

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

مهربان هماوندی<sup>۱</sup>، غلامرضا کیانی<sup>۲\*</sup>، رضا غفارشمر<sup>۳</sup>، رامین اکبری<sup>۴</sup>

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

دریافت: ۹۲/۲/۱۳ پذیرش: ۹۲/۳/۲۶

### 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 ازسوی 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 استدلال محور است. برای رسیدن به این هدف، از گروههای مختلف ذی‌نفع در آزمون، ازجمله داوطلبان آزمون دکتری، مدرسان دوره‌های آمادگی آزمون زبان دکتری، استدان رشته تعلیم و تربیت و آزمون‌سازی زبان با استفاده از روش کیفی و ابزارهای مصاحبه (گروهی و فردی) داده‌ها جمع‌آوری شد. برایه روش کیفی تحلیل محتوا، الگوهای یافتشده در مجموعه داده‌ها براساس قواعد ازپیش‌ تعیین شده بررسی محتوا و با رعایت ویژگی‌های پژوهش ارائه می‌شوند. یافته‌های پژوهش نشان دادند که آزمون فعلی ازنظر شرکت‌کنندگان به دگرگونی‌های گاه اساسی در ابعاد محتوا، شیوه برگزاری و ضریب اختصاصی آزمون نیازمند است. علاوه بر ارائه نتایج، در پایان، پیشنهاداتی ارائه و دلالتها، چالش‌ها و ملزمات اجرایی‌سازی آن‌ها بیان شده‌اند.

**واژگان کلیدی:** آزمون زبان دکتری، بررسی روایی آزمون، ارزشیابی استدلال محور، ابعاد محتوا، ضریب اختصاصی آزمون.

## ۱. مقدمه

روایی آزمون<sup>۱</sup> از دیرباز یکی از کلیدی‌ترین ویژگی‌های آزمون به‌شمار می‌رفت. روایی در گذشته و به‌صورت معمول، از چهار منظر روایی محتوا، معیار، سازه و صوری<sup>۲</sup> می‌شد. پس از ارائه مفهوم وحدت روایی<sup>۳</sup> ساموئل مسیک (۱۹۸۹)، به‌منظور بررسی روایی آزمون و مقبولیت این نظریه از منظر عالمان تعلیم و تربیت، چارچوب‌های متفاوتی برای عملیاتی کردن و شیوه‌مندسازی آن ارائه شد. از سوی دیگر، پیشرفت فناوری، ارتباطات و انواع پایگاه‌های داده مرتبط با رشتۀ‌های مختلف، توانش زبان انگلیسی<sup>۴</sup> را به امری حیاتی در زندگی فردی و دانشگاهی افراد تبدیل کرده است. در این راستا، مقاضیان 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 مقاضیان 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 زمینه‌آزمون‌های زبان انگلیسی پراهمیت در کشور، همچنان وضعيت روبه‌رشدی دیده نمی‌شود و «تقریباً همه آزمون‌های ملی کشور با آزمون‌های استاندارد فاصلۀ بسیار فراوان دارند» (کیانی، ۱۳۸۸: ۲). با تغییر شیوه برگزاری آزمون ورودی دوره‌های دکتری داخل، متمرکز شدن برگزاری آن و عهددار شدن برگزاری آزمون ازسوی سازمان سنجش و آموزش کشور به‌عنوان تنها متولی، شیوه آزمون‌گیری از سطح توانش زبان انگلیسی مقاضیان نیز دستخوش تغییراتی شد؛ تغییراتی چون یکی شدن آزمون‌های متنوع ورودی زبان انگلیسی که در سال‌های پیش از آن برگزار می‌شد، تغییرات در محتوای آزمون‌ها گاه به صورت اساسی و شیوه متفاوت بارمبنده و تعداد سؤالات بخش‌های مختلف محتوای آن. همه این تغییرات در حالی 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. تاکنون پیدا است 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 درزمنینه 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 درحال حاضر به عنوان بخشی از کنکور ورودی دکتری داخل ازسوسی 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

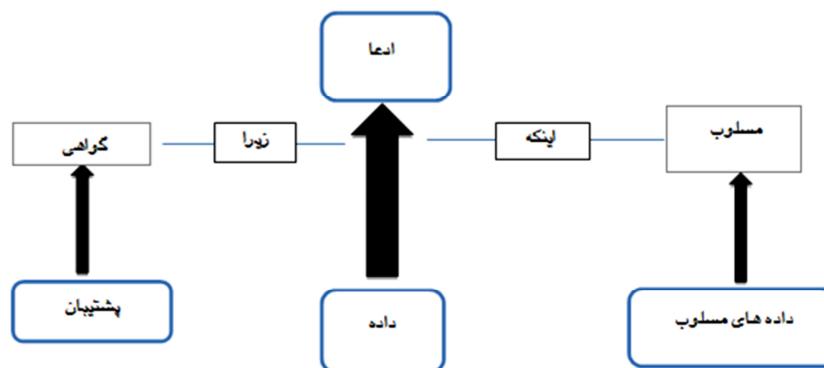
یکی از ویژگی‌هایی که هر آزمون باید داشته باشد تا استاندارد به شمار آید، پایایی<sup>۸</sup> و روایی است. روایی در آزمون‌سازی به صورت معمول عبارت است از «میزان هم‌سنگی بین نمره به دست آمده از آزمون با ویژگی که آزمون مورد نظر برای اندازه‌گیری آن ساخته شده است» (Cronbach & Meehl, 1955: 23). مهم‌ترین عاملی که در ارزیابی یک آزمون باید به آن توجه شود، روایی آزمون است. مقصود از روایی آزمون مناسب بودن، بامعنا بودن و مفید بودن استنbatهای خاصی است که از روی نمرات آزمون انجام می‌شود (کین، ۲۰۰۷). برای تأیید این‌گونه استنbatها، باید شواهدی جمع‌آوری شود که در فرهنگ روان‌سنجی<sup>۹</sup> به آن رواسازی آزمون<sup>۱۰</sup> گفته می‌شود.

### ۱-۲. 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 (۲۰۰۳، ۱۹۵۸) گرفته شده است. بر مبنای نظریه مسیک، هر آزمون تنها شامل یک نشانگر کیفی از روایی آن است که برای به دست آوردن این نشانگر، پژوهشگر باید از روش‌های عمدتاً کیفی به منظور تحلیل

پایایی و مشخصه‌های روایی دیگر استفاده کند و درنهایت، به یک نظر واحد از درجهٔ روایی آزمون برسد. مسیک این نوع روایی را که تنها نوع روایی است، روایی سازه آن آزمون و نظریهٔ خود را وحدت روایی نام نهاد. اگرچه بسیاری از دانشمندان و پژوهشگران روان‌سنجی ابتدا از نظریهٔ مسیک استقبال کردند، این نظریه به‌دلیل شفاف نبودن چگونگی کاربردش در دنیای واقعی، از مرحلهٔ نظریهٔ فراتر نرفت و در آزمونی خاص به‌کار برده نشد. کاری که بک من، به عنوان نظریهٔ پرداز اصلی مدل ارزشیابی استدلال محور انجام داد این بود که در مدل خود، نظریهٔ مسیک را به دنیای واقعی ارزشیابی آموزشی نزدیک کرد و برای اجرای نظریهٔ وحدت روایی، مراحلی کاملاً مشخص تدوین نمود. در همین راستا و به‌منظور عینی ساختن مراحل بررسی سازه آزمون و طراحی چارچوب یک آزمون، بک من از مدل استدلال محور تولمین کمک گرفت. تولمین (۱۹۵۸)، به عنوان 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<sup>۱۰</sup> بر عینیات<sup>۱۱</sup> (داده) مبتنی است. هرچه برای این ادعا بر مبنای گواهی<sup>۱۲</sup> مورد انتظار پشتیبان<sup>۱۳</sup> بیشتری وجود داشته باشد، ادعا قوت می‌گیرد و هرقدرت که برای ادعا مسلوب<sup>۱۴</sup> وجود داشته باشد، ادعا ضعیفتر می‌شود.

ساختار کلی نظریهٔ تولمین به صورت شکل زیر است:

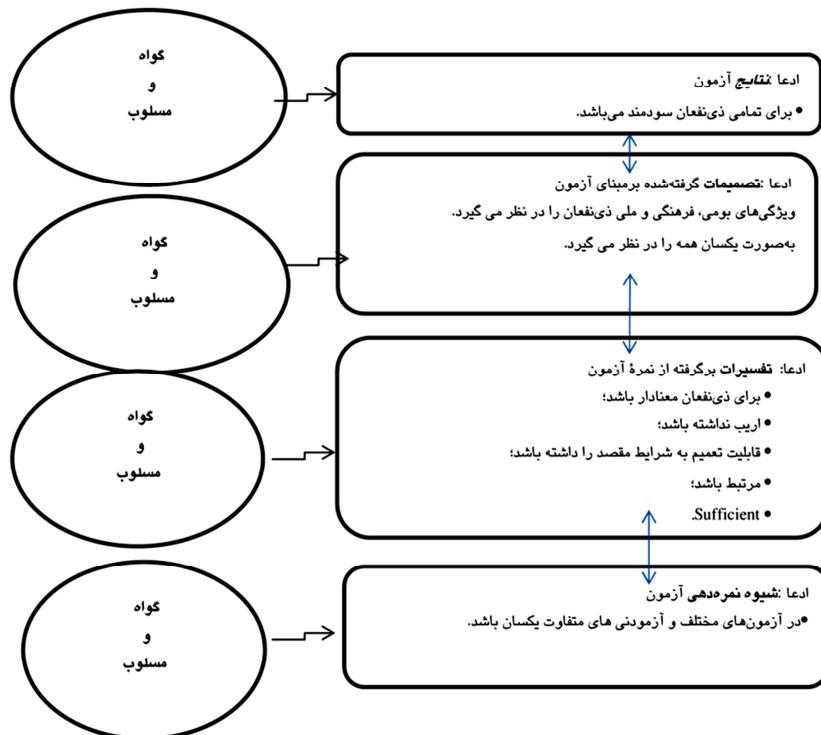


شکل ۱ ساختار استدلال محور تولمین (۲۰۰۳)

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

منفی (مسلوب)، کوشید که ادعای روایی سازه آزمون مورد نظر را ثابت کند. در اینجا، باید به این نکته توجه کنیم که اولاً نظریه وحدت روایی سازه آزمون (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. نکته مهم این است که نتیجه‌گیری نهایی درمورد پذیرش یا رد روایی سازه بر عهده مخاطب و خواننده (نه ضرورتاً پژوهشگر) است.

در این روش، برای رسیدن به هدف نهایی (روایی سازه آزمون)، باید چهار هدف زیرمجموعه در نظر گرفته شود. تلاش می‌شود که برای هریک از این هدف‌ها از میان شرکت‌کنندگان پژوهش، گواهی به منظور پشتیبان و همچنین مسلوبی به عنوان مردودکننده گواهی یافته شود. هریک 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 پالمر (۲۰۱۰)

مهمترین ویژگی این روش توانایی بررسی سازهٔ یک آزمون و ویژگی‌های مرتبط با آن و همچنین به وجود آوردن چارچوب اولیهٔ یک آزمون زبان با توجه به مبناهای نظری این تئوری است. در این روش، تأثیرات مثبت یک آزمون بر آزمودنی‌ها به عنوان سنگ بناء نقطهٔ آغازین آن درنظر گرفته می‌شود؛ بنابراین، اولین ویژگی یک آزمون استاندارد تأثیرگذاری مثبت آن در سطح آزمودنی‌ها، مریبیان، همهٔ ذی‌نفعان و به‌طور کلی جامعه است. قدم‌های بعدی شامل تصمیم‌گیری دربارهٔ نمرهٔ قبولی/ردی، چگونی تفسیر و بررسی نتایج آزمون، شیوهٔ آزمون‌گیری و محتویات آزمون است.

وانگ (۲۰۱۲)، کین (۲۰۱۳) و بک من و چوی (۲۰۱۲) از جملهٔ پژوهشگرانی هستند که تاکنون بر مبنای ارزشیابی استدلال محور پژوهش کرده‌اند. آن‌ها در مطالعاتشان یا در پی بررسی و تحلیل سازهٔ یک آزمون زبان بوده‌اند و یا براساس این شیوهٔ 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 استدلال محور این است که یک آزمون کاملاً استاندارد در یک محیط خاص، در همهٔ محیط‌ها به عنوان آزمون استاندارد به کار نمی‌رود.

### ۳. روش‌شناسی پژوهش

این مطالعه را با رویکرد کیفی<sup>۱۷</sup> منطبق با چارچوب ارزشیابی استدلال محور و با استفاده از روش تحلیل محتوا<sup>۱۸</sup> 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<sup>۱۹</sup> بود. در این پژوهش، معیار انتخاب حجم نمونه رسیدن داده‌ها به سطح اشباع اطلاعاتی<sup>۲۰</sup> بود و زمانی که اطلاعات جدیدی به دست نیامد، نمونه‌گیری به پایان رسید. از بین ذی‌نفعان دیگر این آزمون، ۳ نفر از استادان رشته‌های مختلف دانشگاهی که سابقاً 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. به‌دلیل دسترسی نداشتن به سیاست‌گذاران کلان آزمون دکتری، داده‌ای در رابطه با این گروه جمع‌آوری نشد. ابزار جمع‌آوری داده‌ها برگزاری مصاحبه‌های ساختاریافتاً فردی<sup>۲۰</sup> و

مصاحبه‌های متمرکز گروهی<sup>۲۱</sup> است. برای مصاحبه، ۱۹ سؤال را برپایه چارچوب پیشنهادی بک من و پالمر (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<sup>۲۲</sup>، بررسی و تأیید شد. پس از فرآیند گردآوری داده‌ها، همه مصاحبه‌های ضبط شده 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 پدیدارشده از داده‌های به‌دست‌آمده، اساس بررسی نهایی را تشکیل دادند.

#### ۴. نتایج و یافته‌های پژوهش

چنانکه گفتیم، ارزشیابی استدلال‌محور برمبنای پژوهش کیفی است؛ از این‌رو، پس از هر ادعای مطروحه براساس ارزشیابی استدلال‌محور، تم‌های برجسته گواهی مرتبط را که ممکن است نقش گواه و یا مسلوب داشته باشند، به همراه 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، سال‌های 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 درنظر 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 به شیوه 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 تحت‌نظر سازمان سنجش و در یک روز برگزار شد تا به امسال، هر سال شیوه برگزاری، ضربی 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 ضربی دو و امسال هم که کلاً هم ضربی 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، ضربی آزمون و وجود/عدم وجود نمره منفی برای آزمون زبان بسته 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 الحاظ اهمیت موضوع ضربی‌گذاری 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، ۴ نفر بیان کردند که میان آنچه به عنوان آزمون زبان انگلیسی ورودی دکتری امتحان گرفته می‌شود و مجموعه توانایی‌ها و فعالیت‌های مورد نیاز یک دانشجوی مقطع دکتری تناسب معناداری وجود ندارد. به عنوان 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 به عنوان یک دانشجوی دکتری، داشتن مهارت خواندن و نوشتن در وهله اول و توانایی مکالمه و شنیداری در گام بعدی است؛ چون یک دانشجوی دکتری علاوه بر سفرهای احتمالی خارج برای ارائه مقاله یا فرصت مطالعاتی، به مطالعه منابع خارجی و نوشتمن 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 برمبنای 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 (انجمان‌های 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 برمبنای 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 به‌فرض یک مقاضی با نمره ۴۰ درصد ضرورتاً زبان قوی‌تری از یک فرد با درصد ۳۵ ندارد؛ و گرنه قاعده‌اً اگر فردی در صدهای 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 برمبنای نمره این آزمون چنین بیان کرد: «سال اول خیلی برای زبان مطالعه کردم و نتیجه‌نهایی ۱۷ درصد بود. سال بعد اصلاً برای زبان وقت نگذاشتم و کاملاً شناسی به سوالات جواب دادم و نمره آزمون من ۲۸ 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 سوالات 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 ترتیب بیانگر عدم ارتباط 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. به عنوان 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». مرتب‌بودن پیش‌زمینه محتواهای 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 نخواهد 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 در مقایسه با سال‌های قبل و دروس دیگر داشت، در افراد داوطلب کنکور برای تقویت زبان انگلیسی انگیره 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 در مقایسه با آزمون تخصصی افزایش دهن، 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 بخريم 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 بر عهده مخاطب است، با توجه به یافته‌های این پژوهش به نکات مهمی دست یافته‌یم. نکته نخست این است که تغییرات پی‌درپی و بنیادین این آزمون که به‌نظر بسیاری از شرکت‌کنندگان، براساس شیوه سعی و خطاب‌گزار شده است، علاوه‌بر اینکه در اندازه‌گیری توانش زبان انگلیسی افراد شرکت‌کننده موفق نبوده، نتوانسته است انگیزش مثبتی در یادگیری زبان و به‌دبال 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 واریاش غیرمرتب سازه<sup>۳۳</sup> و خطای اندازه‌گیری<sup>۴۴</sup> (Spanjers, 2007: 80) و به‌دبال آن، عدم پایایی نمرات به‌دست‌آمده از یک آزمون است (بک من و پالمر، ۲۰۱۰). این نکته به این دلیل درخور توجه است که اطمینان داشتن به رابطه بین نمرات کسب شده از این آزمون و سطح واقعی زبان داوطلبان در میان ذی‌نفعان گوناگون، به‌منزله فقدان روایی آزمون نیز تلقی می‌شود (کارلسون و دیگران، ۲۰۰۹).

ارائه یک چارچوب و جدول مختصات آزمون<sup>۵۰</sup> که بر مبنای نیازمنجی از ذی‌نفعان باشد و با نیازهای بومی و امکانات سخت‌افزاری و نرم‌افزاری برگزاری آزمون نیز هماهنگ باشد، از جمله 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 استدلال محور 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، از جمله 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<sup>۳۳</sup> 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 روان‌سننجی پیشنهاد دادند. همچنین، به منظور تقویت جایگاه این درس مهم و افزایش تأثیر آن در انگیزش یادگیری زبان انگلیسی برای دانشجویان و فارغ‌التحصیلان دکتری، بسیاری از استادان روان‌سننجی، ارزشیابی و علوم تربیتی پیشنهاد دادند که ضریب و تعداد سؤالات آزمون زبان دکتری افزایش یابد.

از آنچه گفته شده 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، به کارگیری 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 غفارشمر. (۱۳۹۲). «مقایسه میزان منصفانه بودن انواع آزمون‌های زبان انگلیسی در دوره دبیرستان». *جستارهای زبانی*. 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 بهنظر شما، هیچ تفاوتی در شیوه برگزاری آزمون زبان انگلیسی در شرایط متفاوت (نوبت برگزاری، سنترات 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 به‌دست آمده از نوبت‌های مختلف برگزاری آزمون از نظر محتوایی با یکدیگر پیوستگی دارند؟
۴. آیا بهنظر 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 بهنظر شما، نوشتارهای راهنمای 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 حداقل توانایی ممکن را که در زبان انگلیسی دارند به نمایش بگذارند؟

۸. آیا بهنظر 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 نمره‌دهی آزمون، موارد (بخش‌ها و فعالیت‌ها) و قابلیت‌هایی از توانایی زبان انگلیسی را از لحاظ اهمیت موضوع ضربه‌گذاری 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 درمورد توانایی زبان انگلیسی افراد در اختیار قرار دهد؟ به سخنی دیگر، با توجه به نتایج و تفسیرهای نمرات این آزمون، آیا اطلاعات کافی به تصمیم گیرندگان داده می‌شود تا بتوانند به درستی درمورد کفايت سطح زبانی فرد تصمیم‌گیری کنند؟
۱۲. آیا بهنظر شما، ارزش‌های اجتماعی، آموزشی و مفاد قانونی موجود درمورد جدی بودن تبعات اشتباهات احتمالی در تعیین دقیق توانایی زبان انگلیسی فرد و تصمیمات گرفته‌شده ناشی از آن، حساب شده و دقیقاً از قبل مشخص و برنامه‌ریزی شده هستند؟
۱۳. آیا آزمون‌دهندگان و دیگر ذی‌نفعان متأثر از آزمون کاملاً از شیوه تصمیم‌گیری درمورد نمره‌دهی این آزمون آگاه هستند و این تصمیمات همان‌گونه که ذکر شده‌اند، در نمره‌دهی تأثیر دارند؟
۱۴. آیا بهنظر 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 محترمانه 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 بهنظر شما، نتیجه و نمرات آزمون در موعده مقرر اعلام می‌شود؟
۱۸. آیا بهنظر 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 به‌سود دانشجویان، استادان زبان، دانشگاه‌ها و ... باشد؟
۱۹. آیا تصمیمات گرفته‌شده درمورد ورود/ عدم ورود دانشجویان به مقطع بالاتر بر مبنای این آزمون، برای داوطلبان و همه ذی‌نفعان، سودمند خواهد بود؟