



Validity of Ph.D. Entrance Exams of Engineering in 2018

Soleyman Zolfagharnasab¹, Ali Delavar², Norali Farokhi³, Ehsan Jamali⁴

1. Ph.D. Candidate of Assessment & Measurement, Allameh Tabataba'i University, Iran. Email: salarnik2001@yahoo.com

2. Emeritus Professor, Department of Assessment & Measurement, Faculty of Psychology and Educational Science, Allameh Tabataba'i University, Iran. Email: delavarali@yahoo.com

3. Associate Professor, Department of Assessment & Measurement, Faculty of Psychology and Educational Science, Allameh Tabataba'i University, Iran. Email: farokhinoorali@yahoo.com

4. Assistant Professor, National Organization of Educational Testing (NOET), Iran. Email: ehsanjamali@gmail.com

Article Info ABSTRACT

Article Type:
Research Article

Received:
2020-05-09

Revised:
2020-06-04

Accepted:
2020-07-20

Objective: The quality of a Ph.D. candidate's performance is the most important factor that can affect admission committee members' decision-making and selection of applicants. The primary purpose of this research was to evaluate the predictive criterion validity of subject tests scores x_i in four *engineering* Ph.D. programs on interview measures y_i as the criterion for successful performance Y .

Methods: In this descriptive correlational study, candidates' scores in subject tests x_i and interviews y_i in four engineering Ph.D. programs in 2018 were analyzed based on the Generalizability theory. Depending on their majors, examinees' data included their scores on 3 to 8 subtests in subject tests x_i and 2 composite interview criteria y_i . Data were analyzed using GENOVA software based on a multivariate analysis with a single facet $p \times i^\circ$ design. The relationship between two composite universe scores was illustrated by the bivariate normal distribution in a typical scatter diagram.

Results: The results showed that subject tests scores can predict candidates' scores on the interview criterion as an index for successful performance except in the field of Information Technology. The large number of subtests in subject tests, their instability, and the construct irrelevant variance was among the factors that may influence the accuracy of predictions, especially on the cut scores, for accepting or rejecting candidates.

Conclusion: Accurately predicting the success of an applicant's performance Y based on test scores x_i is an unattainable goal. However, precisely defining the criteria for success y_i in Ph.D. programs, increasing the significance of general English and aptitude tests, adjusting the difficulty level of tests, reducing measurement error in predictive measures x_i , and having a two-step assessment program in which the general language and aptitude tests are administered as a primary screening by NOET and the subject tests and interviews by universities can be effective method to increase the criterion validity of test scores in predicting candidates' successful performance and enhance the chances of admitting more qualified students.

Keywords: Generalizability theory, Predictive validity, Subject test, Cut score, Doctoral programs.

Cite this article: Zolfagharnasab, S., Delavar, A., Farokhi, N. & Jamali, E. (2020). Validity of Entrance Exams of Engineering in 2018. *Educational Measurement and Evaluation Studies*, 10 (30): 50 pages. DOI:



© The Author(s).

Publisher: National Organization of Educational Testing (NOET)



روایی آزمون های ورودی دکتری مهندسی ۱۳۹۷

سلیمان ذوالفقارنسب، علی دلاور، نورعلی فرخی، احسان جمالی

۱. دانشجوی دکتری سنجش و اندازه گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی، ایمیل: salarnik2001@yahoo.com

۲. استاد تمام گروه سنجش و اندازه گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی

۳. دانشیار گروه سنجش و اندازه گیری، دانشکده روانشناسی و علوم تربیتی، دانشگاه علامه طباطبائی

۴. استادیار پژوهشی سازمان سنجش آموزش کشور

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	هدف: عملکرد موفقیت آمیز یک داوطلب دوره دکتری مهمترین عاملی است که در تصمیم گیری اعضای کمیته پذیرش می تواند کمک کند کدام متقاضی را بپذیرند. هدف این پژوهش، بررسی روایی ملاکی-پیش بین نمرات آزمونهای تخصصی X_i در ۴ رشته دکتری مهندسی روی سنجه های مصاحبه Y_i به عنوان ملاک عملکرد موفقیت آمیز Y بوده است.
دریافت: ۱۳۹۹/۰۲/۲۰	روش پژوهش: در این پژوهش توصیفی از نوع همبستگی، رتبه های داوطلبان در آزمون های اختصاصی X_i و مصاحبه Y_i در ۴ رشته دکترای مهندسی در سال ۱۳۹۷ بر اساس تئوری تعمیم پذیری تحلیل شدند. بسته به رشته دانشگاهی داده های آزمون شوندگان دربرگیرنده نمرات آنها روی ۳ تا ۸ خرده آزمون اختصاصی X_i و ۲ نمره ترکیبی ملاک مصاحبه Y_i بود. داده ها با نرم افزار <i>mGENOVA</i> بر اساس یک طرح تک رویه ای $p \times i$ تحلیل شدند. سپس همبستگی بین دو نمره جهانی ترکیبی با توزیع نرمال دو متغیری در نمودار پراکنندگی نشان داده شده.
اصلاح: ۱۳۹۹/۰۳/۱۵	یافته‌ها: نتایج نشان می دهد نمرات آزمون های تخصصی X_i به خوبی می توانند نمرات افراد را در ملاک مصاحبه Y_i به عنوان شاخص عملکرد موفقیت آمیز Y و یا کیفیت یک داوطلب پیش بینی کنند به جز در رشته فناوری اطلاعات؛ تعداد زیاد خرده آزمون های تخصصی، ناپایایی آنها و واریانس ناپسته به سازه از عواملی هستند که می تواند پیش بینی های درست برای پذیرش و یا رد داوطلبان را -بویژه روی نمره برش- تحت تاثیر قرار دهند.
پذیرش: ۱۳۹۹/۰۴/۳۰	نتیجه گیری: اگر چه پیش بینی دقیق عملکرد موفقیت آمیز Y داوطلبان دوره دکتری بوسیله نمرات آزمون ها X_i هدف دست نیافتنی است، اما تعریف بهتر آنچه که ملاک موفقیت Y_i در دوره دکتری است و با اهمیت تر شدن آزمون های عمومی زبان و استعداد X_i و همزمان تعدیل سطوح دشواری این آزمون ها و کاهش خطای اندازه گیری سنجه های پیش بینی کننده X_i در یک برنامه سنجش دو مرحله ای که در آن آزمون های عمومی زبان و استعداد به عنوان غربال اولیه توسط سازمان سنجش برگزار شود و آزمون های اختصاصی و مصاحبه توسط دانشگاهها، می تواند راه بهینه ای باشد برای افزایش روایی نمرات آزمون ها X_i در پیش بینی عملکرد موفقیت آمیز Y داوطلبان و پذیرش دانشجویان باکیفیت تر.
	واژه‌های کلیدی: تئوری تعمیم پذیری، روایی پیش بین، آزمون های تخصصی، نمره برش، برنامه های دکتری

استناد: ذوالفقارنسب، سلیمان، دلاور، علی، فرخی، نورعلی و جمالی، احسان (۱۳۹۹). روایی آزمون های ورودی دکتری مهندسی ۱۳۹۷. مطالعات اندازه‌گیری و ارزشیابی

آموزشی، ۱۰ (۳۰)، ۵۰ صفحه. DOI:



مقدمه

فرایند سنجش و پذیرش^۱ افراد توانمند و کارآمد برای تحصیل در دوره دکتری تخصصی در ایران بسیار پرچالش بوده است. اهمیت مسئله پذیرش افراد توانمند و شایسته به حدی بوده که معیارهای پذیرش $(X=x_1, x_2, \dots, x_n)$ و ملاک‌های $(Y=y_1, y_2, \dots, y_n)$ مهم آن و همچنین شیوه آزمون‌گیری به صورت تستی یا تشریحی برای تدوین الگوی بهینه پذیرش همواره مورد بحث و جدال باشد. به همین دلیل در دو دهه گذشته، تغییرات زیادی در فرایند سنجش و پذیرش دوره دکتری روی داده است. به‌عنوان مثال تا پیش از سال ۱۳۹۰ پذیرش دانشجویان دوره دکتری به وسیله هر دانشگاه و به صورت غیرمتمرکز صورت می‌گرفت. دانشگاه‌ها مکانیزم‌های گوناگونی ایجاد کرده بودند تا داوطلب بر اساس معیارهای مرسوم کمی و کیفی به‌گونه‌ای انتخاب شود که با برنامه‌های ویژه دکتری ارائه شده در دانشگاه آنها هماهنگ باشد. اعضای هیئت علمی هر رشته در هر دانشکده به‌طور مستقل آزمون‌های تخصصی و عمومی تهیه و اجرا می‌کردند و افراد دارای بالاترین نمره در چند برابر ظرفیت به مصاحبه دعوت می‌شدند. در این مرحله، سوابق علمی و پژوهشی داوطلبان به‌عنوان «ملاک عملکرد موفقیت‌آمیز Y » به وسیله یک کمیته پذیرش متشکل از اعضای هیئت علمی گروه‌های آموزشی دانشکده، ارزیابی می‌شد و در پایان بر اساس نمره آزمون‌های تخصصی و عمومی x_i با سنجش‌های مصاحبه y_i مناسب‌ترین داوطلبان با بهترین نمره‌ها و سوابق تحصیلی، پذیرش می‌شدند. عمدتاً سوال آزمون‌های تخصصی و حتی برخی پرسش‌های آزمون‌های عمومی استعداد و زبان به صورت بازپاسخ یا تشریحی بودند و به صورت سهمی نمره‌گذاری می‌شدند و به همین دلیل نمره آزمون برآورد خوبی از توانایی‌های اساسی نوشتاری داوطلبان بود. در تهیه این آزمون‌های x_i تشریحی تلاش می‌شد که بسته به رشته تحصیلی، سطوح بالای تفکر شناختی (وب^۲، ۱۹۹۷) مثل توان تجزیه و تحلیل کردن، استدلال کردن، قضاوت کردن، استنتاج کردن را که از توانایی‌های اساسی داوطلبان دوره دکتری است به شیوه‌های گوناگون بسنجند.

اما نبود الگوی یکسان در روش‌های برگزاری آزمون ورودی براساس مقررات و ضوابط حاکم بر هر دانشگاه، نامشخص بودن و گوناگونی شاخص‌های لازم داوطلبان

^۱. assessment and admission process

^۲. Webb

در رشته‌های یکسان در دانشکده‌های مختلف در فرایند مصاحبه، اعمال سلیقه‌های مختلف در پذیرش دانشجو و رعایت نکردن عدالت آموزشی، مشکلات فراوانی را پیش‌روی داوطلبان ورود به مقطع دکتری قرار داده و به نارضایتی‌های گسترده منجر شده بود (یونسی، ۱۳۹۵ و نورشاهی، ۱۳۸۸). به دلیل اعتراضات زیادی که بر نحوه پذیرش بخش‌های دانشگاهی می‌شد و لزوم شرکت داوطلبان در آزمون‌های مختلف که در دانشگاه‌های متفاوت برگزار می‌شد و افزایش تعداد داوطلبان پذیرش دوره دکتری، تصمیم‌گیرندگان در وزارت علوم به این نتیجه رسیدند که شیوه پذیرش را به صورت نیمه‌متمرکز به سازمان سنجش آموزش کشور بسپارند. با این توضیح می‌توان انتظار داشت که رویکرد وزارت علوم در اجرای نیمه‌متمرکز آزمون‌های تخصصی و عمومی دکتری توسط سازمان سنجش آموزش کشور و سپس انجام مصاحبه توسط دانشگاه‌ها، بتواند در ایجاد موقعیتی استاندارد و کم‌خطا و ایجاد یکپارچگی و در نهایت برقراری عدالت آموزشی به منظور انتخاب شایسته‌ترین افراد، رویکرد میانه‌ای ارائه دهد.

اما تجمیع این رویکرد دوگانه و اعلام نتایج نهایی نیز با مسائلی همراه بوده است؛ یکی از مشکلاتی که در سال‌های گذشته، تصمیم‌گیرندگان با آن روبرو بوده‌اند تغییرات زیادی است که در اعمال وزن و ضریب آزمون تخصصی و عمومی سازمان سنجش آموزش کشور و سهم مصاحبه دانشگاه‌ها روی داده است. به‌عنوان مثال، در سال ۱۳۹۰ بر اساس نمره‌های آزمون تخصصی و عمومی سازمان سنجش آموزش کشور، چند برابر ظرفیت به دانشگاه‌ها اعلام شد، اما پذیرش نهایی بر اساس ۱۰۰ درصد نمره مصاحبه بوده است؛ به عبارتی آزمون سازمان سنجش آموزش کشور تنها به صورت غربال اولیه برای معرفی داوطلبان به بخش‌های دانشگاهی در نظر گرفته شده بود. سپس در سال ۱۳۹۱ پذیرش داوطلبان بر اساس یک‌سوم سهم آزمون تخصصی و عمومی سازمان سنجش آموزش کشور و دو سوم سهم مصاحبه علمی آموزشی دانشگاه‌ها بوده است. در حالی که تغییرات اینچنینی در سال‌های بعد نیز روی داده است. این تغییرات باعث شد که یک شیوه‌نامه برای فرایند مصاحبه و معیارهای آن تهیه شود و هر بخش آن -که از سنجش‌های متفاوتی تشکیل شده- نمره و وزن مشخصی داشته باشد. جدول (۱) تغییرات سال‌های مختلف را نشان می‌دهد.

جدول (۱) تغییرات آزمون دکتری طی سال‌های ۱۳۹۰ تا ۱۳۹۸

سال	تعداد رشته‌ها	تعداد عناوین درس‌ها	وزن یا ضریب در نحوه پذیرش نهایی داوطلبان
۱۳۹۰	۱۰۷	۳۲۱	۱۰۰٪ نمره مصاحبه
۱۳۹۱	۲۸۵	۸۵۵	یک‌سوم نمره آزمون کتبی و دو سوم نمره مصاحبه
۱۳۹۲	۲۸۶	۸۵۸	۵۰٪ نمره آزمون کتبی و ۵۰٪ نمره مصاحبه
۱۳۹۳	۲۶۷	۸۰۱	۳۰٪ آزمون کتبی و ۷۰٪ نمره مصاحبه
۱۳۹۴	۲۵۱	۷۵۳	۳۰٪ آزمون کتبی و ۷۰٪ نمره مصاحبه
۱۳۹۵	۷۷	۲۳۱	در اختیار دانشگاه
*۱۳۹۶	۲۴۴	۷۳۲	۵۰٪ نمره آزمون کتبی و ۵۰٪ نمره مصاحبه در رشته محل‌های آموزشی و پژوهشی
*۱۳۹۷	۲۴۶	۷۳۸	۳۰٪ نمره آزمون کتبی و ۷۰٪ نمره مصاحبه در رشته‌محل‌های پژوهش‌محور
*۱۳۹۸	۲۴۵	۷۳۵	

*از سال ۱۳۹۶ بر اساس قانون مصوب مجلس شورای اسلامی موسوم به «قانون سنجش و پذیرش تحصیلات تکمیلی» سنجش و پذیرش داوطلبان با این ضرایب انجام گرفته است.

از عوامل بحث‌برانگیز دیگر تغییراتی است که در چگونگی تأثیر نمره معدل کارشناسی و کارشناسی ارشد در ساخت نمره کل داوطلب بوده است. در سال‌های اولیه ۸۰ درصد نمره آزمون عمومی و اختصاصی^۱ و ۲۰ معدل معدل تراز شده [با وزن مقطع کارشناسی (۶۰ درصد) و کارشناسی ارشد (۴۰ درصد)] محاسبه شده است و اثر معدل هر دو مقطع در نمره آزمون کتبی از سوی سازمان سنجش آموزش کشور اعمال می‌شد. اما همان‌طور که آشکار شده ارزش یک معدل مثلاً ۱۵ به دست آمده از دانشگاه‌های گوناگون روزانه، با دیگر انواع دانشگاه‌ها تفاوت دارد. به همین دلیل در سال‌های اخیر، اعمال ضریب معدل هر داوطلب را به عهده کمیته پذیرش دانشکده‌ها گذاشته‌اند و در جلسه مصاحبه بر حسب نوع دانشگاه فارغ‌التحصیلی داوطلب یک نمره سهمی به معدل افراد داده می‌شود. در نهایت، اثر معدل در بخش مربوط به نمره‌های مصاحبه اعمال

1. General test battery, subject test battery

می‌شود و نه به‌عنوان یک پارامتر معین و مستقل در کنار سایر معیارها. این «دوگانگی» بین آزمون‌های تخصصی و عمومی سازمان سنجش آموزش کشور با فرایند مصاحبه‌ای که در دانشگاه‌ها صورت می‌گیرد باعث شده بسیاری از دانشگاه‌ها و اعضای هیئت علمی دوباره خواستار اصلاح روند موجود برگزاری آزمون دکتری و پذیرش دانشجویان باشند. از جمله مشکلاتی که مطرح می‌شود ناکارآمدی سؤال‌های چهارگزینه‌ای آزمون‌های تخصصی در اندازه‌گیری دانش عمیق افراد و انتخاب دانشجویی با کیفیت در مقطع دکتری است (یونسی، ۱۳۹۵، صص ۲۳۷ و ۲۳۰).

در سال ۱۳۹۷ در بازتعریفی که از ایجاد دوره دکتری تخصصی در ایران شد هدف آن را به این صورت مشخص کردند: «... تربیت افرادی که با احاطه یافتن به آثار علمی در یک زمینه خاص و آشنا شدن با روش‌های پیشرفته تحقیق و دستیابی به جدیدترین مبانی آموزشی و پژوهشی بتواند با نوآوری در زمینه‌های علمی و تحقیقی در رفع نیازهای کشور و گسترش مرزهای دانش در رشته تخصصی خود مؤثر بوده و به تازه‌هایی در جهان دانش دست یابد» (آیین‌نامه آموزشی دوره دکتری تخصصی، ۱۳۹۷).^۱ با این هدف، هر سال تعداد زیادی داوطلب دوره دکتری خود را آماده رقابت در آزمون‌های تخصصی و عمومی می‌کنند و پس از به دست آوردن نمره‌های حد نصاب در این آزمون‌ها وارد مرحله مصاحبه علمی می‌شوند. در نهایت، افرادی که بیشترین نمره‌ها را در مجموع دو مرحله آزمون و مصاحبه کسب کنند به‌عنوان باکیفیت‌ترین یا موفق‌ترین داوطلبان، پذیرش می‌شوند.

در عمل نمی‌توان از این فرایند به‌ظاهر ساده و در عین حال پیچیده در انتخاب دانشجویی باکیفیت تصویری روشن ارائه کرد. پاسخ به این پرسش که چه معیارهای کمی ($X=x_1, x_2, \dots, x_n$) و ویژگی‌های کیفی به‌عنوان ملاک عملکرد موفقیت‌آمیز ($Y=y_1, y_2, \dots, y_n$) در فرایند انتخاب دانشجوی دکتری وجود دارد تا حد زیادی وابسته است به چگونه تأثیر دادن این معیارهای پیش‌بینی‌کننده X مثل آزمون‌ها x_i و همچنین عملیاتی کردن تعریف واژه کیفیت Y یک داوطلب توسط دانشگاه‌ها؛ عمدتاً برای تعریف کردن متغیرهای کیفی یا ملاک عملکرد موفقیت‌آمیز ($Y=y_1, y_2, \dots, y_n$) هم مجموعه‌ای از معیارهای کمی ($X=x_1, x_2, \dots, x_n$) در نظر گرفته می‌شود (اتیه^۲، ۱۹۹۹). آنهایی که

^۱. برای جزئیات بیشتر به فصل اول آیین‌نامه آموزشی دوره دکتری تخصصی، ۱۳۹۷ نگاه کنید.

^۲. Attiyeh

مسئول برنامه‌های انتخاب و پذیرش افراد برای دوره‌های آموزشی هستند، تمایل دارند که به پیش‌بینی‌کننده‌ها X (یا همان معیارها^۱) بیشتر توجه کنند و ملاک‌های پیشرفت یا عملکرد موفقیت‌آمیز Y را نادیده می‌گیرند. از جمله دلایل آن ماهیت عمدتاً کیفی ملاک به‌عنوان عملکرد موفقیت‌آمیز در آینده است و نمی‌توان در زمانی که داوطلبان به‌طور واقعی پذیرش می‌شوند آنها را شناسایی کرد. در واقع داوطلبان بر اساس «استنباط رابطه^۲» بین معیارهای پذیرش $(X=x_1, x_2, \dots, x_n)$ و سنجه‌های ملاک $(Y=y_1, y_2, \dots, y_n)$ پذیرش می‌شوند. اگر داوطلبان روی معیارهای X پذیرش خوب عمل کرده باشند، استنباط می‌کنیم آنها روی سنجه‌های ملاک Y هم خوب عمل خواهند کرد (هارتنت و ویلینگهام^۳، ۱۹۸۰، ص ۳). این معیارهای X پذیرش از ترکیب‌های مختلفی از سنجه‌های پیش‌بینی‌کننده $(X=x_1, x_2, \dots, x_n)$ مثل آزمون‌ها تشکیل شده است و انتظار می‌رود کارکردهای پیش‌بینی‌کنندگی یا همان روایی پیش‌بین این معیارها $(X=x_1, x_2, \dots, x_n)$ به‌ویژه نمره آزمون‌های ورودی x_i دقیق، باثبات و روا باشد (ویلینگهام^۴، ۱۹۷۶).

به هر حال، به گفته مسیک^۵ (۱۹۸۹) «روایی» یک قضاوت ارزشی یکپارچه است در رابطه با حد و اندازه‌ای که مدارک تجربی و منطق نظری [بتواند] از کفایت و مناسبت استنباط‌ها و اقدام‌هایی که بر اساس نمره‌های آزمون x_i یا روش‌های دیگر ارزشیابی می‌شود، حمایت کند. چون نمره‌های آزمون‌ها x_i برای انتخاب افراد استفاده شده، در این پژوهش مدارک تجربی فراهم شده برای میزان مناسب بودن تصمیم‌هایی که برای رد یا پذیرش افراد بر اساس نمره‌های این آزمون تخصصی x_i گرفته شده است.

بیان مسئله

در برنامه‌های سنجش و پذیرش، کیفیت داوطلب که در عملکرد موفقیت‌آمیز او Y نمایان می‌شود مهم‌ترین عاملی است که در تصمیم‌گیری اعضای کمیته پذیرش می‌تواند کمک کند کدام داوطلب را بپذیرند. در این خصوص، یکی از مباحثی که همیشه همراه با معیارهای کمی x_i پذیرش دانشجوی دکتری بوده مسئله روایی^۶ این معیارها یا تفسیر

1. Focus upon predictors

2. Inferences about the relationship

3. Hartnett & Willingham

4. Willingham

5. Messick

6. validity

این نمره‌ها بوده است. به‌ویژه در رابطه با نمره‌های x_i آزمون‌هایی که برای غربال اولیه داوطلبان اجرا می‌شود، باید بتوانند کیفیت هر داوطلب را برای عملکرد موفقیت‌آمیز در دوره دکتری و همچنین در چشم‌اندازی گسترده‌تر، موفقیت حرفه‌ای Y او را نیز پیش‌بینی کنند. بنابراین با توجه به تعاریف متعددی که بر اساس رشته‌های تخصصی-موضوعی از روایی پیش‌بین یک آزمون x_i شده، آن را «حد و اندازه‌ای تعریف کرده‌اند که یک معیار سنجشی مثل نمره x_i یک آزمون یا چند آزمون ($X=x_1, x_2, \dots, x_n$) بتواند عملکرد موفقیت‌آمیز افراد در یک حوزه y_i یا حوزه‌های گسترده‌تری ($Y=y_1, y_2, \dots, y_n$) را پیش‌بینی کند» (ویلینگهام، ۱۹۷۶؛ مسیک، ۱۹۹۳). به‌عنوان مثال، روایی پیش‌بین آزمون تافل حد و اندازه‌ای است که «نتایج یک سنجش معین با آزمون زبان انگلیسی x_i بتواند عملکرد آینده فرد را در استفاده از زبان انگلیسی y_i در زمینه مورد نظر پیش‌بینی کند» (بچمن و پالمر، ۱۹۹۶). اغلب گفته شده که این مفهوم مرسوم و سنتی از روایی که در آن همبستگی بین یک نمره آزمون x_i و برخی سنجش‌های موفقیت y_i در فعالیت‌های بعدی است، معنای محدودی از روایی را می‌رساند و این حقیقت را باید در نظر داشت که یک نمره در یک آزمون x_i می‌تواند جوانب گوناگون و مفاهیم اجتماعی مختلفی داشته باشد و روایی یک معیار اندازه‌گیری x_i را می‌توان به زمان‌ها و مکان‌های دیگر نیز تعمیم داد.^۱ در این صورت همه این جوانب باید در نظر گرفته شود و برای آنها مدارک و شواهد روایی تهیه شود (مسیک، ۱۹۹۳).

در تعریف کلی‌تری که انجمن تحقیقات آموزشی آمریکا^۲ (AERA)، انجمن روان‌شناسان آمریکا^۴ (APA) و کمیسیون ملی اندازه‌گیری‌های آموزشی^۵ (NCME) (۲۰۱۴) از معنای روایی ارائه کرده‌اند آن را حد و اندازه‌ای می‌دانند که شواهد و تئوری بتواند از تفسیرهای مربوط به نمره آزمون x_i «برای هدفی که به کار رفته» پشتیبانی کند. بنابراین، هدف آزمون‌های پذیرش x_i در ایران با توجه به آیین‌نامه آموزشی دوره دکتری تخصصی (۱۳۹۷) یعنی حد و اندازه‌ای که نمره آزمون‌های تخصصی و عمومی x_i بتواند عملکرد موفقیت‌آمیز Y یک داوطلب برای تولید علم y_i و نوآوری y_i و مواردی

1. Bachman & Palmer

2. validity generalization

3. American Educational Research Association

4. American Psychological Association

5. National Council for Measurement in Education

نظیر اینها را پیش‌بینی کند. حالا فرایند روایی‌یابی دربرگیرنده گردآوری مدارک و شواهد مناسبی برای آزمون‌های تخصصی x_i است تا به کمک آن پایه‌ای علمی و مستحکم برای تفسیر پیشنهاد شده از نمره آزمون - یعنی انتخاب داوطلب باکیفیت - فراهم سازد. بنابراین روایی، تفسیر پیشنهاد شده برای نمره‌های آزمون x_i است نه خود آزمون. اگر بنا باشد که نمره‌های آزمون x_i به بیش از یک روش تفسیر شود (به‌عنوان مثال، شاخصی از سطح توانایی فعلی y_i ، پیش‌بینی عملکرد دانشگاهی y_i ، پیش‌بینی دستاوردهای بعدی y_i و موفقیت حرفه‌ای y_i) برای همه تفسیرهایی که از نمره می‌شود باید شواهد و مدارک روایی تهیه شود (AERA, APA & NCME, 2014).

هدف اولیه این نوشتار بررسی تجربی و در سطح آماری روایی ملاکی - پیش‌بین آزمون‌های تخصصی x_i در پیش‌بینی ملاک دوگانه مصاحبه y_i در چهار رشته دکتری مهندسی در ایران است. اگر تعریف آیین‌نامه آموزشی از ایجاد دوره دکتری تخصصی را مبنای پذیرش دانشجوی دکتری - به‌عنوان یک قاعده - بپذیریم، باید فردی باشد که پس از سال‌ها آموزش و کسب تجربه در زمینه تخصصی $(X=x_1, x_2, \dots, x_n)$ «بتواند نوآوری داشته باشد، تولید علم کند و مرزهای دانش را در جهان گسترش دهد $(Y=y_1, y_2, \dots, y_n)$ ». بدین ترتیب می‌توان رگه‌هایی از عملکرد موفقیت‌آمیز فرد داوطلب را در سنج‌های y_i مصاحبه دید.

در این پژوهش، ملاک عملکرد موفقیت‌آمیز Y سنج‌های y_i مصاحبه است که از آن برای روایی‌یابی آزمون‌های تخصصی x_i دوره دکتری استفاده شده است. به عبارتی از عناصر «هدف دوره دکتری» که در آیین‌نامه آموزشی دوره دکتری تخصصی (۱۳۹۷) ذکر شده، کمک گرفته شده و بخش نخست این تعریف تا فعل «بتواند» را به‌عنوان بخشی که در آن آزمون‌های تخصصی و آزمون‌های عمومی به‌عنوان معیارهای پیش‌بینی‌کننده پذیرش $(X=x_1, x_2, \dots, x_n)$ تجلی می‌یابد، در نظر گرفته‌ایم و از فعل «بتواند» به بعد که عناصری از واژه کیفیت عملکرد موفقیت‌آمیز یک داوطلب در آن نمود پیدا می‌کند به‌عنوان ملاک عملکرد موفقیت‌آمیز $(Y=y_1, y_2, \dots, y_n)$ در نظر گرفته شده است. با توجه به این تعریف می‌توان این‌گونه استنباط کرد که:

تربیت تخصصی افراد x_i احاطه یافتن به علم در یک زمینه خاص x_i ، آشنا شدن با روش‌های پژوهش x_i ، دستیابی به جدیدترین مبانی آموزشی و پژوهشی x_i عناصر معیارهای پیش‌بینی‌کننده $(X=x_1, x_2, \dots, x_n)$ را تشکیل می‌دهند که در آزمون‌های اختصاصی و عمومی نمود و تجلی می‌یابند.

بتواند نوآوری کند y_i ، بتواند نیازهای کشور را رفع کند y_i ، بتواند مرزهای دانش را گسترش دهد y_i و بتواند به تازه‌هایی در جهان دست یابد y_i عناصر ملاک $(Y=y_1, y_2, \dots, y_n)$ را تشکیل می‌دهند و در سنجه‌های مصاحبه نمود و تجلی می‌یابند و می‌توان سنجه‌های مصاحبه را به‌عنوان «ملاک عملکرد موفقیت‌آمیز Y » یک داوطلب به شمار آورد.

چون تعریف آیین‌نامه آموزشی از ایجاد دوره دکتری به‌طور ضمنی در برگیرنده بسیاری از شاخص‌های علوم و فناوری است و این تعریف با نظام‌های نوآورانه هماهنگ است، هدف دیگر از اجرای این پژوهش، نشان دادن فرایند سنجش و پذیرش در یک چهارچوب تئوریک است تا بتوان این فرایند را با ساختارهای توسعه ملی هماهنگ و هدفمند کرد.

به هر حال، از آنجا که عمده‌ترین شکل روایی یک آزمون معمولاً به صورت روایی محتوایی، روایی ملاکی و روایی سازه بررسی می‌شود. در این نوشتار، نخست «روایی محتوایی» به صورت گذرا بحث شده، سپس «روایی ملاکی» و فرایند روایی‌یابی ملاکی آمده و در دل این نوع روایی‌یابی، یک الگوی تئوریک با جزئیات توضیح داده شده است. سپس در ادامه همین بخش، این الگوی تئوریک مبنایی قرار داده شده برای معرفی و توجیه ساختار «روایی سازه» سنجه‌ها و آزمون‌هایی که برای پذیرش داوطلبان دکتری در ایران به کار می‌رود. اما در این پژوهش تنها بر اساس شواهد تجربی (رگرسیون) روایی ملاکی آزمون‌های تخصصی x_i چهار رشته مختلف مهندسی در پیش‌بینی سنجه‌های مصاحبه y_i گزارش شده است.

در برنامه‌های سنجش و پذیرش دکتری، چون علاوه بر آزمون‌های اختصاصی از آزمون زبان انگلیسی و آزمون استعداد نیز استفاده می‌شود و تهیه‌کنندگان این آزمون‌ها عمدتاً از روی آزمون سوابق تحصیلات تکمیلی^۱ (GRE) الگوبرداری می‌کنند، برای درک بهتر مفاهیم از آزمون GRE و مقیاس‌های آن نام برده شده است. همچنین چگونگی استفاده از این آزمون و هدف خرده‌آزمون‌های آن در پیش‌بینی عملکرد موفقیت‌آمیز توصیف شده است. این آزمون به وسیله خدمات آزمون‌گیری آموزشی^۲ (ETS) تهیه می‌شود و دست کم برای ۹۳ درصد برنامه‌های سنجش و پذیرش دکتری و ۸۱ درصد

1. Graduate Record Examinations (GRE)

2. Educational Testing Service

برنامه‌های پذیرش کارشناسی ارشد نمره‌های آن لازم است. بیشتر پژوهش‌های روایی‌یابی در ادبیات پژوهشی جهان روی این آزمون انجام گرفته است (کانسل، هزلت و وانس، ۲۰۰۱).

هدف دیگر، نشان دادن کارایی آزمون‌های تخصصی x_i در پیش‌بینی درصد افراد موفق در بین مصاحبه‌شوندگان با استفاده از مدل تیلور و راسل^۲ (۱۹۳۹) است. هدف از ارائه این مدل، نشان دادن اهمیت فرایند روایی‌یابی آزمون‌هاست؛ هنگامی که شناس افراد موفق و ناموفق بر اساس ملاک y_i به نسبت ۵۰/۵۰ باشد، اگر از آزمونی استفاده شود که حتی ضریب روایی متوسطی دارد می‌تواند کمک کند درصد افراد موفق نسبت به افراد ناموفق بر اساس نمره برش در بین کسانی که انتخاب خواهند شد، افزایش پیدا کند. نمودارهای خط رگرسیون (یا پیش‌بینی) ملاک مصاحبه y_i از روی آزمون‌های اختصاصی x_i برای بازنمایی دیداری نیز رسم شده است. این نمودارها نشان می‌دهند اگر توزیع نمرات افراد روی چندین پیش‌بینی‌کننده (مثل آزمون‌ها x_i) و چندین ملاک (مثل مصاحبه y_i) به توزیع نرمال دو متغیره $N(0, \sigma^2)$ تبدیل شوند، کمیته‌های پذیرش این چهار رشته چگونه برای انتخاب افراد باکیفیت بر اساس نمره برش آزمون‌های اختصاصی عمل کرده‌اند.

پرسش‌های پژوهش عبارت‌اند از: ۱- آیا نمره‌های آزمون‌های تخصصی به‌عنوان پیش‌بینی‌کننده x_i ارتباط قابل قبول و منطقی با سنجه‌های مصاحبه به‌عنوان ملاکی که تا حدودی کیفیت Y یک داوطلب موفق (مثل نوآوری و تولید علم) در آینده را پیش‌بینی می‌کنند، دارند یا نه؟ ۲- آیا این دو رویکرد متفاوت- آزمون نیمه‌متمرکز سازمان سنجش و مصاحبه کمیته پذیرش دانشگاه‌ها در ارتباط با یکدیگر توانسته‌اند در نهایت باکیفیت‌ترین و موفق‌ترین افراد را پذیرش کنند؟

اهمیت و ضرورت پژوهش

این پژوهش روایی‌یابی از آن جهت اهمیت دارد که از رویکرد روان‌سنجی محض و ارائه صرف داده‌های آماری در رابطه با آزمون‌ها پا فراتر گذاشته شده و با ارائه یک الگوی تئوریک نشان می‌دهد چگونه سنجه‌هایی که در فرایند سنجش و پذیرش دکتری

1. Kuncel, Hezlett & Ones

2. Taylor & Russell

به کار می‌روند با توجه به تعریف آیین‌نامه آموزشی دوره دکتری تخصصی (۱۳۹۷) می‌توانند با سازه‌های گسترده‌تری در چهارچوب‌های اقتصادی به‌طور نظری ارتباط داشته باشند (به نمودار ۱ نگاه کنید). از دیدگاه نظری گروه مهندسی، نوک پیکان توسعه ملی هستند. این گروه از لحاظ خلق فناوری، ثبت اختراعات و انتشارات علمی پیشرو بوده و مهم‌ترین عنصر در انتقال علوم و فناوری هستند. بی‌شک نوآوری متکی به خلق دانش به وسیله یک نظام آموزشی خوب و استاندارد است. برتری یافتن به چالش‌های قرن ۲۱ و توسعه فناورانه، نیازمند پرورش افراد مستعد در نظام آموزشی به‌ویژه دوره دکتری است. در حقیقت، کارکردهای اقتصادی و عملکرد صنعتی کشورهای در حال توسعه به پرورش دانشمندان و مهندسان در تحصیلات تکمیلی متکی بوده است. در آینده ای نه‌چندان دور، توان نوآورانه کشورها در تجاری‌سازی محصولات، تولیدات و خدمات از اصلی‌ترین اهرم‌های رقابتی در جهانی‌سازی و افزایش ثروت خواهد بود. هزینه روی آموزش و تربیت پژوهشگران و کارکنان با مهارت و باکیفیت از عمده‌ترین عواملی است که تعیین می‌کنند پژوهش‌های علمی می‌تواند به پیشرفت علمی و نوآوری تبدیل شود. از این گذشته، سرمایه انسانی باکیفیت عامل کلیدی در پذیرش فناوری‌های جدید و خلق نوآوری است (ذوالفقارنسب و یادگارزاده، ۱۳۹۱). هر چند بحث اخیر ممکن است تا حدودی از مبحث روایی‌یابی فاصله گرفته باشد اما توجهی است برای انتخاب گروه‌های مهندسی با توجه به تعریف آیین‌نامه آموزشی در ایجاد دوره دکتری. به هر حال، در فرایند پذیرش با آزمون‌ها، ضریب روایی r_{xy} به‌طور مستقیم با ارزش اقتصادی یا سودمندی آزمون x_i و روش سنجشی در ارتباط و همسو است. استفاده از روش‌هایی که موجب افزایش روایی پیش‌بین ابزارهای اندازه‌گیری x_i می‌شود به افزایش اساسی در بروندادها، افزایش ارزش پولی بروندادها و افزایش مهارت‌های مرتبط با شغل در گنجینه نیروی انسانی منجر می‌شود (اشمیت و همکاران، ۱۹۷۹، ص ۶۱۱). چون پذیرش دانشجویان دوره دکتری همانند به کارگیری افراد برای مشاغل خاص است، باید گروه‌های آموزشی ابتدا دوره دکتری را تجزیه و تحلیل و سپس ملاک $(Y=y_1, y_2, \dots, y_n)$ عملکرد موفق و سنجش‌های y_i آن را به‌درستی تعریف کنند. این تجزیه و تحلیل‌ها، هدف‌ها و ویژگی‌های دوره دکتری را توصیف می‌کند و مراحل، الزامات یا آزمون‌هایی که دانشجو برای به پایان رساندن موفقیت‌آمیز دوره باید بگذراند را مشخص می‌کند. سپس بر این اساس می‌توان شایستگی‌های مطلوب در پذیرش دانشجویان دوره دکتری و الزامات ورود به این دوره‌ها را مشخص کرد (فرجی

ده‌سرخ، آراسته، بغدادچی، نوه‌ابراهیم، عبدالهی، ۱۳۹۴، ص ۹۳). به همین دلیل، بررسی روایی این برنامه سنجش و پذیرش گام به گام در این الگوی تئوریک توضیح داده شده است. این الگو در نمودار (۱) می‌تواند راهنمایی برای برنامه‌های سنجش و پذیرش در ایران باشد.

پیشینه پژوهش

در کشورهای مختلف، تعریف واژه کیفیت (عملکرد موفقیت‌آمیز Y) بسته به شرایط نظام آموزشی آنها ممکن است متفاوت باشد. به‌عنوان مثال، در پژوهشی که اتیه (۱۹۹۹) روی ۳۱ مؤسسه پذیرش دانشجوی دکتری در آمریکا اجرا کرد متغیر کیفی Y را با عنوان «توان ماندگاری x_i یک داوطلب پذیرفته شده در دوره دکتری و به پایان رساندن موفقیت‌آمیز دوره با ملاک ارائه رساله دکتری x_i عملیاتی کرد». او برای این متغیر کیفی و تأثیر آنها بر ماندگاری دانشجویان پذیرفته شده، چهار گروه از متغیرهای کمی x_i را بررسی کرد. این پیش‌بینی‌کننده‌ها ($X = x_1, x_2, \dots, x_n$) عبارت بودند از: الف) ویژگی‌های مربوط به مؤسسه‌ای که برنامه دکتری ارائه می‌کند؛ به این معنی که هرچه دانشکده یا دانشگاه مورد نظر در تولید علم و ثبت اختراعات و اکتشافات در رده بالاتری قرار داشته باشد دوره دکتری نیز سخت‌تر و به دنبال آن ماندگاری و فارغ‌التحصیلی در آن مشکل‌تر است؛ ب) نمره‌های آزمون‌های پیشرفت تحصیلی و خرده‌آزمون‌های GRE؛ ج) میزان و تداوم کمک‌های مالی به دانشجو؛ د) ویژگی‌های جمعیت‌شناختی دانشجویان مثل جنسیت، قومیت، دانشجوی خارجی یا داخلی بودن. در پژوهش نظرسنجی والپل، بارتن، کانی و جکنتال^۱ (۲۰۰۲) بر اساس دیدگاه ۱۰۰ نفر از اعضای هیئت علمی و سرپرستان ۲۰ دانشکده تحصیلات تکمیلی، «کیفیت Y » را مواردی همچون دوره‌های گذرانده شده x_i و رتبه‌های به دست آمده x_i از سوی دانشجو و همین‌طور حسن انتخاب دانشکده کارشناسی و ارشد x_i ، استانداردهایی که دانشجو به دست آورده x_i ، کیفیت اعضای هیئت علمی دانشکده‌هایی که فرد در آن تحصیل کرده x_i و سختی برنامه درسی x_i در رشته تخصصی دانشجو را تعریف کرده‌اند. اما مصاحبه‌شوندگان که عمدتاً اعضای هیئت علمی و سرپرست گروه‌ها بوده‌اند با این امر موافق بودند که در نهایت این اطلاعات ذهنی و اغلب دسترس‌ناپذیر هستند. همچنین گفته‌اند که آنها نمره‌های آزمون عمومی

1. Walpole, Burton, Kanyi & Jackenthal

GRE را به‌عنوان یک استاندارد عینی x_i می‌نگرند که می‌توان آن را برای قضاوت ذهنی در رابطه با کیفیت Y دوره کارشناسی که فرد گذرانده است، به کار برد (وال پُل و همکاران، ۲۰۰۲). با این حال، بزرگی اندازه ارتباط بین نمره‌های x_i آزمون‌هایی مثل GRE و یک بُعد از عملکرد y_i در سطح دکتری وابسته به این است که این عملکرد y_i چگونه تعریف شده و چگونه اندازه‌گیری شده است و به‌طور ویژه، توانایی‌های شناختی^۱ تا چه اندازه این عملکرد y_i را تعیین می‌کنند (کانسل و همکاران، ۲۰۰۱، ص ۱۶۳). همچنین، کانسل و همکاران (۲۰۰۱) در پژوهشی فراتحلیل که روی ۸ معیار x_i پذیرش مختلف در ۱۷۵۳ نمونه متفاوت گردآوری کرده بودند، دریافتند که نمره‌های x_i آزمون GRE و نمره‌های x_i معدل دوره کارشناسی در پیش‌بینی نمره‌های معدل y_i دوره دکتری، نمره‌های y_i آزمون جامع، میزان ارجاع y_i به مقاله‌های دانشجویان دوره دکتری و رتبه‌بندی y_i دانشجویان به وسیله اعضای هیئت علمی به‌طور کل پیش‌بینی‌کننده‌های روایی هستند. همچنین، نمرات x_i آزمون GRE با میزان دستیابی به درجه دکتری y_i و مولد بودن y_i تحقیقاتی دانشجوی همبستگی دارد.

در مطالعه‌ای کیفی که فرجی ده‌سرخ، آراسته، نوه‌ابراهیم و عبدالهی (۱۳۹۴) روی شناسایی ملاک‌های y_i پذیرش دانشجویان دوره دکتری اجرا کردند، دریافتند که بهترین روش بهبود انتخاب دانشجویان تحصیلات تکمیلی و همچنین اولین الزام در طراحی نظام پذیرش، «تعریف و تدوین ملاک‌های عملکرد خوب یا موفقیت‌آمیز ($Y=y_1, y_2, \dots, y_n$) است» و سپس نحوه اندازه‌گیری این عملکرد خوب است. هنگامی که موفقیت و سنج‌ها و مقیاس‌های y_i آن به‌درستی تعریف نشوند نمی‌توان پیش‌بینی‌کننده‌ها x_i را شناسایی و بررسی کرد. آنها با تحلیل محتوای مصاحبه از ۱۴ عضو هیئت علمی و سیاست‌گذاران حوزه پذیرش به این نتیجه رسیدند که از دید صاحب‌نظران سه مقوله کلی تسلط حرفه‌ای (با ۳ مقوله فرعی)، عملکرد حرفه‌ای (با ۴ مقوله فرعی) و شایستگی فردی (با ۶ مقوله فرعی) را می‌توان به‌عنوان ملاک‌های y_i موفقیت در نظر گرفت و نهایتاً داوطلب موفق Y در دوره دکتری کسی است که بر حوزه تخصصی خود تسلط دارد، عملکرد حرفه‌ای چشمگیری در دوره دکتری از خود نشان می‌دهد و از نظر شخصیتی ویژگی‌های برجسته‌ای دارد. در مطالعه تطبیقی دیگر فرجی ده‌سرخ و همکاران (۱۳۹۵) روی پیش‌بینی‌کننده‌های X موفقیت، نتیجه گرفتند در پیش‌بینی

1. cognitive ability

عملکرد موفقیت‌آمیز Y افراد داوطلب دکتری تکیه صرف بر متغیرهای شناختی X مثل سطح دانش فرد x_i ، آزمون‌ها x_i ، نمره‌های پیشرفت تحصیلی و معدل x_i توان پیش‌بینی عملکرد موفقیت‌آمیز Y را کاهش می‌دهد و ممکن است پذیرش سوگیری داشته باشد. بنابراین، علاوه بر متغیرهای شناختی آکادمیک یا سنتی باید متغیرهای غیرشناختی x_i همچون انگیزه، شخصیت، علاقه، اخلاقیات، توانایی‌های نوشتاری، توانایی برقراری رابطه و تعهد به‌عنوان متغیرهای پیش‌بین x_i نیز مورد توجه قرار بگیرد. به عبارتی برای بهبود پیش‌بینی باید از مدل تک‌بعدی انتخاب داوطلبان مبتنی بر فاکتورهای شناختی فاصله گرفته و به سمت مدل‌های چندبعدی انتخاب که از عوامل غیرشناختی نیز استفاده می‌کنند، حرکت کنیم. فرجی ده‌سرخ و همکاران (۱۳۹۵، ص ۲۶) به تفاوت مهم دیگری در رابطه با سیستم سنجش و پذیرش ایران با دانشگاه‌های برتر اشاره کرده‌اند و آن مشخص نبودن پایایی و روایی سنج‌ها و معیارهای x_i پذیرش است. آنها بیان کرده‌اند که روایی و پایایی مقیاس‌ها x_i و الزاماتی که دانشگاه‌های برتر استفاده می‌کنند در مطالعات مختلف کم‌ویش تأیید شده است. اما شواهدی دال بر روایی و پایایی آزمون کتبی (آزمون استعداد تحصیلی، آزمون زبان انگلیسی و آزمون اختصاصی) در ایران در دسترس نیست؛ بنابراین نمی‌توان ادعا کرد که آزمون کتبی x_i می‌تواند عملکرد تحصیلی y_i داوطلب پذیرفته شده در دوره دکتری را پیش‌بینی کند.

فرایند و مراحل روایی‌یابی آزمون x_i

الف) روایی محتوایی

پیش از بررسی روایی ملاکی باید شواهد مربوط به روایی محتوایی یک آزمون x_i توسط سازندگان آن بررسی شود. در ایران، شواهدی که برای روایی محتوایی در نظر گرفته می‌شوند بیشتر معیارهای ضمنی^۱ هستند که بر اساس سرفصل تصویب شده برای درس مربوطه، تهیه‌کنندگان آزمون تلاش می‌کنند تا نمونه‌ای از پرسش‌ها را انتخاب کنند که محتوای سرفصل‌ها را پوشش دهد. دستیابی به شواهد «روایی محتوایی» هنگامی لازم است که تهیه‌کنندگان آزمون می‌خواهند چگونگی عملکرد یک فرد را در جهانی X از شرایطی^۲ بسنجند که آزمون x_i مورد نظر تنها نماینده یا نمونه‌ای از آن شرایط است

1. implied criterion

2. universe of situations

(ویلینگهام، ۱۹۷۶، ص ۷). بنابراین، آزمون x_i یک نمونه از رفتار انسان را در موقعیت‌های استاندارد به دست می‌دهد. از مفروضه زیربنایی نمونه‌گیری این است که ویژگی‌ها و مشخصه‌های رفتارهای مورد نظر در مجموعه کل رفتارها $(X=x_1, x_2, \dots, x_n)$ ، در این نمونه x_i بازتاب یافته است (فوگارتی^۱، ۲۰۰۸). اگر هر نمونه x_j دیگری از رفتار انسان «در آن زمینه» به تصادف انتخاب شود باید هم با نمونه پیشین x_i و هم با مجموعه کل رفتارهای مورد نظر همبستگی $r_{x_i x_j}$ داشته باشد و این همبستگی $r_{x_i x_j}$ اساس و پایه مطالعات روایی‌یابی در برنامه‌های سنجش و پذیرش است.

ب) روایی ملاکی

سوی نقش ضریب همبستگی $r_{x_i x_j}$ در روایی‌یابی محتوای یک آزمون، این ضریب در همه زمینه‌های آزمون‌گیری‌های روان‌شناختی به‌ویژه در روایی‌یابی ملاکی اهمیت اساسی دارد. این امر از این حقیقت ناشی می‌شود که نمونه رفتار به دست آمده با اجرای آزمون x_i ، اغلب همان رفتاری که می‌خواهیم اندازه بگیریم y_i نیست اما با آن ارتباط قوی دارد. بنابراین انتخاب آزمونی x_i که یک داوطلب لازم است آن را بگذراند معمولاً دربرگیرنده تکالیف و پرسش‌هایی است که ممکن است فرد در شغل و حرفه مربوطه هرگز با آن روبه‌رو نشود. اما آنچه درباره آزمون‌های پذیرش می‌دانیم این است که عملکرد در این آزمون‌ها x_i با عملکرد واقعی در شغل y_i مربوطه همبستگی دارد. کسی که نمره بالایی روی یک آزمون به دست می‌آورد احتمالاً روی آن شغل نیز به خوبی عمل می‌کند و برعکس (فوگارتی، ۲۰۰۸). در برخی از تصمیم‌گیری‌ها برای انتخاب افراد هم نمره آزمون x_i و هم نمره عملکرد شغلی^۲ y_i برای یک گروه از کارکنان و مهندسان وجود دارد. بنابراین همبستگی $r_{x_i y_i}$ بین این دو را می‌توان با عنوان روایی ملاکی-هم‌زمان محاسبه کرد. از آن زمان به بعد رفتاری که از کل محتوای X به وسیله آزمون x_i نمونه‌گیری شده، $X = x_i$ ، به‌عنوان پیش‌بینی‌کننده نمونه‌ای از عملکرد $Y = y_i$ واقعی در شغل مربوطه است و از آن می‌توان برای پیش‌بینی موفقیت هم گروه آزمون‌شوندگان فعلی که تازه قرار است انتخاب شوند و هم گروه‌هایی از افراد که قرار است بعدها انتخاب شوند نیز استفاده کرد. این کار مستلزم گردآوری شواهد و مدارکی است که نشان دهد آزمون x_i فعلی بدون اینکه نمره‌های مربوط به عملکرد $Y = y_i$

1. Fogarty

2. job performance

واقعی وجود داشته باشد، اما می‌تواند آن را پیش‌بینی کند. این فرایند اساس روایی‌یابی ملاکی-پیش‌بین یک آزمون x_i است؛ افزون بر این، بررسی شواهد و مدارک مربوط به روایی ملاکی-پیش‌بین یک آزمون هنگامی ضرورت دارد که از روی نمره x_i یک فرد بخواهیم در مورد برخی دیگر از متغیرهایی که ملاک $Y = y_i$ نامیده می‌شوند-مثل موفقیت تحصیلی یا موفقیت حرفه‌ای در آینده-استنباط کنیم. روایی ملاکی-پیش‌بین حد و مرزی را نشان می‌دهد که بر اساس عملکرد آزمون‌شونده در آزمون‌های x_i فعلی، سطح عملکرد آینده فرد در ملاک y_i را بتوان پیش‌بینی کرد (ویلینگهام، ۱۹۷۶).

ج) فرایند روایی‌یابی ملاکی پیش‌بین یک آزمون x_i

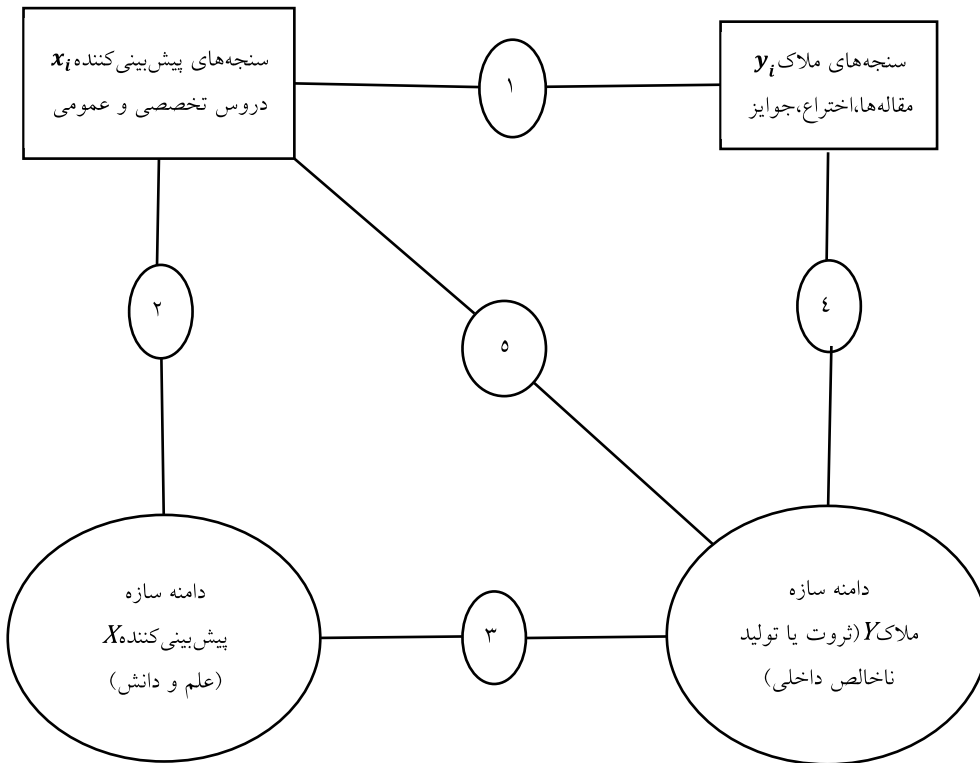
در بسیاری از کاربردهای آزمون در موقعیت‌های استخدام یا پذیرش، استنباط اساسی که از روی نمره‌های آزمون x_i می‌شود، پیش‌بینی است: کمیته پذیرش دانشکده می‌خواهد از روی نتایج آزمون x_i برخی از رفتارهای شغلی مهندسان یا دستاوردها و موفقیت‌های شغلی y_i آنها را در آینده پیش‌بینی کند. در فرایند روایی‌یابی یک ابزار اندازه‌گیری مثل آزمون‌ها x_i اغلب باید با تحلیل شغل (در اینجا مهندسی) آغاز شود. یعنی تحلیل همان عناصری که ملاک Y را تشکیل می‌دهند. همان‌طور که گفته شد، در این فرایند اطلاعات راجع به وظایف و تکالیف مربوط به شغل مهندسی، دانش، مهارت‌ها و توانایی‌ها، مسئولیت‌های مهندسان و هر اطلاعات اساسی دیگری که برای عملکرد موفق y_i در یک جایگاه شغلی یا سازمانی لازم باشد، گردآوری می‌شود (روگلبرگ^۱، ۲۰۰۷). این اطلاعات، مبنایی تجربی فراهم می‌کند برای مواردی همچون الف) فرضیه‌سازی در رابطه با ویژگی‌های افرادی که پیش‌بینی می‌شود در آن شغل کارکرد y_i موفقیت‌آمیزی خواهند داشت؛ ب) برای تهیه استانداردها و معیارهای سنجش عملکرد y_i شغلی؛ ج) برای تفسیر آنچه منظور از عملکرد موفق y_i شغلی در شغل مورد نظر است (AERA, APA & NCME, 2014). تهیه‌کنندگان آزمون‌هایی که برای پذیرش کارکنان یا پذیرش داوطلبان آزمون می‌سازند برای تعیین اینکه کدام بخش از توانایی‌ها، علائق و شخصیت x_i با موفقیت y_i فرد ارتباط دارد باید جستجوی کاملی در پیشینه انجام دهند. هنگامی که اینها مشخص شدند، فرایند تهیه آزمون آغاز می‌شود (فوگارتی، ۲۰۰۸).

1. Rogelberg

توجیه تئوریک الگوی روابط بین پیش‌بینی‌کننده‌ها X و ملاک Y

اگرچه در همه فعالیت‌هایی که برای روایی‌یابی X صورت می‌گیرد، راهبردهای متفاوتی برای گردآوری مدارک ممکن است به کار رود اما استنباطی که از روی نمره‌های x_i مدنظر است و از آن پشتیبانی می‌شود این است که نمرات آزمون $X = x_i$ را می‌توان برای پیش‌بینی رفتارهای شغلی در آینده $Y = y_i$ به کار برد. در موقعیت‌های پذیرش و استخدام فرایند روایی‌یابی این آزمون‌ها x_i در برگزیده گردآوری و ارزیابی مدارک و شواهدی است که نشان دهد آیا این استنباط درست و پابرجا است یا به چالش کشیده شده است. اما در بیشتر مواقع، راهبردهای روایی‌یابی آزمون یا هر ابزار اندازه‌گیری x_i می‌تواند در برگزیده روابط تجربی مثل رگرسیون بین متغیر پیش‌بینی‌کننده x_i با ملاک y_i نباشد. برای پشتیبانی از این استنباط روش‌های گوناگون روایی‌یابی را می‌توان به کار برد. بنابراین برای ایجاد این استنباط آینده‌نگر^۱ باید به دو بعد توجه کرد: یکی پیش‌بینی‌کننده‌ها ($X = x_1, x_2, \dots, x_n$) و دیگری ملاک‌ها ($Y = y_1, y_2, \dots, y_n$) که عبارت است از رفتار شغلی یا دستاورد یا پیامد مورد نظر. به‌طور عملی راه‌های بسیاری برای ارتباط دادن این حوزه‌ها هست، از جمله پنج رابطه‌ای که در نمودار (۱) نشان داده شده است. در این نمودار بین دامنه سازه پیش‌بینی‌کننده X و سنجه پیش‌بینی‌کننده x_i که از آن نمونه‌گیری شده و همچنین بین دامنه سازه ملاک Y و سنجه ملاک y_i که از آن نمونه‌گیری شده، تفاوت وجود دارد. دامنه پیش‌بینی‌کننده X بر اساس مشخص کردن مجموعه‌ای از رفتارها، دانش‌ها، مهارت‌ها، توانایی‌ها، صفات، علایق و ارزش‌ها تعریف می‌شود که عمدتاً پنهان هستند اما می‌توان هر یک را با عنوان برجسب سازه خاصی به صورت سنجه ($X = x_1, x_2, \dots, x_n$) عملیاتی کرد و به آزمون گذاشت (به‌عنوان مثال، مهارت عددی و محاسباتی، تحلیل منطقی، استنباط کلامی، سرعت تایپ کردن، تعهد یا وظیفه‌شناسی).

^۱. predictive inference



نمودار (۱) روابط جایگزین^۱ بین سنجه پیش‌بینی کننده و سنجه ملاک (برگرفته از استانداردهایی برای آزمون‌گیری آموزشی و روان‌شناختی، ۲۰۱۴، ص ۱۷۲، با تغییرات)

همچنین، دامنه سازه ملاک Y ، مجموعه‌ای از رفتارهای شغلی یا دستاوردهای کاری را مشخص می‌کند که بسته به شرایط اینها هم عمدتاً پنهان هستند اما می‌توان هر کدام را با عنوان برجسب یا سازه خاصی تعریف کرد و برای آنها سنجه‌های چندگانه‌ای به صورت $(Y=y_1, y_2, \dots, y_n)$ عملیاتی و سپس اندازه‌گیری کرد (مثل عملکرد در تکالیف اساسی شغل مهندسی همچون تولید و نوآوری، گسترش مرزهای دانش، کار تیمی، حضور به موقع، میزان فروش یا عملکرد کلی در شغل). سنجه‌های پیش‌بینی کننده x_i و سنجه‌های ملاک y_i باید به گونه‌ای نمونه‌گیری شوند تا بتوان جایگاه یا موقعیت فرد را

¹. alternative

در آن خصیصه‌های پنهانی که در دامنه سازه پیش‌بینی‌کننده X یا در دامنه سازه ملاک Y است را به‌خوبی ارزیابی کرد.

روش‌های گوناگونی برای بررسی این روابط وجود دارد به‌عنوان مثال، این نمودار تعداد استنباط‌هایی را نشان می‌دهد که می‌توان در مورد روابطی که عموماً در روایی‌یابی یک آزمون یا ابزار اندازه‌گیری x_i مورد نظر و مطلوب هستند، انجام داد (AERA, APA & NCME, 2014).

الف) نخستین ارتباط (که با برچسب ۱ در نمودار است) بین نمره سنجه پیش‌بینی‌کننده x_i و نمره سنجه ملاک y_i است. این استنباط را می‌توان از طریق آزمون تجربی روابط بین سنجه‌های این دو یعنی «رگرسیون» به آزمون گذاشت.

ب) دومین و چهارمین رابطه (که با ۲ و ۴ برچسب زده شده) از لحاظ مفهومی یکی هستند. در هر دو، رابطه‌ای که بین یک سنجه x_i یا y_i عملیاتی شده^۱ است را با دامنه آن سازه X یا Y مورد نظر به آزمون می‌گذارند؛ این کار با بررسی «ارتباط موضوعی داشتن» سنجه‌های پیش‌بینی‌کننده x_i و سنجه‌های ملاک y_i ، با دامنه سازه خودشان صورت می‌گیرد و بر اساس تحلیل‌های منطقی، قضاوت کارشناسان، یا با شیوه‌های آماری که در آن میزان ارتباط آنها با سنجه‌های همگرا و واگرایی-که از لحاظ مفهومی مشابه یا متفاوت با سنجه مورد نظر هستند-بررسی می‌شود.

ج) رابطه ۳ دربرگیرنده ارتباط بین دامنه سازه پیش‌بینی‌کننده X و دامنه سازه ملاک Y است. این «رابطه استنباطی»^۲ است و بر اساس تحلیل‌های تئوریک و منطقی بنا می‌شود. این کار عموماً با بررسی سیستماتیک و منظم از محتوای شغل و قضاوت کارشناسان طراحی و رسم می‌شود به گونه‌ای که ویژگی‌ها و مشخصه‌های افراد^۳ که به وسیله ابزارهای اندازه‌گیری مثل آزمون‌ها ($X = x_i$) عملیاتی شده، به عملکرد شغلی موفقیت‌آمیز آنها-که با مشاهده و ثبت رفتار آنها ($Y = y_i$) در دوران کاری است-پیوند داده می‌شود. به‌عنوان مثال، نمی‌توان از یک مؤسسه که در آن افراد متخصص آموزش زبان انگلیسی هستند، انتظار داشت نرم‌افزارهای آموزشی تولید کنند. یا در سطح ملی می‌توان این کار را بر اساس ارزیابی سیستماتیک و منظم از محتوای سازه علم و دانش X که در دانشگاه‌ها و حتی نظام متوسطه آموزش داده می‌شود و ارتباط آن با دامنه سازه

1. operational measure

2. inferred linkage

3. individual characteristic

ثروت Y و تولید ناخالص داخلی^۱ (GDP) انجام داد. به گونه‌ای که بتوان با مدارک و شواهد، ویژگی‌ها و مشخصه‌های سازه علم و دانش را به ساختار تولید ثروت پیوند داد. به‌عنوان مثال، بررسی محتوای آموزشی در رشته‌های مختلف و قابلیت آنها در تبدیل شدن به کالاها و خدمات تکنولوژیک.

د) رابطه ۵ ارتباط مستقیم از سنجه‌های پیش‌بینی‌کننده x_i به دامنه سازه ملاک Y را به آزمون می‌گذارد. این کار با بررسی شواهد «روایی محتوای» آزمون ارزیابی می‌شود. در برخی از رشته‌ها، سنجه‌های پیش‌بینی‌کننده x_i را می‌توان به‌عنوان نمونه‌ای از دامنه سازه ملاک Y مورد نظر به‌خوبی طراحی کرد؛ براین اساس در روایی‌یابی بر پایه رابطه ۵، هم‌ریختی^۲ بین این سنجه x_i و دامنه سازه Y می‌تواند شواهد و مدارک سراسری را برای رابطه ۵ فراهم کند. ایجاد رابطه ۵ به این شکل تا حد زیادی نمایانگر رویکردی است که در استانداردهای یادگیری یا همان استانداردهای محتوایی از آن به‌عنوان «شواهد روایی بر اساس محتوای آزمون» اشاره می‌شود. آزمون‌هایی که در آنها داوطلبان برای شغل نجات غریق عملیات نجات انجام می‌دهند، یا در آن داوطلبان برای شغل تاپیست، یک متن را تایپ و ویرایش می‌کنند یا برای استخدام رانندگان متبحر آزمون شهری رانندگی گرفته می‌شود، نمونه‌هایی هستند از محتوای آزمون‌هایی که مبنایی را برای روایی‌یابی بر اساس رابطه ۵ فراهم می‌کنند. به عبارتی، ارتباط مستقیم بین سنجه پیش‌بینی‌کننده x_i و دامنه سازه ملاک Y را می‌توان به‌خوبی مشاهده کرد (AERA, APA, & NCME, 2014).

به‌طور خلاصه و به‌عنوان «یک قاعده کلی در روایی‌یابی»، پیش‌نیاز اولیه برای به‌کاربردن یک آزمون یا یک سنجه پیش‌بینی‌کننده x_i برای انتخاب افراد این است که بتوان استنباطات مربوط به روابط^۳ بین این سنجه پیش‌بینی‌کننده x_i و دامنه سازه ملاک Y یعنی رابطه ۵ را ایجاد کرد. در همه فرایندهای روایی‌یابی باید شواهد و مدارکی تهیه شوند که نشان دهد پیش‌بینی‌کننده‌های x_i با دامنه سازه ملاک Y ارتباط دارد. همان‌طور که این نمودار نشان می‌دهد راهبردهای چندگانه‌ای برای ایجاد این رابطه مهم وجود دارد. اما این روابط از سه حالت خارج نیستند؛

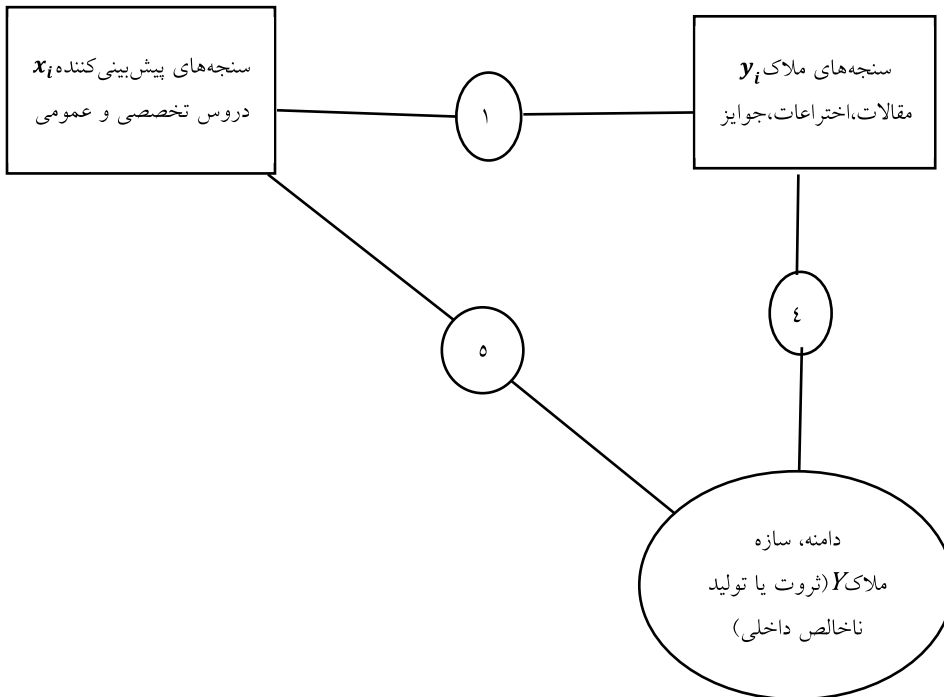
1. Gross Domestic Product

2. isomorphism

3. inferences concerning the linkage

راهبرد یکم: مستقیم از طریق رابطه ۵ است؛ هنگامی که آزمون x_i به گونه ای طراحی شده که نمونه‌ای باشد از دامنه سازه ملاک Y ، شواهد مربوط به روایی را می‌توان از طریق رابطه ۵ مستقیم ایجاد کرد.

راهبرد دوم: ارتباط یک سنجه پیش‌بینی‌کننده x_i با دامنه سازه ملاک Y مبتنی بر روابط ۱ و ۴: یعنی ایجاد کردن یک رابطه تجربی^۱ بین سنجه‌های پیش‌بینی‌کننده x_i و سنجه‌های ملاک y_i هم‌زمان فراهم سازی شواهدی از کفایت^۲ که در آن نشان داده شود سنجه ملاک y_i نمایانگر دامنه سازه ملاک Y است (به نمودار ۲ نگاه کنید).



نمودار (۲) شواهد روایی برای بررسی ارتباط سنجه x_i با دامنه ملاک Y (رابطه ۵) بر مبنای رابطه

۱ و ۴

1. empirical link

2. adequacy

شواهد مربوط به ارتباط سنجه ملاک y_i با دامنه سازه ملاک Y عموماً بر اساس تحلیل مشاغل صورت می‌گیرد، اگرچه در برخی موارد ارتباط بین سنجه y_i و دامنه Y چنان سراسر است که بدون تحلیل شغل هم مرتبط بودن^۱ آنها آشکار است. به‌عنوان مثال، هنگامی که سنجه مورد نظر حضور و غیاب یا گردش مالی y_i به‌عنوان شاخص افزایش تولید و درآمد Y در سطح شرکت‌ها و مؤسسات بررسی می‌شود. در سطح ملی می‌توان ارتباط تولیدات علمی یا جوایز بین‌المللی را به‌عنوان سنجه ملاک y_i ، با تحلیل و ارزیابی ساختار تولید ناخالص داخلی به‌عنوان دامنه سازه ملاک Y بررسی کرد. در برنامه‌های توسعه‌ای، کمیت و کیفیت مقاله‌های علمی y_i حاصل از نوآوری یا دریافت جوایز بین‌المللی را به‌عنوان شاخص‌های علوم و تکنولوژی Y ارزیابی می‌کنند.

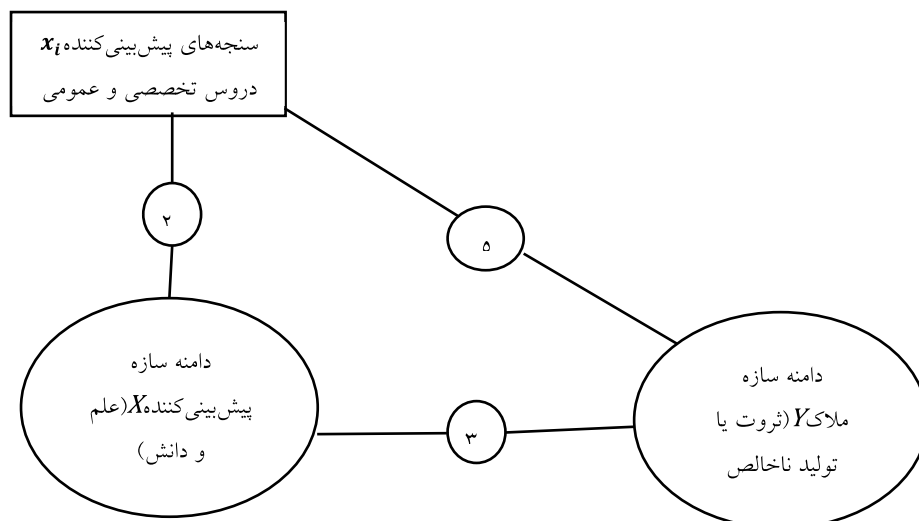
راهبرد سوم: برای مرتبط ساختن نمره‌های پیش‌بینی‌کننده x_i با دامنه سازه ملاک Y دربرگیرنده جفت رابطه ۲ و ۳ است. این رابطه متمرکز است بر فراهم کردن شواهدی که نشان دهد سنجه پیش‌بینی‌کننده x_i برای نشان دادن دامنه سازه پیش‌بینی‌کننده X کفایت دارد. هم‌زمان شواهد و مدارکی تهیه شود که نشان دهد بین دامنه سازه پیش‌بینی‌کننده X و دامنه سازه ملاک Y رابطه وجود دارد (ارتباط ۳).

اگر بخواهیم این رابطه اخیر را بسیار ساده و با مثال تبیین کنیم با توجه به اینکه زبان انگلیسی زبان تولید علم (مقاله y_i) است فرض بگیریم آزمون زبان انگلیسی x_i سنجه‌ای است که از دامنه علم و دانش ($X=x_1, x_2, \dots, x_n$) به گونه‌ای نمونه‌گیری شده که نماینده دامنه و محتوای علم و دانش $X=x_i$ مورد نظر است و به‌خوبی این کفایت را دارند که توانمندی افراد داوطلب در زمینه علم و دانش انگلیسی را اندازه بگیرد (پیدا کردن شواهد روایی محتوایی رابطه ۲). سپس شواهد و مدارکی برای رابطه (۳) تهیه شود که نشان دهند نمونه آزمون x_i دانش زبان انگلیسی با بخشی از شاخص‌های علوم و تکنولوژی y_i (مثل انتشارات علمی) که در دامنه سازه ملاک ($Y=y_1, y_2, \dots, y_n$) قرار دارد مرتبط است و در چارچوب شاخص‌های علوم و تکنولوژی، برون‌داد علمی^۲ یا تولید مقاله y_i از سنجه‌هایی است که می‌تواند به‌عنوان شاخص تحقیق و توسعه تعیین‌کننده افزایش تولید ناخالص داخلی به‌عنوان دامنه سازه ملاک Y باشند (برای مطالعه شاخص‌های علوم و فناوری نگاه کنید به گودین^۳، ۲۰۰۱، ۲۱).

1. relevance

2. scientific output measure

3. Godin



نمودار (۳) شواهد روایی برای بررسی ارتباط سنج x_i با دامنه ملاک Y (رابطه ۵) بر مبنای رابطه ۲ و ۳

بنابراین در تهیه شواهد و مدارک مربوط به «روایی‌یابی سازه» آزمون زبان انگلیسی باید به‌طور خلاصه این سه گام برداشته شود: نخست شواهدی مربوط به اینکه پرسش‌های آزمون x_i نمونه‌ای باکفایت از دامنه دانش زبان X است و آن را به‌خوبی پوشش داده و اندازه می‌گیرد (روایی محتوایی). دوم، با افزایش مهارت داوطلبان در زبان انگلیسی x_i احتمالاً انتشارات علمی y_i آنها نیز افزایش می‌یابد (روایی ملاکی). سوم، بر اساس شواهد و مدارک در این چارچوب تئوریک، آشکار شده که این دامنه از دانش زبان انگلیسی $X = x_i$ با رشد علوم و تکنولوژی $Y = y_i$ که تعیین‌کننده تولید ناخالص داخلی است در ارتباط (۳) است.

همان‌طور که پیشتر بیان شد، تنها یک راه مستقیم معمول برای ایجاد این روابط وجود ندارد. این روابط می‌تواند دربرگیرنده رسته‌ای^۱ از شواهد و مدارک باشند که در مفهوم سازی و تصویرسازی^۲ اولیه از فرایند روایی‌یابی با عنوان «روایی سازه» طبقه‌بندی شده‌اند. ترکیب چند رسته از شواهد و مدارک مثل قضاوت کارشناسان در رابطه با

1. line

2. conceptualization

خصیصه‌ها و صفاتی که موفقیت شغلی را پیش‌بینی می‌کنند، استنباط‌هایی^۱ که از تحلیل شاخص‌های بسیار مهم عملکرد شغلی اثربخش یا بی‌اثر به دست می‌آید، روش‌های مصاحبه و مشاهده و تحلیل‌های آماری می‌توانند از این استنباط پشتیبانی کنند که آیا دامنه سازه‌های پیش‌بینی‌کننده X با دامنه سازه ملاک Y ارتباط دارد؟ یا در سطح ملی، آیا محتوای علوم و تکنولوژی X باعث افزایش تولید و نوآوری Y و به دنبال آن افزایش GDP می‌شوند یا نه؟ (روایی سازه). سپس سنجه‌های x_i مربوط به سازه‌های پیش‌بینی‌کننده X را می‌توان انتخاب یا تهیه کرد و ارتباط بین آنها را با کمک رسته‌های متفاوتی از مدارک و شواهد ایجاد کرد که در بالا بحث شد.

کدام نوع از روایی اهمیت بیشتری دارد؟

همه انواع روش‌های روایی‌یابی اهمیت زیادی دارند اما بسته به هدف آزمون x_i ممکن است اهمیت یک نوع روایی‌یابی از دیگری بیشتر شود. اگر هدف، انتخاب کارکنان خوب باشد روایی‌یابی پیش‌بین آزمون از همه مهم‌تر است. اگر هدف، ارزیابی عملکرد افراد در یک دوره یا یک برنامه آموزشی است، روایی‌یابی محتوای آزمون بسیار اهمیت دارد (فوگارتی، ۲۰۰۸). روایی محتوای آزمون‌هایی بیشتر ارزیابی می‌شوند که مهارت یا دانش^۲ را در یک زمینه خاص اندازه می‌گیرند (ویلینگهام، ۱۹۷۶). اگر هدف تهیه مدل‌هایی باشد که بخواهد نشان دهد بر اساس یک تئوری، سازه‌های مختلف چگونه با یکدیگر ارتباط دارند، روایی سازه اهمیت دارد. در برخی موقعیت‌ها هم لازم است که نشان داده شود همه انواع روایی برآورده شده است (فوگارتی، ۲۰۰۸). بنابراین برای ارزیابی رابطه بین یک سنجه پیش‌بینی‌کننده x_i و دامنه سازه ملاک Y مورد نظر، تنها یک روش برتر وجود ندارد بلکه، منابع چندگانه‌ای از داده‌ها و رسته‌ای از شواهد و مدارک را می‌توان فراهم کرد. به هر حال، در بررسی روایی و ایجاد این رابطه‌ها کاربر آزمون باید ویژگی‌ها و مشخصه‌های موقعیت آزمون‌گیری را در نظر داشته باشد و برای فرض ارتباط بین یک سنجه یا پیش‌بینی‌کننده x_i و دامنه سازه ملاک Y قضاوت حرفه‌ای را به کارگیرد (AERA, APA & NCME, 2014).

1. inferences

2. skill or knowledge

انتخاب ملاک $Y=y_i$ و پیش‌بینی‌کننده‌ها $X=x_i$ در پذیرش

انتخاب درست یک ملاک y_i قابل‌سنجش^۱ و پایا برای مطالعات روایی‌یابی ملاکی یک آزمون x_i نقش اساسی دارد. همان‌طور که گفته شد، در فرایند روایی‌یابی ملاکی یک آزمون x_i ، کارایی آن را در پیش‌بینی عملکرد موفقیت‌آمیز y_i افراد در یک فعالیت معین بررسی می‌کنند. برای این کار همبستگی نمره‌های آزمون x_i را با برخی از ملاک‌های واقعی عملکرد y_i برآورد می‌کنند. در هر برنامه آموزشی، یک سؤال مهم همیشگی این است که چگونه باید ملاک عملکرد موفقیت‌آمیز y_i را تعریف کرد. این امر به‌عنوان یکی از مباحث مهم در بررسی‌های مربوط به روایی آزمون‌ها x_i با عنوان «چالش ملاک^۲» همیشه مطرح بوده است (هارتنت و ویلینگهام، ۱۹۸۰ ص ۱). ملاک عملکرد موفق y_i در برنامه‌های سنجش و پذیرش باید سنجه‌هایی باشند که عمدتاً شاخص موفقیت دانشگاهی است و در پذیرش داوطلب در نظر گرفته می‌شوند، به‌گونه‌ای که بتوان همه دانشجویان را بر اساس آن مقایسه کرد (شاو و همکاران^۳، ۲۰۱۶). همچنین با هدف‌هایی که در آیین‌نامه‌های پذیرش آمده (مثل، گسترش مرزهای دانش، نوآوری و تولید علم) ارتباط داشته باشد. در فرایند روایی‌یابی یک پیش‌بینی‌کننده مثل آزمون x_i ، اختلاف نظرهایی هم وجود دارد در رابطه با اینکه آیا ملاک y_i باید یک سنجه کوتاه‌مدت باشد مثل میزان موفقیت در ترم اول دانشگاه یا یک ملاک میان‌مدت مثل آزمون جامع یا معدل در پایان دوره تحصیلی دکتری یا یک ملاک درازمدت باشد مثل موفقیت شغلی، گسترش مرزهای دانش و نوآوری.

در زنجیره پیش‌بینی‌ها، سنجه‌هایی x_i هم وجود دارند که به‌عنوان سنجه‌های میانی^۴ شناخته می‌شوند (مثل نمره‌های آزمون جامع یا نمره‌های پیشرفت تحصیلی خاص). سنجه‌های پیش‌بینی‌کننده x_i می‌توانند به پیش‌بینی موفقیت y_i یک فرد در تحصیلات تکمیلی کمک کنند و هم‌زمان یک ارتباط منطقی با یک ملاک y_i در درازمدت^۵ نیز داشته باشند. یک گام مهم در بررسی روابط منطقی بین سنجه‌ها در زنجیره پیش‌بینی‌ها، تهیه یک مفهوم بهتر از معیار x_i میانی موفقیت و شناسایی این

1. quantifiable

2. Criterion problem

3. Shaw et al

4. intermediate criteria

5. long-range criteria

معیارهای میانی در تحصیلات تکمیلی است. این سنجه‌های میانی اگر به صورت آزمون باشند باید روایی سازه^۱ داشته باشند؛ یعنی بتواند بیانگر یک توانایی باشد که اهمیت آن در تعیین یکی از چند نوع مختلف موفقیت فرد در درازمدت، قابل اثبات باشد (ویلینگهام، ۱۹۷۶، ص ۱۷-۱۴).

بر اساس نتایج بررسی‌های زوویک^۲ (۱۹۹۱، ۲۰۰۶، ۲۰۰۷) معیارهای عمومی x_i پذیرش، که در برگزیده نمره‌های آزمون و نمره عملکرد دوره‌های پیشین تحصیلی داوطلب است (مثل معدل کارشناسی)، می‌تواند پیش‌بینی‌کننده‌های خوبی برای موفقیت y_i سال‌های اولیه در تحصیلات تکمیلی (مثلاً ارشد) باشند. اما انتخاب دانشجویان برای دوره دکتر مبتنی است بر شناسایی کسانی که توان خود را برای موفقیت در سال‌های بعدی دکتری نیز نشان بدهند. به همین دلیل نمره‌ها و «دستاوردهای موفقیت‌آمیز» در کارشناسی ارشد می‌توانند به‌عنوان یک معیار میانی در نظر گرفته شوند. عملکرد در دانشکده کارشناسی ارشد باید در فعالیت‌های بعدی مرتبط با شغل هم بتواند موفقیت را پیش‌بینی کند. اما از آنجایی که موفقیت نهایی در شغل بر بسیاری از شرایط و مهارت‌های توسعه‌یافته مبتنی است و لزوماً بر عملکرد در دوره کارشناسی ارشد یا پیش‌بینی‌کننده‌های اولیه وابسته نیست، بنابراین در نظر گرفتن آزمون پذیرش دکتری به‌عنوان پیش‌بینی‌کننده موفقیت شغلی آینده زیاد مناسب نیست (ویلینگهام، ۱۹۷۶). یک آزمون تنها نمونه‌ای کوچک از مهارت‌های داوطلب است و گستره آزمون‌ها محدود است. از سوی دیگر، موفقیت شغلی به عوامل زیادی وابسته است و تنها به صلاحیت علمی وابسته نیست. از این رو می‌توان گفت که آزمون‌ها x_i «گه گاه» می‌توانند موفقیت شغلی y_i را پیش‌بینی کنند (زوویک، ۲۰۰۶). برخی پژوهش‌ها نشان داده‌اند حتی شناسایی اولیه و تعریف پیش‌بینی‌کننده‌های x_i موفقیت یک فرد در مراحل دوره دکتری، به‌ویژه موفقیت در مرحله رساله هم بسیار سخت است و پذیرش یک فرد علاوه بر پیش‌بینی‌کننده‌های x_i چندگانه نیازمند بررسی، شور و مشورت کمیته‌های پذیرش است (لوو و مولر^۳، ۲۰۱۴). از لحاظ آماری نیز ملاک y_i باید بر اساس مجموعه‌ای از داده‌های معنی‌دار روی یک پیوستار که دست کم چهارارزشی باشد قابل توصیف باشد. بهتر است از ملاک‌هایی که به صورت دوارزشی مثل رد یا قبول، موفق ناموفق یا ۰ و ۱ نمره‌گذاری

1. construct validity

2. Zwick

3. Louw & Muller

شده‌اند، خودداری شود. اگر عملکرد شغلی y_i به صورت موفق و ناموفق گروه‌بندی شود اطلاعات مربوط به سطوح پیوستار عملکرد در هر گروه از دست می‌رود، به‌عنوان مثال همه کسانی که در گروه موفق قرار می‌گیرند فرض بر این است که ارزش برابری دارند چه آنها که عالی بوده‌اند و چه آنها که به سختی از نمره برش عبور کرده‌اند (اشمیت و همکاران^۱، ۱۹۷۹، ص ۶۱۰).

کارکرد آزمون GRE به‌عنوان معیار x_i در پیش‌بینی عملکرد موفق Y

آزمون عمومی GRE یکی از معیارهای x_i مهمی بوده است که برای پیش‌بینی موفقیت در آینده y_i از آن بسیار استفاده شده است. این آزمون از سه بخش مختلف کمی، تحلیلی و کلامی تشکیل شده است. هر بخش آن، گونه‌ای از «توانایی شناختی عمومی^۲» سطح بالا را اندازه می‌گیرد (هانتز و هانتز^۳، ۱۹۸۴، ص ۸۱)؛ به‌عنوان مثال در بخش کمی که ترکیبی است از مسئله‌های کمیت‌های تجریدی^۴، مقایسه‌های کمی^۵ و تفسیر داده‌ها از داوطلبان می‌خواهند حل مسئله کنند. در بخش مقیاس تحلیلی آن که دربرگیرنده پرسش‌های استدلال تحلیلی^۶ و استدلال منطقی^۷ است، داوطلبان باید اطلاعات را ترکیب و یکپارچه^۸ کنند و روابط پیچیده بین اطلاعات پراکنده را حل و فصل کنند^۹ (آزمون‌های استعداد در ایران از روی این دو بخش شبیه‌سازی می‌شود با این تفاوت که بخش تحلیلی GRE انشایی است اما آزمون استعداد در ایران چهارگزینه‌ای است). بخش کلامی GRE نیز به صورت چهارگزینه‌ای است و دربرگیرنده قیاس^{۱۰}، تضاد^{۱۱}، تکمیل جمله و مسئله‌های خواندن و درک مطلب است (و شاید معادل آزمون‌های زبان انگلیسی در ایران باشد).

1. Schmidt et al

2. general cognitive ability

3. Hunter & Hunter

4. discrete quantitative

5. quantitative comparison

6. analytical reasoning

7. logical reasoning

8. synthesis

9. resolve

10. analogy

11. antonym

نتایج پژوهش کانسل و همکاران (۲۰۰۱) نشان داده است بین نمره‌های x_i بخش‌های عمومی آزمون GRE (مثل بخش کمی، تحلیلی و کلامی) با برخی عملکردهای دانشگاهی y_i درازمدت (مثل اجرای پروژه، تولید علم و خلق نوآوری) ارتباط زیادی است. به‌عنوان مثال، خواندن و خلاصه کردن یک متن^۱ که در خرده‌آزمون تحلیلی GRE است با برخی از عملکردهای دانشگاهی مثل نوشتن مقاله y_i همبستگی دارد. بنابراین سنجه‌های عملکرد موفقیت‌آمیزی y_i که برای انجام آن «بار شناختی بالایی» لازم است و نیاز به تفکر پیشرفته دارند را می‌توان با بخش عمومی GRE (بخش کمی، تحلیلی و کلامی) پیش‌بینی کرد. به‌طور کلی، هر چقدر کارکرد شغلی از پیچیدگی شناختی بالایی برخوردار باشد به همان نسبت برای انجام موفقیت‌آمیز آن کار، توانایی‌های شناختی عمومی پیچیده تری لازم است (همبستگی مثبت نمره‌های x_i بخش عمومی GRE و نمره‌های x_i آزمون‌های استعداد با عملکرد موفق y_i در مشاغل پیچیده). هر چقدر پیچیدگی شغل مورد نظر از لحاظ شناختی کاهش می‌یابد به همان نسبت برای انجام کار، توانایی‌های روانی-حرکتی^۳ افزایش می‌یابد.

در برخی رشته‌ها آزمون موضوعی یا تخصصی GRE نیز وجود دارد که برای پذیرش، نمره آن الزامی است. آزمون‌های موضوعی GRE مشابه آزمون‌های تخصصی در ایران است و دانش کسب شده مربوط به رشته تخصصی افراد (مثل زیست‌شناسی، شیمی یا روان‌شناسی و نظایر آن) را اندازه می‌گیرد. به شرط برابر بودن چیزهای دیگر، انتظار این است که دانشجویی که در آزمون‌های تخصصی x_i خوب بوده و با «دانش شغلی» بالایی وارد تحصیلات تکمیلی می‌شود کارکرد شغلی y_i بهتری نیز داشته باشد نسبت به کسی که در آزمون‌های تخصصی بد بوده و دانش شغلی کمتری دارد. دانشجویانی که دانش شغلی بالاتری دارند برای یکپارچه کردن دانش تخصصی مرتبط با رشته و افزایش یادگیری در چهارچوب بهتری قرار می‌گیرند، بدون هیچ آموزش بیشتری عملکرد بهتری هم در کلاس و هم در آزمون‌های جامع دارند، رساله دکتری بهتری می‌نویسند، به‌عنوان مربی یا همکار پژوهشی کارکرد بهتری دارند و پژوهش بهتری اجرا می‌کنند تا دانشجویانی با سطح دانش شغلی پایین‌تر. بسیاری از دانشگاه‌ها از آزمون جامع x_i برای ارزیابی و استنباط در رابطه با میزان دانش شغلی y_i دانشجویان

1. passage

2. cognitively loaded performance measures

3. psychomotor ability

تحصیلات تکمیلی نیز استفاده می‌کنند. به‌عنوان مثال، برای عملکرد حرفه‌ای موفق به‌عنوان روان‌شناس یا بیولوژیست، دانش روان‌شناسی یا بیولوژی لازم است؛ از این رو انتظار می‌رود بین نمره‌های آزمون موضوعی GRE و نمره‌های آزمون‌های تخصصی این رشته‌ها که دانش شغلی را اندازه می‌گیرند همبستگی‌های بزرگ‌تری باشد نسبت به نمره‌های آزمون عمومی GRE (کانسل و همکاران، ۲۰۰۱، ۱۶۴-۱۶۳).

در ادامه این نوشتار، شواهد روایی ملاکی آزمون‌های تخصصی x_i با نمره‌های چندگانه در بخش مصاحبه y_i بررسی می‌شود. این شواهد در سطح آماری است؛ بر اساس تئوری تعمیم‌پذیری، نمره‌های جهانی^۱ حاصل از چندین (۳ تا ۸) خرده‌آزمون تخصصی و نمره‌های جهانی حاصل از دو سنج ترکیبی مصاحبه برآورد شده است. رابطه این دو توزیع نرمال به صورت نمودارهای خطی برای چهار رشته مهندسی رسم شده است. این نمودارها نشان می‌دهند چگونه نمره ترکیبی چندین خرده‌آزمون تخصصی می‌تواند نمره ترکیبی حاصل از ملاک دوگانه مصاحبه را پیش‌بینی کند. آزمون‌های عمومی زبان و استعداد به دلیل همبستگی‌های کانونی صفر یا منفی با دیگر متغیرهای x_i و y_i قابل بررسی نبوده‌اند (به جدول ۴ نگاه کنید).

روش پژوهش

روش این پژوهش از نوع توصیفی-همبستگی است. جامعه آن ۵۰ رشته تخصصی دکتری در سال ۱۳۹۷ است که به صورت نمونه‌گیری تغییرات بیشینه^۲ (بورگ، گال و بورگ، ۱۹۴۲؛ ترجمه نصر و همکاران، ۱۳۸۶، ص ۳۹۱) تنها ۴ رشته انتخاب شد. دلیل نمونه‌گیری به این شیوه آن بود که این ۴ رشته اگر چه در گروه مهندسی بوده‌اند اما هم از لحاظ تعداد درس‌های تخصصی x_i و هم از لحاظ تعداد دانشگاه‌هایی که برنامه دکتری این رشته‌ها را ارائه می‌کرده‌اند، بسیار تنوع داشتند. درس‌های تخصصی آنها کمینه ۳ و بیشینه ۸ درس بود: رشته برق الکترونیک با ۳ درس تخصصی در ۳۰ دانشگاه، رشته هوش مصنوعی با ۳ درس تخصصی در ۱۵ دانشگاه، رشته برق قدرت با ۴ درس تخصصی در ۳۷ دانشگاه و رشته فناوری اطلاعات با ۸ درس تخصصی در ۶ دانشگاه

1. Universe score

2. Maximum variation sampling

دانشجوی دکتری می‌پذیرفته‌اند. هر یک از ۵۰ رشته دیگر خصوصیات شبیه به این چهار رشته داشته‌اند و نتایج دیگر رشته‌ها هم در بازه نتایج این چهار رشته قرار می‌گیرد. آزمون اختصاصی x_i همه آنها تقریباً دارای ۴۵ سؤال بوده که بسته به رشته‌های مختلف دارای سه تا هشت خرده‌آزمون تفکیک شده و وزن اسمی یا ضریب این ۴۵ سؤال در برنامه سنجش و پذیرش ۰/۳ بوده است.

آزمون‌های عمومی آنها دربرگیرنده یک آزمون ۳۰ سؤالی زبان انگلیسی و یک آزمون ۳۰ سؤالی استعداد تحصیلی است که جمعاً وزن اسمی برابر با ۰/۲ دارند (به جدول ۲ و ۳ نگاه کنید).

برنامه مصاحبه از داوطلبان بر اساس یک فهرست که عموماً ۱۱ سنجه دارد در کمیته پذیرش دانشکده‌ها انجام گرفته است. داوران، این ۱۱ سنجه را تبدیل به دو نمره y_i و z_i می‌کنند؛ نمره بخش نخست z_i که وزنی برابر با ۰/۳ دارد، از ملاک‌های علمی و سنجش علمی تشکیل شده است (مثل مصاحبه تخصصی و نظر استادان کمیته پذیرش که عمدتاً به‌طور ذهنی سنجه‌های غیرشناختی و شخصیت فرد را نمره‌گذاری می‌کنند و آزمون شفاهی یا کتبی داخلی [احتمالاً]). نمره بخش دوم y_i که وزنی برابر با ۰/۲ دارد، از معیارهای مربوط به سوابق آموزشی و پژوهشی و فناوری تشکیل شده است (مثل مقاله‌ها، ثبت اختراعات، شرکت در جشنواره‌های علمی معتبر، تألیف یا ترجمه کتاب، امتیاز پایان‌نامه کارشناسی ارشد، پژوهش‌های اجرا شده و مانند آن). همچنین معدل کارشناسی و ارشد با توجه به کیفیت دانشگاه محل تحصیل داوطلب در این بخش به وسیله کمیته داوری، وزن‌دهی و محاسبه می‌شود.

نمره‌های هر درس تخصصی و درس‌های عمومی (زبان و استعداد) که ضریب تصحیح حدس از آن کم شده همچنین دو نمره مربوط به مصاحبه، از سازمان سنجش دریافت شده است. درس‌های تخصصی $(X=x_1, x_2, \dots, x_n)$ به‌عنوان معیارهای پیش‌بینی‌کننده و دو نمره مصاحبه $(Y=y_1, y_2)$ ملاک دوگانه این پژوهش هستند. پس از وزن‌دهی آنها، همبستگی‌های کانونی خرده‌آزمون‌ها x_i با سنجه‌های دوگانه مصاحبه y_i همچنین امید نمره‌های مشاهده شده هر فرد روی خرده‌آزمون‌های چندگانه و نمره‌های مصاحبه برآورد شده و تبدیل به دو نمره جهانی $E(X)$ و $E(Y)$ شده‌اند. داده‌ها با نرم‌افزار mGENOVA و بر اساس یک طرح تک‌رویه‌ای $p^* \times i^1$ چندمتغیری تحلیل شده‌اند

¹. single facet $p^* \times i^1$ design

(برنان^۱، ۲۰۰۱، ۲۰۰۳)؛ به عبارتی، داده‌های ثانویه مربوط به رتبه‌های هر فرد روی درس‌های تخصصی (۳ تا ۸ نمره = x_i) و مصاحبه (۲ نمره = y_i) پس از وزن‌دهی تحلیل شده و به «توزیع نرمال دو متغیری» $E(X)$ و $E(Y)$ تبدیل شده‌اند و فرض بر این است که نمره‌های آزمون‌های تخصصی x_i کفایت لازم برای پیش‌بینی نمره‌های سنجه‌های مصاحبه y_i (شاید به‌عنوان سنجه‌هایی از موفقیت در دوره دکتری یا سنجه‌ای درازمدت برای تولید علم و نوآوری) را دارد. در الگوی تحلیل مبتنی بر تئوری تعمیم‌پذیری می‌توان نمره‌های چندگانه x_i و y_i را بر اساس وزن‌های پیشین ترکیب کرد و هم‌زمان لایه‌های خطا را روی نمره‌های هر بخش و تعامل این خطاها را با یکدیگر مشخص کرد. از طرفی می‌توان نتایج تحلیل‌ها را به جهانی از نمره‌های آزمون‌ها و مصاحبه‌های مشابه و سال‌های پیش و پس از این تاریخ در همان رشته - اگر که نمره‌های آزمون‌ها و مصاحبه به همین صورت تهیه شده‌اند - تعمیم داد. هنگامی که لازم است فرضاً نمره ۸ خرده‌آزمون (با ضریب ۰/۳) به‌علاوه نمره مصاحبه علمی و سنجش علمی (با ضریب ۰/۳) و نمره سوابق آموزشی و پژوهشی و فناوری (با ضریب ۰/۲) با یکدیگر ترکیب شوند و یک یا چند نمره جهانی $E(X)$ و $E(Y)$ ساخته شود و پارامترهای هر خرده‌آزمون و نمره کل هر بخش و نمره کل نهایی بررسی شود به یک مدل و تئوری فراتر از تئوری کلاسیک نیاز هست تا بتوان همه جوانب این فرایند را بررسی کرد. تئوری کلاسیک از سرجمع ساده نمره‌های x_i و y_i استفاده می‌کند و توان وزن‌دهی و سپس تهیه امید نمره‌های مشاهده شده $E(X)$ و $E(Y)$ و ترکیب و پیش‌بینی نمره‌های چندگانه را ندارد. تئوری تعمیم‌پذیری هم بر اساس مقیاس نمره‌های مشاهده شده و هم بر اساس توزیع نرمال استاندارد شده $N(0, \sigma^2)$ نمره‌های جهانی ترکیبی را پیش‌بینی و برآورد می‌کند^۲.

نتایج و یافته‌ها

در جدول (۲) مواد آزمون نیمه‌متمرکز و مصاحبه و وزن اسمی هر یک ارائه شده است. دو سنجه معدل کارشناسی و کارشناسی ارشد همان‌طور که گفته شد در بخش سوابق آموزشی و پژوهشی و فناوری نمره‌گذاری می‌شوند و نه به‌عنوان دو سنجه جداگانه.

^۱. Brennan

^۲. برخی توضیحات اضافی بنا به درخواست داوران به مقاله اضافه شده است.

جدول (۲) سنجه‌های پذیرش داوطلبان دکتری سال ۱۳۹۷ و سهم % آنها در نمره کل ترکیبی

نمره کل (۱۰۰٪)	مرحله مصاحبه و بررسی سوابق داوطلبان توسط کمیته پذیرش دانشگاه (۵۰٪)		معدل		آزمون متمرکز توسط سازمان سنجش (۵۰٪)			مواد آزمون
	سوابق علمی و پژوهشی و فناوری (۶ سنجه)	مصاحبه علمی و سنجش علمی (۳ سنجه)	کارشناسی ارشد (۱ معیار)	کارشناسی (۱ معیار)	آزمون استعداد تحصیلی (۱)	آزمون زبان (۱)	آزمون درس‌های تخصصی (۳ تا ۸ درس)	
(۱/۰۰)	(۰/۲)	(۰/۳)	بخشی از سوابق آموزشی و پژوهشی و فناوری است. با توجه به هشت نوع دانشگاه فارغ‌التحصیلی وزن آن نامشخص است.		(۰/۱)	(۰/۱)	(۰/۳)	وزن اسمی یا ضرایب
					۰/۲ با هم وزن‌دهی می‌شوند			

در جدول (۳) آماره‌های توصیفی بخش‌های مختلف آزمون‌های تخصصی و عمومی و مصاحبه آمده است. همان‌طور که پیشتر گفته شد نمره‌های آزمون‌های تخصصی به‌عنوان پیش‌بینی‌کننده x_i و دو سنجه مصاحبه به‌عنوان ملاک y_i به کار رفته است. آزمون‌های استعداد و زبان به‌عنوان پیش‌بینی‌کننده x_i همبستگی‌های کانونی و ضرایب شبه پایایی مناسبی نداشته‌اند. همبستگی‌های کانونی چندمتغیره مبتنی بر نمره‌های مشاهده شده آنها در جدول (۴) آمده است. به‌عنوان مثال، این دو درس در رشته برق قدرت، ضریب تعمیم‌پذیری $E\hat{p}_8^2 = 0/086$ و ضریب اعتمادپذیری $\hat{\Phi} = 0/07$ دارد. به همین دلیل در تحلیل‌ها کنار گذاشته شده‌اند. همچنین نتایج روایی‌یابی آزمون‌های تخصصی در این پژوهش بر اساس مدل تیلور و راسل (۱۹۳۹) تفسیر شده است. اشمیت و همکاران (۱۹۷۹، ص ۶۱۰) معتقد بودند که این مدل تصویر بسیار واقع‌بینانه‌ای به دست می‌دهد از ارزش یک ابزار اندازه‌گیری مثل آزمون که برای انتخاب افراد به کار

می‌رود. این مدل نشان می‌دهد یک آزمون که ضریب روایی متوسطی دارد می‌تواند به‌طور اساسی درصد افراد موفق را بین کسانی که انتخاب شده‌اند، افزایش دهد حتی هنگامی که نسبت پذیرش^۱ (یا همان درصد پذیرفته‌شدگان نهایی) کم است. در مدل تیلور و راسل باید برای پذیرش و عدم پذیرش نقطه برش نمرات معین شود و سپس بر اساس این نقطه برش میزان خطاها بررسی شوند.

¹. selection ratio

جدول (۳) آماره های توصیفی و ضرایب روایی ملاکی-پیش‌بین و درصد پیش‌بینی‌های موفق و ناموفق

نام رشته (تعداد دانشگاه) و تعداد درس‌های تخصصی x_i	میانگین آزمون‌های تخصصی \bar{X}	میانگین آزمون‌های استعداد و زبان \bar{X}	میانگین ملاک مصاحبه \bar{Y}	ضریب همبستگی نرمال دومتغیره $r_{x_i y_j}$	نسبت پذیرش شده به مصاحبه شده یا نسبت پایه	نمره برش آزمون‌های تخصصی صدک ام p	٪ موفق
(ضریب)	۰/۳	۰/۲	۰/۵				۱/۰۰
برق قدرت (۳۷) دانشگاه) $x_i = 4$	۱۹۴۲/۱۲ ۲۰۷۲/۱۱ ۱۲۹۷/۲۳ ۳۳۸۴/۱۳	۱۷۱۵/۰۳ ۹۰۵/۹۲	۱۵۰۴/۲۸ ۵۵۳/۴۳	۰/۹۱	$\frac{۱۷۸}{۴۶۰} = 39$	$x_p = 61$	۰/۹۷ ۰/۰۳
برق الکترونیک (۳۰) دانشگاه) $x_i = 3$	۱۹۳۹/۱۴ ۲۱۳۶/۲۴ ۲۲۴۲/۴۷	۱۸۸۸/۸۵ ۱۲۲۹/۸۲	۱۷۰۶/۸۸ ۶۵۶/۹۵	۰/۹۲	$\frac{۱۳۵}{۳۸۱} = 35$	$x_p = 65$	۰/۷۸ ۰/۲۲
هوش مصنوعی (۱۵) دانشگاه) $x_i = 3$	۱۹۸۹/۷۲ ۲۴۳۱/۳۲ ۲۱۲۷/۴۶	۲۰۲۱/۶۰ ۱۸۲۸/۹۱	۱۲۷۰/۳۳ ۶۵۹/۴۳	۰/۸۹	$\frac{۶۸}{۲۱۱} = 32$	$x_p = 68$	۰/۹۱ ۰/۰۹
فناوری اطلاعات (۵ دانشگاه) $x_i = 8$	۱۱۲۳/۲۰ ۹۴۳/۱۱ ۱۴۷۶/۲۷ ۱۳۴۵/۴۴ ۷۸۶/۰۱ ۱۲۷۶/۲۵ ۳۶۸۵/۸۱ ۳۲۶۶/۷۴	۱۸۲۵/۴۸ ۱۴۴۲/۹۸	۱۲۵۰/۱۴ ۶۱۱/۸۵	۰/۲۲	$\frac{۲۲}{۷۰} = 31$	$x_p = 69$	۰/۶۳ ۰/۳۷

نکته: نمره برش بر مبنای صدک p مبتنی بر نسبت پذیرش (٪) قرار داده شده است.

در یک صفحه بُرداری تقاطع نمره برش برای آزمون‌های تخصصی و مصاحبه، «به‌طور نظری» صدک ۵۰ام توزیع نمره‌های جهانی استاندارد برابر با میانگین $z = 0$ است؛ یعنی محدوده‌ای که ۵۰٪ - ۵۰٪ در هر دو توزیع x_i و y_i در بالای آن است. «در عمل» و در این پژوهش نمره برش مصاحبه صدک ۵۰ام است و فرض ما بر این است که تقریباً همه افراد مصاحبه شوند شانس برابری برای پذیرش دارند، اما نمره برش آزمون‌های تخصصی برابر با نسبت پذیرش (٪) یا صدک p ام^۱ قرار داده شده است که n ٪ قبولی در بالای آن قرار دارد یعنی همان صدکی که قرار است افراد مساوی یا بالاتر از آن پذیرش شوند؛ به‌عنوان مثال، رشته فناوری اطلاعات به ۳۱٪ افراد مصاحبه‌شونده پذیرش نهایی می‌دهد. بنابراین نمره صدکی برابر با $x_p = 69$ آزمون اختصاصی "در توزیع داده‌ها"، که ۳۱٪ افراد در بالای آن قرار دارند، نمره برش آزمون‌های تخصصی x_i این رشته بوده است. بر اساس این نمره برش، کارایی آزمون‌های تخصصی در تفکیک افراد موفق و ناموفق روی ملاک مصاحبه y_i بررسی شده است.

در اینجا بار دیگر پرسش نخست این پژوهش تکرار می‌شود مبنی بر اینکه که «آیا نمره‌های آزمون‌های تخصصی به‌عنوان پیش‌بینی‌کننده x_i ارتباط قابل قبول و منطقی با سنج‌های مصاحبه به‌عنوان ملاکی که تا حدودی کیفیت عملکرد موفق یک داوطلب (نوآوری و تولید علم) در آینده را پیش‌بینی می‌کنند، دارند یا نه»؟

با توجه به ستون پنجم جدول (۳) ضریب همبستگی دومتغیره نمره جهانی آزمون‌های اختصاصی $E(X)$ با نمره جهانی مصاحبه $E(Y)$ برای رشته برق قدرت برابر با ۰/۹۱ صدم ($R^2 = 0/81$)، برای رشته برق الکترونیک ۰/۹۲ صدم ($R^2 = 0/85$) برای رشته هوش مصنوعی ۰/۸۹ صدم ($R^2 = 0/80$) و برای رشته فناوری اطلاعات برابر با ۰/۲۲ صدم ($R^2 = 0/05$) است. بنابراین تنها در رشته فناوری اطلاعات ضریب همبستگی آزمون تخصصی $E(X)$ با ملاک مصاحبه $E(Y)$ برابر با ۰/۲۲ صدم است و آزمون‌های تخصصی این رشته نتوانسته بیش از ۰/۰۵ صدم واریانس ملاک مصاحبه به‌عنوان ملاک موفقیت را برآورد کند.

برای پاسخ‌گویی به پرسش دوم این پژوهش مبنی بر اینکه «آیا این دو رویکرد متفاوت - آزمون نیمه‌متمرکز سازمان سنجش و مصاحبه کمیته پذیرش دانشگاه‌ها در

1. percentile

ارتباط با یکدیگر توانسته‌اند در نهایت باکیفیت‌ترین و موفق‌ترین افراد را مشخص و پذیرش کنند؟^۱

نمودارهایی برای توزیع پذیرش داوطلبان رسم شده که ارتباط نمره آزمون تخصصی با مصاحبه را نشان می‌دهند. این دو نمره بر اساس میانگین توزیع استاندارد $z = 0$ در تقاطع با یکدیگر چهار بخش را مشخص کرده‌اند؛

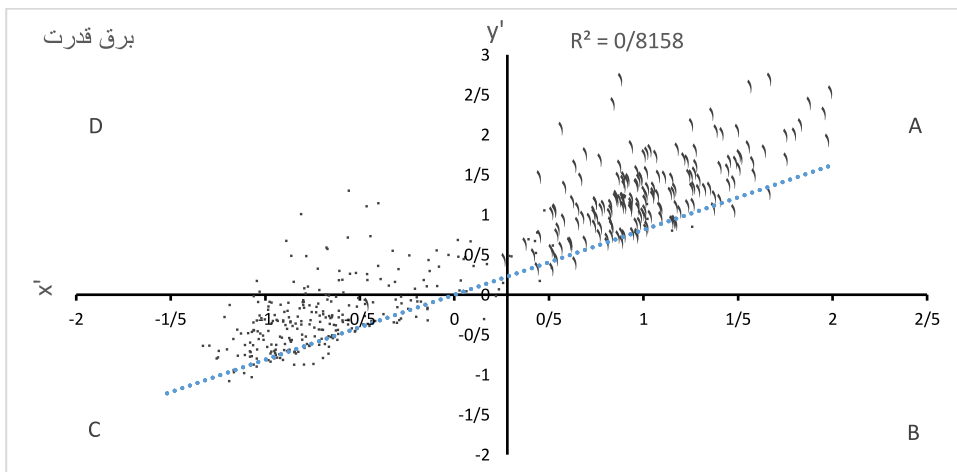
ناحیه A کسانی را نشان می‌دهد که هم در آزمون x_i و هم در مصاحبه y_i نمره بالاتر از میانگین توزیع جمعیت دریافت کرده‌اند. ناحیه B کسانی را نشان می‌دهد که در آزمون x_i نمره بالایی دریافت کرده‌اند ولی در مصاحبه y_i نمره پایین‌تر از میانگین کسانی را نشان می‌دهد که هم در آزمون x_i و هم در مصاحبه y_i نمره پایین‌تر از میانگین توزیع جمعیت دریافت کرده‌اند و ناحیه D کسانی را نشان می‌دهد که در مصاحبه y_i نمره‌های بالای اما در آزمون x_i نمره‌های پایین‌تر از توزیع جمعیت دریافت کرده‌اند.

هنگامی که نمره برش آزمون‌های تخصصی به صورت خط x' و نمره برش مصاحبه به صورت خط y' تعیین می‌شود، مفید بودن یک آزمون یا سنج x_i نسبت به دیگر سنج‌ها در پیش‌بینی یک ملاک y_i معین را می‌توان با مقایسه نسبت $\frac{A}{A+B}$ هر آزمون ارائه کرد و کارایی آزمون یا سنج‌های x_i مختلف را در پیش‌بینی ملاک y_i با یکدیگر مقایسه کرد (تیلور و راسل، ۱۹۳۹).

فرض کنید نصف گروه یعنی ۵۰٪ مصاحبه شوندگان، کسانی هستند که می‌توان آنها را به‌عنوان رضایت‌بخش (موفق) در نظر بگیریم. به عبارت دیگر میزان پایه^۱ یعنی درصدی از داوطلبان که در مصاحبه y_i بدون استفاده از آزمون-موفق می‌شوند و محدوده $A + D$ برابر با ۵۰٪ از کل این محدوده است و بنابراین خط y' در میانگین $z = 0$ یا میانه توزیع نمره‌های ملاک y_i قرار داده شده است. برای روایی‌یابی آزمون تخصصی x_i برق‌گرایش قدرت که ضریب روایی آن $0/91 =$ است، موقعیت پذیرش به‌گونه‌ای است که کمیته پذیرش باید تنها ۳۹٪ از افرادی را بردارد که برای مصاحبه دعوت شده‌اند. نقطه برش روی خط x' به‌گونه‌ای قرار داده شده که محدوده $A+B$ در برگرفته نسبت پذیرش ۳۹ درصدی باشد. این ارزش برابر با آن مقدار صدکی است که ۳۹ درصد کل ناحیه بالای توزیع استاندارد را دربر می‌گیرد. نتایج نشان می‌دهند در چنین شرایطی برای رشته مهندسی برق-قدرت ۹۷٪ از کسانی که انتخاب شده‌اند، با

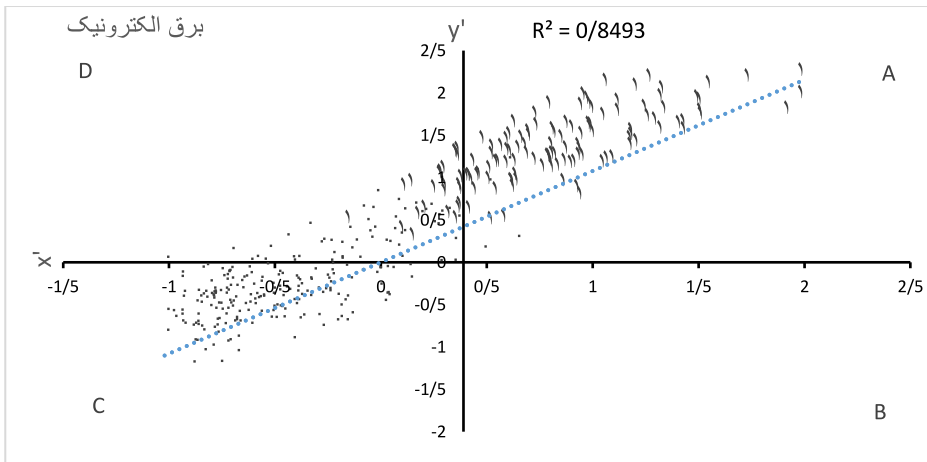
1. base ratio

عملکرد Y رضایت‌بخش و موفق طبقه‌بندی می‌شوند و آزمون‌های تخصصی تنها ۳٪ آنها را نتوانسته تشخیص دهد. این مقادیر برای رشته برق الکترونیک ۷۸٪ رضایت‌بخش و موفق و آزمون‌های تخصصی ۲۲٪ آنها را تشخیص نداده‌اند. برای رشته هوش مصنوعی ۹۱٪ موفق و رضایت‌بخش و آزمون‌های تخصصی ۹٪ آنها را تشخیص نداده و برای رشته فناوری اطلاعات ۶۳٪ موفق و رضایت‌بخش و آزمون‌های تخصصی آنها نتوانسته ۳۷٪ آنها را در پیش‌بینی بر اساس ملاک مصاحبه y_i موفق تشخیص دهد. در ادامه، نمودارهای همبستگی توزیع نرمال دو متغیره نمره‌های جهانی آزمون‌های تخصصی x_i و مصاحبه y_i این چهار رشته آمده است. خطوط راه راه که در نمودار آمده جایگاه افراد پذیرفته شده هستند با توجه به نمره‌های جهانی آنها بر اساس توزیع نرمال دو متغیری $E(X)$ و $E(Y)$. این نمودارها نشان می‌دهند که چگونه نمره‌های کمیته‌های پذیرش در مصاحبه و نمره‌های تهیه شده از سوی سازمان سنجش در تعامل با یکدیگر انتخاب نهایی داوطلبان را معین کرده است.

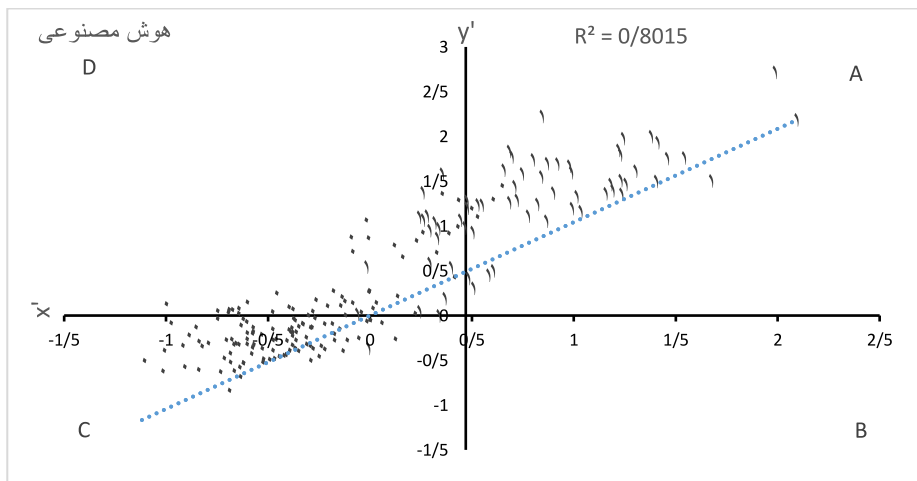


نمودار (۴) همبستگی توزیع نرمال دو متغیره نمره‌های جهانی آزمون‌های تخصصی و مصاحبه رشته

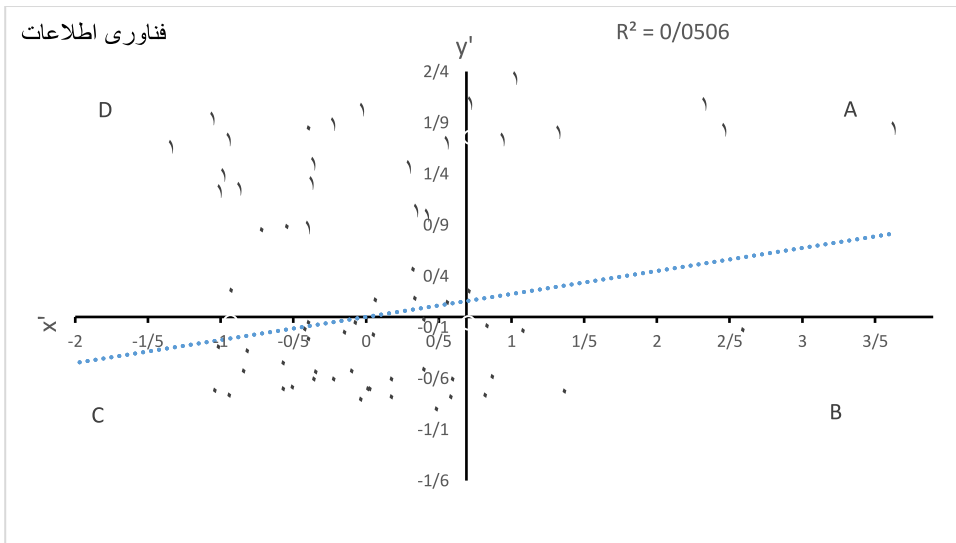
برق قدرت با نمره برش صدکی $x_p=61$ و $y_p=50$



نمودار (۵) همبستگی توزیع نرمال دومتغیره نمره‌های جهانی آزمون‌های تخصصی و مصاحبه رشته
برق الکترونیک با نمره برش صدکی $x_p=65$ و $y_p=50$



نمودار (۶) همبستگی توزیع نرمال دومتغیره نمره‌های جهانی آزمون‌های تخصصی و مصاحبه رشته
هوش مصنوعی با نمره برش صدکی $x_p=68$ و $y_p=50$



نمودار (۷) همبستگی توزیع نرمال دومتغیره نمره‌های جهانی آزمون‌های تخصصی و مصاحبه رشته فناوری اطلاعات با نمره برش صدکی $x_p=69$ و $y_p=50$

لازم به یادآوری است که ملاک عملکرد موفقیت‌آمیز Y رشته فناوری اطلاعات بر اساس نمره جهانی ترکیبی $E(X)$ آزمون‌ها قابل پیش‌بینی نبوده‌اند. به همین دلیل بر اساس ترکیب توزیع استاندارد نمره‌های مشاهده شده x روی ۸ آزمون تخصصی رتبه‌بندی شده‌اند. عملاً آزمون تخصصی آنها در انتخاب افراد بی‌فایده بوده و عمدتاً براساس نمره‌های مصاحبه پذیرفته شده‌اند. با توجه به بردار x' در نمودار ۷ حتی آنها که روی ۸ خرده‌آزمون تخصصی نمره استاندارد تقریباً $-1/5$ دارند پذیرفته شده‌اند. در شرایط یکسان، اگر ضریب روایی صفر یا یک باشد چه رخ می‌دهد؟ اگر ضریب روایی $r_{xy} = 0$ باشد آشکارا همان نسبت‌های پایه از افراد ناموفق و موفق را به دست می‌آوریم که آزمون را اصلاً به کار نمی‌بردیم. به‌عنوان مثال، در گروه فناوری اطلاعات ضریب روایی آن $r_{xy} = 0.22$ است و خط رگرسیون در نمودار ۷ تقریباً پنج درصد واریانس مشترک ($R^2 = 0/05$) پیش‌بینی ملاک y از روی x_i را نشان می‌دهد. بر این اساس، احتمالاً از ۲۲ نفری که در نهایت پذیرش شده‌اند به طور تصادفی ۳۱٪ (یعنی ۶ تا ۷ نفر) موفق بودند و ۶۹٪ (یعنی ۱۵ نفر) آنها ناموفق. به این معنی که کمترین روایی

ممکن، به انتخاب افرادی منجر می‌شود که تنها ۳۱٪ از ۲۲ نفر آنها به‌عنوان موفق طبقه‌بندی می‌شوند.

اگر ضریب روایی $r_{xy} = 1$ بود، در این شرایط همه ۲۲ نفر انتخاب شده موفق طبقه‌بندی می‌شدند. به عبارتی بالاترین روایی ممکن، به انتخاب افرادی منجر می‌شود که ۱۰۰٪ یا همه ۲۲ نفر آنها به‌عنوان موفق در بخش A نمودار ۷ طبقه‌بندی می‌شدند. در واقع مقدار $\frac{A}{A+B}$ هنگامی به کار می‌رود که برای انتخاب افراد علاوه بر سنجه‌های y_i از آزمون نیز استفاده شده و قرار است این آزمون x_i روایی‌یابی شود. در الگوی تیلور و راسل (۱۹۳۹) از ارزش دیگری نیز استفاده می‌شود که عبارت است از ناحیه $\frac{A+D}{A+B+C+D}$. رشته فناوری اطلاعات از این ناحیه استفاده کرده است. این مقدار تنها ارزش‌های مبتنی بر ملاک y_i را در بر می‌گیرد بدون استفاده از آزمون x_i افرادی که در این محدوده قرار می‌گیرند به‌عنوان نسبت مصاحبه‌شوندگانی است که در ملاک y_i به‌عنوان رضایت‌بخش در نظر گرفته شده‌اند^۱. یعنی نسبت گروهی که در حال حاضر بیشترین نمره را در ملاک مصاحبه y_i به دست آورده‌اند و آزمون x_i توان پیش‌بینی ندارد. نمره‌های آزمون‌های تخصصی این رشته به دلیل تعداد زیاد درس‌های تخصصی (۸ خرده‌آزمون) کارایی چندانی نداشته است. ضرایب تعمیم‌پذیری آن کم و بسیاری از داوطلبان در این آزمون‌های تخصصی نمره منفی دریافت کرده‌اند. ضریب تعمیم‌پذیری (شبه پایایی) درس‌های تخصصی آن $0/90 -$ درس‌های عمومی $0/26 -$ و مصاحبه y_i برابر با $0/60$ است. ضریب تعمیم‌پذیری نمره جهانی ترکیبی این رشته $0/697 -$ است که برای یک برنامه سنجش، ضریب پایایی منفی مناسب نیست (به جدول ۴ نگاه کنید). این تحلیل‌ها نشان می‌دهند در رشته فناوری اطلاعات با اینکه ۸ خرده‌آزمون تخصصی داشته‌اند کمیته پذیرش آنها عملاً از ملاک مصاحبه y_i برای پذیرش استفاده کرده است. با توجه به جدول (۴) در رشته مهندسی برق قدرت، همبستگی کانونی نمره چهار خرده‌آزمون اختصاصی x_i با دو نمره مصاحبه y_i برابر با $0/41$ و ضریب همبستگی کانونی نمره دو آزمون عمومی (زبان و استعداد) با دو نمره مصاحبه برابر با $0/05$ است. در دیگر رشته‌ها نیز همبستگی کانونی نمره آزمون عمومی (زبان و استعداد) با دیگر سنجه‌ها مناسب نبود. همچنین ضرایب شبه پایایی آنها یا منفی یا صفر بوده‌اند. در واقع ضریب همبستگی چندمتغیره این سنجه‌ها که بر اساس ماتریس همبستگی کانونی و بر

1. proportion of employees considered satisfactory

اساس تحلیل چندمتغیری با طرح تک‌رویه‌ای $p \times i^\circ$ محاسبه شده، بسیار کمتر از ارزش‌های محاسبه شده در نمودارهاست. چون ضرایب همبستگی نرمال دومتغیره r_{xy} مبتنی بر ترکیب نمره‌های جهانی خرده‌آزمون‌های اختصاصی مختلف $E(X)$ و مصاحبه $E(Y)$ مبنای ترسیم نمودارها بوده است.

جدول (۴) ضرایب همبستگی کانونی آزمون‌ها و ضرایب شبه پایایی

ضریب اعتمادپذیری $\hat{\Phi}$	ضریب تعمیم‌پذیری $E\hat{p}_\phi^2$	مصاحبه $y_i = 2$	آزمون عمومی (زبان و استعداد) $x_j = 2$	آزمون تخصصی x_i	
۰/۱۹	۰/۲۱۹			۱	آزمون تخصصی
۰/۰۷	۰/۰۸۶		۱	۰/۰۵	آزمون عمومی (زبان و استعداد)
۰/۳۷۷	۰/۵۴۶	۱	۰/۲۶	۰/۴۱	مصاحبه
۰/۳۵	۰/۴۰				ضریب کل نمره جهانی ترکیبی
۰/۳۱	۰/۳۱			۱	آزمون تخصصی
-۰/۰۷۷	-۰/۰۸۵		۱	۰/۱۷	آزمون عمومی (زبان و استعداد)
۰/۳۸	۰/۵۸	۱	۰/۲۲	۰/۳۰۷	مصاحبه
۰/۴۳	۰/۴۶				ضریب کل نمره جهانی ترکیبی
۰/۲۵۶	۰/۲۵۹			۱	آزمون تخصصی
-۰/۲۲	-۰/۲۲		۱	۰/۰۲۹	آزمون عمومی (زبان و استعداد)
۰/۴۴۷	۰/۵۳۶	۱	۰/۲۶۶	۰/۲۸	مصاحبه
۰/۳۱	۰/۳۲				ضریب کل نمره جهانی ترکیبی
-۰/۶۲۷	-۰/۹۰			۱	آزمون تخصصی
-۰/۲۵۹	-۰/۲۶		۱	-۰/۲۰	آزمون عمومی (زبان و استعداد)
۰/۴۸۷	۰/۶۰	۱	۰/۲۵	۰/۲۲	مصاحبه
-۰/۵۱۵	-۰/۶۹۷				ضریب کل نمره جهانی ترکیبی

بحث و نتیجه‌گیری

گرچه هدف اولیه برنامه‌های سنجش و پذیرش دکتری، پیش‌بینی درست عملکرد موفقیت‌آمیز حرفه‌ای و دستاوردهای بعدی افراد است، اما در پیش‌بینی این ملاک‌ها y_i تا اندازه‌ای محدودیت وجود دارد. برای بسیاری از جنبه‌های رفتاری انسان دستیابی به نمونه‌ای معرف از x_i یا y_i که بتواند آن سازه‌های X یا Y را به‌خوبی پوشش دهد معمولاً بسیار دشوار است. در پیش‌بینی عملکرد موفقیت‌آمیز y_i یک داوطلب، سازه‌هایی همانند شخصیت، هوش، انگیزش، علائق، ارزش‌ها و دانش را نمی‌توان مستقیماً مشاهده کرد و بی‌نهایت پیچیده هستند (فوگارتی، ۲۰۰۸، ص ۳). همچنین پیش‌بینی دقیق موفقیت y_i یک فرد پذیرفته شده بر اساس آزمون‌های چندگانه x_i و حتی با در دست داشتن ملاک‌های چندگانه y_i ، هدفی دست‌نیافتنی است. همان‌طور که نتایج نشان داده‌اند در رشته فناوری اطلاعات حتی با داشتن معیارهای چندگانه (۸ آزمون تخصصی) نمرات x_i آنها روایی پیش‌بین مناسبی نداشت. یعنی مدارک و شواهد تجربی حاصل از نتایج تحلیل‌ها و ضرایب همبستگی متغیرها نتوانست کفایت و مناسبت این آزمون‌های چندگانه x_i را در پیش‌بینی عملکرد موفقیت‌آمیز داوطلبان این رشته روی ملاک y_i مصاحبه تأیید کند. از دلایل آن می‌تواند دشواری زیاد سؤال‌ها، تعدد منابع و کتاب‌های درسی مختلف برای خواندن ۸ آزمون تخصصی متفاوت و همچنین تنوع داوطلبان از لحاظ سابقه تحصیلی و مدارک متفاوت کارشناسی ارشد و نظایر اینها باشد که نمره درس‌های اختصاصی قابلیت پیش‌بینی نداشته‌اند. یکی از راه‌های بهینه ساختن پارامترهای چنین آزمون‌هایی آن است که تنها یک منبع مشخص برای هر درس معین کنند. با این کار اگر چه داوطلبان باید ۸ درس متفاوت بخوانند اما تنها از منابع معینی استفاده می‌کنند. برای رشته‌هایی که تعداد درس‌های اختصاصی کمی در آزمون دارند (مثلاً ۳ درس) تعدد زیاد منابع درسی احتمالاً نمی‌تواند مشکل‌ساز باشد. همچنین نمره‌گذاری خرده‌آزمون‌ها با کم کردن ضریب تصحیح حدس مشکلات زیادی در برآورد پارامترهای آزمون و جامعه داوطلبان ایجاد می‌کند.

از سوی دیگر، عملکرد موفقیت‌آمیز در موقعیت‌های شغلی و دانشگاهی نه تنها تحت تأثیر بسیاری از عوامل سازمانی و فراسازمانی از جمله سرپرست، استاد، همکار، آموزش‌های رسمی و غیررسمی، طراحی نظام آموزشی و طراحی مشاغل، ساختارهای دانشگاهی و نظام‌های سازمانی، مسئولیت‌های خانوادگی و نظایر آن است. بلکه تحت تأثیر بسیاری از ویژگی‌های شخصیتی، از جمله دانش، مهارت‌ها، توانایی‌ها، شخصیت و

نگرش نسبت به تحصیل و کار و نظایر آن نیز قرار می‌گیرد. بنابراین تکیه بر یک خصیصه یا نمونه‌گیری از یک ویژگی x_i یا y_i ، تنها می‌تواند یک پیش‌بینی ناقص و ناکامل از عملکرد موفقیت‌آمیز به دست دهد. حتی سامانه‌های گزینشی پیچیده تنها می‌توانند بر مجموعه‌ای از سازه‌ها تمرکز کنند که فرض می‌شود برای آن شغل ضروری هستند تا اینکه بر همه خصیصه‌هایی که می‌توانند بر رفتار شغلی موفقیت‌آمیز اثر بگذارند.

در پیش‌بینی عملکرد y_i موفقیت‌آمیز، ارتباط داشتن سنجه‌ها، دقت سنجه‌ها، کاستی‌ها و پایایی سنجه‌ها ملاحظات دیگری هستند که باید رعایت شوند. به‌عنوان مثال **معدل کارشناسی و ارشد** از سنجه‌های x_i مهمی هستند که برای پیش‌بینی عملکرد y_i در تحصیلات تکمیلی به گستردگی از آنها استفاده می‌شود. اما این پیش‌بینی‌کننده x_i دارای معایب و مزایایی است. ایراد آن این است که استانداردهای آموزشی تحصیلات تکمیلی می‌توانند در دانشکده‌های مختلف و حتی بین اعضای هیئت علمی که یک دوره یا ترم مشابه را آموزش می‌دهند، بسیار متفاوت باشد (کانسل و همکاران، ۲۰۰۱)؛ به عبارتی ارزش یک معدل مثلاً ۱۵ به دست آمده از دانشگاه‌های گوناگون روزانه، با دیگر انواع دانشگاه‌ها (۸ نوع دانشگاه مختلف در ایران) تفاوت اساسی دارد. همچنین دامنه نمره‌های معدل کوتاه است و عملاً نمره معدل کارشناسی و کارشناسی ارشد به‌عنوان متغیر مستقل، پیش‌بینی‌کننده‌های مناسبی نیستند. این نمره‌ها به دلیل تورم در بالای توزیع نمره‌ها، چولگی چپ و محدود بودن دامنه آن در ترکیب با دیگر پیش‌بینی‌کننده‌ها، توانایی ایجاد واریانس واقعی بسیار کمی دارند و نمی‌توانند به خوبی افراد توانمند را از یکدیگر جدا کنند. خوبی آن این است که معدل کارشناسی و ارشد، کارکرد طولانی‌مدت، دانش کسب شده، تلاش، پایداری و توانایی را اندازه می‌گیرد. همچنین با موفقیت پس از مدرسه^۱ ارتباط دارد (اتیه، ۱۹۹۹، ص ۳۵ و ۳۶). معدل نقش خود را در غربال اولیه افراد، برای ارائه مدرکی دال بر صلاحیت داوطلب شدن ایفا می‌کنند. چون معدل‌های بالای ۱۲ در لیسانس و ۱۴ در فوق لیسانس شرط الزامی برای شرکت در برنامه سنجش و پذیرش دکتری است.

در رابطه با **خرده‌آزمون‌های استعداد^۲ و زبان** x_i این آزمون‌ها نیز همانند بخش کمی و کلامی GRE عمدتاً آزمون‌های توانایی‌های شناختی عمومی هستند تا اینکه سنجه‌هایی

1. post-school success

2. aptitude test

در رابطه با دانش شغلی x_i یا علایق، پایداری و انگیزه (کانسل و همکاران، ۲۰۰۱). شکل سؤال‌های این آزمون‌های استعداد دربرگیرنده گستره‌ای از سؤال‌هایی با هدف بیرون کشیدن پاسخ‌های آگاهی‌دهنده از داوطلبان است و عمدتاً پاسخ‌گویی به آنها ربطی به موضوعات آموزشی در دوران تحصیل ندارد (کرایتون، ۲۰۰۳). در ایران، آزمون استعداد عمدتاً به دلیل دشواری زیاد و وزن کم در مجموعه آزمون رها می‌شوند. پاسخ‌گویی به سؤال‌های آن بسیار وقت‌گیر است. همچنین نمره‌گذاری هر سؤال با کم کردن ضریب تصحیح حدس، این خرده‌آزمون‌ها را به مقیاسی کاملاً ناکارآمد تبدیل کرده است. دشواری بالای سؤال‌ها و روی آوردن آزمون‌شوندگان به حدس‌شناسی باعث شده این مقیاس بازتابی از خطای اندازه‌گیری باشد یا به دلیل رها کردن و بی‌پاسخ گذاشتن سؤال‌های آن آماره‌هایی مثل ضریب تمیز، پایایی و دشواری این مقیاس‌ها قابل بررسی نباشند. از سوی دیگر، تهیه و توسعه‌دهندگان این خرده‌آزمون عمدتاً نمی‌دانند در تهیه تکالیف این آزمون چه ریزمهارت‌ها یا فرایندهای شناختی را هدف‌گذاری کرده‌اند. ساختار چهارگزینه‌ای سؤال‌های آن در بهترین صورت آزمون‌شونده را ملزم به یادآوری یا بازشناسی یا استنباط ضمنی روی چهارگزینه می‌کند. همچنین بر خلاف گذشته در سال‌های اخیر، آزمون عمومی زبان انگلیسی به‌عنوان یک معیار الزامی برای پذیرش نیست. بی‌اهمیت شدن این آزمون باعث شده همه داوطلبان دوره دکتری نسبت به گذشته با پیشینه زبان انگلیسی ضعیف‌تری وارد این دوره‌ها بشوند. پیامد آن به دلیل ضعف در «زبان بین‌المللی تولید علم» احتمالاً کاهش توان تولید علم پذیرفته‌شدگان خواهد بود.

به هر حال، در فرایند روایی‌یابی آزمون‌های ورودی در سطح تحصیلات تکمیلی مسائل و مشکلات روش‌شناختی و مفهومی گوناگونی وجود دارد. از جمله اینکه (۱) برخی از خطاهای اندازه‌گیری همیشه هستند. حتی در مورد آزمون‌های x_i خوش ساخت و سنجه‌های ملاکی y_i که به‌خوبی عملیاتی شده‌اند (زوویک، ۲۰۰۶)؛ (۲) بنابر آنچه گفته شد، تنها با افزایش بهبود پیش‌بینی‌کننده‌ها x_i احتمالاً کارایی آنها در پیش‌بینی^۱ ملاک y_i افزایش نمی‌یابد؛ و (۳) امید اصلی برای افزایش اثربخشی پیش‌بینی‌کننده‌های x_i در تحصیلات تکمیلی مبتنی بر تعریف بهتر سنجه‌های موفقیت y_i است. به‌عنوان مثال، سنجه‌های پایا تر y_i آنهایی هستند که با توجه به هدف‌های آموزشی (مثل نوآوری

1. efficiency of prediction

و گسترش مرزهای دانش) تفاوت‌های آشکارتری بین داوطلبان ایجاد می‌کنند (ویلینگهام، ۱۹۷۶).

یکی از محدودیت‌هایی که در برنامه سنجش و پذیرش دوره دکتری در رابطه با روایی‌یابی آزمون‌های تخصصی x_i در پیش‌بینی $Y = y_i$ وجود داشته، واریانس (نامرتبب یا) نایسته به سازه^۱ سنجه‌های y_i ملاک است که در دو نمره مصاحبه خلاصه می‌شوند. یازده سنجه‌ای که ملاک y_i را تشکیل می‌دهند به دو نمره، یکی نمره مصاحبه علمی و سنجش علمی (با ۳ سنجه) و دیگری نمره سوابق آموزشی و پژوهشی و فناوری (با ۶ سنجه بعلاوه ۲ نمره معدل) تفکیک شده‌اند. اما مشخص نیست که نخست، خرده‌آزمون‌های تخصصی x_i دقیقاً با کدام بخش از سنجه‌های یازده‌گانه ملاک y_i واریانس مشترک دارد و می‌تواند به‌خوبی آن را پیش‌بینی کند. دوم این یازده سنجه ممکن است آلوده به متغیرهای دیگری (مثل سخت‌گیری یا آسان‌گیری مصاحبه‌گران) نیز باشد. بنابراین نمره‌ای که یک داوطلب روی ملاک y_i کسب می‌کند ممکن است دقیقاً بازتاب توانایی واقعی او روی سازه $Y = y_i$ مورد نظر (مثل موفقیت، تولید علم و نوآوری) نباشد. اگر یک سنجه مصاحبه y_i دربرگیرنده واریانس (نامرتبب یا) نایسته به سازه باشد نمره‌های به دست آمده از آنها سودار^۲ هستند (تامپسون^۳، ۲۰۱۸). بنابراین بسیار سخت است که بتوان استفاده از نمره‌های ملاک y_i را بدون در نظر گرفتن ملاحظات و محدودیت‌های آن برای تصمیم‌گیری توجیه کرد. محدودیت دیگر ناشی از ضریب جریمه حدس در نمره‌گذاری خرده‌آزمون‌های تخصصی بوده است. در این پژوهش برای به دست آوردن پارامترهای مناسب از آزمون‌ها به ناچار بسیاری از افراد با نمره‌های منفی از تحلیل‌ها کنار گذاشته شده‌اند. از دیگر محدودیت‌های این پژوهش نبود پیشینه مدون در رابطه با برنامه سنجش و پذیرش دانشگاه‌ها تا پیش از سال ۱۳۹۰ بوده است.

پیشنهادها

پذیرش دانشجوی دکتری، تعهدی سنگین برای سرپرست آینده‌نگر، کمیته پذیرش و همین‌طور برای دانشگاه است. بهتر است یک درخواست دکتری را از بسیاری جهات

1. construct irrelevant variance

2. biased

3. Thompson

دقیقاً مثل یک درخواست شغل در نظر بگیرند که در آن باید یک داوطلب آینده‌دار پذیرفته شود (OTAGO, ۲۰۱۱). بر این اساس، چهار پیشنهاد کلی مطرح می‌شود. نخست، با توجه به چند بعدی بودن عملکرد موفقیت‌آمیز Y در تحصیلات تکمیلی باید از چندین ملاک عملکرد موفقیت‌آمیز y_i استفاده شود تا تصویر کامل‌تری از موفقیت نیز ارائه شود (فرجی ده‌سرخ و همکاران، ۱۳۹۴، ص ۷۵-۷۴). در این رابطه، پیشنهاد می‌شود سنج‌های فرم مصاحبه به‌عنوان ملاک دست کم به سه بُعد $(Y=y_1, y_2, y_3)$ تفکیک شوند و هر مقوله آن بر اساس یک چارچوب منطقی به این سه بعد اختصاص داده شود. به‌ویژه سنج‌های مربوط به سوابق پژوهشی داوطلبان از سوی کمیته‌های پذیرش، با دقت بیشتری مطالعه و نمره‌گذاری شوند. پیشنهاد دوم، تعدیل و کاهش سطوح دشواری آزمون عمومی استعداد و هم‌زمان اهمیت بیشتر دادن به آزمون عمومی زبان پیش از پذیرش نهایی داوطلبان است. پیشنهاد سوم، نمره‌گذاری سؤال‌های آزمون‌ها به صورت ۰ به پاسخ غلط و ۱ به پاسخ درست و بدون جریمه حدس است. به دلیل تعداد کمتر شرکت‌کنندگان دکتری نسبت به آزمون‌های ارشد و سراسری و تعداد نسبتاً زیاد سؤال این آزمون‌ها امکان مشاهده نمره‌های همانند بسیار کم است و نمره‌گذاری سؤال‌های بدون جریمه کاملاً امکان‌پذیر است. از دیدگاه تجربی (احتمالات) می‌توان ثابت کرد هنگامی که به افراد اطلاع داده می‌شود حدس صحیح سه امتیاز و حدس غلط یک امتیاز منفی دارد عملاً به دلیل سختی سؤال‌ها و هم‌زمان سرنوشت‌ساز بودن آن، آزمون‌شوندگان به اضافه کردن این مقدار واریانس خطا تشویق می‌شوند که باعث کاهش پایایی و به دنبال آن کاهش روایی آزمون می‌شود. به هر حال، راه میانه این است که از سرجمع ساده نمره سؤال‌ها (صفر و یک) برای نمره‌دهی به آزمون استفاده شود. سرجمع ساده نمره سؤال‌ها نسبت به نمره‌گذاری فرمولی واریانس خطای کمتری تولید می‌کند (ذوالفقارنسب، خدایی و یادگارزاده، ۱۳۹۱). پیشنهاد چهارم، ایجاد یک برنامه سنجش و پذیرش دو مرحله‌ای است که در آن سازمان سنجش برای غربال اولیه داوطلبان آزمون عمومی (استعداد و زبان) اجرا کند و مثل آزمون‌های GRE و TOEFL حتی در صورت امکان، بخش‌های نوشتاری و شنیداری به آزمون عمومی زبان اضافه شود. سپس آزمون تخصصی و مصاحبه توسط دانشگاه‌ها اجرا شود و در نهایت، پذیرش نهایی با ترکیب این دو نمره به دانشگاه‌ها محول شود. این کار مستلزم یک تقسیم کار بهینه بین سازمان سنجش و دانشگاه‌ها است. به گونه‌ای که سطوح آزمون‌های عمومی زبان و استعداد برای رشته‌های مختلف تعدیل شود و تأکید آزمون‌های تخصصی دانشگاه‌ها بر سؤال‌ها انشایی

باشد. به هر حال، همان‌طور که نتایج رشته فناوری اطلاعات نشان می‌دهد حتی با اجرای آزمون در سازمان سنجش این کمیته‌های پذیرش بخش‌های دانشکده‌ها هستند که در نهایت مشخص کرده‌اند چه کسانی باید پذیرش بگیرند. همه افرادی که در مصاحبه نمره بالایی گرفته‌اند حتی اگر در آزمون تخصصی ۱/۵- نمره انحراف استاندارد پایین‌تر از توزیع نرمال باشند!

در نهایت، تهیه کنندگان آزمون‌ها، سازمان سنجش و دانشکده‌ها باید اطمینان حاصل کنند که نمره‌های آزمون‌ها و معیارهای x_i دیگری که برای پذیرش یا رد افراد به کار می‌برند، برای پیش‌بینی عملکرد Y داوطلب کیفیت لازم به‌ویژه پایایی و روایی را داشته باشد. چون این نمره‌ها به‌علاوه معیارهای دیگری که برای پذیرش به کار می‌روند در تعیین سرنوشت افراد یک جامعه نقش اساسی دارند. اطمینان از پارمترهای آماری هر یک از آنها باید بخش جدایی‌ناپذیر یک برنامه سنجش و پذیرش به‌ویژه در دوره دکتری باشد (ویلینگهام، ۱۹۷۶). همچنین فرایند روایی‌یابی آزمون‌های ورودی در سطح تحصیلات تکمیلی نباید متوقف شود. اگر در شرایطی شیوه‌های آزمون تغییر پیدا کرد همچنان باید همه جنبه‌های روایی و پایایی آزمون تا هنگامی که از آن استفاده می‌شود بازبینی و کنترل شود (فوگارتی، ۲۰۰۸).

منابع

- آلن، مری جی. و یین، وندی ام. (۱۹۷۹). *مقدمه‌ای بر نظریه‌های اندازه‌گیری (روان‌سنجی)*؛ ترجمه علی دلاور (۱۳۷۴). تهران: انتشارات سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت).
- ذوالفقارنسب، سلیمان؛ خدایی، ابراهیم؛ یادگارزاده، غلامرضا (۱۳۹۱). وزن‌دهی بهینه به سؤال‌ها و خرده‌آزمون‌های ورودی برای ساخت نمره کل ترکیبی. *فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی*، ۳(۴)، ۷۹-۱۰۴.
- ذوالفقارنسب، سلیمان؛ یادگارزاده، غلامرضا (۱۳۹۱). چهارچوب سامانه نوآوری و کارکرد آن در رشد و توسعه ملی: رویکردی بر پایه نشانگرهای سازمان ملل متحد. *رهیافت*، ۵۱، ۴۱-۵۰.
- فرجی ده‌سرخ، حاتم؛ آراسته، حمیدرضا؛ نوه‌ابراهیم، عبدالرحیم؛ عبداللهی، بیژن (۱۳۹۴). شناسایی ملاک پذیرش دانشجویان دوره دکتری: مطالعه کیفی. *فصلنامه پژوهش و برنامه‌ریزی در آموزش عالی*، ۲۱(۴)، ۷۱-۹۷.
- فرجی ده‌سرخ، حاتم؛ آراسته، حمیدرضا؛ بغدادچی، رضا؛ نوه‌ابراهیم، عبدالرحیم؛ عبداللهی، بیژن (۱۳۹۵). سیستم پذیرش دکتری در ایران و دانشگاه‌های برتر: یک مطالعه تطبیقی. *مجله علوم تربیتی دانشگاه شهید چمران اهواز*، ۶(۲)، ۳۴-۵.
- وزارت بهداشت، درمان، و آموزش پزشکی (۱۳۹۷). آیین‌نامه آموزشی دوره دکتری تخصصی (Ph.D.). مصوب شصت و نهمین جلسه شورای عالی برنامه ریزی علوم پزشکی: [http://satim.tums.ac.ir/app/webroot/upload/files/PhD\(2\).pdf](http://satim.tums.ac.ir/app/webroot/upload/files/PhD(2).pdf)
- American Educational Research Association. American Psychological Association & National Council for Measurement in Education [AERA, APA & NCME] (2014). *The Standards for Educational and Psychological Testing*. Washington.
- Attiyeh, G. M. (1999). Determinants of persistence of graduate students in Ph. D. programs. *ETS Research Report Series*, 1999(1), i-43.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.
- Brennan, R. L. (2003). *Coefficients and Indices in Generalizability Theory*. (CASMA Research Report No.1). Iowa City: Center for Advanced Studies in Measurement and Assessment, the University of Iowa.

- Brennan, R. L. (2001). *Generalizability Theory, Iowa Testing Programs*. University of Iowa. Springer-Verlag, New York.
- Fogarty, G. J. (2008). Principles and applications of educational and psychological testing. In *Adult educational psychology* (pp. 351-383). Brill Sense.
- Godin, B. (2003). The emergence of S&T indicators: why did governments supplement statistics with indicators? *Research Policy*, 32(4), 679-691.
- Hartnett, R. T., & Willingham, W. W. (1980). The criterion problem: What measure of success in graduate education? *Applied Psychological Measurement*, 4(3), 281-291.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological bulletin*, 96(1), 72.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychological bulletin*, 127(1), 162.
- Louw, J., & Muller, J. (2014). *A literature review on models of the PhD*. Centre for Higher Education Trust (CHET). Available online at: <https://www.chet.org.za/papers/literature-review-models-phd>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-104). New York, NY: American Council on education and Macmillan.
- Messick, S. (1993). Foundations of validity: Meaning and consequences in psychological assessment. *ETS Research Report Series*, 1993(2), i-18.
- Otago university (2011). *Selecting high quality PhD candidates seven tips for prospective supervisors*. Graduate Research Services. New Zealand. Available at: <https://www.otago.ac.nz/graduate-research/study/otago029265.pdf>
- Rogelberg, S. G. (Ed.). (2007). *Encyclopedia of industrial and organizational psychology* (Vol. 1). Sage.
- Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology*, 64(6), 609.
- Shaw, E. J., Marini, J. P., Beard, J., Shmueli, D., Young, L., & Ng, H. (2016). *The Redesigned SAT® Pilot Predictive Validity Study: A First Look*. Research Report 2016-1. College Board.
- Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of applied psychology*, 23(5), 565.

- Thompson, W. J. (2018). The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation.
- Walpole, M., Burton, N. W., Kanyi, K., & Jackenthal, A. (2002). Selecting successful graduate students: in-depth interviews with GRE® users. *ETS research report series, 2002(1)*, I-29.
- Willingham, W. W. (1976). Validity and the Graduate Record Examinations Program.
- Zwick, R. (1991). *An Analysis of Graduate School Careers in Three Universities: Differences in Attainment Patterns across Academic Programs and Demographic Groups*. GRE Board Report No. 86-21P
- Zwick, R. (2006). Higher education admissions testing. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.), pp. 647-679 Westport, CT: American Council on Education/Praeger.
- Zwick, R. (2007). *College admission testing*. National Association for College Admission Counseling, 1-44.