۱۰). اصول اساسی این روش از نظریه وحدت روایی مسیک (۱۹۸۹) و ساختار استدلال تولمین (۱۹۵۸، ۲۰۰۳) گرفته شده است. بر مبنای نظریه مسیک، هر آزمون تنها شامل یک نشانگر کیفی از روایی آن است که برای به‌دست آوردن این نشانگر، پژوهشگر باید از روش‌های عمدتاً کیفی به منظور تحلیل

پایایی و مشخصه‌های روایی دیگر استفاده کند و در نهایت، به یک نظر واحد از درجه‌ی روایی آزمون برسد. مسیک این نوع روایی را که تنها نوع روایی است، روایی سازه‌ی آن آزمون و نظریه‌ی خود را وحدت روایی نام نهاد. اگرچه بسیاری از دانشمندان و پژوهشگران روان‌سنجی ابتدا از نظریه‌ی مسیک استقبال کردند، این نظریه به دلیل شفاف نبودن چگونگی کاربردش در دنیای واقعی، از مرحله‌ی نظریه‌ی فراتر نرفت و در آزمونی خاص به کار برده نشد. کاری که بک من، به‌عنوان نظریه‌پرداز اصلی مدل ارزشیابی استدلال‌محور انجام داد این بود که در مدل خود، نظریه‌ی مسیک را به دنیای واقعی ارزشیابی آموزشی نزدیک کرد و برای اجرای نظریه‌ی وحدت روایی، مراحل کاملاً مشخص تدوین نمود. در همین راستا و به‌منظور عینی ساختن مراحل بررسی سازه‌ی آزمون و طراحی چارچوب یک آزمون، بک من از مدل استدلال‌محور تولمین کمک گرفت. تولمین (۱۹۵۸)، به‌عنوان یک منطق‌دان، برای اینکه نشان دهد که هر امر عینی چقدر به واقعیت نزدیک است، یک ساختار را تعریف کرد. در این ساختار، هر امر عینی برآمده از یک ادعا^{۱۱} بر عینیات^{۱۲} (داده) مبتنی است. هرچه برای این ادعا بر مبنای گواهی^{۱۳} مورد انتظار پشتیبان^{۱۴} بیشتری وجود داشته باشد، ادعا قوت می‌گیرد و هر قدر که برای ادعا مسلوب^{۱۵} وجود داشته باشد، ادعا ضعیف‌تر می‌شود. ساختار کلی نظریه‌ی تولمین به‌صورت شکل زیر است:

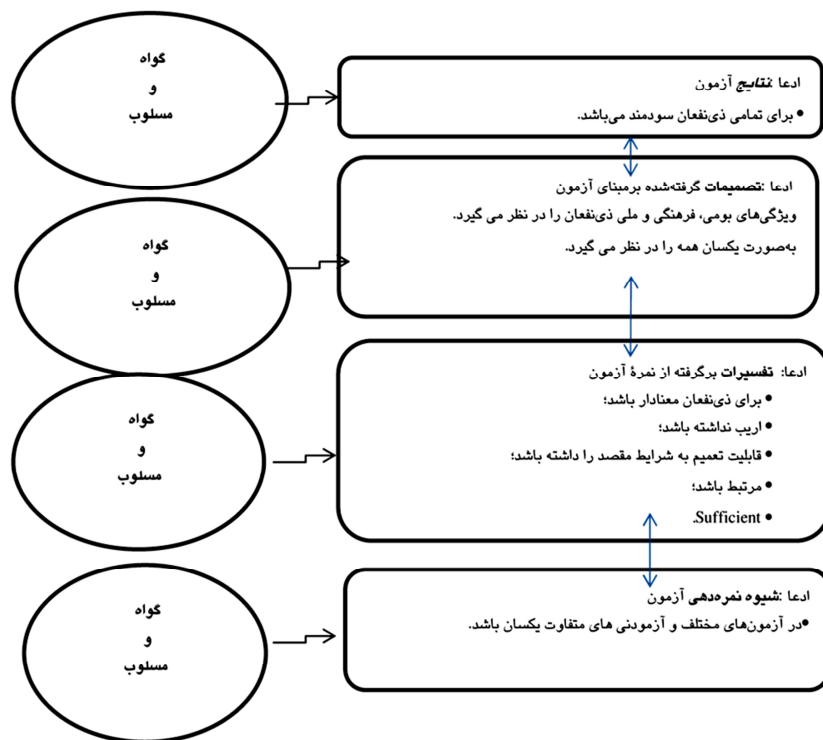


شکل ۱ ساختار استدلال‌محور تولمین (۲۰۰۲)

کاری که بک من (۲۰۰۵) انجام داد این بود که به‌منظور بررسی سازه‌ی یک آزمون، ابتدا آن را به‌عنوان ادعا مطرح کرد و سپس با استفاده از نظرات مثبت (گواه) در مقابل نظرات

منفی (مسلوب)، کوشید که ادعای روایی سازهٔ آزمون مورد نظر را ثابت کند. در اینجا، باید به این نکته توجه کنیم که اولاً نظریهٔ وحدت روایی سازهٔ مسیک (۱۹۸۹) در اینجا به عنوان ادعای مطلوب هر آزمون در نظر گرفته می‌شود و ثانیاً نوع داده‌های جمع‌آوری شده عموماً به صورت کیفی و نرم است. نکتهٔ مهم این است که نتیجه‌گیری نهایی در مورد پذیرش یا رد روایی سازه برعهدهٔ مخاطب و خواننده (نه ضرورتاً پژوهشگر) است.

در این روش، برای رسیدن به هدف نهایی (روایی سازهٔ آزمون)، باید چهار هدف زیرمجموعه در نظر گرفته شود. تلاش می‌شود که برای هر یک از این هدف‌ها از میان شرکت‌کنندگان پژوهش، گواهی به منظور پشتیبان و همچنین مسلوبی به عنوان مردودکنندهٔ گواهی یافته شود. هر یک از این چهار هدف نیز به چند زیرمجموعه تقسیم‌بندی می‌شود. در شکل زیر، به صورت شماتیک این بحث نمایش داده شده است.



شکل ۲ ساختار ارزشیابی استدلال محور بک من و پالم (۲۰۱۰)

مهم‌ترین ویژگی این روش توانایی بررسی سازه یک آزمون و ویژگی‌های مرتبط با آن و همچنین به وجود آوردن چارچوب اولیه یک آزمون زبان با توجه به مبناهای نظری این تئوری است. در این روش، تأثیرات مثبت یک آزمون بر آزمودنی‌ها به‌عنوان سنگ بنا و نقطه آغازین آن در نظر گرفته می‌شود؛ بنابراین، اولین ویژگی یک آزمون استاندارد تأثیرگذاری مثبت آن در سطح آزمودنی‌ها، مریبان، همه‌ذی‌نفعان و به‌طور کلی جامعه است. قدم‌های بعدی شامل تصمیم‌گیری دربارهٔ نمرهٔ قبولی/ردی، چگونگی تفسیر و بررسی نتایج آزمون، شیوهٔ آزمون‌گیری و محتویات آزمون است.

وانگ (۲۰۱۲)، کین (۲۰۱۳) و بک من و چوی (۲۰۱۲) از جمله پژوهشگرانی هستند که تاکنون بر مبنای ارزشیابی استدلال‌محور پژوهش کرده‌اند. آن‌ها در مطالعاتشان یا در پی بررسی و تحلیل سازه یک آزمون زبان بوده‌اند و یا براساس این شیوه، چارچوب جدیدی را برای یک آزمون طراحی کرده‌اند. یکی از مزایای اصلی این شیوهٔ طراحی چارچوب آزمون توجه ویژه به نیازهای بومی و فرهنگی آزمودنی‌ها با استفاده از تحلیل نیاز آن‌ها است؛ به سخن دیگر، یکی از اصول اساسی ارزشیابی استدلال‌محور این است که یک آزمون کاملاً استاندارد در یک محیط خاص، در همهٔ محیط‌ها به‌عنوان آزمون استاندارد به‌کار نمی‌رود.

۳. روش‌شناسی پژوهش

این مطالعه را با رویکرد کیفی^{۱۱} منطبق با چارچوب ارزشیابی استدلال‌محور و با استفاده از روش تحلیل محتوا^{۱۲} انجام دادیم. در این مطالعه، از ۱۸ نفر داوطلب آزمون دکتری و ۹ نفر دانشجوی دورهٔ دکتری که در سال‌های گذشته در این آزمون شرکت کرده بودند، مصاحبه کردیم. شیوهٔ نمونه‌گیری نیز براساس انتخاب هدفمند^{۱۸} بود. در این پژوهش، معیار انتخاب حجم نمونه رسیدن داده‌ها به سطح اشباع اطلاعاتی^{۱۹} بود و زمانی‌که اطلاعات جدیدی به‌دست نیامد، نمونه‌گیری به پایان رسید. از بین ذی‌نفعان دیگر این آزمون، ۳ نفر از استادان رشته‌های مختلف دانشگاهی که سابقهٔ تدریس و هدایت دانشجویان دورهٔ دکتری را داشتند، ۲ نفر از استادان رشتهٔ تعلیم و تربیت و ارزشیابی آزمون و ۲ نفر از مدرسان دوره‌های آمادگی آزمون زبان دکتری در این مطالعه شرکت کردند. به دلیل دسترسی نداشتن به سیاست‌گذاران کلان آزمون دکتری، داده‌ای در رابطه با این گروه جمع‌آوری نشد. ابزار جمع‌آوری داده‌ها برگزاری مصاحبه‌های ساختاریافتهٔ فردی^{۲۰} و

مصاحبه‌های متمرکز گروهی^{۲۱} است. برای مصاحبه، ۱۹ سؤال را برپایه چارچوب پیشنهادی بک من و پالم (۲۰۱۰) برای طراحی ابزار جمع‌آوری داده‌ها در مطالعات استدلال‌محور طرح و برای اطمینان از فهم آن‌ها، به فارسی ترجمه کردیم (پیوست). به‌منظور بررسی روایی ترجمه سؤالات از انگلیسی به فارسی، برگردان همه سؤالات توسط دو مترجم حرفه‌ای و با استفاده از شیوه برگردان مفهومی^{۲۲}، بررسی و تأیید شد. پس از فرآیند گردآوری داده‌ها، همه مصاحبه‌های ضبط‌شده را روی کاغذ پیاده و مرور کردیم. سپس، نکات مشترک با درون‌مایه‌های یکسان را استخراج و فهرستی از آن‌ها تهیه کردیم. پس از بررسی میزان شباهت‌ها و تفاوت‌های آن‌ها، نکاتی را که بیشترین شباهت را از لحاظ مفهومی با یکدیگر داشتند در یک طبقه قرار دادیم که یک تم واحد را تشکیل دادند. تم‌های پدیدار شده از داده‌های به‌دست‌آمده، اساس بررسی نهایی را تشکیل دادند.

۴. نتایج و یافته‌های پژوهش

چنانکه گفتیم، ارزشیابی استدلال‌محور بر مبنای پژوهش کیفی است؛ از این رو، پس از هر ادعای مطروحه بر اساس ارزشیابی استدلال‌محور، تم‌های برجسته گواهی مرتبط را که ممکن است نقش گواه و یا مسلوب داشته باشند، به‌همراه نمونه ارائه کردیم.

ادعای اول، شیوه نمره‌دهی آزمون

گواهی مطلوب ادعا: هیچ تفاوتی در شیوه برگزاری آزمون زبان انگلیسی در شرایط متفاوت (نوبت‌های متفاوت برگزاری، سنوات مختلف و ...) و برای داوطلبان گوناگون وجود ندارد. همچنین، شیوه نمره‌دهی پیش از برگزاری آزمون با جزئیات بیان شده است و جزئیات در اعلام نمره عیناً در نظر گرفته شده‌اند.

درون‌مایه (تم) برجسته‌ای که از داده‌ها استخراج شد، نشان می‌داد که تقریباً هیچ ثباتی در برگزاری آزمون زبان انگلیسی عمومی در چند سالی که از برگزاری آزمون به‌شیوه متمرکز می‌گذرد، وجود ندارد. یکی از آزمون‌دهندگان سابق که در زمان مصاحبه، دانشجوی سال اول دوره دکتری بود، نظر خود را چنین بیان کرد: «از سال اولی که آزمون تحت نظر سازمان سنجش و در یک روز برگزار شد تا به امسال، هر سال شیوه برگزاری، ضریب آزمون و مفاد آن تغییر می‌کند. در یکی از سال‌ها، صد سؤال چندگزینه‌ای با نمره منفی بود، سال بعد شصت سؤال بدون نمره منفی و با ضریب دو و امسال هم که کلاً هم ضریب آزمون و هم تعداد سؤالات نصف شده‌اند».

مشابه این نظر را یکی از مدرسان آمادگی آزمون زبان دکتری نیز این‌گونه بیان کرد: «به‌مدت چهار سال و در طی برگزاری چهار کنکور، چهار شیوه کاملاً متفاوت از آزمون دیده شده است. فکر نمی‌کنم هیچ‌جای دنیا این شیوه متحول کردن محتوا و برگزاری آزمون سابقه داشته باشد؛ آن هم برای جذب دانشجوی مقطع دکتری». در مورد شیوه نمره‌گذاری آزمون، درون‌مایه اصلی سخنان افراد شرکت‌کننده این بود که سازمان برگزارکننده آزمون تنها به ارائه اطلاعات کلی در مورد تعداد سؤالات، ضریب آزمون و وجود/عدم وجود نمره منفی برای آزمون زبان بسنده می‌کند و افراد داوطلب تا زمان برگزاری آزمون از داشتن اطلاعات جزئی‌تر در مورد بخش‌های زبانی مورد امتحان، تعداد سؤالات در هر بخش و میزان دشواری آزمون زبان محروم هستند.

ادعای دوم، استنباط دقیق از نمره اکتسابی از آزمون

گواهی مطلوب ادعا: شیوه نمره‌دهی آزمون موارد (بخش‌ها و فعالیت‌ها) و قابلیت‌هایی از توانایی زبان انگلیسی را از لحاظ اهمیت موضوع ضریب‌گذاری کرده است که با مهارت‌های زبان انگلیسی که از یک دانشجوی دوره دکتری انتظار می‌رود، بیشترین هماهنگی و ارتباط را دارد.

از میان ۲۷ دانشجوی مصاحبه‌شونده، ۲۴ نفر بیان کردند که میان آنچه به‌عنوان آزمون زبان انگلیسی ورودی دکتری امتحان گرفته می‌شود و مجموعه توانایی‌ها و فعالیت‌های مورد نیاز یک دانشجوی مقطع دکتری تناسب معناداری وجود ندارد. به‌عنوان نمونه، یکی از دانشجویان سال اول دکتری که دوبار نیز در این آزمون شرکت کرده بود، بیان کرد: «هیچ‌کس از فرد دانشجوی دکتری سطح لغت و یا قوانین گرامر را سؤال نمی‌کند؛ بلکه نیاز من به‌عنوان یک دانشجوی دکتری، داشتن مهارت خواندن و نوشتار در وهله اول و توانایی مکالمه و شنیداری در گام بعدی است؛ چون یک دانشجوی دکتری علاوه‌بر سفرهای احتمالی خارج برای ارائه مقاله یا فرصت مطالعاتی، به مطالعه منابع خارجی و نوشتن مقاله ISI نیاز دارد».

یکی دیگر از مصاحبه‌شوندگان که دوبار در این آزمون شرکت کرده بود، در اظهارنظری کمی متفاوت بیان کرد: «آزمون درک مطلب نسبتاً خوب است، لغت و گرامر هم که فرد باید بلد باشد تا بتواند انگلیسی صحبت کند یا متن بخواند. فقط اگر بخش مکالمه هم جزئی از آزمون بود، خیلی بهتر می‌شد».

ادعای سوم، میزان منصفانه بودن تصمیماتی که بر مبنای این آزمون گرفته می‌شود

گواهی مطلوب ادعا: همه آزمون‌دهندگان به اطلاعات مرتبط با شیوه برگزاری و محتوای

آزمون دسترسی یکسانی داشته‌اند و تصمیمات گرفته‌شده بر مبنای این آزمون برای همه شرکت‌کنندگان یکسان است.

در رابطه با این گواهی، دو درون‌مایه اصلی به دست آمد. گروهی معتقد بودند که همه اطلاعات مرتبط با برگزاری، محتوا و ضریب آزمون از سوی سازمان برگزارکننده (سازمان سنجش) به نحوی مقتضی از طریق وبسایت سازمان و جراید و مطبوعات دیگر اطلاع‌رسانی شده است. به عنوان نمونه، یکی از داوطلبان آزمون بیان کرد: «همه اطلاعات درون سایت سازمان سنجش وجود دارد. نکته خاصی هم وجود ندارد که فراتر از این اطلاعات نیاز باشد». از سوی دیگر، گروهی از داوطلبان آزمون معتقد بودند که اطلاعات مورد نیاز آزمون به صورت مناسبی برای همه افراد اعلام نمی‌شود. یکی دیگر از داوطلبان با تأکید بر این نکته، نظر خود را چنین بیان کرد: «سایت سازمان سنجش اطلاعات خیلی کلی و تاحدی مشخص را ارائه می‌دهد. من داوطلب در خیلی از موارد، نکات مرتبط با آزمون را از یکی از سایت‌هایی که به کنکور دکتری مربوط است و همچنین فروم (انجمن‌های مجازی) مرتبط با کنکور دکتری به دست می‌آوردم. تازه، ما در تهران زندگی می‌کنیم، افراد مناطق دورافتاده و محروم شاید وضعیتشان کاملاً بدتر هم باشد».

در رابطه با میزان منصفانه بودن تصمیمات گرفته‌شده بر مبنای این آزمون، تم برآمده از نظرات افراد این بود که این آزمون معیار خوبی برای قضاوت و تصمیم‌گیری در مورد سطح توانش زبان انگلیسی افراد نیست و به همین دلیل، ابزار مناسبی برای رسیدن به این هدف به شمار نمی‌رود. یکی از استادان دانشگاه که بارها عضو جلسه مصاحبه دکتری بوده است، بیان کرد: «به هیچ عنوان نمره زبان کنکور دکتری معیاری اصولی برای قضاوت در مورد دانش زبانی افراد نیست. البته منظور این است که به فرض یک متقاضی با نمره ۴۰ درصد ضرورتاً زبان قوی‌تری از یک فرد با درصد ۳۵ ندارد؛ وگرنه قاعدتاً اگر فردی درصدهای بالایی چون ۷۰ یا ۸۰ داشته باشد، به احتمال زیاد زبان انگلیسی قوی دارد و بالعکس». یکی از دانشجویان سال نخست دکتری که دوبار در این آزمون شرکت کرده بود، نظر خود را در مورد منصفانه بودن قضاوت بر مبنای نمره این آزمون چنین بیان کرد: «سال اول خیلی برای زبان مطالعه کردم و نتیجه نهایی ۱۷ درصد بود. سال بعد اصلاً برای زبان وقت نگذاشتم و کاملاً شانسی به سؤالات جواب دادم و نمره آزمون من ۳۸ درصد شد. حالا به نظر شما کدام یکی از این دو نمره سطح واقعی زبان من را نشان می‌دهد؟».

یکی دیگر از ویژگی‌های آزمون منصفانه، عدم اریب سؤالات آزمون به سود یا به زیان

افراد با پیشینه‌های تحصیلی، قومی، نژادی و جنسیتی متفاوت بود (کیانی و دیگران، ۱۳۹۲). در این مورد، درون‌مایه اصلی نتایج بیانگر عدم اریب آزمون در موارد یادشده بود. تنها گروهی از داوطلبان آزمون معتقد بودند که پاسخگویی سؤالات در بخش متن مستلزم داشتن پیش‌زمینه در زمینه‌های غیرمرتبط با آزمون بوده است. به‌عنوان نمونه، یکی از شرکت‌کنندگان آزمون رشته زبان و ادبیات فارسی معتقد بود: «آزمون متن بخش زبان کنکور امسال در زمینه‌های فلسفی و تاریخی بود و برای کسی که در این موارد اطلاع خاصی نداشت، فهمیدن متن و جواب دادن به سؤالات خیلی سخت بود». مرتبط نبودن پیش‌زمینه محتوای هر آزمون با تخصص مورد سنجش از داوطلبان عاملی منفی در درجه منصفانه بودن آزمون تلقی می‌شود.

ادعای چهارم، نتایج مثبت یا منفی آزمون برای تمامی ذی‌نفعان و ابعاد اجتماع در کوتاه‌مدت، میان‌مدت و بلندمدت

گواهی مطلوب ادعا: نتایج آزمون برای همه ذی‌نفعان سودمند است و این آزمون می‌تواند باعث ایجاد روند مثبت در یادگیری و تدریس زبان انگلیسی شود.

در داده‌ها، دو تم اصلی مرتبط با این گواهی دیده شد. گروهی از پرسش‌شوندگان بیشتر عقیده داشتند که برگزاری آزمون بدین شیوه، تأثیر مثبت و چشمگیری چه در ابعاد اجتماعی و چه از نظر آموزش و یادگیری زبان انگلیسی خواهد داشت. یکی از مدرسان دوره‌های آمادگی آزمون زبان دکتری معتقد بود: «به این دلیل که این تست ضریب و تعداد سؤالات پایینی در کنکور سال گذشته درمقایسه با سال‌های قبل و دروس دیگر داشت، در افراد داوطلب کنکور برای تقویت زبان انگلیسی انگیزه ایجاد نمی‌کند». یکی از استادان رشته علوم تربیتی که درس زبان تخصصی کارشناسی ارشد را تدریس می‌کند، بیان کرد: «تاکنون این آزمون هیچ تأثیر قابل مشاهده‌ای در سطح علاقه و یادگیری زبان دانشجویان من نداشته است. حداقل اگر ضریب آزمون را درمقایسه با آزمون تخصصی افزایش دهند، احتمال افزایش انگیزش برای تقویت زبان افرادی که داوطلب آزمون دکتری هستند وجود دارد».

از سوی دیگر، گروه دیگری از شرکت‌کنندگان این مطالعه معتقد بودند که این آزمون تأثیر مثبت، نه چشمگیر در یادگیری زبان انگلیسی دارد. یکی از داوطلبان آزمون که برای اولین بار در آزمون شرکت کرده بود، گفت: «من و بسیاری دیگر از دوستان داوطلب شرکت در آزمون تصمیم گرفتیم که برای کنکور، منابع احتمالی زبان را بخیریم و مطالعه کنیم [...] در هر صورت، زبان هم بخش مهمی از سه درس اصلی کنکور دکتری است».

۵. جمع‌بندی و پیشنهادات

هرچند در مطالعات کیفی، نتیجه‌گیری نهایی برعهده مخاطب است، با توجه به یافته‌های این پژوهش به نکات مهمی دست یافتیم. نکته نخست این است که تغییرات پی‌درپی و بنیادین این آزمون که به‌نظر بسیاری از شرکت‌کنندگان، براساس شیوه سعی و خطا برگزار شده است، علاوه‌بر اینکه در اندازه‌گیری توانش زبان انگلیسی افراد شرکت‌کننده موفق نبوده، نتوانسته است انگیزش مثبتی در یادگیری زبان و به‌دنبال آن، ارتقای سطح معلومات زبانی افراد داوطلب آزمون ایجاد کند. اینکه چرا یک آزمون توانش زبان با این درجه اهمیت، در دوره‌های متوالی اجرا نتوانسته است به یک ثبات (حتی نسبی) برسد، بسیار قابل بحث است؛ اما نباید این نکته را فراموش کنیم که نبودن ثبات در برگزاری یک آزمون، از اصلی‌ترین عوامل واریانس غیرمرتبط سازه^{۲۳} و خطای اندازه‌گیری^{۲۴} (Spanjers, 2007) و به‌دنبال آن، عدم پایایی نمرات به‌دست‌آمده از یک آزمون است (بک من و پالمر، ۸۰) و به‌دنبال آن، این نکته به این دلیل درخور توجه است که اطمینان نداشتن به رابطه بین نمرات کسب‌شده از این آزمون و سطح واقعی زبان داوطلبان درمیان ذی‌نفعان گوناگون، به‌منزله فقدان روایی آزمون نیز تلقی می‌شود (کارلسون و دیگران، ۲۰۰۹).

ارائه یک چارچوب و جدول مختصات آزمون^{۲۵} که بر مبنای نیازسنجی از ذی‌نفعان باشد و با نیازهای بومی و امکانات سخت‌افزاری و نرم‌افزاری برگزاری آزمون نیز هماهنگ باشد، ازجمله پیشنهادات موجود در این زمینه است. بررسی قیاسی شیوه برگزاری آزمون‌های مشابه در مناطق دیگر دنیا و استفاده و استنباط از تجربیات به‌دست‌آمده نیز ممکن است در بسیاری از موارد، راهنمای اجرای موفق و باثبات آزمون (بدون نیاز به انجام سعی و خطای مجدد) باشد.

از موارد دیگری که براساس چارچوب ارزشیابی استدلال‌محور به‌دست آمد، فقدان تأثیرگذاری مثبت این آزمون بر روند تدریس و یادگیری زبان انگلیسی در محیط‌های آموزشی و دانشگاهی است. همچنین، ارائه نشدن اطلاعات کافی، قابل قبول و تفسیرپذیر در مورد سطح زبان انگلیسی داوطلبان ورود به مقطع دکتری به افراد تصمیم‌گیرنده در مورد پذیرش/عدم پذیرش دانشجویان در مقطع دکتری، ازجمله موارد مهم و قابل ذکری است که نیاز به بازنگری در محتوا و شیوه برگزاری آزمون فعلی را دوچندان می‌کند. در این زمینه، پیشنهاد می‌کنیم که چارچوب آزمون بر مبنای دامنه استفاده زبان در شرایط مقصد^{۲۶} متناسب با مستلزمات دوره‌های دکتری آماده شود. همچنین، شفافیت و اطلاع‌رسانی مناسب در مورد

بخش‌ها و موارد آزمون زبان ممکن است که به تأثیرگذاری مثبت، تفسیرپذیری و افزایش اعتماد به آزمون زبان دکتری کمک شایانی کند (Saif, 2006: 30).

تغییر در محتوای آزمون و تا حد امکان نزدیک کردن آن به نیازهای واقعی دانشجویان در مقطع دکتری، مانند خواندن متون دانشگاهی مرتبط با رشته تحصیلی (نه متون عمومی)، توانش نگارش با زبانی علمی و مهارت بیان از طریق گفتار، از جمله مواردی بودند که بسیاری از داوطلبان کنونی آزمون، آزمون‌دهندگان سابق که موفق شدند به مقطع دکتری راه یابند و تعدادی از استادان آزمون‌سازی و روان‌سنجی پیشنهاد دادند. همچنین، به‌منظور تقویت جایگاه این درس مهم و افزایش تأثیر آن در انگیزش یادگیری زبان انگلیسی برای دانشجویان و فارغ‌التحصیلان دکتری، بسیاری از استادان روان‌سنجی، ارزشیابی و علوم تربیتی پیشنهاد دادند که ضریب و تعداد سؤالات آزمون زبان دکتری افزایش یابد.

از آنچه گفتیم نتیجه می‌گیریم که با توجه به اینکه آزمون‌های فعلی که برای سنجش توانایی زبان انگلیسی داوطلبان ورود به بالاترین مقطع تحصیلی در کشور برگزار می‌شوند نقاط ضعف و قوت گوناگونی دارند، شاید بهترین شیوه، بهینه‌سازی تدریجی آزمون بر مبنای استفاده از نیازها و نظرات ذی‌نفعان، به‌کارگیری جدیدترین شیوه‌های ارزشیابی آموزشی و البته تصمیم‌گیری درست و مقتضی سیاست‌گذاران و برگزارکنندگان آزمون باشد.

۶. پی‌نوشت‌ها

1. test validity
2. content validity
3. criterion-related validity
4. construct validity
5. face validity
6. unitary concept of validity
7. language proficiency
8. reliability
9. psychometrics
10. test validation
11. claim
12. data
13. warrant
14. backing
15. rebuttal
16. qualitative approach

17. content analysis
18. purposeful sampling
19. data saturation
20. structured interview
21. focus group
22. conceptual translation method
23. construct irrelevant variance
24. measurement error
25. table of specifications
26. target language use domain

۷. منابع

- کیانی، غلامرضا؛ مهدوی، بابک و رضا غفارثمر. (۱۳۹۲). «مقایسه میزان منصفانه بودن انواع آزمون‌های زبان انگلیسی در دوره دبیرستان». *جستارهای زبانی*. د. ۴. ش ۱. صص ۱۸۱-۲۰۲.
- کیانی، غلامرضا. (۱۳۸۸). «استانداردسازی آزمون زبان انگلیسی داوطلبان مقطع دکترا». *طرح کاربردی*. دانشگاه تربیت مدرس تهران.

References:

- Bachman, L. F. & I. Choi (2012). Review of pearson test of English academic: Building an assessment use argument. *Language Testing*, 29(4), Pp. 603–619.
- Bachman, L. F. & A. S. Palmer (2010). *Language Assessment in Practice*. Oxford University Press: Oxford
- Carlson, N. R.; D. Heth; H. Miller; J. Donahoe & G. N. Martin (2009). *Psychology: The Science of Behavior*. Pearson.
- Cronbach, L. J. & P. E. Meehl (1955). “Construct validity in psychological tests”. *Psychological Bulletin*. 52(4). P. 281
- Kane, M. (2004). “Certification Testing as an Illustration of Argument-Based Validation”. *Measurement: Interdisciplinary Research and Perspectives*. 2(3). Pp. 135- 170.
- Kane, M. T. (2013). “Validating the interpretations and uses of test scores”. *Journal of Educational Measurement*. 50(1). Pp. 1-73.
- Kiany, Gh.R. (2008). *Standardization of the English Test of the Ph.D. Entrance Examination*. An Applied Plan. Tarbiat Modares University [In Persian].

- Kiany, Gh.R.; B. Mahdavy & R. Ghaffar Samar (2013). Comparing the fairness of high school English exams in Iran. *Language Related Research*. 4 (1). Pp.181-202 [In Persian].
- Messick, S. (1989). "Meaning and values in test validation: The science and ethics of assessment". *Educational Researcher*. 18(2). Pp. 5-11.
- Saif, S. (2006). "Aiming for positive washback: A case study of international teaching assistants". *Language Testing*. 23(1). Pp. 1-34.
- Spanjers, D. M. (2007). *Cognitive Engagement as a Predictor of Achievement*. ProQuest.
- Toulmin, S. E. (1957•2003). *The Uses of Argument*. Cambridge University Press.
- Wang, H. (2010). *Investigating the Justifiability of an Additional Test Use: An Application of Assessment Use Argument to English as a Foreign Language*. Unpublished PhD dissertation. UCLA.

۸. پیوست

سؤالات مصاحبه ساختاریافته

۱. آیا به نظر شما، هیچ تفاوتی در شیوه برگزاری آزمون زبان انگلیسی در شرایط متفاوت (نوبت برگزاری، سنوات مختلف و ...) و برای داوطلبان گوناگون وجود دارد؟
۲. آیا شیوه نمره‌دهی پیش از برگزاری آزمون، با جزئیات بیان شده است و جزئیات در اعلام نمره، عیناً لحاظ شده‌اند؟
۳. آیا نمرات بخش‌های مختلف آزمون انگلیسی (واژگان، گرامر، متن و ...) و نمرات به‌دست‌آمده از نوبت‌های مختلف برگزاری آزمون از نظر محتوایی با یکدیگر پیوستگی دارند؟
۴. آیا به نظر شما، گروه‌های مختلف افراد (رشته تحصیلی، قومی، نژادی، زبانی و ...) از آزمون زبان عمومی نتایج نامتناسبی دریافت می‌کنند؟
۵. آیا اساس بخش‌های (مهارت‌ها) آزمون براساس زبان انگلیسی واقعی مورد استفاده در محیط‌های آکادمیک و دانشگاهی است؟
۶. آیا به نظر شما، نوشتارهای راهنما در بخش‌های مختلف آزمون به‌وضوح ویژگی‌های فعالیت خواسته‌شده از داوطلب آزمون را بیان می‌کنند؟
۷. آیا شیوه برگزاری آزمون به شرکت‌کنندگان آزمون این قابلیت را می‌دهد که بتوانند حداکثر توانایی ممکن را که در زبان انگلیسی دارند به نمایش بگذارند؟

۸. آیا به نظر شما، بخش‌ها و فعالیت‌های آزمون محتوایی را شامل می‌شود که از جهات فرهنگی، محتوایی یا زبان‌شناختی برای برخی از آزمون‌دهندگان توهین‌آمیز محسوب شود؟
۹. آیا با همه آزمون‌دهندگان به صورت یکسان و غیرتبعیض‌آمیز، در تمامی فرآیندهای برگزاری آزمون برخورد می‌شود؟
۱۰. آیا شیوه نمردهی آزمون، موارد (بخش‌ها و فعالیت‌ها) و قابلیت‌هایی از توانایی زبان انگلیسی را از لحاظ اهمیت موضوع ضریب‌گذاری کرده است که با مهارت‌های زبان انگلیسی که از یک دانشجوی دوره دکتری انتظار می‌رود، بیشترین هماهنگی و ارتباط را داشته باشد؟
۱۱. آیا تفسیر برآمده از این ارزشیابی می‌تواند اطلاعات کافی را برای تصمیم‌گیری در مورد توانایی زبان انگلیسی افراد در اختیار قرار دهد؟ به سخنی دیگر، با توجه به نتایج و تفسیرهای نمرات این آزمون، آیا اطلاعات کافی به تصمیم‌گیرندگان داده می‌شود تا بتوانند به درستی در مورد کفایت سطح زبانی فرد تصمیم‌گیری کنند؟
۱۲. آیا به نظر شما، ارزش‌های اجتماعی، آموزشی و مفاد قانونی موجود در مورد جدی بودن تبعات اشتباهات احتمالی در تعیین دقیق توانایی زبان انگلیسی فرد و تصمیمات گرفته‌شده ناشی از آن، حساب‌شده و دقیقاً از قبل مشخص و برنامه‌ریزی شده هستند؟
۱۳. آیا آزمون‌دهندگان و دیگر ذی‌نفعان متأثر از آزمون کاملاً از شیوه تصمیم‌گیری در مورد نمردهی این آزمون آگاه هستند و این تصمیمات همان‌گونه که ذکر شده‌اند، در نمردهی تأثیر دارند؟
۱۴. آیا به نظر شما، نتایج و تأثیرات این آزمون (هرچه که باشد) برای هر گروه از ذی‌نفعان (آزمون‌دهندگان و خانواده‌هایشان، دانشگاه‌ها، سازمان‌هایی که فرد قبول در آزمون در آن‌ها شاغل خواهد شد، آزمون‌گیرندگان و ...) مفید است؟
۱۵. آیا نمرات و نتایج این آزمون کاملاً به صورت محرمانه برای فرد آزمون‌دهنده اعلام می‌شود؟
۱۶. آیا آزمون‌دهندگان و دیگر افراد تحت‌تأثیر آزمون کاملاً در جریان شیوه تصمیم‌گیری بر مبنای این آزمون هستند و تصمیمات آن‌گونه که قبلاً برای آن‌ها بیان شده، گرفته می‌شود؟
۱۷. آیا به نظر شما، نتیجه و نمرات آزمون در موعد مقرر اعلام می‌شود؟
۱۸. آیا به نظر شما، این آزمون باعث تأثیرگذاری مثبت بر روند تدریس و یادگیری زبان انگلیسی در محیط‌های آموزشی و دانشگاهی می‌شود؟ در این صورت، آیا استفاده از این آزمون ممکن است که به سود دانشجویان، استادان زبان، دانشگاه‌ها و ... باشد؟
۱۹. آیا تصمیمات گرفته‌شده در مورد ورود/عدم ورود دانشجویان به مقطع بالاتر بر مبنای این آزمون، برای داوطلبان و همه ذی‌نفعان، سودمند خواهد بود؟