

آزمون (آزمایشی، میدانی، ...)

نتیجه آن، میزان یادگیری، صلاحیت‌ها و توانایی‌های افراد در آزمون، با مقایسه کمی به دست آید (خرقانی و سلسله، 1388).

Testing (Pilot, Field, ...)

در طراحی آزمون، گاهی آموزش‌گران و طراحان نخستین آزمون شناسایی شده، که به نظر می‌رسد میزان فراگیری آموزش‌های عرضه شده را اندازه‌گیری می‌کند، انتخاب می‌کنند و با انتقادهایی مواجه شوند مبنی بر اینکه؛ که آزمون برای سنجش ساختار، نامعتبر و یا برای نمونه پژوهش نامناسب است. باید توجه داشت که معمولاً برای آزمون، هیچ مورد خاصی از آزمون کامل نیست، بلکه با بررسی همه شرایط و بر اساس مقتضیات تشخیص داده شده از سوی طراحان و اجراکنندگان آزمون، متناسب‌ترین و بهترین ترکیب از آزمون به کار می‌رود. این امر به‌ویژه درباره فراگیران در نظام آموزش علمی-کاربردی، (با توجه به تنوع موارد آموزشی با هدف افزایش و تثبیت قابلیت‌های کاربردی و مهارتی در فراگیران) اهمیتی ویژه دارد و از این رو آشنایی و تسلط کامل بر چگونگی اعمال آزمون در آموزش برای همه دست‌اندرکاران نظام آموزش علمی-کاربردی، لازم است.

مفهوم آزمون در آموزش‌های علمی-کاربردی

در واژه‌نامه تخصصی آموزش و پرورش، واژه «آزمون» معادل واژه‌های Exam, Test, Examination در نظر گرفته شده است. آزمون، ابزاری برای جمع‌آوری داده‌ها درباره‌ی افراد و پدیده‌ها است. با این ابزار می‌توان اطلاعاتی درباره استعدادها، پیشرفت فعالیت‌ها و جنبه‌های گوناگون شخصیت افراد به دست آورد. آزمون را می‌توان بخشی از اندازه‌گیری یا سنجش یک فرایند آموزشی دانست که نتایج آن معمولاً به صورت کمی و عددی نشان داده می‌شود (English, 1992)؛ به طوری که بر مبنای آن بتوان به تحلیل اثربخشی فعالیت‌های آموزشی پیرامون ارتقاء دانش، بینش، نگرش و مهارت‌های فراگیران پرداخت. بر این اساس، آزمون مقیاسی برای سنجش اثربخشی (به عنوان ملاک دستیابی به اهداف از پیش تعیین شده در آموزش) تلقی می‌شود. بنابراین، به طور خاص در «آموزش‌های علمی-کاربردی» که تأکید آن بر عملیات و کاربرد است، آزمون را می‌توان ابزار استاندارد و علمی دانست که در آن، آزمودنی‌ها (فراگیران) با تعدادی از پرسش‌های همخوان و هماهنگ به انجام فعالیت‌های عملی مشابه می‌پردازند تا در

آموزش‌های علمی-کاربردی، آموزش‌هایی است که با هدف ارتقاء و انتقال دانش کار، ایجاد مهارت‌ها، افزایش بهره‌وری، به‌هنگام کردن و ارتقاء معلومات و تجارب شاغلان، رشد استعدادهای بارز و به‌فعلیت‌درآوردن استعدادها، نهفته، برای عهده‌دارشدن مشاغل و حرفه‌های گوناگون انجام می‌شود، تا توانایی افراد را برای انجام دادن کاری که به آنان محول می‌شود، به سطح مطلوب برساند. شورای عالی انقلاب فرهنگی، آموزش‌های علمی-کاربردی را آموزش‌هایی تعریف می‌کند که به نام آموزش‌های فنی و حرفه‌ای، بیرون از نظام آموزش رسمی برگزار می‌شود. همچنین آموزش‌هایی که در قالب نظام آموزش رسمی کشور در یکی از مقاطع سلسله مراتب آموزشی اجرا و به دریافت مدرک تحصیلی منجر می‌شود. بدیهی است که، این اهداف فقط در مقطعی خاص یا دوره‌ای کوتاه تأمین نمی‌شوند و ناگزیر از سطح کاردانی فراتر می‌رود و به سطح کارشناسی و کارشناسی‌ارشد و حتی دوره‌های دکترای حرفه‌ای می‌رسد (بینقی و سعیدی‌رضوانی، 1380). بر اساس این تعریف و ویژگی‌های آن و با توجه به تعریف پیشین آزمون، برای اثربخشی آموزش‌های علمی-کاربردی، می‌توان انواعی از آزمون مانند، آزمون‌های اختصاصی، استعداد، انطباق، کلامی، غیرکلامی، ورودی، مرحله‌ای، پایانی، پیشرفت تحصیلی، توانایی، صلاحیت، استخدامی، تخصصی، شفاهی، عملی، عملکردی و ... طراحی و اجرا کرد.

انواع آزمون‌ها در یک طبقه‌بندی، شامل «آزمون‌های کتبی» با پرسش‌های (تشریحی، کوتاه پاسخ، کامل کردنی (تکمیلی)، چندگزینه‌ای (ترجیحی)، جورکردن (تطبیقی)، گزارش نویسی (مانند سمینار، مسأله مخصوص، پایان‌نامه، رساله و ...)؛ «آزمون‌های عملی» (گزارش کار عملیاتی، ثبت طولانی هدفمند ساختاریافته فردی/گروهی)؛ «آزمون‌های مشاهده‌ای» (گزارش آموزش‌گر، چک‌لیست‌ها)، مقیاس‌های امتیازدهی)؛ «کارپوشه‌های ثبتی عملکرد» (یادداشت‌های روزانه، کارپوشه، یادداشت‌های مرحله‌ای، گزارش هم‌تایان،

و ملاک‌های شایستگی آن، از این منظر کدام است؟

گام‌های اساسی در طراحی و ساخت آزمون متناسب با آموزش‌های علمی-کاربردی

ساخت آزمون جدید، به‌ویژه از دیدگاه فلسفه حاکم بر آموزش‌های علمی-کاربردی، فرایندی دشوار است و نیاز به دانش چشمگیری دارد. بنابراین، لازم است قبل از ساخت آزمون، نخست مطمئن شد که در آن حیطه، از پیش آزمون مناسبی در دسترس نباشد. گام‌های اساسی در ساخت آزمون در قاب (1) نشان داده شده است (Osterlind, 1998).

بهترین شیوه برای جلب و افزایش همکاری فراگیران مشغول به تحصیل در نظام آموزش علمی-کاربردی، این است که پس از تکمیل آزمون، هدف آزمون و یافته‌های آن را برای فراگیران آشکار و بازخورد کامل و دقیقی به آنها عرضه کرد. با این تمهید، حس کنجکاوی آنها برانگیخته شده و برای مشارکت در فرایند آزمون تمایل می‌یابند.

خودارزیابی، خودگزارش‌دهی؛ «آزمون‌های شفاهی»، «آزمون‌های میدانی» و «آزمون‌های تجربی» (تجربه فردی/گروهی) می‌شود (سیف، 1379) که بر اساس فلسفه آموزش‌های علمی-کاربردی (با تأکید بر کاربرد و عمل) می‌توان، بنا به مقتضای زمان، مکان، محیط، هدف و ... از یکی یا ترکیبی از آنها بهره گرفت. اما در آموزش‌های علمی-کاربردی در تفاوت با دیگر آموزش‌ها، نوع آزمون نه تنها نسخه‌ای تقلیدی و مشابه با سایر فعالیت‌های متعارف آموزشی نیست، بلکه متناسب با هدف‌های علمی-کاربردی است. این نکته‌ای است که اغلب در آزمون از عملکرد فراگیران در آموزش‌های علمی-کاربردی فراموش شده یا به آن توجه نشده است. در این مسیر باید بیشتر به آزمون‌های میدانی، تجربی، عملیاتی مشاهده‌ای، گزارش‌نویسی و به‌ویژه خودارزیابی فراگیران توجه شود.

در این باره پرسش‌هایی مطرح می‌شود، از جمله: چه گام‌هایی باید برای طراحی و تدارک آزمون متناسب با فعالیت‌های آموزش علمی-کاربردی برداشت؟ آزمون خوب

قاب 1. گام‌های اساسی در ساخت آزمون

- گام 1. **تعریف سازه‌ای که باید اندازه‌گیری شود:** درباره سازه‌ای که آزمون آن را اندازه‌گیری می‌کند، به دقت تعمق کنید. آیا برای این سازه یک مبنای نظری وجود دارد؟
- گام 2. **تعریف جامعه مورد نظر:** ویژگی‌های فراگیران مورد نظر را باید در بسیاری از تصمیم‌گیری‌های مربوط به ساخت آزمون به حساب آورد. بنابراین، نیاز است به طور مشروح، ویژگی‌های دقیق جامعه‌ی فراگیران تعریف شود.
- گام 3. **مرور آزمون‌های مربوط:** آزمون‌های دیگری که سازه مشابهی را اندازه‌گیری می‌کنند، باید مرور شود تا درباره مفاهیمی همچون قالب‌بندی آزمون و روش‌های اعتبارسنجی آن ایده‌هایی به دست آید.
- گام 4. **پدید آوردن نمونه اولیه:** ویرایش اول آزمون را که نمونه اول نامیده می‌شود، تهیه کنید. منابع انتشار یافته متعددی وجود دارد که شما را در نوشتن پرسش‌ها راهنمایی می‌کند.
- گام 5. **ارزشیابی پیش‌نمونه:** مروری انتقادی از پیش‌نمونه و اندازه‌گیری سازه‌های آن، که افراد صاحب‌نظر در ساختن آزمون انجام داده‌اند، به عمل آید. سپس اجرای مقدماتی آزمون انجام شود. برای این کار باید نمونه‌ای از جامعه فراگیران مورد نظر را برگزیده، سپس تحلیل پرسش‌ها روی داده‌هایی که از این نمونه حاصل شده است، انجام شود.
- گام 6. **مرور آزمون:** به بازنگری نمونه اول آزمون پرداخته و اجرای مقدماتی دیگری را روی نسخه تجدیدنظر شده انجام دهید. این سیر باید تداوم یابد، به طوری که تا اصلاح نهایی، ممکن است چندین بار، این کار تکرار شود.
- گام 7. **گردآوری داده‌ها برای پایایی و روایی آزمون:** نشانه‌هایی برای حمایت از روایی و پایایی آزمون به دست آورید.

که معرف جامعه فراگیران است.

- **روایی و پایایی آزمون:** آزمون‌های خوب، آزمون‌هایی هستند که نمره‌هایی پایا (قابل اعتماد) از آنها حاصل می‌شود و استنباط از این آزمون‌ها دارای روایی (اعتبار) بالایی است. معیارهای آزمون در آموزش و پرورش (که در سال 1985م تدوین شده)، مرجع قابل استنادی برای بررسی روایی آزمون و دیگر موضوعات مربوط به آزمون‌هاست (American Educational Research Association, 1999). این مجموعه معیارها را کمیته مشترک آزمون‌های روان‌شناسی و آموزش و پرورش تدوین کردند، که متشکل از نمایندگان انجمن روان‌شناسی آمریکا، انجمن پژوهش‌های آموزش و پرورش آمریکا و شورای ملی سنجش در آموزش و پرورش است. از دیدگاه این معیارهای تدوین شده، روایی به معنی استنباط‌هایی است که برپایه آزمون‌ها انجام می‌شود و نشانگر میزان مفید بودن و تناسب آزمون است. این تعریف، این حقیقت را تأیید می‌کند که نمره‌های آزمون‌ها را نمی‌توان به خودی خود معتبر یا بی‌اعتبار دانست، بلکه اعتبار در واقع به دلیل استنباطی است که فراگیران از نمره دارند (Ibid).

در این باره، نشانه‌هایی (معیارهایی) برای بررسی اعتبار استنباط‌ها از نمره‌های آزمون عرضه شده است. یکی از آنها «نشانه مربوط به سازه (اعتبار سازه)» است که بر اساس ساختار نظری درباره ماهیت رفتار آدمی، مانند خودپنداره، شیوه یادگیری، درون‌گرایی و انگیزه پیشرفت تفسیر می‌شود. این مفاهیم هرچند به طور مستقیم قابل مشاهده نیستند، اما بر رفتار فراگیران آثار قابل مشاهده‌ای دارند. بنابراین، میزان دارایی آزمونی خاص در اعتبار سازه است؛ یعنی توان سنجش سازه‌ای که ادعای اندازه‌گیری آن را دارد (Messick, 1989).

معیار دیگر «نشانه مربوط به محتوا (اعتبار محتوا)» است که به میزان معرفی محتوا یا حوزه مفهومی با نمره‌های به دست آمده در آزمون اشاره دارد؛ که این نمره‌ها ادعای اندازه‌گیری آن را دارند. گاهی روایی محتوا با روایی صوری اشتباه گرفته می‌شود. روایی صوری بررسی ذهنی و ایجابی

راهبردهای مفید و متعددی در این مسیر برای جلب همکاری و مشارکت فعال فراگیران در آزمون وجود دارد. برای مثال، می‌توان به هر آزمودنی عددی به‌عنوان کد نسبت داد. همچنین محیط فیزیکی راحتی با در نظر گرفتن حالت‌های ذهنی و بدنی آزمودندگان فراهم کرد (Moreland et al., 1995). هرگونه اقدام در آشنایی با اجزاء و مراحل آزمون و کاهش استرس و نگرانی فراگیران، در اجرای هرچه بهتر آزمون مؤثر بوده و نتایج بهتری به‌همراه دارد.

ویژگی‌های آزمون خوب متناسب با آموزش علمی-کاربردی

ملاک انتخاب آزمون

ویژگی‌های لازم برای آزمون متناسب با آموزش‌های علمی-کاربردی به شرح زیر است:

- **عینیت:** افرادی که آزمون را اجرا و یا نمره‌گذاری می‌کنند، معمولاً دارای سوگیری هستند. عینیت آزمون یعنی، افراد، تحت تأثیر این سوگیری‌ها قرار نگیرند. مهم‌ترین علامت کیفیت آزمون، مهار هر چه بیشتر خطاهای فردی در سنجش است. در این میان درجه عینیت آزمون‌های چندگزینه‌ای بالاست، زیرا از نظر شیوه اجرا، خودآزما هستند و کل نمره‌ها با به‌کارگیری کلید نمره‌گذاری به دست می‌آید، که درباره آن توافق کامل وجود دارد.

- **شرایط تراز شده برای اجرا و نمره‌گذاری:** آزمون باید در شرایط تراز شده اجرا و نمره‌گذاری شود. زیرا، وجود این شرایط به افزایش عینیت آزمون می‌انجامد.

- **داده‌های هنجار:** نمره‌های آزمون عینی ذاتاً قابل تفسیر نیست. آنها معمولاً با امری بیرون از آزمون تفسیر می‌شوند. این امر ممکن است ملاکی یا هنجاری باشد. در تفسیر معیار ملاکی، نمره‌ها بر اساس معیار عملکرد مطلق تفسیر می‌شوند. در تفسیر هنجاری، نمره‌ها نسبت به عملکرد افراد گروهی معین تفسیر می‌شوند. برای تدوین آزمون خوب هنجاری، آزمون باید در نمونه‌ای بزرگ از افرادی اجرا شود،

این است که، به جای اینکه از فراگیران آزمون به عمل آید، به طور تصادفی به آنها نمره اختصاص یابد (Cole & Moss, 1989). هرچند وجود پایایی برای اعتبار شرطی لازم شمرده می‌شود، اما به این معنا نیست که نمره‌های آزمون دارای پایایی مناسب، همواره به استنباط معتبر از نمره‌ها منجر شود؛ زیرا می‌توان ابزاری ساخت که واحدهای مقیاس به‌کار برده‌شده در آن سازگاری (پایایی) سطح بالایی داشته باشند، اما مفهوم و معنای این واحدهای مقیاسی روشن نباشد. به بیان دیگر، نمی‌توان استنباط‌های معتبری از این نمره‌ها به دست آورد، هرچند که آنها دارای پایایی باشند. در این رابطه، می‌توان، تحلیل از پایایی را با در نظر گرفتن عواملی که سبب خطای اندازه‌گیری می‌شوند، فراتر برد. به برخی از این عوامل در زیر اشاره شده است:

- پرسش‌های آزمون تنها نمونه‌ای از کل حوزه پرسش‌های احتمالی هستند که می‌توانند برای معرفی توانایی شناختی، صفت شخصیتی، نگرش، یا ساختارهای دیگری که اندازه‌گیری می‌شوند، به‌کار روند. اگر پرسش‌های گوناگون آزمون از نظر اینکه چگونه از حوزه ساختار مورد نظر نمونه‌گیری شده‌اند، متعادل نشده باشند (یعنی به طور منطقی و متناسب از همه موارد آموزش داده شده، نباشند)، خطای اندازه‌گیری پدید می‌آید.

- اجراکنندگان آزمون ممکن است سبب بروز خطای اندازه‌گیری شوند؛ به این معنی که آزمون را به شکلی هماهنگ اجرا نکنند.

- اگر رویه‌های نمره‌دهی به طور یکسان و هماهنگ انجام نشود، منجر به خطای اندازه‌گیری می‌شود.

- شرایط آزمون باید نرمال باشد. برای مثال اگر مکان آزمون بسیار گرم، سرد یا شلوغ باشد، سبب شود آزمودنی‌ها در آزمون به شکل غیرعادی عمل کنند.

- تغییر در احساسات آزمودنی‌ها (مانند بیماری در روز آزمون) ممکن است منجر به عملکرد غیر عادی شود (Cannell, 1988).

با توجه این مفروضات رویکردهای متفاوتی برای برآورد پایایی نمره‌های آزمون تدوین گردیده که چهار نوع آن در

از پرسش‌های آزمون است، مبنی بر اینکه: آیا این پرسش‌ها محتوایی را که آزمون ادعای اندازه‌گیری آن را دارد، می‌پوشانند یا خیر (Nevo, 1985)؟

معیار دیگر «نشانه مربوط به پیش‌بینی (اعتبار پیش‌بین)» است. این معیار، به درجه میزان تأیید پیش‌بینی‌های انجام‌شده با یک آزمون از طریق مشاهده رفتارهای بعدی فراگیرانی مربوط می‌شود که آزمون برای آنها انجام شده است. چون اعتبار پیش‌بین مستلزم داشتن ملاکی صریح است، آن را به‌منزله نوعی اعتبار ملاکی معرفی می‌کنند. نوع دیگر اعتبار ملاکی، اعتبار همزمان نامیده می‌شود (Simmer, 1989).

معیار مهم دیگر، «نشانه مربوط به همزمانی (اعتبار همزمان)» است که دومین نوع اعتبار ملاکی است. اعتبار همزمانی را می‌توان به‌منزله میزان تناظر نمره‌های افراد در آزمون جدید با آزمونی تثبیت‌شده تعریف کرد، که همان ساختار را زمانی کوتاه پیش یا پس از آزمون جدید سنجیده است. آزمون تثبیت‌شده، ملاکی است که آزمون جدید بر اساس آن اعتباریابی می‌شود (Gall, 1969).

«نشانه مربوط به نتیجه (اعتبار نتیجه)»، ملاک مهم دیگری در تعیین روایی آزمون محسوب می‌شود. چهار دلیلی که از ادعاهای اعتبار حمایت می‌کند و در بالا توصیف شد، بر معنا و مفهوم نمره‌های آزمون متمرکز است. اما باید دانست نمره‌های آزمون افزون‌بر معنای آنها حاوی چیزی هستند که آن را اعتبار نتیجه باید نامید (Helmstadter, 1964). این نوع اعتبار به این واقعیت اشاره دارد که هم نمره‌های آزمون، هم نظریه و باورهای که ساختار بر آن استوار است و نیز زبانی که با آن بر ساختار موردنظر تأکید می‌شود، در بردارنده ارزش‌های خاصی است و هنگامی که برای تصمیم‌گیری درباره اشخاص به کار می‌رود دارای نتایج یا بار ارزشی است.

همواره، مطلوب آن است که پایایی نمره‌های آزمون به‌کار گرفته‌شده در حد بالایی باشد. در واقع، از آزمون‌هایی که نمره‌های آن دارای پایایی کمی باشد، نمی‌توان استنباط معتبری داشت، زیرا همگی دارای خطای اندازه‌گیری هستند و بنابراین، نمره واقعی در این حالت وجود ندارد و مانند

آزمون (آزمایشی، میدانی، ...)

که ضریب همسانی دورنی نامیده می‌شود. این ضریب نشان‌دهنده پایایی نیمی از آزمون است. روش هم‌ارزی منطقی روش دیگری برای برآورد همسانی دورنی است. پرسش‌های آزمون به‌طور انفرادی با یکی از فرمول‌های متعددی که می‌توان آن‌ها را به‌کار برد، تحلیل می‌شود. در این میان، فرمول کودر ریچاردسون بیشتر از بقیه فرمول‌ها به‌کار می‌رود (Richardson & Kuder, 1939).

- پایایی میان آزمون‌کنندگان: برای به‌دست آوردن نمره‌های آزمون به افرادی برای اجرای آزمون و به برخی دیگر برای نمره‌گذاری آن نیاز است. هر دو نوع آزمون‌کننده که یکی مجری آزمون و دیگری تصحیح‌کننده آزمون است، در صورتی که نتوانند، به‌دقت رویه‌های اجرا و تصحیح نمره‌های آزمون را انجام دهند، ممکن است، سبب ورود خطای اندازه‌گیری در نمره‌های آزمون شوند. با اجرای آزمون با چند آزمون‌کننده روی نمونه‌ای از فراگیران و سپس محاسبه همبستگی نمره‌های حاصل شده از سوی آنها می‌توان، اندازه خطای مربوط به اجرای آزمون را سنجید. درباره رویه نمره‌گذاری آزمون و خطای حاصل از آن نیز می‌توان به همین صورت عمل کرد (Shavelson & Webb, 1991).

در پایان باید گفت که انتخاب هر آزمونی به‌تنهایی مزایا و کاستی‌هایی دارد. از این رو با تعیین تعدادی از آزمون‌های قابل حصول و جمع‌آوری اطلاعات کامل از آنها، پیش از انتخاب و اجرای یکی از آزمون‌ها، می‌توان، این کاستی‌ها را برطرف کرد. در این زمینه پرسش‌های زیر، قبل از اجرای آزمون برای جست‌وجوی اطلاعات مربوط، راه‌گشا است (Haladna et al., 1991):

- آیا شواهدی وجود دارد که آزمون (برای موردی که قرار است به‌کار گرفته شود)، روا (معتبر) و پایا (قابل اعتماد) باشد؟
- آیا سطح خوانایی آزمون برای نمونه موردنظر فراگیران مناسب است؟
- آیا آزمون، در محدوده‌ی زمانی موردنظر قابل اجراست؟
- آیا هنجارهای آزمون جامعه‌ای از فراگیران، مشابه فراگیرانی است که آزمودنی‌ها از آن انتخاب شده‌اند؟

زیر عرضه شده است.

- پایایی شکل همتا: رویکردی برای برآورد پایایی نمره‌های آزمون است که در آن شکل خاص آزمونی که اجرا شده است، بررسی می‌گردد. برای مثال، دو آموزش‌گر هر کدام آزمون خود را تدوین کرده باشند، اما هدف آنها اندازه‌گیری مفهوم مشابهی باشد. هر آزمون تعداد یکسانی پرسش دارد؛ اما سبک و محتوای پرسش‌ها با یکدیگر متفاوت است. در این حالت، خطای اندازه‌گیری در نمره‌های حقیقی برآورد شده فراگیران، در آن سازه که آزمون‌ها برای اندازه‌گیری آن طراحی شده‌اند، وجود خواهد داشت. این خطاها را می‌توان با تعیین پایایی همتا برآورد کرد. همچنین، می‌توان این کار را با محاسبه ضریب همبستگی انجام داد. این ضریب همبستگی، ضریب هم‌ارزی نیز خوانده می‌شود که همبستگی نمره‌های فراگیران را بین شکل‌های موازی آزمون محاسبه می‌کند. پایایی شکل همتا اینک کمتر به‌کار می‌رود، زیرا برای ساختن شکل همتای آزمون زمان و هزینه زیادی باید صرف کرد (Berk, 1986).

- پایایی بازآزمایی: رویکردی برای برآورد پایایی نمره‌هایی آزمون است که در آن، موقعیت و شرایط اجرای آزمون مورد بازیابی قرار می‌گیرد. برای تعیین این نوع پایایی، باید ضریب همبستگی را محاسبه کرد که در اینجا ضریب ثبات نامیده می‌شود و در آن همبستگی بین نمره‌های فراگیران در همان ابزار به‌کار رفته، در دو موقعیت گوناگون محاسبه می‌شود. این نوع تعیین پایایی آزمون‌ها بیشتر در مواقعی به‌کار می‌رود که شکل‌های همتای آزمون‌ها در دسترس نباشند. مهم‌ترین مسأله در محاسبه پایایی بازآزمایی، تعیین فاصله‌ی زمان مناسب بین دو اجرای آزمون است (Porter, 1985).

- همسانی دورنی: همسانی دورنی رویکردی برای برآورد پایایی نمره‌های آزمون است که در آن پرسش‌های آزمون به‌طور مجزا مورد بررسی قرار می‌گیرد. روش‌های متعددی برای بررسی همسانی دورنی آزمون استفاده می‌شود که در هر کدام از این روش‌ها، تحلیل نمره آزمودنی‌ها فقط در یک بار اجرای آزمون انجام می‌شود. روشی برای برآورد همسانی دورنی، محاسبه ضریب همبستگی بین دو نیمه آزمون است

(همان).

در پایان برای آزمون، توجه به همه شرایط و انتخاب متناسب‌ترین و بهترین ترکیب از آزمون، به‌ویژه درباره فراگیران در نظام آموزش علمی-کاربردی لازم است، زیرا در این آموزش‌ها، تنوع موارد آموزشی با هدف افزایش و تثبیت قابلیت‌های کاربردی و مهارتی در فراگیران اهمیت دارد.

کتاب‌شناسی

بینقی، ت؛ سعیدی‌رضوانی، م. (1380). ارزیابی اثربخشی برنامه‌های آموزشی علمی-کاربردی در پاسخ به نیاز کارفرمایان. (طرح پژوهشی، ویراست 1، شماره 39) تهران: مؤسسه پژوهش و برنامه‌ریزی آموزش عالی. خرقانی، س؛ سلسله، م. (1388). آموزش - علمی کاربردی: نظام آموزشی مجاور یا رکنی اساسی در آموزش عالی؟. کنفرانس آموزش مهندسی در 1404، تهران: دانشگاه تهران، فرهنگستان علوم. سیف، ع. (1379). روش‌های اندازه‌گیری و ارزشیابی آموزشی. چاپ پنجم. تهران: نشر دوران.

American Educational Research Association. (1999). American Psychological Association, & National Council on Measurement in Education. *Standards for educational and psychological testing*.

Cannell, J. J. (1988). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average. *Educational Measurement: issues and practice*, 7(2), 5-9.

Cole, N. S., & Moss, P. A. (1989). *Bias in test use*. In R. L. Linn (Ed.), *The American Council on Education/Macmillan series on higher education. Educational measurement* (p. 201-219). Macmillan Publishing Co, Inc; American Council on Education.

Engilish, F. W. (1992). *Deciding What to Teach and Test: Developing, aligning, and leading the curriculum* Newbury Park, Ca: Educational Measurement. (3re Ed.,) Pp. 201-219. New York: Macmillan.

Erk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational research*, 56(1), 137-172.

Gall, M. D. (1969). The relationship between masculinity-feminity and manifest anxiety. *Journal of Clinical Psychology*, 25(3), 294-295.

Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.

Helmestadter, G. C. (1964). *Principles of Psychological Measurement*, Englewood Cliffs, Nj: Prentice Hall.

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.

Moreland, K. L., Eyde, L. D., Robertson, G. J., Primoff, E. S., & Most, R. B. (1995). Assessment of test user qualifications: A research-based measurement procedure.

American Psychologist, 50(1), 14-23.

Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, 22(4), 287-293.

Osterlind, S. J. (1998). What Is constructing test items? (pp. 1-16). Springer Netherlands.

Porter, A. (1985). Content Determinants: An Overview, Presented at *Annual Meeting Of The American Educational Research Association*, San Francisco. (Eric Document Reproduction Service No. Ed 274 510).

Richardson, M. W., & Kuder, G. F. (1939). The calculation of test reliability coefficients based on the method of rational equivalence. *Journal of Educational Psychology*, 30(9), 681-687.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer* (Vol. 1). Sage.

Simmer, M. L. (1989). Predictive validity of an abbreviated version of the Printing Performance School Readiness Test. *Journal of School Psychology*, 27(2), 189-195

فاطمه شیدایی (دانشگاه تربیت مدرس)

مسعود بیژنی (عضو هیئت علمی دانشگاه تربیت مدرس)

ناصر ولی‌زاده (دانشگاه شیراز)

معادل‌ها

Predictive validity	اعتبار پیش‌بین
Construct validity	اعتبار سازه
Content validity	اعتبار محتوا
Consequential validity	اعتبار نتیجه
Concurrent validity	اعتبار همزمان
Value laden	بار ارزشی
Reliability	پایایی
Test retest reliability	پایایی بازآزمایی
Alternate form reliability	پایایی شکل همتا
Inter tester reliability	پایایی میان آزمون‌کنندگان
Self-administered	خودآزما
Validity	روایی
Face validity	روایی صوری
Method of rational equivalence	روش هم‌ارزی منطقی
Bias	سوگیری
Standard condition	شرایط ترازشده
Coefficient of internal consistency	ضریب همسانی دورنی
Objectivity	عینیت
Kuder Richardson formula	فرمول کورد ریچاردسون
Internal consistency	همسانی درونی