



بررسی و آسیب‌شناسی آزمون‌های بسندگی زبان انگلیسی وابسته به وزارت عتف و دانشگاه‌های ایران و نیازسنجی زبانی ذی‌نفعان*

رضا پیش‌قدم*

(نویسنده مسئول)

استاد آموزش زبان انگلیسی و روان‌شناسی تربیتی، دانشگاه فردوسی مشهد، مشهد، ایران
pishghadam@um.ac.ir



شیمای ابراهیمی**

استادیار آموزش زبان فارسی، دانشگاه فردوسی مشهد، مشهد، ایران
Email: shimaabrahimi@um.ac.ir



شقایق شایسته***

استادیار آموزش زبان انگلیسی، دانشگاه فردوسی مشهد، مشهد، ایران
Email: shayesteh@um.ac.ir



سحر طباطبائی فارانی****

دکتری آموزش زبان انگلیسی، دانشگاه فردوسی مشهد، مشهد، ایران
Email: sahar.tabatabaee1983@gmail.com



هانیه جاجرمی*****

استادیار آموزش زبان انگلیسی، مؤسسه آموزش عالی بهار، مشهد، ایران
Email: hjajarmi@gmail.com



چکیده

پژوهش پیش‌رو در تلاش است که با نیازسنجی مخاطبان (دانش‌آموختگان و دانشجویان مقطع دکتری) به بررسی و آسیب‌شناسی آزمون‌های بسندگی زبان انگلیسی وابسته به وزارت عتف و دانشگاه‌های ایران بپردازد و پس از آن آزمون استاندارد و یکسان را به‌عنوان آزمون واحد پیشنهاد دهد. بدین منظور پس از موضوع‌بندی پاسخ‌های کیفی که از طریق مصاحبه نیمه‌ساختاریافته با ۳۰ نفر اعضای هیأت علمی (۱۹ زن و ۱۱ مرد) و ۳۵ نفر دانش‌آموختگان و دانشجویان مقطع دکتری (۱۳ زن و ۲۲ مرد) از شهرها و رشته‌های گوناگون در دانشگاه‌های سراسر ایران گرفته شد، پرسشنامه‌ای محقق‌ساخته تدوین شد و به‌جهت اعتباریابی پرسشنامه طراحی شده، ابتدا از ۲۰۰ نفر دانش‌آموخته یا دانشجوی مقطع دکتری (۱۱۷ مرد و ۸۳ زن) دانشگاه‌های سراسر ایران از رشته‌های گوناگون خواسته شد به پرسشنامه پاسخ دهند و پس از آن نیز ۴۴۲ نفر (۲۲۳ مرد و ۲۱۹ زن) شامل دانش‌آموخته دکتری (۵۷ نفر)، دانشجوی دکتری (۳۲۰ نفر) و عضو هیأت علمی (۶۵ نفر) به پرسشنامه اعتباریابی شده پاسخ دادند. بنابراین جامعه آماری در مجموع ۶۴۲ نفر بوده است که نظر خود را در باره آسیب‌شناسی آزمون‌های بسندگی زبان انگلیسی به‌لحاظ ۴ معیار روانی، پایایی، تأثیر و عدالت آزمون بیان کرده‌اند. یافته‌های پژوهش نشان می‌دهد که شرکت‌کنندگان رضایت چندانی از آزمون‌های مذکور ندارند و باور دارند که این آزمون‌ها براساس بریده‌هایی از آزمون‌های بین‌المللی تافل و آیلتس قدیمی تهیه می‌شوند و ملاحظات فرهنگی و بومی در آن‌ها رعایت چندانی نمی‌شوند. بنابراین، در پایان، برای ارتقای کیفی آزمون‌های بسندگی زبان انگلیسی در ایران پیشنهادهایی ارائه شده است.

اطلاعات مقاله

تاریخ ارسال: ۱۳۹۹/۱۱/۰۵
تاریخ پذیرش: ۱۳۹۹/۱۱/۲۶
تاریخ انتشار: زمستان ۱۳۹۹
نوع مقاله: علمی پژوهشی

کلید واژگان:

آسیب‌شناسی، آزمون‌های
بسندگی زبان انگلیسی،
مؤلفه‌های فرهنگی-بومی،
وزارت عتف

شناسه دیجیتال DOI: 10.22059/jflr.2021.317539.801 کلیه حقوق محفوظ است ۱۳۹۹

پیش‌قدم، رضا، ابراهیمی، شیمای، شایسته، شقایق، طباطبائی فارانی، سحر، جاجرمی، هانیه. (۱۳۹۹). بررسی و آسیب‌شناسی آزمون‌های بسندگی زبان انگلیسی وابسته به وزارت عتف و دانشگاه‌های ایران و نیازسنجی زبانی ذی‌نفعان. *پژوهش‌های زبان‌شناختی در زبان‌های خارجی*, ۱۰ (۱)، ۴-۶۸۶-۷۰۵.
Pishghadam, Reza, Ebrahimi, Shayesteh, Shaghayegh, Tabatabaee Farani, Sahar, Jajarmi, Haniyeh (2021). Evaluating and Examining the Problems of Iran's Ministry of Higher Education and Universities' English Language Proficiency Tests and Stakeholders' Language Needs Analysis. *Journal of Foreign Language Research*, 10 (4), 686-705. DOI: 10.22059/jflr.2021.317539.801

- * این مقاله برگرفته از طرح مورد حمایت وزارت علوم، تحقیقات و فناوری می‌باشد.
- * رضا پیش‌قدم، استاد آموزش زبان انگلیسی و روان‌شناسی تربیتی دانشگاه فردوسی مشهد است. حوزه پژوهشی مورد علاقه وی عصب‌روان‌شناسی، آموزش زبان، جامعه‌شناسی آموزش زبان و شناخت اجتماعی می‌باشد.
- ** شیمای ابراهیمی استادیار آموزش زبان فارسی دانشگاه فردوسی مشهد است. حوزه پژوهشی وی روان‌شناسی آموزش زبان، جامعه‌شناسی آموزش زبان و آموزش زبان فارسی به غیرفارسی‌زبان می‌باشد.
- *** شقایق شایسته استادیار آموزش زبان انگلیسی دانشگاه فردوسی مشهد می‌باشد و حوزه پژوهشی ایشان روان‌شناسی، عصب‌شناسی و آموزش زبان است.
- **** سحر طباطبائی فارانی دکتری آموزش زبان انگلیسی دانشگاه فردوسی مشهد است. وی نگارنده مقالاتی در حوزه آموزش زبان، عصب‌شناسی و روان‌شناسی زبان در نشریات معتبر علمی ملی و بین‌المللی می‌باشد.
- ***** هانیه جاجرمی استادیار آموزش زبان انگلیسی مؤسسه آموزش عالی بهار است. ایشان مقالات خود را در حوزه عصب‌شناسی، روان‌شناسی و جامعه‌شناسی زبان در نشریات معتبر علمی بین‌المللی و ملی به‌چاپ رسانده است.



Evaluating and Examining the Problems of Iran’s Ministry of Higher Education and Universities’ English Language Proficiency Tests and Stakeholders’ Language Needs Analysis



Reza Pishghdam*

(corresponding author)

Professor in TEFL & Educational Psychology, Ferdowsi University of Mashhad, Mashhad, Iran
pishghadam@um.ac.ir



Shima Ebrahimi**

Assistant Professor of Persian Language Teaching, Ferdowsi University of Mashhad, Mashhad, Iran
Email: shimaebrahimi@um.ac.ir



Shaghayegh Shayesteh***

Assistant Professor of ELT, Ferdowsi University of Mashhad, Mashhad, Iran
Email: shayesteh@um.ac.ir



Sahar Tabatabaee Farani****

Ph.D. in TEFL, Ferdowsi University of Mashhad, Mashhad, Iran
Email: sahar.tabatabaee1983@gmail.com



Haniyeh Jajarmi*****

Assistant Professor of ELT, Bahar Institute of Higher Education, Mashhad, Iran
Email: hjajarmi@gmail.com

ABSTRACT

This study aims to explore English language proficiency tests of Iran’s Ministry of Higher Education and universities by analyzing the needs of the participants (PhD candidates and graduates) and then propose a standard and uniform test. To this end, after thematizing the qualitative answers obtained through semi-structured interviews with 30 faculty members (19 females and 11 males) and 35 PhD candidates and graduates (13 females and 22 males) from different cities and majors in universities throughout Iran, a researcher-made questionnaire was developed. To validate the designed questionnaire, first 200 PhD graduates and candidates (117 males and 83 females), studying various majors in universities across Iran, answered the questionnaire (convenience sampling). Then 442 participants (223 males and 219 females) including PhD graduates (N= 57), PhD candidates (N= 320), and faculty members (N= 65) answered the validated questionnaire. Totally, 642 participants expressed their views on the problems of English language proficiency tests in terms of four criteria: validity, reliability, impact, and fairness of the test. The findings show that the participants are not satisfied with these tests and believe that these tests are designed based on the old versions of TOEFL and IELTS international tests and cultural and local considerations are overlooked. Finally, some suggestions are made to improve the quality of English language proficiency tests in Iran.

DOI: 10.22059/jflr.2021.317539.801

© 2021 All rights reserved.

ARTICLE INFO

Article history:

Received:
24th, January, 2021

Accepted:
14th, February, 2021

Available online:
Winter 2021

Keywords:

examining the problems, English language proficiency tests, local-cultural components, Ministry of Higher Education

Pishghdam, Reza, Ebrahimi, Shayesteh, Shaghayegh, Tabatabaee Farani, Sahar, Jajarmi, Haniyeh (2021). Evaluating and Examining the Problems of Iran’s Ministry of Higher Education and Universities’ English Language Proficiency Tests and Stakeholders’ Language Needs Analysis. *Journal of Foreign Language Research*, 10 (4), 686-705. DOI: 10.22059/jflr.2021.317539.801

- * Reza Pishghdam is a professor in TEFL & educational psychology at Ferdowsi University of Mashhad. His research interests are neuro-psychological and socio-psychological aspects of language education, and social cognition.
- ** Shima Ebrahimi is an assistant professor of Persian language teaching at Ferdowsi University of Mashhad. Her areas of interests are educational psychology, sociology of language education, and teaching Persian to the speakers of other languages.
- *** Shaghayegh Shayesteh is an assistant professor of language education at Ferdowsi University of Mashhad. Her research interests are psychology and neuropsychology of language education.
- **** Sahar Tabatabaee Farani, a PhD holder in TEFL, has graduated from Ferdowsi University of Mashhad. Her research interest and focus are sociology, psychology, and neuropsychology of language education.
- ***** Haniyeh Jajarmi is an assistant professor of ELT. She has been engaged in interdisciplinary research within neuroscience, psychology of language, and language education. Her current research focuses on the neuro-psychology and socio-psychology of language education.

۱. مقدمه

آزمون، روشی برای اندازه‌گیری توانایی و دانش فرد در زمینه‌ای مشخص است و تأثیر بسزایی در رشد و ارتقای کیفیت آموزش دارد (هرمن، دری‌فاس و گلان، ۱۹۹۰) (Herman, Dreyfus, & Golan, 1990) و روایی (Validity) (درجه‌ای از صحت نتایج مطالعه) و پایایی (Reliability) (درجه‌ای از یکسان‌بودن نتایج در طول زمان معین و تحت شرایط مشابه) آن امری اجتناب‌ناپذیر است (دون و همکاران، ۲۰۰۷) (DeVon et al, 2007). هنگامی که آزمون از روایی و پایایی لازم برخوردار باشد، می‌تواند با سنجش و ارزیابی صحیح از توانمندی‌های افراد، مسیر درست آموزش و یادگیری را نیز هموار کند و آن‌گاه با اطمینان می‌توان نتایج را استفاده کرد.

از سوی دیگر، نتایج آزمون‌ها همراه با تأثیرات زیادی (تأثیر آزمون بر آموزش) است که اثر بازخورد آزمون (Washback) نامیده می‌شود و در چندسال اخیر، درباره آثار و تبعات آزمون‌های زبان انگلیسی پژوهش‌های گسترده‌ای انجام شده است تا ارتباط یادگیری و این تأثیرات را نشان دهند (روی چان، ۲۰۲۰) (Roy Chan, 2020). آزمون بسندگی (Proficiency Test) یکی از انواع آزمون‌های زبان است که میزان پیشرفت زبان مخاطبان و توانایی زبانی آنان را در همه مهارت‌های زبانی فارغ از زمان، مکان و میزان آموزش می‌سنجد و به وی کارنامه‌ای حاوی سطح زبانی در هر مهارت و سطح کلی زبانی ارائه می‌دهد (عباسی، ۱۳۹۳).

یکی از آزمون‌های ملی مهمی که در کشور ایران برگزار می‌شود، آزمون بسندگی زبان انگلیسی ویژه متقاضیان دکتری است که توسط وزارت عتف یا در سطح دانشگاه‌های داخل کشور ایران انجام می‌شود. با توجه به جایگاه بین‌المللی زبان انگلیسی، هدف از آزمون بسندگی زبان انگلیسی در دانشگاه‌ها، سنجش بسندگی زبان دانشجویان تحصیلات تکمیلی برای ادامه تحصیل در مقطع دکتری و ارائه مجوز ورود یا ادامه تحصیل به این دانشجویان می‌باشد. فرض بر

این است که دانشجویان مقطع تحصیلی دکتری باید در خواندن و نوشتن متون تخصصی توانایی کافی داشته باشند و بتوانند مقالات، رساله‌ها و نوشته‌هایی که به زبان انگلیسی نوشته شده‌اند را بخوانند یا بتوانند به زبان انگلیسی متن تخصصی یا مقاله بنویسند. در شرایطی ایده‌آل‌تر حتی آنان باید بتوانند نتایج پژوهش‌های خود را به انگلیسی در همایش‌های بین‌المللی به صورت نوشتاری یا گفتاری ارائه دهند. توانش‌های زبانی که اغلب در چنین آزمون‌هایی بررسی می‌شوند، در چارچوب مهارت‌های اصلی شنیداری، گفتاری، خوانداری و نوشتاری تعریف شده‌اند که میزان تسلط افراد به این چهار مهارت نشان‌دهنده تسلط مهارت‌های زبانی آنان می‌باشد. البته باید توجه داشت در این آزمون‌ها توانایی زبانی کلی (Overall language ability) دانشجویان شامل میزان دانش، میزان توانایی در تشخیص اجزای زبانی و میزان توانایی در استفاده صحیح صورت زبانی مدنظر قرار می‌گیرد (فرهادی، جعفرپور و بیرجندی، ۲۰۰۷) (Farhady, Jafarpur, & Birjandi, 2007) و به نظر می‌رسد آزمون‌های بسندگی گزینه مناسبی برای اطمینان از میزان تسلط مهارت زبانی دانشجویان به‌شمار می‌روند. این آزمون‌ها تلاش دارند با تأکید بر مهارت‌های چهارگانه زبان‌آموزی، نوع سؤالات را با آزمون‌های زبان انگلیسی بین‌المللی مانند آزمون‌های تافل (TOEFL) و آیلتس (IELTS) (نسخه کاغذی) شبیه‌سازی کنند و تعداد زیادی از آن‌ها تنها بخش محدودی از زبان (اغلب واژگان، دستور و درک مطلب (شنیداری و خوانداری) را مدنظر قرار می‌دهند و بخش‌های مهم‌تر مانند مهارت‌های گفتاری و نوشتاری اغلب مورد غفلت واقع می‌شوند. در حقیقت می‌توان گفت این آزمون‌ها نگاه سیستمی به زبان ندارند و با مثله‌کردن آن یادگیری مکانیکی را تقویت می‌کنند. بنابراین، به نظر می‌رسد طراحی آزمون معیار و استاندارد که تمامی مهارت‌ها را با توجه به رشته‌گرایش تخصصی فرد بسنجد، الزامات است و آسیب‌شناسی آزمون‌های بسندگی موجود راه را برای تحقق این مهم هموار می‌سازد. بنابراین، پژوهش حاضر با هدف آسیب‌شناسی و روایی‌سنجی آزمون‌های

بسندگی زبان انگلیسی به بررسی این آزمون‌ها می‌پردازد و با بررسی دیدگاه شرکت‌کنندگان در مورد روایی، پایایی، تأثیر و عدالت آزمون‌های بسندگی زبان انگلیسی، پیشنهادهای برای طراحی سؤالات استاندارد و یکپارچه‌سازی آزمون‌های بسندگی زبان انگلیسی ارائه می‌دهد.

۲. پیشینه پژوهش

انجمن اروپا (۲۰۰۱) بسندگی زبان را در قالب چارچوب ارجاعی مشترک اروپا (The Common European Framework of Reference (CEFR)) تعریف می‌کند. بر اساس این تعریف، توانش به معنای تمام توانایی‌ها و مهارت‌هایی است که فرد برای برقراری ارتباط به آن‌ها نیاز دارد و به دو بخش عمومی و ارتباطی تقسیم می‌شود. بخش عمومی به معنای دانش و مهارت مستقل از زبان برای امر یادگیری است و بخش ارتباطی به معنای استفاده از زبان است و از منظر توانش زبانی، اجتماعی - زبانی و عمل‌گرا ارزیابی می‌شود. چارچوب ارجاعی مشترک اروپا شش سطح بسندگی (A1, A2, B1, B2, C1, C2) را توصیف می‌کند که به ترتیب از A1 تا C2 جایگزین سطوح مبتدی، متوسط و پیشرفته شده‌اند (هارش، ۲۰۱۷) (Harsch, 2017).

طی سال‌های گذشته، رویکردهای متفاوتی نسبت به بسندگی زبان وجود داشته است. به‌عنوان نمونه، در دهه شصت میلادی، لادو (Lado, 1961) و کرول (Carroll, 1961) در قالب مدل مهارت - مؤلفه (Skill-Component Model) بسندگی زبانی را شامل مهارت‌های زبانی (شنیداری، گفتاری، خوانداری و نوشتاری) و مؤلفه‌های دانش زبانی (دستور زبان، واژگان، آواشناسی/خط‌شناسی) می‌دانستند. این رویکرد که بر مبنای ساختارگرایی (Structuralism) استوار بود منجر به طراحی آزمون‌های مجزا شد که تا مدت‌ها در زمینه سنجش بسندگی زبان به کار می‌رفتند (مطلب‌زاده و بقائی‌مقدم، ۲۰۱۱) (Motallebzadeh, & Baghaee, 2011). (Moghaddam, 2011) اما ایراداتی نیز به این مدل وارد بود؛ به‌عنوان مثال، مشخص نبود که مهارت‌های زبان

و مؤلفه‌های دانش چگونه با یکدیگر مرتبط‌اند. همچنین، مشخص نبود که مهارت‌ها فقط نمایانگر مؤلفه‌های دانش زبانی بودند یا اینکه از لحاظ کیفی متفاوت از مؤلفه‌ها بودند. محدودیت دیگر مدل مهارت - مؤلفه این بود که بافت گفتاری و موقعیتی استفاده از زبان را مشخص نمی‌کرد (Backman, 1990).

روور و مک‌نامارا (۲۰۰۶) (Roever & McNamara, 2006) در نقد مدل بکمن (۱۹۹۰) توضیح توانش زبانی از ابعاد کاربردی، فرهنگی و تعصب (Bias) آزمون سخن گفتند و بعد اجتماعی آزمون زبان را مورد تأکید قرار دادند. ابعاد عملگرا در این رویکرد به معنی ارزیابی استفاده از زبان در موقعیت‌های مشابه واقعیت است و در نظر گرفتن ابعاد فرهنگی و تعصب آزمون در این مدل، به این معناست که گروهی از شرکت‌کننده‌ها به صورت ناعادلانه از محتوای آزمون سود ببرند و به نفعشان باشد، در صورتی که سایرین از این ویژگی بهره‌مند نباشند. همچنین، آنان در قالب تعصب آزمون و در نظر گرفتن نکات فرهنگی در آزمون، از تأثیرات آن بر زبان‌آموزان صحبت کردند و بر نیاز بیشتر به پژوهش‌هایی در زمینه تأثیرات اجتماعی آزمون‌های زبان تأکید کردند.

بکمن و پالمر (۲۰۱۰) (Bachman & Palmer, 2010) نیز، دامنه گسترده‌تری برای بررسی سامانه‌های ارزیابی توانایی زبانی در نظر گرفتند. نظریه استدلال سودمندی ارزیابی (Assessment Use Argument) بکمن و پالمر نه تنها برای ارزیابی آزمون‌های زبان، بلکه برای بررسی هر نوع سنجشی به کار می‌رود. افزون بر این، آنان بر قابلیت استدلال سودمندی ارزیابی تأکید داشتند و معتقد بودند که این مدل از چندین ادعا ساخته شده است که ارتباط بین عملکرد شرکت‌کننده، نمره حاصل از آزمون، تفسیر توانایی مورد سنجش، تصمیماتی که بر اساس نمره و تفسیر گرفته می‌شوند و پیامدهای ارزیابی و تصمیمات اتخاذ شده را برقرار می‌کند (کین، ۲۰۱۱) (Kane, 2011).

به‌منظور تأکید بر سودمندی ارزیابی و تفاسیر حاصل از ارزیابی، این مدل چهار ادعا را مطرح می‌کند: (۱)

پیامدهای آزمون و تصمیمات مربوط به آن‌ها سودمندند؛ (۲) تصمیمات با در نظر گرفتن ارزش‌ها و تساوی افراد جامعه هدف و بر اساس تفاسیر آزمون اتخاذ می‌شوند؛ (۳) تفاسیر مربوط به توانایی زبانی، مفاهیمی همچون معناداری، بی‌طرفی، تعمیم‌پذیری، ارتباط و کفایت را شامل می‌شوند؛ (۴) نمرات ارزیابی از ثبات (پایایی) برخوردارند (کلارک-گارکا، ۲۰۱۰) (Clark-Gareca, 2010). در ادعای سوم، منظور از معناداری این است که سازه مورد سنجش برای همه شرکت‌کننده‌ها یک‌معنا را داشته باشد؛ بی‌طرفی به آن معناست که ساختار و محتوای آزمون به سمت گروه خاصی سوگیری نداشته باشد؛ تعمیم‌پذیری بدین معناست

که نتایج حاصل از آزمون به محیط بیرون از آزمون هم قابل تعمیم باشد؛ منظور از ارتباط این است که آیا محتوای آزمون همان سازه مدنظر را می‌سنجد؛ و کفایت به این معناست که آیا آزمون اطلاعات کافی برای تصمیم‌گیری فراهم می‌کند یا خیر (بکمن و پالمر، ۲۰۱۰) (Bachman & Palmer, 2010). با توجه به تعاریف گوناگون بسندگی زبان و اهدافی که هر یک از این تعاریف دنبال می‌کنند، آزمون‌هایی برای سنجش سطح بسندگی زبان‌های گوناگون طراحی شده‌اند که در ادامه مشخصات آزمون، آزمون‌های بسندگی زبان انگلیسی در ایران به شرح جدول (۱) ذکر می‌شود:

جدول ۱: مشخصات آزمون‌های بسندگی زبان انگلیسی در ایران

ردیف	نام برگزارکننده	نام آزمون	نوع آزمون	تعداد سؤالات	نوع مهارت مورد سنجش					
					واژگان	دستور زبان	شنیداری	خوبداری	نوشتاری	گفتاری
۱	وزارت علوم، تحقیقات و فناوری	MSRT	چهارگزینه‌ای	۱۰۰	✓	✓	✓	✓	---	---
۲	سازمان سنجش آموزش کشور	TOLIMO	چهارگزینه‌ای و تشریحی	۱۴۱	---	✓	✓	✓	✓	---
۳	وزارت بهداشت	MHLE	چهارگزینه‌ای	۱۰۰	✓	✓	✓	✓	---	---
۴	دانشگاه آزاد اسلامی	EPT	چهارگزینه‌ای	۱۰۰	✓	✓	---	---	---	---
۵	دانشگاه تهران	UTEPT	چهارگزینه‌ای	۱۰۰	✓	✓	---	---	---	---
۶	دانشگاه تربیت مدرس	TMUE	چهارگزینه‌ای	۱۰۰	✓	✓	---	---	---	---
۷	دانشگاه فردوسی مشهد	TELP	چهارگزینه‌ای	۱۰۰	---	✓	✓	---	---	---
۸	دانشگاه اصفهان	UIEPT	چهارگزینه‌ای	۸۰	✓	✓	---	---	---	---
۹	دانشگاه شهید چمران اهواز	SCU	چهارگزینه‌ای	۱۰۰	✓	✓	---	---	---	---
۱۰	دانشگاه پیام نور	ETPNU	چهارگزینه‌ای	۱۰۰	---	✓	✓	---	---	---
۱۱	دانشگاه خوارزمی	KELT	چهارگزینه‌ای	۱۰۰	✓	✓	---	---	---	---
۱۲	دانشگاه ارومیه	UUEPT	چهارگزینه‌ای	۱۰۰	---	✓	✓	---	---	---
۱۳	دانشگاه صنعتی شیراز	SELT	چهارگزینه‌ای	۱۰۰	✓	✓	✓	---	---	---
۱۴	دانشگاه رازی	RULPT	چهارگزینه‌ای	۱۰۰	✓	✓	---	---	---	---
۱۵	دانشگاه شهرکرد	SKELT	چهارگزینه‌ای	۱۰۰	✓	✓	✓	✓	---	---

همانگونه که مشاهده می‌شود، به جز آزمون تولیمو که افزون بر پرسش‌های چهارگزینه‌ای، دارای یک پرسش تشریحی است، دیگر آزمون‌های بسندگی تنها به صورت

چهارگزینه‌ای هستند و شمار این پرسش‌ها ۱۰۰ سؤال است. بنابر داده‌های یادشده درباره آزمون‌های بسندگی زبان انگلیسی در ایران و با توجه به جدول (۱)، این آزمون‌ها با

فرض این که مهارت نوشتن پیچیده‌ترین مهارت زبانی است و افراد را دچار چالش می‌کند (نونان، ۱۹۸۹). (Nunan, 1989) به این مهارت توجه زیادی نمی‌کنند. اما باید به این مهم توجه داشت که فهم و تفکر از طریق نوشتن رشد می‌کند (بازرمن، لیتل، و بیثل، ۲۰۰۵). (Bazerman, Little, & Bethel, 2005) و با سنجش این مهارت می‌توان میزان فهم و ادراک مخاطبان را بیشتر ارزیابی کرد. پیش‌قدم و ابراهیمی (۱۳۹۷) معتقدند مهارت نوشتن از چنان اهمیتی برخوردار است که می‌تواند نقش قابل توجهی در پیشرفت توانایی‌های زبانی، درک و فهم، برقراری ارتباط با دیگران و حتی بیان احساسات نسبت به زبان داشته باشد و مشاهده می‌شود که برخلاف اهمیت‌های اشاره شده، به این مهارت توجه زیادی نشده است و دانشجویان همچنان در مهارت نوشتاری زبان انگلیسی دارای ضعف بسیارند.

همانگونه که جدول (۱) نشان می‌دهد، تنها آزمون تولیمو که توسط سازمان سنجش کشور برگزار می‌شود، مهارت نوشتاری را می‌سنجد و سایر مراکز در سنجش خود هیچ توجهی به این مهارت ندارند. در آزمون تولیمو مهارت نوشتاری در دفترچه‌ای جداگانه و به صورت ۱ سؤال تشریحی مورد بررسی قرار می‌گیرد و داوطلب ۳۰ دقیقه زمان دارد، پاسخ را در همان دفترچه بنویسد. بایسته یادآوری است که در این آزمون نیز، بخش نوشتاری نمره‌ای جداگانه و تکمیلی دارد که در کنار نمره کل و در کارنامه درج می‌شود. آزمون تولیمو درباره وارد کردن مهارت نوشتاری و شیوه اعلام نمره آن مشابه آزمون‌های تافل و آیلتس عمل می‌کند. از سوی دیگر، مهارت گفتاری، مهارتی است که در همه مراکز مغفول مانده است. در این میان، همه مراکز (۱۵ مرکز) درصد زیادی از تمرکز خود را بر روی مهارت خوانداری گذاشته‌اند و در همه آن‌ها به دستور زبان (۱۵ مرکز) توجه شده است. واژگان (۱۰ مرکز) و مهارت شنیداری (۸ مرکز) در اولویت بعدی این مراکزند. بایسته یادآوری است ادعای مراکز اشاره شده بر این است که سطح سؤالات آنان در حد آزمون تافل یا آیلتس است، اما باید خاطر نشان ساخت که در آزمون‌های بین‌المللی تافل

و آیلتس به هر چهار مهارت به صورت جداگانه پرداخته می‌شود و به هر مهارت نیز نمره‌ای جداگانه تعلق می‌گیرد که توانایی زبانی فرد را می‌سنجد (تری و ویلسون، ۲۰۰۴) (Terry & Wilson, 2004).

به نظر می‌رسد با توجه به شرایط موجود آزمون‌های بسندگی زبان انگلیسی داخلی، در واقع سطح زبان انگلیسی افراد هیچ‌گونه پیشرفتی نمی‌کند و چنانچه آنان برای آزمون آماده شوند، دانش آنان در سطح حفظیات دستوری و واژگانی باقی می‌ماند و هدف تنها کسب نمره از آزمون است. این امر در حالی است که مهارت‌های اصلی که در دوره دکتری اهمیت زیادی دارد، مهارت‌های گفتاری و نوشتاری می‌باشند. بنابراین، از آنجایی که هیچ یک از آزمون‌های بسندگی زبان انگلیسی در ایران به همه چهار مهارت اصلی، واژگان و دستور توجه نکرده‌اند، می‌توان گفت از روایی لازم برخوردار نیستند و به نظر می‌رسد تأثیری منفی بر روی جامعه پیرامون زبانی و مخاطب می‌گذارند. از این رو، آسیب‌شناسی آزمون‌های موجود و پیشنهاد آزمون استاندارد و هنجاریابی شده که با نگاه سامانه‌ای یکپارچه و همسو با مؤلفه‌های فرهنگی-بومی تهیه شده باشد، از بایسته‌هایی است که در پژوهش حاضر بدان پرداخته می‌شود.

۳. روش‌شناسی پژوهش

جامعه آماری

این پژوهش از نوع ترکیبی (کیفی-کمی) است و نگارندگان با مصاحبه کیفی و تحلیل کمی داده‌ها به آسیب‌شناسی آزمون‌های بسندگی زبان انگلیسی وابسته به وزارت عتف و دانشگاه‌های ایران و نیازسنجی زبانی ذی‌نفعان براساس مؤلفه‌های فرهنگی-بومی پرداخته‌اند. در مرحله کیفی، به منظور انجام مصاحبه، جامعه آماری متشکل از ۳۰ نفر اعضای هیأت علمی (۱۹ زن و ۱۱ مرد) و ۳۵ نفر دانش‌آموختگان و دانشجویان مقطع دکتری (۱۳ زن و ۲۲ مرد) از شهرها و رشته‌های گوناگون در دانشگاه‌های سراسر ایران بودند که در یک یا چند آزمون بسندگی زبان

انگلیسی از سال‌های گذشته تاکنون شرکت کرده‌اند. بایسته یادآوری است که شمار شرکت‌کنندگان در بخش کیفی، از پیش تعیین نشده و از روش اشباع داده استفاده شد. یعنی مصاحبه با شرکت‌کنندگان تا بدان‌جا پیش‌رفت که نتایج به اشباع رسید و پاسخ جدید دیگری حاصل نشد. در قسمت کمی پژوهش به جهت اعتباریابی پرسشنامه طراحی شده ابتدا از ۲۰۰ نفر دانش‌آموخته یا دانشجوی مقطع دکتری (۱۱۷ مرد و ۸۳ زن) دانشگاه‌های سراسر ایران از رشته‌های گوناگون خواسته شد به پرسشنامه پاسخ دهند و پس از آن نیز ۴۴۲ نفر (۲۲۳ مرد و ۲۱۹ زن) شامل دانش‌آموخته دکتری (۵۷ نفر)، دانشجوی دکتری (۳۲۰ نفر) و عضو هیأت علمی (۶۵ نفر) به پرسشنامه اعتباریابی شده پاسخ دادند. بنابراین جامعه آماری بخش کمی، در مجموع ۶۴۲ نفر بوده است.

ابزار پژوهش

ابزار پژوهش برای گردآوری داده‌های کیفی، مصاحبه شفاهی نیمه‌ساختاریافته است که از طریق مصاحبه و گفت‌وگوی غیررسمی با شرکت‌کنندگان مرحله کیفی انجام گرفته است. به این معنا که مصاحبه‌کنندگان، چارچوبی مشخص برای خود در نظر داشته و در طول مصاحبه به آن پرداخته‌اند؛ اما در عین حال اجازه انعطاف در مصاحبه را نیز داشته‌اند. پرسش‌ها از پیش طراحی و به‌منظور کسب اطلاعات دقیق از مصاحبه‌شوندگان پرسیده شدند. این پرسش‌ها به شرح ذیل می‌باشند:

۱. طی سال‌های گذشته در کدام یک از آزمون‌های بسندگی زبان انگلیسی شرکت کرده‌اید؟
۲. به نظر شما آیا آزمون‌های بسندگی زبان انگلیسی مهارت‌های زبانی را به درستی می‌سنجند؟
۳. آیا از محتوای سؤالات آزمون رضایت دارید؟
۴. آیا شرایط برگزاری آزمون، مکان و زمان آن از نظر شما مناسب است؟
۵. به نظر شما این آزمون چه تأثیراتی در زندگی آکادمیک و حرفه‌ای شما دارد؟
۶. نقاط قوت و ضعف این آزمون‌ها از نظر شما

چيست؟

۷. پیشنهاد شما برای ارتقای سطح کیفی این آزمون‌ها چیست؟
در بخش کمی نیز براساس موضوع‌بندی به‌دست‌آمده از داده‌های مصاحبه کیفی، پرسشنامه‌ای محقق‌ساخته شامل ۲۵ سؤال طراحی شد و به‌منظور بررسی روایی این پرسشنامه از روش بررسی عاملی تأییدی (Confirmatory Factor Analysis) و آزمون تک عاملی هارمن (Harman's Single Factor Test) استفاده شد (پیوست ۱).

گردآوری داده‌ها

برای گردآوری داده‌های بخش کیفی، از شرکت‌کنندگان خواسته شد تا نظر خود را درباره آزمون‌های بسندگی زبان انگلیسی بیان کنند. زمان مصاحبه‌ها به‌طور متوسط برای هر فرد ۴۵ دقیقه در نظر گرفته شد و به‌صورت کلی در مدت زمان ۲ ماه (دی‌ماه ۱۳۹۸ تا اسفندماه ۱۳۹۸) به طول انجامید. بایسته یادآوری است پیش از شروع مصاحبه و پرسیدن پرسش‌ها، مصاحبه‌کنندگان مفهوم بسندگی را برای افرادی که از آن اطلاعی نداشتند، توضیح دادند. پس از گردآوری داده‌های موردنیاز از شرکت‌کنندگان در مصاحبه، پژوهشگران پاسخ‌های ضبط‌شده را بر روی کاغذ پیاده کردند و به بررسی کیفی داده‌های گردآوری شده و موضوع‌بندی آنان پرداختند. گوبا و لینکلن (۱۹۹۴) برای رسیدن به روایی و پایایی در بررسی کیفی چهار شاخص اعتبار (Credibility)، اعتماد (Dependability)، تأییدپذیری (Confirmability) و انتقال‌پذیری (Transferability) را به‌عنوان استحکام داده‌های کیفی پژوهش در نظر می‌گیرند (محسن‌پور، ۱۳۹۰). بنابراین، در این پژوهش برای کسب اعتبار زمان کافی برای به‌اشباع‌رسیدن داده‌ها و مشاهدات پیگیر و مصاحبه با افراد منتخب، در نظر گرفته شده است و بررسی کیفی صحت داده‌های پژوهش توسط دو استاد تمام آموزش زبان انگلیسی، یک استادیار آموزش زبان فارسی، دو استادیار آموزش زبان انگلیسی و یک دانش‌آموخته دکتری آموزش زبان انگلیسی صورت پذیرفت. پس از مرحله اشباع

داده‌ها، براساس نظرات اعلام‌شده دسته‌بندی‌هایی صورت گرفت و اعتبار داخلی تحلیل محتوا از طریق روایی صوری ارزیابی شد. بدین‌معنا که پاسخ‌های دریافتی از مصاحبه و کدهای استخراج‌شده به‌مشارکت‌کنندگان توسط نگارندگان بررسی شد و آنان دربارهٔ صحت و سقم آن‌ها اظهارنظر می‌کردند و در صورت داشتن هرگونه مغایرت، مراتب مورد توجه و بررسی قرار می‌گرفت. پس از این ارزیابی، پاسخ‌هایی که بیشترین فراوانی را داشتند، به‌عنوان دستهٔ اصلی در نظر گرفته شدند. این دسته‌ها عبارت بودند از: (۱)

روایی (۲) پایایی (۳) تأثیر و (۴) عدالت که هر یک از آنان دارای زیرمجموعه‌هایی است. در ادامه این دسته‌بندی به‌عنوان دسته‌بندی نهایی طرح در نظر گرفته شد و براساس آن ۲۴ سؤال (هر قسمت شامل ۶ سؤال کمی) و یک سؤال کیفی در قالب انتقادات و پیشنهادها طراحی شد که برای گردآوری داده‌های کمی پژوهش از آن بهره گرفته شده است (پیوست ۱). مجموع موضوع‌بندی‌های بدست‌آمده و زیرمجموعه‌های آن به‌شرح جدول (۲) است.

جدول ۲: موضوع‌بندی‌های بدست‌آمده از پاسخ‌های کیفی افراد مصاحبه‌شونده

ردیف	زیرمجموعه
۱	روایی آزمون
	سنجش توانایی زبانی شرکت‌کننده
	مطابقت با معیارهای فرهنگی و بومی
۲	پایایی آزمون
	سنجش تمامی مهارت‌های زبانی
	سنجش کلیت زبانی
۳	تأثیر آزمون
	شرایط برگزاری آزمون
	محتوای آزمون
	یادگیری زبان
	روان و جسم
۴	عدالت آزمون
	هزینهٔ مالی
	شغل حرفه‌ای
	توجه به نیازهای واقعی شرکت‌کننده
	رسیدگی به اعتراضات
	سوگیری
	نمرهٔ معیار
کلاس‌های آمادگی آزمون	
منصفانه بودن	
در دسترس بودن نمونه سؤالات	

پس از موضوع‌بندی پاسخ‌های کیفی، پرسشنامه‌ای محقق‌ساخته تدوین شد و در روند تأیید روایی محتوایی و صوری آن، ابتدا توسط کارشناسان مورد بررسی قرار گرفت. پس از آن پرسشنامه در اختیار ۲۰۰ نفر دانش‌آموخته و دانشجوی دکتری و اعضای هیأت علمی از رشته‌های گوناگون سراسر کشور قرار گرفت و پرسشنامهٔ مذکور با استفاده از روش بررسی عملی تأییدی و آزمون تک عملی هارمن مورد تأیید قرار گرفت. سپس به‌منظور

بررسی و تحلیل کمی پرسشنامه، پرسشنامهٔ مذکور از طریق گوگل فرم در اختیار ۴۴۲ نفر (۲۲۳ مرد و ۲۱۹ زن) شامل دانش‌آموختهٔ دکتری (۵۷ نفر)، دانشجوی دکتری (۳۲۰ نفر) و عضو هیأت علمی (۶۵ نفر) قرار گرفت. پس از جمع‌آوری داده‌ها در این مرحله، تحلیل داده‌ها با استفاده از نرم افزار اس. پی. اس. اس. (نسخهٔ ۲۱) صورت گرفت. روش آماری مورد استفاده در این پژوهش تحلیل واریانس با اندازه‌گیری‌های مکرر بود.

۴. نتایج و بررسی

الف: بررسی پاسخ‌های کیفی

پس از بررسی پاسخ‌های کیفی شرکت‌کنندگان در پاسخ به این سؤال، موضوع‌بندی به دست آمده مربوط به سؤال مذکور در ۴ دسته روایی، پایایی، تأثیر و عدالت آزمون قابل دسته‌بندی است و شرکت‌کنندگان راجع به هریک از این موارد پاسخ‌هایی داده‌اند که دیدگاه آنان را به صورت کلی نسبت به آزمون‌های بسندگی مشخص

روایی آزمون

روایی بدین معناست که روش یا ابزار به کار رفته تا چه حد می‌تواند خصوصیت موردنظر را درست اندازه‌گیری کند. با توجه به پاسخ‌های دریافتی مصاحبه‌شوندگان بر این اعتقادند که آزمون‌های بسندگی زبان انگلیسی به دلایلی که در ادامه عنوان می‌شوند، مهارت‌های زبانی را به درستی نمی‌سنجند و از روایی کافی برخوردار نیستند.

جدول ۳: زیرمجموعه‌های مربوط به روایی آزمون

ردیف	زیرمجموعه
۱	آزمون مطابق با سنجش توانایی زبانی شرکت‌کننده نیست (۹۳٪)
۲	عدم مطابقت با معیارهای فرهنگی و بومی (۹۴٪)
۳	عدم سنجش تمامی مهارت‌های زبانی (۸۹٪)
	عدم توجه به مهارت گفتاری
۴	عدم سنجش کلیت زبانی (۱۰۰٪)
	عدم توجه به مهارت نوشتاری

پایایی آزمون

به پاسخ‌های دریافتی زیرمجموعه‌های مربوط به پایایی آزمون و نظرات شرکت‌کنندگان در مورد آن با ذکر درصد به شرح جدول (۴) می‌باشند:

منظور از پایایی این است که اگر آزمون را در چند زمان متفاوت و در یک جمعیت از آن استفاده کنیم، در نتیجه به دست آمده اختلاف چندانی مشاهده نشود. با توجه

جدول ۴: زیرمجموعه‌های مربوط به پایایی آزمون

ردیف	زیرمجموعه	
۱	شرایط برگزاری آزمون مناسب نیست (۳۷٪)	
	مدت زمان برگزاری مناسب نیست (۷۹٪)	
۲	مکان برگزاری مناسب نیست (۵۶٪)	
	تعداد سؤالات آزمون مناسب نیست (۲۱٪)	
	آزمون توجه به رشته گرایش خاص دارد (۸۴٪)	
	سوگیری در طراحی سؤال دارد (۴۷٪)	
	درجه سختی	طراحی توسط افراد نامتخصص (۲۸٪)
		عدم وجود چارچوب مشخص (۶۱٪)
		دشواری بخش خواندن و درک مطلب (۹۴٪)
		دشواری بخش دستور (۴۳٪)
	گزینه‌های مربوط به آزمون	اشتباه در سؤالات (۵۷٪)
		اشتباه در گزینه‌ها (۶۴٪)
تصادفی انتخاب کردن گزینه‌ها (۷۵٪)		

تأثیر آزمون

زندگی تحصیلی و حرفه‌ای او دارند. باتوجه به همه این موارد نگارندگان تأثیر آزمون را به زیرمجموعه‌های ذیل تقسیم کرده‌اند:

افزون بر روایی و پایایی آزمون، تعدادی از پاسخ‌ها مربوط به تأثیری هستند که آزمون‌ها در یادگیری مخاطب و

جدول ۵: زیرمجموعه‌های مربوط به تأثیر آزمون

ردیف	زیرمجموعه
۱	آزمون بر یادگیری زبان تأثیری ندارد (۸۳٪).
۲	آزمون تأثیر منفی بر روان و جسم دارد (۹۱٪).
	ایجاد استرس و اضطراب تپش قلب، سردرد، معده‌درد و ...
۳	هزینه مالی آزمون مناسب نیست (۶۹٪).
	هزینه آزمون بالا است.
	هزینه شرکت در کلاس‌های آمادگی آزمون بالا است. برگزاری آزمون در شهرهای بزرگ هزینه رفت و آمد را به‌دنبال دارد.
۴	آزمون تأثیری بر شغل حرفه‌ای ندارد (۷۹٪).
۵	آزمون به توجه به نیازهای واقعی شرکت‌کننده ندارد (۱۰۰٪).
	به مهارت مقاله‌نویسی توجهی ندارد. به مهارت سخنرانی در همایش‌ها و ... توجهی نمی‌شود.

عدالت آزمون

به این پاسخ‌ها، مواردی که در دسته عدالت آزمون قرار می‌گیرند در جدول (۶) خلاصه شده‌اند.

افزون بر موارد اشاره‌شده، یکی از مواردی که دغدغه اکثر مصاحبه‌شوندگان بود عدالت در آزمون است. با توجه

جدول ۶: زیرمجموعه‌های مربوط به عدالت آزمون

ردیف	زیرمجموعه
۱	نسبت به سؤال و گزینه نادرست نحوه رسیدگی به اعتراضات منصفانه نیست (۲۱٪).
	نسبت ب نمره اعلام شده نسبت به سهمیه‌های اعلامی
	سوگیری آزمون به سمت رشته‌گرایش خاص (۶۳٪)
۳	نمره معیار آزمون مناسب نیست (۷۳٪).
	نمره آزمون نمره قابل قبول برای دانشگاه
۴	شرکت در کلاس‌های آمادگی آزمون تنها برای برخی امکان‌پذیر است و این امر منصفانه نیست (۹۲٪)
۵	منصفانه طراحی نشدن سؤالات (۷۲٪).
	در نظر گرفتن رشته تحصیلی درجه دشواری متعادل
۶	در دسترس نبودن نمونه سؤالات آزمون (۶۵٪).

طراحی نشده‌اند و تقلید صرف از آزمون‌های بین‌المللی تافل و آیلتس (نسخه کاغذی) اند، موافق بودند و نظر خود

اکثر مصاحبه‌شوندگان با این موضوع که آزمون‌های داخلی مطابق با استانداردهای فرهنگی و بومی ایرانیان

را اینگونه بیان کردند: «آزمونی که تماماً تلاش دارد کپی از آزمون‌های بین‌المللی باشد، طبیعی است که مطابق با نیازهای فرهنگی داخل ایران نیست»؛ «البته باید به طراحان سؤالات حق داد که سؤالات براساس نیازهای فرهنگی داخلی طراحی نشوند. جدا از اینکه نیازهای فرهنگی در سنجش دانش زبانی اهمیت چندانی ندارد، تلاش دارند سؤالات را به همان سطح آزمون‌های بین‌المللی نزدیک سازند تا استانداردتر به نظر برسد»؛ «دانشگاه‌ها از سال‌های قبل تلاش کردند برای خود آزمونی جداگانه طراحی کنند ولی وقتی همه آزمون‌ها را کنار هم قرار می‌دهیم باز هم می‌بینیم تقلیدی از همان آزمون‌های تافل و آیلتس هستند که کمی ساده‌سازی شدند»؛ «متأسفانه کپی کردن از آزمون‌های بین‌المللی در سطحی ناقص صورت گرفته است و از کیفیت این آزمون‌ها کاسته است»؛ «به‌نظر من سؤالات باید بومی‌سازی و ایرانی‌سازی بشوند. اینکه سؤالات مربوط به درک مطلب مربوط به موضوعی در زبان انگلیسی می‌شود درک مطلب دقیق را نمی‌سنجد و منطبق با فضای داخلی کشور نیست». در این میان تعدادی از افراد با اینکه تقلیدی

بودن آزمون‌ها را مزیت آن‌ها می‌دانستند، نظر خود را اینگونه بیان کردند: «من فکر می‌کنم این آزمون‌ها باید مطابق با استانداردهای جهانی طراحی شوند و اینکه بتوانند تا حد ممکن تقلیدی از آزمون‌های بین‌المللی معتبر باشند خیلی خوب است. در واقع با هزینه‌ای بسیار کمتر می‌توان در آزمونی داخلی شبیه به آزمون‌های بین‌المللی شرکت کرد و به هدف موردنظر که همان ارائه مدرک است هم دست یافت». در نگاه و باور افراد، این آزمون‌ها براساس بریده‌هایی از آزمون‌های بین‌المللی تافل و آیلتس قدیمی تهیه می‌شوند و با توجه به این که آزمون‌ها اغلب بار فرهنگی دارند، ملاحظات فرهنگی و بومی در آن‌ها رعایت نمی‌شوند.

ب: بررسی یافته‌های کمی

آمار توصیفی برای پرسشنامه طراحی شده به همراه سن شرکت‌کنندگان در جدول (۷) نمایش داده شده است. در این جدول کمترین مقدار میانگین مربوط به «تأثیر» (۲,۴۱) و بیشترین مقدار آن مربوط به «پایایی» (۲,۹۵) است.

جدول ۷: آمار توصیفی پرسشنامه طراحی شده به همراه سن شرکت‌کنندگان

سازه پرسشنامه	تعداد	کمترین مقدار	بیشترین مقدار	میانگین	انحراف از معیار
روایی	۲۰۰	۱,۰۰	۵,۰۰	۲,۵۰	۰,۹۲
پایایی	۲۰۰	۱,۰۰	۵,۰۰	۲,۹۵	۰,۸۵
تأثیر	۲۰۰	۱,۱۷	۵,۰۰	۲,۴۱	۰,۸۷
عدالت	۲۰۰	۱,۰۰	۴,۶۷	۲,۸۰	۰,۶۷
سن	۲۰۰	۲۳	۵۸	۳۴,۳۷	۶,۱۹

(Kurtosis) در بازه ۲- و ۲ قرار گرفتند که بیانگر نرمال بودن توزیع داده‌های پرسشنامه است.

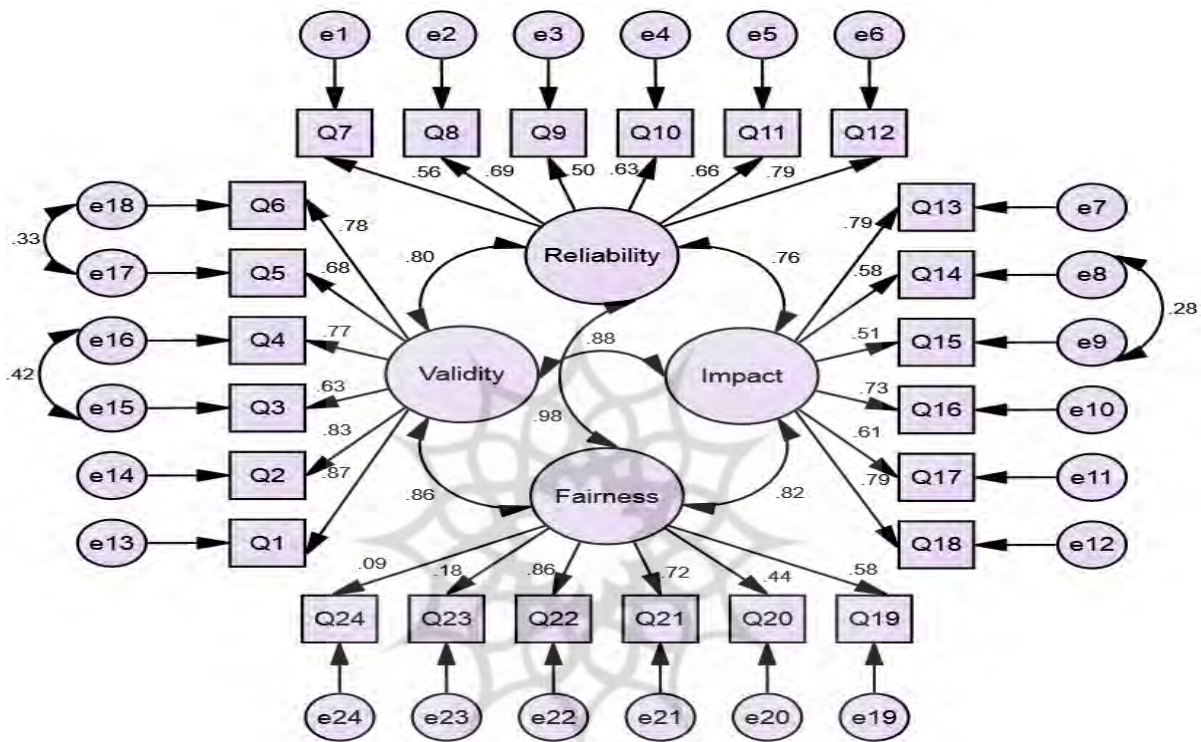
پیش از بررسی روایی پرسشنامه، نرمال بودن و یکسانی واریانس متغیرها مورد بررسی قرار گرفت. مطابق جدول ۸، مقادیر چولگی (Skewness) و کشیدگی

جدول ۸: نتایج آزمون نرمال بودن برای داده‌های پرسشنامه طراحی شده

پرسشنامه	چولگی	کشیدگی
	۰,۲۹	-۰,۰۴

از آنجا که تمام داده‌ها با استفاده از یک پرسشنامه گردآوری شده بود، برای جلوگیری از خطای واریانس روش مشترک (CMV)، آزمون تک عاملی هارمن اجرا شد. نتایج آزمون هارمن نشان داد که نخستین عامل به دست آمده میزان ۲۵ درصد کل واریانس را تبیین می‌کند. بنابراین می‌توان نتیجه گرفت که خطای روش مشترک نمی‌تواند

مشکلی در این پژوهش ایجاد کند. پس از آزمون هارمن، روایی پرسشنامه از راه بررسی عاملی تأییدی بررسی شد. شکل (۱) مدل بررسی عاملی تأییدی برای پرسشنامه طراحی شده را نشان می‌دهد. همان‌طور که مشاهده می‌شود، پرسشنامه از ۴ خرده‌مقیاس (روایی، پایایی، تأثیر و عدالت) تشکیل شده که هر یک توسط ۶ آیتم اندازه‌گیری می‌شوند.



شکل ۱: مدل تحلیل عاملی تأییدی برای پرسشنامه طراحی شده

جدول (۹) شاخص‌های نیکویی برازش (Goodness of fit indices) را ارائه می‌کند. شاخص‌های گزارش شده (نسبت مجذور کای به درجه آزادی χ^2/df)، برازندگی تطبیقی (CFI)، تاکر لوئیس (TLI)، ریشه خطای میانگین مجذورات تقریبی (RSMEA) و ریشه استاندارد باقیمانده مجذور میانگین (SRMR) برازش کلی مدل را تأیید می‌کنند. به منظور برازش کامل مدل با داده‌ها، بر اساس

نظر آلمن (۲۰۰۱) (Ulman, 2001) مطلوب آن است که نسبت مجذور کای به درجه آزادی زیر ۳ باشد. همچنین، به گفته براون و کیودک (۱۹۹۳) (Browne & Cudeck, 1993)، شاخص‌های برازندگی تطبیقی و تاکر لوئیس باید بالای ۰,۹۰ و مقدار ریشه خطای میانگین مجذورات تقریبی از ۰,۸ کوچکتر باشد.

جدول ۹: شاخص‌های نیکویی برازش مدل

نسبت مجذور کای به درجه آزادی	درجه آزادی	برازندگی تطبیقی	تاکر لوئیس	RMSEA ریشه خطای میانگین مجزورات تقریبی	ریشه استاندارد باقیمانده مجذور میانگین
۱,۹۵	۲۴۳	۰,۹۱	۰,۹۰	۰,۰۶	۰,۰۵

همان‌طور که مشاهده می‌شود، شاخص‌ها در سطح مطلوبی قرار دارند و مدل، بدون حذف هیچ آیتمی، از برازش خوبی برخوردار است. بنابراین می‌توان اذعان داشت که روایی پرسشنامه طراحی شده مورد تأیید قرار

گرفته است. پس از تکمیل اعتبارسنجی، پایایی کل پرسشنامه و ۴ سازه آن توسط آزمون آلفای کرونباخ بررسی شد (جدول ۱۰) بر اساس نتایج، تمام مقادیر بالای ۰,۷۰ هستند که از لحاظ آماری مطلوب است.

جدول ۱۰: پایایی کل و هریک از ۴ سازه پرسشنامه

تعداد گویه‌ها	آلفای کرونباخ	کل پرسشنامه
۲۴	۰,۹۳	سازه ۱ (روایی)
۶	۰,۹۰	سازه ۲ (پایایی)
۶	۰,۸۱	سازه ۳ (تأثیر)
۶	۰,۸۱	سازه ۴ (عدالت)
۶	۰,۷۰	

پیش از بررسی، نرمال بودن داده‌های گردآوری شده مورد بررسی قرار گرفت. مطابق جدول ۱۱، مقادیر چولگی

و کشیدگی «روایی»، «پایایی»، «تأثیر» و «عدالت» در بازه ۲- و ۲ قرار گرفتند که بیانگر نرمال بودن توزیع داده‌هاست.

جدول ۱۱: نتایج آزمون نرمال بودن برای ۴ سازه پرسشنامه

سازه‌های پرسشنامه	چولگی	کشیدگی
روایی	۰,۶۱	-۰,۱۳
پایایی	-۰,۱۱	-۰,۵۱
تأثیر	۰,۷۱	۰,۲۴
عدالت	-۰,۱۵	۰,۰۰

با هدف بررسی معناداری تفاوت‌های موجود بین میانگین‌های ۴ سازه پرسشنامه، آزمون بررسی واریانس با اندازه‌گیری‌های مکرر استفاده شد. همانگونه که در جدول ۱۲ مشاهده می‌شود، نتایج حاکی از آن است که تفاوت معناداری بین میانگین روایی (۲,۴۸)، پایایی (۲,۸۸)، تأثیر (۲,۴۰) و عدالت (۲,۷۵) وجود دارد ($F(3, 1323) = 97.21, p = 0.00, \eta^2 = 0.18, 1-\beta = 1.00$). در این

واکاوی، اندازه اثر بر اساس مقدار مربع اتای جزئی، زیاد است. همچنین مقدار توان مشاهده شده نشان می‌دهد که حجم نمونه برای نتیجه‌گیری نهایی مطلوب بوده و نتایج از توان بالایی برخوردارند. بر اساس یافته‌های آماری، مقدار توان مشاهده شده (که بر پایه اندازه تأثیر است) بین ۰ تا ۱ است که هر چه این مقدار به عدد ۱ نزدیک‌تر باشد، یافته‌ها قدرتمندترند.

جدول ۱۲: نتایج واریانس با اندازه‌گیری‌های مکرر برای مقایسه تفاوت میانگین‌های ۴ سازه پرسشنامه

منبع	مجموع مربعات	درجه آزادی	میانگین مربعات	تحلیل واریانس	سطح معناداری	مربع اتای جزئی	توان مشاهده شده
مفروضه کرویت	۶۸,۲۴	۳	۲۲,۷۴	۹۷,۲۱	۰,۰۰	۰,۱۸	۱,۰۰
سازه‌های پرسشنامه							
گرین هوس-گایزر	۶۸,۲۴	۲,۷۵	۲۴,۸۰	۹۷,۲۱	۰,۰۰	۰,۱۸	۱,۰۰
هیون-فلت	۶۸,۲۴	۲,۷۷	۲۴,۶۳	۹۷,۲۱	۰,۰۰	۰,۱۸	۱,۰۰
حد پایین	۶۸,۲۴	۱,۰۰	۶۸,۲۴	۹۷,۲۱	۰,۰۰	۰,۱۸	۱,۰۰

از آنجا که میانگین ۴ سازه روایی، پایایی، تأثیر و زوجی انجام شد تا محل تفاوت مشخص شود (جدول عدالت به صورت معناداری باهم متفاوت است، مقایسه ۱۳).

جدول ۱۳: مقایسه زوجی ۴ سازه پرسشنامه

آزمون I	آزمون J	اختلاف میانگین (I-J)	خطای معیار	سطح معناداری
۱	۲	-۰,۴۰*	۰,۳۵	۰,۰۰
	۳	۰,۰۸*	۰,۳۰	۰,۰۴
	۴	-۰,۲۶*	۰,۳۳	۰,۰۰
	۱	۰,۴۰*	۰,۳۵	۰,۰۰
۲	۳	۰,۴۸*	۰,۳۶	۰,۰۰
	۴	۰,۱۳*	۰,۲۹	۰,۰۰
۳	۱	-۰,۰۸*	۰,۳۰	۰,۰۴
	۲	-۰,۴۸*	۰,۳۶	۰,۰۰
	۴	-۰,۳۴*	۰,۳۱	۰,۰۰
۴	۱	۰,۲۶*	۰,۳۳	۰,۰۰
	۲	-۰,۱۳*	۰,۲۹	۰,۰۰
	۳	۰,۳۴*	۰,۳۱	۰,۰۰

* اختلاف میانگین در سطح ۰,۰۵ معنادار است.

در کل، نتایج حاصل از تحلیل واریانس با اندازه‌گیری‌های مکرر و مقایسه زوجی ۴ سازه روایی، پایایی، تأثیر و عدالت حاکی از آن است که:

تعدیل مقایسه چندگانه:
روش بن فرونی

$$(۲,۸۸) \text{ پایایی} < (۲,۷۵) \text{ عدالت} < (۲,۴۸) \text{ روایی} < (۲,۴۰) \text{ تأثیر}$$

۵. بحث و نتیجه‌گیری

آزمون، ابزاری است که برای سنجش پیشرفت تحصیلی مخاطبان در موقعیت‌های آموزشی به‌کار گرفته می‌شود. در فرایند طرح پرسش‌های آزمون، باید اصولی را در نظر گرفت تا اهداف آموزشی محقق شود. یکی از آزمون‌هایی که نقش پراهمیتی در زمینه ارزیابی سطح زبانی دارد، آزمون‌های بسندگی زبان انگلیسی است که در حالت طبیعی باید توانش زبانی، ارتباطی و توانایی کاربردی کردن اطلاعات راجع به زبان انگلیسی مخاطبان را بسنجد. از آنجا که نتایج این آزمون‌ها در عرصه آموزشی مانند پیش‌شرط آزمون جامع دکتری برای رشته‌های گوناگون است، در تعیین سرنوشت افراد نقش بسزایی ایفا می‌کند. بیشتر آزمون‌های سنجش زبان نقطه هدف خود را پایان دوره عمومی یا پایه و بر اساس اصطلاح رایج پیشرفته قرار می‌دهند. ویژگی این سطح عبارت است از اینکه شرکت‌کننده دامنه کافی از واژگان، اصطلاحات، تعابیر، دستور زبان، مهارت‌های چهارگانه شنیداری و فهم مربوط به آن، سخن‌گفتن، خواندن و فهم خوانشی و نگارش را در نزدیک‌ترین سطح به‌گوشوران اصلی آن زبان دارا می‌باشد. با توجه به این عوامل، نمی‌توان از اهمیت آزمون‌های بسندگی زبان انگلیسی غافل ماند.

در بررسی دیدگاه شرکت‌کنندگان و یافته‌های اجمالی پژوهش به‌طور کلی مشاهده شد که در کشور ایران نگاه جدی و تخصصی نسبت به آزمون‌های بسندگی زبان انگلیسی وجود ندارد و این آزمون‌ها از روایی کافی برخوردار نیستند. با توجه به اینکه روایی آزمون صحت نتایج را می‌سنجد و نشان می‌دهد آزمون تا چه اندازه چیزی را که قصد اندازه‌گیری آن را داشته است به‌درستی سنجیده است (محمدبیگی، محمدصالحی و گل، ۱۳۹۳). چنانچه آزمونی از روایی کافی برخوردار نباشد، نمی‌توان انتظار داشت که سنجش سطح زبانی مخاطب به‌درستی تحقق پذیرد. بنابراین، میزان تأثیر این آزمون‌ها نیز مورد تردید است.

برپایه نظریه انگلیسی‌های جهانی (World Englishes) مطرح‌شده توسط کچرو (۱۹۸۶) (Kachru, 1986)، پذیرش زبان‌های محلی و احیای دانش بومی در آموزش از اهمیت زیادی برخوردار است. براساس این مفهوم، انواع مختلفی از انگلیسی می‌تواند از راه بومی‌سازی (تطبیق

فرهنگ انگلیسی‌زبان با فرهنگ بومی) و پیوندزنی (ترکیب فرهنگ انگلیسی‌زبان با فرهنگ بومی) و فرهنگ‌پذیری (پذیرش فرهنگ انگلیسی‌زبان) وجود داشته باشد (توپاس، ۲۰۰۴) (Tupas, 2004) که به‌نظر می‌رسد هر سه مورد می‌تواند در طراحی آزمون‌های بسندگی زبان انگلیسی مدنظر قرار بگیرد. از این‌رو، در راستای رعایت الزامات و موازین استانداردهای مدیریت یکپارچه کیفیت آزمون‌های بسندگی زبان انگلیسی، لازم است آزمون‌های معیار و استاندارد مطابق با نیازهای فرهنگی و بومی ذی‌نفعان از سوی وزارت عتف طراحی شود و دستورالعمل اجرای آن نیز به دانشگاه‌ها ابلاغ شود.

براساس یافته‌های پژوهش به‌نظر می‌رسد دانشگاه‌های متولی این آزمون‌ها، صرفاً به چشم رفع تکلیف به آن‌ها می‌نگرند که باید حتماً برگزار شوند و دانشجو موظف است مدرک و نمره مربوط به آن را به دانشگاه خود ارائه دهد. ولی باید به این مهم توجه داشت که درخواست مدرک از دانشجو بدون فراهم‌کردن زیرساخت‌های کافی، پاک‌کردن صورت مسئله است. پرواضح است شرکت‌کننده در نهایت در زمان کم و فشرده صرفاً به‌دلیل ارائه مدرک در مدت زمان اندک برای آزمون مطالعه کرده و بعد از آزمون عمده مطالب را فراموش خواهد کرد؛ زیرا یادگیری مطالب صرفاً به‌صورت طوطی‌وار و سطحی است و مطالب تنها برای مدت محدودی در حافظه شرکت‌کننده ثبت خواهد شد. از این‌رو، پس از مدت زمان اندکی مطالب نیز به‌راحتی به‌فراموشی سپرده خواهند شد. همانگونه که اشاره شد، آزمون استاندارد از روایی و پایایی لازم برخوردار است و باید به‌لحاظ سنجش کلیت زبانی و سطح زبانی مخاطبان استاندارد طراحی شود. از سوی دیگر محتوای سؤالات و گزینه‌ها و عدم وجود اشتباه در آنان از ضروریات است؛ اما یافته‌های پژوهش نشان داد در شماری از پرسش‌ها ممکن است پرسش‌ها و گزینه‌ها در حد سنجش اطلاعات جزئی و کم‌اهمیت طراحی شوند به‌گونه‌ای که از شدت ساده‌بودن مخاطب را به‌اشتباه بیندازند و در پاره‌ای موارد برعکس، ممکن است آنچنان دشوار باشند که مخاطب را دچار سردرگمی کنند. در چنین مواردی ممکن است داوطلب به‌صورت تصادفی تصمیم بگیرد که یکی از گزینه‌ها را انتخاب کند.

به‌صورت کلی، از آن‌جایی که ارزیابی نقش مهمی در نظام

آموزشی دارد، افزون بر روش‌های ثابت و ایستا که معتقدند نمره شرکت‌کننده نمایانگر ظرفیت کامل ذهنی، اطلاعات و مهارت‌های وی است (غفوریان و عاشوری، ۱۳۹۶)، سنجش و ارزیابی پویا نیز از اهمیت بسیاری برخوردار است. این نوع ارزیابی، دو فرایند آموزش و ارزیابی را به‌عنوان فعالیتی ترکیبی در راستای درک توانایی‌های فراگیران و حمایت فعالانه در جهت ارتقای این توانایی مدنظر قرار می‌دهد (پوهنر، ۲۰۰۸) (Poehner, 2008). بنابراین، مهارت‌های در حال پیشرفت متقاضیان، بهتر نشان داده می‌شود. این امر منجر به ارائه تصویری کامل از توانایی‌ها و تفاوت‌های پنهان متقاضیان است و با درنظرگرفتن ویژگی‌های فردی دانشجویان جایگزین مناسبی برای ارزیابی ایستا است. در چنین شرایطی، ارزیابی پویا مکملی برای رسیدن به شناخت بهتر نقاط ضعف و قوت دانشجویان است و می‌تواند به آنان در ارتقای سطح کیفی زبانی کمک کند (پوهنر، ۲۰۰۹) (Poehner, 2009).

پرواضح است که میزان آمادگی قبلی داوطلبان و شرایط برگزاری آزمون نیز از اهمیت بسیاری برخوردار است و کارکرد و تفسیر نتایج آزمون را تحت تأثیر قرار می‌دهد. بنابراین، این شرایط باید تا حد امکان استاندارد و برای داوطلبان یکسان باشد (کلمن، ۱۹۹۸) (Colemen, 1998).

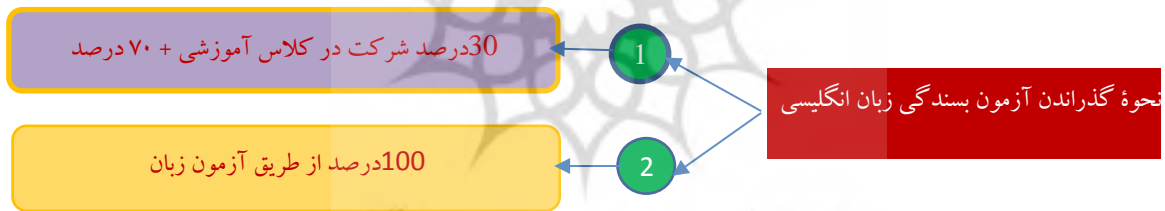
محمدی‌روزبهنانی (۱۳۸۵) معتقد است رضایت‌مندی داوطلبان از جلسه آزمون به آرامش جلسه، شرایط محیطی، تجهیزات استقرار داوطلب، اطلاع‌رسانی قبل و به‌هنگام اجرای آزمون، هدایت داوطلب، دشواری آزمون و میزان آمادگی پیشین داوطلب بستگی دارد و مجموعه این عوامل مبنایی بر استانداردسازی جلسات آزمون هستند که در دسته پایایی آزمون قرار می‌گیرند. همانگونه که نتایج پژوهش نشان می‌دهد، شرکت‌کنندگان از شرایط برگزاری آزمون نسبتاً رضایت دارند اما از محتوای آزمون‌ها و درجه سختی آنان رضایت چندانی ندارند. درواقع برخلاف ادعای برگزارکنندگان این آزمون‌ها که تلاش دارند با تأکید بر مهارت‌های چهارگانه زبان‌آموزی، نوع سؤالات را با آزمون‌های بین‌المللی شبیه‌سازی کنند، تعداد زیادی از این آزمون‌ها تنها بخش محدودی از زبان (اغلب واژگان، دستور و درک مطلب) را مدنظر قرار می‌دهند و بخش‌های مهم‌تر مانند مهارت‌های گفتاری و نوشتاری اغلب مورد غفلت واقع می‌شوند. پرواضح است که در چنین شرایطی داوطلبان تنها

خود را برای تقویت مهارت‌های دستوری، واژگانی و درک مطلب آماده می‌کنند و به سایر مهارت‌ها توجه زیادی نمی‌کنند. براین مبنا به‌نظر می‌رسد که آزمون‌های بسندگی زبان انگلیسی و زیربخش‌های آن مانند بخش شنیداری، واژگان، دستور، درک مطلب و غیره نباید از یکدیگر تفکیک شوند. به‌عبارت دیگر، نگاه به‌آزمون‌سازی باید آکادمیک، ارگانیک و یکپارچه باشد و تقطیع بخش‌ها نباید صورت پذیرد. از این‌رو، پیشنهاد می‌شود خواندن و ترجمه، شنیدن و گفتن، و شنیدن و نوشتن با یکدیگر مورد ارزیابی قرار گیرند و چنانچه قرار است واژگان و دستور نیز لحاظ شود می‌تواند در دل همین پرسش‌ها لحاظ شود.

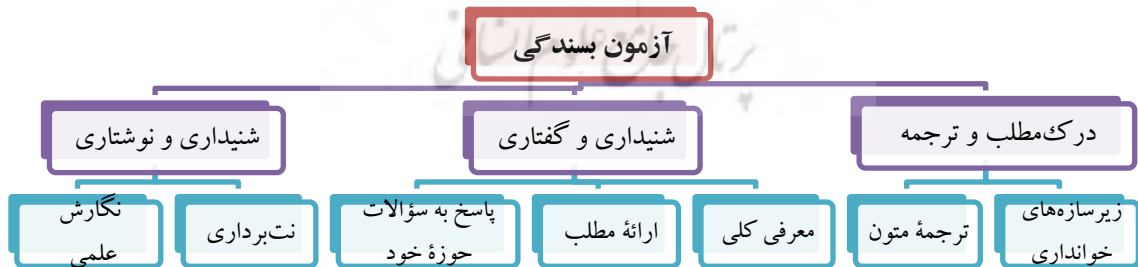
به‌طور کلی، با توجه به‌نتایج بدست آمده از آزمون‌های فعلی مشاهده می‌شود که این آزمون‌ها نه تنها تأثیر مثبتی بر میزان یادگیری و کارکرد مخاطب ندارند بلکه اثر بازخورد آزمون آنان از نوع منفی است. در طراحی سؤالات آزمون باید به این مهم توجه شود که اهداف و انگیزه اصلی از برگزاری این آزمون‌ها چیست و متولیان، مخاطبان و ذی‌نفعان چه کسانی‌اند؟ از این‌رو، با نتایج بدست‌آمده از نتایج این پژوهش می‌توان آزمونی استاندارد با نگاهی سیستمی و همسو با مؤلفه‌های فرهنگی-بومی پیشنهاد داد که در روند آزمون‌های مذکور تغییر ایجاد کند. براساس یافته‌های بدست‌آمده از پژوهش پیشنهادهای مطرح‌شده در راستای ارتقای سطح کیفی آزمون‌های بسندگی زبان انگلیسی به‌اختصار به‌شرح زیر است:

۱. بهتر است نگاه به‌آزمون‌سازی حرفه‌ای شود و مرکز سنجشی در هر دانشگاه برای پایش و هنجاریابی آزمون‌ها ایجاد شود.
۲. چنانچه مرکز سنجش واحدی برای پایش و هنجاریابی آزمون‌ها وجود داشته باشد، آزمون‌های بسندگی زبان انگلیسی که در دانشگاه‌ها برگزار می‌شوند، تحت نظارت وزارت عتف قرار گیرد و در صورتی که استاندارد بودند، تأیید شوند.
۳. پیشنهاد می‌شود تا حد امکان آزمون به‌صورت خوشه‌ای و براساس گروه‌های آموزشی برگزار شود. پرواضح است که ماهیت هریک از رشته‌های تحصیلی با یکدیگر گوناگون است و نمی‌توان آزمونی واحد طراحی کرد که در آن سوگیری وجود نداشته باشد.

۴. باید تلاش شود در آزمون‌های بسندگی زبان انگلیسی، همگی مهارت‌های زبانی سنجیده شوند و به مهارت‌های گفتاری و نوشتاری نیز در کنار سایر مهارت‌ها نگریده شود.
 ۵. باید تلاش شود آزمون‌ها براساس فلسفه «انگلیسی‌های جهانی» و فرهنگ ایرانی به زبان انگلیسی طراحی شوند.
 ۶. پرسش‌ها از حالت چهارگزینه‌ای صرف و حافظه‌محور خارج شوند.
 ۷. تفکیک و تقطیع زبان به دستور و واژگان علمی نیست و پیشنهاد نمی‌شود.
 ۸. آزمون‌های وزارت عتف به شکل اینترنتی و مجازی در تمام دانشگاه‌های متقاضی برگزار شوند.
 ۹. امکان برگزاری آزمون به صورت اینترنتی و آنلاین فراهم باشد.
 ۱۰. برای برقراری عدالت آموزشی و آشنایی با آزمون نمونه پرسش‌ها در وبگاه و در اختیار داوطلبان قرار گیرند.
 ۱۱. از آنجایی که تهیه آزمون استاندارد پرهزینه است،
- پیشنهاد می‌شود وزارتخانه و دانشگاه‌ها یارانه‌ای را به آزمون‌ها اختصاص دهند و به هزینه دریافتی از دانشجویان برای شرکت در آزمون اکتفا نشود.
۱۲. تلاش شود آزمون‌ها براساس مکتب «آزمون برای یادگیری» به شکل پویا طراحی شوند.
 ۱۳. بانک پرسش‌های آزمون باید پیوسته مورد پایش و کنترل کیفی قرار بگیرد.
 ۱۴. به دانشجویان اجازه داده شود در صورت تمایل ۳۰ درصد نمره را از طریق شرکت در کلاس‌های استاندارد یادگیری زبان دانشگاه و ۷۰ درصد را از طریق شرکت در آزمون کسب کنند.
 ۱۵. تعداد شعبه‌های برگزاری آزمون افزایش یابد تا از حجم سفرهای بین شهری کاسته و به عدالت آموزشی نزدیک شود.
 ۱۶. اجزای آزمون شامل الف: درک مطلب و ترجمه ب: شنیداری و گفتاری پ: شنیداری و نوشتاری باشد.
- به صورت کلی، پیشنهاد اصلی پژوهش حاضر برگزاری آزمون به صورت ذیل است:



شکل ۲: پیشنهاد نحوه برگزاری آزمون بسندگی زبان انگلیسی



شکل ۳: ساختار آزمون پیشنهادی آزمون‌های بسندگی زبان انگلیسی براساس مهارت‌های چهارگانه

می‌توان برحسب نیاز متقاضیان هر رشته دسته‌بندی متفاوتی از آزمون‌ها ارائه داد. در برخی شرایط می‌توان برای برخی از رشته‌ها، آزمون را به صورت شفاهی برگزار کرد. بنابراین، نیازسنجی دقیق مخاطبان براساس رشته تحصیلی می‌تواند از جمله موضوعات قابل توجه برای پژوهش‌های آینده باشد.

با توجه به اهمیت طراحی سؤالات براساس نیازهای بومی متقاضیان پژوهشگران می‌تواند با بررسی جامعه‌شناختی دقیق‌تر ویژگی‌های فرهنگی جامعه ایران را بررسی کنند و در تدوین پرسش‌های آزمون مدنظر قرار دهند. از سوی دیگر برای تخصصی‌کردن آزمون‌های بسندگی زبان انگلیسی

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. UK, Oxford: Oxford University Press.
- Bazerman, C., Little, J., & Bethel, L. (2005). *Reference guide to writing across the curriculum*. US, South Carolina, Anderson: Parlor Press LLC.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Carroll, J. B. (1961). The nature of data, or how to choose a correlation coefficient. *Psychometrika*, 26(4), 347-372.
- Clark-Gareca, B. (2010). Language assessment in practice (Book Review). *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 10(2), 41-46.
- Coleman, A. L. (1998). Excellence and equity in education: High standards for high-stakes tests. *Virginia Journal of Social & the Law*, 6(10), 81-114.
- DeVon, H. A., Block, M. E., Moyle Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D.J., et al. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship*, 39(2), 155-64.
- Farhady, H., Jafarpur, A., & Birjandi, P. (2007). *Testing language skills: From theory to practice*. Tehran: SAMT.
- Harsch, C. (2017). Proficiency. *ELT Journal*, 71(2), 250-253.
- Herman, J., Dreyfus, J., & Golan, Sh. (1990). *The effects of testing on teaching and learning*. Los Angeles, CA, National Center for Research on Evaluation, Standards, and Student Testing.
- Kachru, B. B. (1986). *The alchemy of English: The spread functions and models of non-native Englishes*. Oxford, UK: Pergamon.
- Kane, M. T. (2011). Validating score interpretations and uses. *Language Testing*, 29(1), 3-17.
- Lado, R. (1961). *Language testing*. New York, NY: McGraw-Hill.
- Motallebzadeh, K., & Baghaee Moghaddam, P. (2011). Models of language proficiency: A reflection on the construct of language ability. *Iranian Journal of Language Testing*, 1(1), 42-48.
- Nunan, D. (1989). *Second language teaching & learning*. Boston: Heinle & Heinle.
- Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Berlin: Springer.

- Poehner, M. E. (2009). Group dynamic assessment: Mediation for the L2 classroom. *TESOL Quarterly*, 43(3), 471-491.
- Roever, C., & McNamara, T. (2006). Language testing: The social dimension. *International Journal of Applied Linguistics*, 16(2), 242-258.
- Roy Chan, K. L. (2020). Washback in education: A critical review and its implications for language teachers. *Journal of Foreign Language Education and Technology*, 5(1), 108-124.
- Terry, M. & Wilson, J. (2004). *Focus on academic skills for IELTS*. London, England: Pearson Education Limited.
- Tupas, T.R.F. (2004). The politics of Philippine English: Neocolonialism, global politics and the problem of postcolonialism. *World Englishes*, 23(1), 47-58.
- Ulman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L.S. Fidell (Eds.), *Using multivariate statistics* (4th ed., pp. 653-771). Needham Heights, MA: Allyn & Bacon.
- پیش‌قدم، رضا و ابراهیمی، شیما (۱۳۹۷). نسبت حسی و تأثیر آن بر مهارت نوشتاری زبان‌آموزان غیرفارسی‌زبان. *جستارهای زبانی*، ۹(۶)، ۲۱۳-۲۴۰.
- عباسی، زهرا (اسفندماه ۱۳۹۳). تهیه و تدوین آزمون بسندگی زبان فارسی. نخستین همایش آموزش زبان فارسی، تهران: دانشگاه تربیت مدرس.
- غفوریان، محسن و عاشوری، محمد (۱۳۹۶). دانش‌آموزان دیرآموز: ارزیابی پویا، ویژگی‌ها، شناسایی، شیوه‌های تدریس و بهبود ظرفیت یادگیری. *تعلیم و تربیت استثنائی*، ۱۷(۴)، ۵۷-۶۴.
- محسن‌پور، محدثه (۱۳۹۰). ارزیابی داده‌های کیفی. *بیهقی*، ۱۶(۲)، ۵۰-۵۵.
- محمدیگی، ابوالفضل، محمدصالحی، نرگس و گل، محمدعلی (۱۳۹۳). روایی و پایایی ابزارها و روش‌های مختلف اندازه‌گیری آنها در پژوهش‌های کاربردی در سلامت. *مجله دانشگاه علوم پزشکی رفسنجان*، ۱۳(۱۲)، ۱۱۵۳-۱۱۶۹.
- محمدی‌روزبهرانی، کیانوش (۱۳۸۵). مقدمه‌ای بر استانداردسازی اجرای آزمون‌ها: بررسی شرایط برگزاری آزمون‌های سراسری ورود به دانشگاه‌ها از راه سنجش رضایتمندی شرکت‌کنندگان. *پژوهش و برنامه‌ریزی در آموزش عالی*، ۳(۴۱)، ۱-۱۲.

پیوست ۱: پرسشنامه آسب‌شناسی آزمون‌های بسندگی زبان انگلیسی

لطفاً در مورد آزمونی که شرکت کرده‌اید، به پرسش‌های زیر پاسخ دهید:

سؤال	کاملاً موافقم	موافقم	نظری ندارم	مخالقم	کاملاً مخالفم
۱ آزمون توانایی زبانی شرکت‌کننده را به‌درستی می‌سنجد.					
۲ متناسب با نیازهای واقعی متقاضیان طراحی شده است.					
۳ مطابق با معیارهای فرهنگی کشور و به‌دور از تقلید محض از آزمون‌های بین‌المللی طراحی شده است.					
۴ براساس نیازهای بومی کشور در راستای استفاده و انتشار دانش تهیه شده است.					
۵ مهارت‌های اصلی زبان را می‌سنجد.					
۶ کلیت دانش زبانی را به‌درستی می‌سنجد.					
۷ شرایط برگزاری آزمون استاندارد است.					
۸ گزینه‌های مربوط به سؤالات مناسب طراحی شده‌اند.					
۹ مکان برگزاری آزمون مناسب است.					
۱۰ مدت زمان برگزاری آزمون مناسب است.					
۱۱ تعداد سؤالات مناسب است.					
۱۲ درجه سختی آزمون مناسب است.					
۱۳ این آزمون در یادگیری زبان تأثیر مثبت دارد.					
۱۴ شرکت در این آزمون‌ها در دوره دکتری ایجاد اضطراب نمی‌کند.					
۱۵ شرکت و قبولی در این آزمون هزینه مالی زیادی تحمیل نمی‌کند.					
۱۶ موفقیت در آن می‌تواند در شغل حرفه‌ای من تأثیر مثبت داشته باشد.					
۱۷ آزمون حافظه‌محور نیست.					
۱۸ موفقیت در آن به‌رفع نیازهای تخصصی دوره دکتری (مقاله و سخنرانی) کمک می‌کند.					
۱۹ از چگونگی پاسخگویی و رسیدگی به اعتراضات رضایت دارم.					
۲۰ محتوای آزمون به‌سمت رشته گرایش خاصی سوگیری ندارد.					
۲۱ نمره معیار قبولی به‌درستی انتخاب شده است.					
۲۲ آزمون منصفانه طراحی شده است.					
۲۳ آزمون به‌نفع افرادی است که دسترسی به‌کلاس‌های آمادگی آزمون و تست‌زنی دارند و بدون یادگیری نمره می‌آورند.					
۲۴ نمونه سؤالات آزمون از قبل در اختیار متقاضیان قرار می‌گیرد.					

۲۵- نقاط ضعف و قوت آزمون‌های بسندگی زبان چیست و چه پیشنهادهایی برای ارتقای سطح آن‌ها دارید؟