

اعتبار سنجی آزمون نهایی زبان فارسی مدرسه المهدی (عج) جامعه المصطفی

زهرا سمائی

کارشناسی ارشد دانشگاه فردوسی مشهد

دکتر عطیه کامیابی گل

(استادیار گروه زبان و ادبیات فارسی و گروه زبان‌شناسی دانشگاه فردوسی مشهد)

چکیده

در بحث آموزش زبان، همواره در کنار آموزش، ارزشیابی آموخته‌ها نیز مطرح بوده است. این ارزشیابی نیازمند طراحی دقیق آزمون و برگزاری آن به شیوه استاندارد است تا بتوان به سطح حقیقی توانایی‌های زبان‌آموز پی برد. با وجود اینکه مراکز بسیاری در داخل و خارج ایران مشغول تدریس زبان فارسی به غیرفارسی‌زبانان هستند، متأسفانه هنوز هیچ آزمون معیار و استاندارد سطح بندی شده‌ای برای سنجش توانایی فارسی-آموزان وجود ندارد. هر مرکز آموزش زبان فارسی به غیرفارسی‌زبانان (آزفا)، با توجه به نوع منابع آموزشی‌ای که در اختیار دارد و آن‌ها را در امر آموزش به کار می‌برد، آزمون‌هایی طراحی کرده و استفاده می‌کند. جامعه المصطفی العالمیه یکی از این مراکز آزفا است. با توجه به نقش مهم و پررنگ جامعه المصطفی در امر گسترش زبان فارسی در جهان و همین‌طور اهمیت مدرسه المهدی (عج) به عنوان یکی از مهم‌ترین مراکز آزفای المصطفی، ارزیابی یکی از آزمون‌های جامع این مرکز به لحاظ داشتن استاندارد کافی، هدف این مقاله است. آزمون‌هایی که در واقع نشان می‌دهد که فارسی‌آموز پس از طی دوره جامع آموزش فارسی از نظر مهارت زبانی در چه سطحی قرار دارد. بررسی روایی از طریق تحلیل سؤالی نشان داد که سؤالات آزمون از سطح دشواری مناسبی برخوردار است اما ضریب تشخیص بالایی ندارد و نیاز است به این قسمت بیشتر توجه شود. پایایی این آزمون نیز به روش کودر- ریچاردسون سنجیده شد و آزمون از پایایی متوسطی برخوردار بود.

کلیدواژه‌ها: آزمون سازی، پایایی، روایی، آزمون زبان فارسی، جامعه المصطفی

العالمیه.

مقدمه

جامعه المصطفی العالمیه نهادی علمی و بین المللی با هویت حوزه‌ای است که با هدف گسترش علوم اسلامی، انسانی و اجتماعی، با رویکرد آموزشی، پژوهشی و تربیتی، اهتمام دارد که انبوهی از داوطلبان را از سراسر گیتی تحت پوشش قرار دهد و فرصت کم نظیری برای تشنگان زلال معارف اسلامی و هدایت قرآنی در سراسر جهان فراهم آورد تا ضمن تربیت مجتهدان، عالمان و متخصصان پارسا و متعهد، به تبیین، تولید و تعمیق تفکرات و اندیشه‌های اسلامی و نشر و ترویج اسلام ناب محمدی همت گمارد. این نهاد علمی بین‌المللی بیش از ۵۰۰۰۰ دانش پژوه مرد و زن را از ۱۲۲ ملیت، تحت تعلیم و تربیت خود داشته که تاکنون، بیش از ۲۵۰۰۰ نفر از آنان دانش آموخته شده‌اند. در جامعه المصطفی العالمیه رشته‌های اسلامی فقه، اصول فقه، قرآن، حدیث، فلسفه، عرفان، اخلاق، کلام و تاریخ اسلام و در حوزه علوم انسانی نیز رشته‌ها و گرایش‌هایی همچون علوم تربیتی، حقوق، روان‌شناسی، اقتصاد، جامعه‌شناسی، علوم سیاسی، بانک‌داری، ارتباطات، مدیریت و زبان و ادبیات زبان‌های فارسی، عربی، انگلیسی، فرانسوی و روسی، با رویکرد اسلامی ارائه می‌شود (برگرفته از سایت جامعه المصطفی).

همچنین جامعه المصطفی العالمیه یکی از بزرگ‌ترین سازمان‌هایی است که در بخش آموزش زبان فارسی در داخل و خارج ایران فعال است و در گسترش زبان فارسی در جهان نقش بسزائی دارد. هر ساله افراد زیادی از سراسر دنیا جذب این مرکز می‌شوند و پس از گذراندن دوره ۶ ماهه آموزش زبان فارسی و گذراندن یک‌ترم ۴ ماهه آموزش معارف اسلامی، وارد دانشگاه می‌شوند و در مقطع کارشناسی، ارشد و دکتری ادامه تحصیل می‌دهند. اغلب این فارسی آموزان برای تحصیل در رشته‌های علوم اسلامی و دینی به ایران می‌آیند. جامعه المصطفی بیش از ۸۵ نمایندگی در کشورهای مختلف جهان دارد و در شهرهای بزرگ ایران از قبیل، تهران، قم، مشهد، اصفهان، کاشان، گرگان، تبریز و غیره نیز نمایندگی دارد. منابع و محتوای آموزشی این مراکز، تولید داخلی خود جامعه المصطفی است که انتشارات بین‌المللی جامعه المصطفی آن‌ها را منتشر کرده است. ۶ کتاب با عنوان «آموزش فارسی به فارسی» در سه سطح مقدماتی، میانه و پیشرفته، در این

مجموعه مقالات همایش ملی الزامات آموزش زبان و ادبیات فارسی... دانشگاه فردوسی مشهد، آبان ۱۳۹۹

مراکز تدریس می‌شود. جامعه المصطفی برای ارزشیابی فارسی آموزان خود، یک آزمون نهایی داخلی دارد که بر اساس متون آموزشی تدریس شده در آن مرکز، طراحی شده است.

یکی از مهم‌ترین وظایفی که مدرسین با آن مواجه هستند، ارزیابی زبان آموزان است. این وظیفه مهم نیاز به مهارت‌های قابل توجهی دارد زیرا این آزمون‌ها به عنوان یک شاخص کلیدی و حساس برای ارتقاء زبان آموزان به کار می‌روند. مسلماً کار آزمون با نوشتن سؤالات به پایان نمی‌رسد، بلکه پس از آن باید به دنبال گردآوری شواهد برای اطمینان از کیفیت سؤال بود. برای چنین هدفی طرح آزمون باید یک فرایند دقیق نظام‌مند را برای جمع‌آوری اطلاعات صحیح به انجام رسانده و بر پایه آن اطلاعات قضاوت نماید. هدف چنین قضاوت‌هایی صرفاً شناسایی منابع احتمالی خطا در اندازه‌گیری است. بدین منظور بعد از اجرای آزمون و تعیین نمرات آن، تجزیه و تحلیل سؤال روش مناسبی برای قضاوت در مورد کیفیت آن به ویژه در آزمون‌های چندگزینه‌ای است. یکی از راه‌های تحلیل این آزمون‌ها، به دست آوردن روایی و پایایی آزمون هست. زیرا از ویژگی‌های یک آزمون موفق و استاندارد داشتن روایی و پایایی مناسب است. دو فاکتور مهم در بحث روایی سنجی ضریب تشخیص (Item Discrimination) و درجه دشواری (Item Difficulty) سؤالات است. جامعه المصطفی به عنوان یکی از مراکز آرفای فعال، همواره سعی کرده که در آزمون‌هایش از سؤالات استاندارد بهره بگیرد. پژوهش حاضر آزمون سری هفتم از سری آزمون‌های نهایی این مرکز را از نظر پایایی و همچنین درجه دشواری و ضریب تشخیص مورد بررسی و تحلیل قرار داده است. پژوهش با استفاده از تحلیل آزمون (Item Analysis) به این دو ویژگی مهم در حوزه روایی آزمون پرداخته است. پایایی آزمون نیز با روش کودر-ریچاردسون و با فرمول KR-21 ارزیابی شده است و به نتایج معناداری در این حوزه دست یافته است.

۱- پیشینه

در حوزه آزمون‌سازی و ارزشیابی درسی تاکنون پژوهش‌ها و تحقیقات زیادی انجام شده است. بلوم، هستینگر و ماداوس، (۱۹۷۱) درباره شیوه‌های آزمون‌گیری و آزمون‌های

مجموعه مقالات همایش ملی الزامات آموزش زبان و ادبیات فارسی... دانشگاه فردوسی مشهد، آبان ۱۳۹۹

درسی تحقیقاتی انجام دادند و مطالب مهمی را ارائه کردند. جعفر پور (۱۳۷۵) در کتابی به نام «مقدمه‌ای بر آزمون شناسی زبان» مبانی آزمون شناسی را به خوبی تبیین کرده است. فرهادی، بیرجندی و جعفرپور (۱۹۹۴) در کتاب «آزمون در زبان انگلیسی نظریه‌ها و کاربردها» مباحث آزمون شناسی زبان و انواع تحلیل آزمون‌ها را بررسی نموده‌اند. مقالاتی نیز در این موضوع کار شده است اما بیشتر مقالات درباره آزمون بسندگی زبان فارسی بوده است. قنصولی (۱۳۸۹) نمونه‌ای از آزمون بسندگی در زبان فارسی را طراحی کرده است. عباسی (۱۳۹۳) یک نمونه آزمون بسندگی معیار زبان فارسی را همانند دو آزمون تافل و آیلتس، تهیه و تنظیم کرده است که فارسی آموز را در همه مهارت‌های زبانی می‌سنجد و سطح کلی زبانی وی را نشان می‌دهد. سرافراز یزدی (۱۳۹۷) در پایان‌نامه خویش، اعتبار سنجی و روایی سنجی آزمون تعیین سطح مرکز آموزش زبان فارسی به غیر فاسی‌زبانان دانشگاه فردوسی مشهد را مورد ارزیابی قرار داده و در نتیجه حاصل از پژوهش ضریب تشخیص آزمون ضعیف ارزیابی شده است. پس از تحقیق و بررسی این نتیجه حاصل شد که تا به حال هیچ آزمونی از آزمون‌های مراکز جامعه المصطفی، به‌طور تخصصی و علمی ارزشیابی نشده است. این پژوهش به مدرسین این مرکز در تهیه آزمون‌های بهتر کمک می‌کند و همین‌طور در حوزه آزمون‌سازی زبان، به مدرسان دیگر مراکز آموزش زبان فارسی اطلاعات علمی و کاربردی می‌دهد.

۲- انواع آزمون‌های زبان:

آزمون‌های زبان بطور کلی در دو دسته آزمون‌های ارزیابی عملکرد (Evaluation of Attainment) و آزمون‌های پیش‌بینی کننده (Prognostic) طبقه بندی می‌شوند که هر کدام دارای سه زیر شاخه هستند (فرهادی، جعفرپور، و بیرجندی؛ ۲۰۰۴).

• آزمون‌های ارزیابی عملکرد:

هدف اصلی آزمون‌های ارزیابی عملکرد، تعیین چیزی است که آزمودنی، توانایی انجام آن را دارد و به سه دسته تقسیم می‌شوند: آزمون پیشرفت یا دستاورد (Achievement)، مهارت (Proficiency) و معلومات (Knowledge)

الف) آزمون پیشرفت:

آزمون پیشرفت برای سنجیدن میزان یادگیری دانشجو از مجموعه خاصی از مطالب آموزشی طرح می‌شوند و بر پایه برنامه آموزشی ویژه‌ای تدوین می‌گردد. این آزمون‌ها در محیط‌های آموزشی نقش حیاتی ایفا می‌کنند. مانند آزمون‌های میان‌ترم و پایان‌ترم (فرهادی، جعفرپور، و بیرجندی؛ ۲۰۰۴).

ب) آزمون مهارت تخصصی زبان:

آزمون‌های مهارت برای سنجیدن توانایی کلی زبان فراگیران به کار می‌روند. آزمون‌های مهارت برای سنجیدن میزان دانشی که یادگیرنده در طی آموزش زبان کسب کرده است. میزان توانایی‌اش در بخشهای زبان و درجه‌ای که شخص قادر به کاربرد عملی از دانش زبانی‌اش است، طراحی می‌شوند (بکمن، ۱۹۹۵). مثل آزمون انگلیسی تافل

ج) آزمون دانش یا معلومات:

در آزمون دانش، میزان آشنایی آزمودنی با شیوه اندیشه و آداب و رسوم و عادت‌های اهل زبان را می‌سنجد. آزمون‌های معلومات در مواقعی بکار می‌روند که محیط آموزش، زبانی غیر از زبان مادری فراگیران است. در این گونه موارد، زبان دوم به عنوان زبان آزمون‌ها برای سنجش معلومات داوطلبان در مواردی غیر از خود زبان استفاده می‌شود (براون، ۲۰۰۴). به عنوان مثال، آزمون فیزیک یا روانشناسی که به زبان انگلیسی نوشته شده به عنوان آزمون معلومات به حساب می‌آید.

• آزمون‌های پیش بینی کننده:

هدف از آزمون‌های پیش بینی کننده اخذ تصمیمات منطقی درباره موفقیت‌های آینده داوطلبان براساس توانایی‌های فعلی‌شان است. این آزمون‌ها به منظور پیش بینی کردن جریان عملی در آینده در مورد داوطلبان بکار برده می‌شوند. آزمون‌های پیش بینی شامل: آزمون‌های گزینشی (Selection Tests)، تعیین سطح (Placement) و استعداد (Aptitude) می‌شود.

الف) آزمون‌های گزینشی:

آزمون‌های گزینشی فراهم ساختن اطلاعاتی هست که براساس آن پذیرش یا عدم پذیرش داوطلبان در برنامه‌ای خاص تعیین می‌شود. برنامه لزوماً مرتبط با یک برنامه آموزشی نیست مثل آزمون اخذ گواهینامه رانندگی (فرهادی، جعفرپور، و بیرجندی؛ ۲۰۰۴).

ب) آزمون تعیین سطح:

آزمون‌های تعیین سطح به منظور تعیین مناسب‌ترین روش آموزشی برای داوطلبان بکار می‌روند. در آزمون‌های تعیین سطح قبولی یا مردودی وجود ندارد. هدف آزمون‌های تعیین سطح صرفاً سنجیدن توانایی‌های داوطلب در دنبال کردن مسیر خاصی از یادگیری زبان است (فولچر، ۲۰۱۳).

ج) آزمون استعداد:

آزمون‌های استعداد به منظور پیش بینی کردن موفقیت‌های داوطلبان در دستیابی به اهداف خاصی در آینده بکار می‌روند. به منظور شرکت در آزمون استعداد، لزومی ندارد داوطلب از موضوع آزمون دانش قبلی داشته باشد. این آزمون‌ها کاربردهای زیادی دارند مثلاً برای تعیین رشته تحصیلی دانشجویان یا پیش بینی شغل مناسب برای داوطلبان مفید هستند (فولچر، ۲۰۱۳).

آزمونی که در این پژوهش بررسی گردید، جزء دسته آزمون‌های ارزیابی عملکرد و شاخه آزمون پیشرفت یا دستاورد بود. آزمون مورد بررسی یکی از آزمون‌های سری هفتم از سری آزمون‌های جامع مدرسه المهدی (عج) که شامل ۵۰ سؤال در دو حوزه واژگان و دستور زبان بود. تعداد ۳۳ نفر آزمودنی داوطلب خانم در آن شرکت کردند.

۳- نحوه برگزاری آزمون و اجرای پژوهش:

پس از انتخاب یکی از آزمون‌های جامع مدرسه المهدی (عج) (نماینده‌گی المصطفی در قم، بخش مرکزی آموزش زبان فارسی)، سوالهای آزمون بررسی شد. بعد از نامه نگاری با

مجموعه مقالات همایش ملی الزامات آموزش زبان و ادبیات فارسی... دانشگاه فردوسی مشهد، آبان ۱۳۹۹

ریاست محترم جامعه المصطفی، نمایندگی خراسان و کسب اجازه برای حضور در محل اسکان فارسی آموزان، به دلیل شیوع بیماری کرونا و محدودیت در رفت و آمدها، شرایط برای برگزاری آزمون مهیا شد. آزمون در محل سالن مطالعه خوابگاه با حضور ۳۳ نفر از فارسی آموزان سطح پیشرفته برگزار گردید. فارسی آموزان از ملیت های هندوستان، پاکستان و آذربایجان و همه خانم بودند. مدت زمان برگزاری آزمون ۴۵ دقیقه بود. شرایط آزمون طبق استانداردهای موجود شامل مکانی با نور کافی، سکوت لازم، میز و صندلی راحت برای نشستن، دمای مناسب و غیره تنظیم شد (بکمن، ۱۹۹۵؛ فرهادی، جعفرپور، و بیرجندی؛ ۲۰۰۴، فولچر، ۲۰۱۳؛ کوهن، ۱۹۹۴؛ هیوز، ۲۰۱۲). قبل از آزمون با فارسی آموزان صحبت شد و سعی بر آن شد تا با برقراری ارتباط صمیمی از میزان اضطراب آن ها کاسته شود (هیوز، ۲۰۱۲؛ فالچر، ۲۰۱۳). بعد از آزمون، برگه ها تصحیح شد، سپس طبق فرمول درجه دشواری و درجه دشواری سوالها محاسبه شد. پایایی آزمون نیز با فرمول (KR-21) اندازه گیری شد و سپس نتایج نهایی در جدولی ثبت گردید.

۴- تحلیل آزمون (Item Analysis) چیست و چگونه انجام می شود؟

آزمون، وسیله ای است که مدرس را در تصمیم گیری های بعدی آموزشی یاری می دهد (هیوز، ۲۰۱۲). پس از اینکه آزمونی برگزار شد و نمره های آن محاسبه شدند، لازم است مدرس، نتایج را تفسیر کند. از آنجا که نمره های آزمودنی ها، خود به خود توانایی واقعی آزمودنی را نشان نمی دهند؛ ارزش هر آزمون، بسته به داوری های منطقی در مورد نمره های آزمون ها خواهد بود. نمره های آزمون ها، تنها تخمینی از کارکرد آزمودنی ها هستند که ممکن است روز به روز، تغییر کنند (فالچر، ۲۰۱۳). به دیگر سخن، یک نمره، نشانه ۱۰۰ درصد توانایی واقعی آزمودنی نیست به همین دلیل، نمره ها را باید با دقت ویژه ای تحلیل و تفسیر کرد. در مورد تحلیل آزمون، اولین چیزی که مهم به نظر می رسد، کمی بودن پرسش های عینی در آزمون است. نکته دیگر اینکه ماهیت آزمون باید بصورت چند گزینه ای باشد زیرا تحلیل آزمون برای آزمون های تشریحی و کیفی که نمرات بر اساس دیدگاه و نظر مصحح مشخص می شود، قابل اجرا و پیاده سازی نیست (جعفر پور، ۱۳۷۵).

مجموعه مقالات همایش ملی الزامات آموزش زبان و ادبیات فارسی... دانشگاه فردوسی مشهد، آبان ۱۳۹۹

تفسیر سؤالات آزمون دو مورد مهم را به ما نشان می‌دهد. اول اینکه هر سؤال چقدر دشوار است و درجه دشواری (Item Difficulty) آزمون را به ما نشان می‌دهد. مورد دوم اینکه آیا سؤال توانایی تشخیص دادن یا تمیز دادن بین زبان آموز قوی و ضعیف را دارد یا نه، که به آن درجه تشخیص (Item Discrimination) می‌گویند.

• **درجه دشواری ((Item Difficulty چیست و چگونه تعیین می‌شود؟**

درجه دشواری سؤال به ما نشان می‌دهد که چند نفر از زبان آموزان می‌توانند یک سؤال را به خوبی پاسخ بدهند و برای همین برخی معتقدند که درجه دشواری باید به درجه سادگی تغییر نام دهد چراکه ما به درصدی از زبان آموزان که پاسخ صحیح داده‌اند، اشاره می‌کنیم.

فرمول محاسبه درجه دشواری سؤالات آزمون:

پاسخ‌های صحیح گروه A + پاسخ‌های صحیح گروه B

N (تعداد کل)

تحلیل درجه دشواری: با تحلیل درجه دشواری، بازه قابل قبولی که نشان بدهد سؤالات ما نه خیلی سخت و نه خیلی ساده است، به دست می‌آید. برای رسیدن به این مطلوب از فرمول و آماری که بررسی کردیم، کمک گرفتیم. اگر نتیجه به دست آمده بیشتر از ۸۰ درصد باشد، نشانگر ساده بودن سؤال است. نتیجه زیر ۳۰ درصد نیز بیانگر دشواری آن است و در نهایت نتیجه بین ۳۰ تا ۸۰ درصد، قابل قبول است. (فرهادی و دیگران، ۱۳۹۶)

• **ضریب تشخیص (Item Discrimination) چیست و چگونه تعیین می‌شود؟**

ضریب تشخیص یا تمیز سؤال به این معناست که آزمون تا چه اندازه بین زبان آموزان قوی و ضعیف، تفاوت قائل شده است.

فرمول محاسبه درجه تشخیص سؤالات آزمون:

پاسخ‌های صحیح گروه A - پاسخ‌های صحیح گروه B

N (تعداد کل)

تحلیل ضریب تشخیص بدین گونه است که عدد بدست آمده اگر ۱۰ درصد یا کمتر از آن باشد، به این معناست که درجه تشخیص سؤال ضعیف است و در نتیجه غیر قابل قبول است. اگر نتیجه از ۱۵ درصد به بالا باشد نتیجه قابل قبول است. پس سؤال مورد نظر توانسته قوی را از ضعیف تشخیص بدهد و اگر نتیجه بین ۱۰ تا ۱۵ درصد باشد، ضریب تشخیص مرزی است.

• **واریانس (Variance) و انحراف معیار (Deviation score)**

یکی از راه‌های تشخیص استاندارد بودن یا نبودن آزمون، بررسی فاکتورهای واریانس و انحراف معیار نمرات است. اندازه تغییر پذیری که در بیشتر واکاویهای آماری به کار می‌رود، واریانس نام دارد. انحراف معیار، میانگین فاصله تمامی نمره‌ها را از میانگین نشان می‌دهد (فرهادی و دیگران، ۱۹۹۴).

بازه نمرات بدست آمده در این آزمون بین ۱۸-۴۶ است. واریانس نمرات ۳۹/۲۹ و انحراف معیار آزمون ۶/۲۶ بدست آمد.

$$\text{فرمول واریانس} = \sum \frac{(X-\bar{X})^2}{N-1}$$

$$\text{واریانس نمرات} = 39/29$$

$$\text{فرمول انحراف معیار} = \sqrt{\sum \frac{(X-\bar{X})^2}{N-1}}$$

$$\text{انحراف معیار نمرات} = 6/26$$

همانطور که گفته شد؛ انحراف معیار، میانگین فاصله تمام نمره‌ها را از میانگین نشان می‌دهد. هنگامی که گفته می‌شود، آزمونی چندین انحراف معیار بالا یا زیر میانگین قرار

مجموعه مقالات همایش ملی الزامات آموزش زبان و ادبیات فارسی... دانشگاه فردوسی مشهد، آبان ۱۳۹۹

دارد، اشاره به درصد نمره‌هایی در مجموعه است که زیر آن نمره قرار گرفته‌اند (جعفرپور، ۱۳۷۵) محاسبه انحراف معیار، میزان دقت آزمون در تمیز زبان‌آموزان از یکدیگر را نشان می‌دهد. انحراف معیار هر آزمون نه تنها پراکندگی نسبی نمرات را در مقایسه با منحنی طبیعی نشان می‌دهد بلکه مقایسه عددی را نیز با آزمون‌های مشابه ممکن می‌سازد.

« جدول تحلیل سوالات »										
10	9	8	7	6	5	4	3	2	1	شماره سوال
72%	65%	77%	68%	54%	72%	100%	90%	90%	94%	درجه دشواری
قابل قبول	قابل قبول	قابل قبول	قابل قبول	قابل قبول	قابل قبول	ساده	ساده	ساده	ساده	نتیجه
0	13%	22%	22%	18%	9%	0	0	9%	4%	ضریب تشخیص
ضعیف	مرزی	قابل قبول	قابل قبول	قابل قبول	ضعیف	ضعیف	ضعیف	ضعیف	ضعیف	نتیجه
20	19	18	17	16	15	14	13	12	11	شماره سوال
45%	50%	72%	63%	22%	45%	68%	90%	81%	86%	درجه دشواری
قابل قبول	قابل قبول	قابل قبول	قابل قبول	دشوار	قابل قبول	قابل قبول	ساده	ساده	ساده	نتیجه
18%	13%	0	18%	13%	0	22%	9%	18%	13%	ضریب تشخیص
قابل قبول	مرزی	ضعیف	قابل قبول	مرزی	ضعیف	قابل قبول	ضعیف	قابل قبول	مرزی	نتیجه
30	29	28	27	26	25	24	23	22	21	شماره سوال
86%	86%	72%	40%	72%	50%	77%	18%	95%	90%	درجه دشواری
ساده	ساده	قابل قبول	قابل قبول	قابل قبول	قابل قبول	قابل قبول	دشوار	ساده	ساده	نتیجه
13%	13%	18%	22%	9%	4%	22%	9%	4%	9%	ضریب تشخیص
مرزی	مرزی	قابل قبول	قابل قبول	ضعیف	ضعیف	قابل قبول	ضعیف	ضعیف	ضعیف	نتیجه
40	39	38	37	36	35	34	33	32	31	شماره سوال
77%	63%	54%	63%	63%	27%	68%	68%	77%	86%	درجه دشواری
قابل قبول	قابل قبول	قابل قبول	قابل قبول	قابل قبول	دشوار	قابل قبول	قابل قبول	قابل قبول	ساده	نتیجه
13%	9%	9%	27%	18%	27%	22%	22%	22%	13%	ضریب تشخیص
مرزی	ضعیف	ضعیف	قابل قبول	قابل قبول	قابل قبول	قابل قبول	قابل قبول	قابل قبول	مرزی	نتیجه
50	49	48	47	46	45	44	43	42	41	شماره سوال

درجه دشواری	72%	72%	90%	40%	45%	86%	77%	72%	40%	86%
نتیجه	قابل قبول	قابل قبول	ساده	قابل قبول	قابل قبول	ساده	قابل قبول	قابل قبول	قابل قبول	ساده
ضریب تشخیص	22%	22%	9%	22%	36%	13%	13%	18%	4%	4%
نتیجه	قابل قبول	قابل قبول	ضعیف	قابل قبول	قابل قبول	مرزی	مرزی	قابل قبول	ضعیف	ضعیف

● تحلیل داده‌های بخش روایی آزمون:

همانطور که گفته شد در تحلیل درجه دشواری اگر نتیجه به دست آمده بیشتر از ۸۰ درصد باشد نشانگر ساده بودن سؤال است. طبق جدول شماره ۱، از ۵۰ تا سؤال آزمون مورد نظر ۱۵ سؤال درجه دشواری پایینی داشتند و درصد آنها بیشتر از ۸۰ درصد بودند. اگر رقم به دست آمده بین ۳۰ تا ۸۰ درصد باشد، یعنی دشواری سؤال قابل قبول است. در این آزمون ۳۲ سؤال از ۵۰ سؤال رتبه قابل قبول را به خود اختصاص دادند. اما اگر نتیجه به دست آمده زیر ۳۰ درصد باشد، یعنی آن سؤال بسیار دشوار است. از این ۵۰ سؤال، ۳ سؤال دشوار بود. با توجه به کل نتایج به دست آمده، می‌توانیم نتیجه‌گیری کنیم که آزمون مورد نظر از درجه دشواری قابل قبولی برخوردار بوده است.

اما در تحلیل ضریب تشخیص گفته شد که اگر نتیجه به دست آمده کمتر از ۱۰ درصد باشد یعنی درجه تشخیص این سؤال بسیار ضعیف است. در این آزمون درجه تشخیص ۱۹ سؤال زیر ۱۰ درصد بود و این سوال‌ها، توانایی کافی برای تمیز فارسی آموز قوی از ضعیف را نداشتند. ۱۰ سؤال بین ۱۰ تا ۱۵ درصد بودند و درجه تشخیص و تمیز آنها مرزی بود. ۲۱ سؤال از ۵۰ سؤال ضریب تشخیص قابل قبولی داشتند و بالای ۱۵ درصد قدرت تشخیص داشتند.

۵- پایایی آزمون: (Reliability)

پایایی به معنی درستی و همسانی نتایج است. آزمونی که در تکرار اجرای آن، نتایج همسانی به دست آید، دارای پایایی لازم است. حقیقت آن است که ابزاری یا آزمونی

وجود ندارد که بدون خطا باشد و پایایی؛ یعنی تعیین خطای اندازه گیری که به وسیله آن می توان مشخص کرد نمره آزمودنی تا چه اندازه تحت تأثیر عوامل نامربوط و تصادفی (Random) قابل نوسان است (بکمن، ۱۹۹۵). پایایی یک آزمون، نشانگر آن است که تا چه اندازه تفاوت های فردی، واقعی یا آلوده به خطا هستند (کراب، ۱۹۹۴). یعنی میزان پایایی آزمون به ما نشان می دهد که چه مقدار از کل واریانس یک نمره، نتیجه خطاهای تصادفی و چه مقدار از آن، نشانه توانایی یا دانش آزمودنی است (جعفر پور، ۱۳۷۵). ضریب قابلیت اعتماد یا پایایی نشانگر آن است که تا چه اندازه ابزار اندازه گیری ویژگی های با ثبات آزمودنی یا ویژگی های متغیر و موقتی آن را می سنجد. لازم به ذکر است که پایایی در یک آزمون می تواند از موقعیتی به موقعیت دیگر کاملاً متفاوت باشد. برای محاسبه ضریب قابلیت اعتماد ابزارهای اندازه گیری شیوه های مختلفی وجود دارد. از جمله الف) بازآزمایی، ب) همتاسازی، پ) روش تصنیف، ت) دو نیمه کردن، ث) روش کودر- ریچاردسون، ج) الفای کروناخ (عباس زادگان، و فتوت (۱۳۸۴). در قسمت بعدی پژوهش، پایایی آزمون بر اساس روش کودر- ریچاردسون و فرمول (KR-21) محاسبه و ارزیابی شد. تمام اطلاعات نمرات در جداول مربوطه موجود است و فرمول این متد نیز در بخش مربوطه آمده است. در این بخش تمام محاسبات آماری بطور دقیق انجام شده است و داده ها در بخش تحلیل و نتیجه گیری، بررسی شده اند.

• پایایی با روش کودر - ریچاردسون (KR-21 Method)

این روش نسبت به روش های دیگر مزیتی دارد، برخلاف روش های آزمون - بازآزمون، دو نیمه و موازی که ممکن است در اجرا خطاهای زیادی داشته باشیم این روش از یک فرمول کاملاً مشخص آماری استفاده می کند و احتمال خطا را کاهش می دهد. طراحان این فرمول که دو نفر به نام های کودر (Kuder) و ریچاردسون (Richardson)، دو آماردان معروف بودند؛ این روش را نسبتاً جزء ساده ترین و علمی ترین روش ها دانستند (تراب، ۱۹۹۴؛ گلن، ۲۰۱۶). برای محاسبه پایایی با این روش، به دو فاکتور میانگین نمرات و

مجموعه مقالات همایش ملی الزامات آموزش زبان و ادبیات فارسی... دانشگاه فردوسی مشهد، آبان ۱۳۹۹

واریانس نیاز داریم. در روش KR-21 همبستگی بین نمرات مهم است (فرهادی و دیگران، ۱۳۹۶؛ کراب، ۱۹۹۴، گلن، ۲۰۱۶). محاسبه پایایی مبتنی بر واریانس ماده‌هاست و فرض بر این است که ماده‌ها از اهمیت یکسانی برخوردار هستند و هیچگونه وابستگی به هم ندارند اگر ماده‌ها به نسبت مساوی، امتیاز نداشته باشند، پایایی به دست آمده، کمتر از واقعیت است و اگر ماده‌ها واحدهای مستقل نباشند، پایایی به دست آمده، بیشتر از واقعیت است. نتیجه به دست آمده نشان می‌دهد پاسخ نهایی همبستگی حدود ۰/۷۳ است که نشانگر پایایی متوسط است (درنیه، ۲۰۱۱).

$$(KR - 21)r = \left[\frac{K}{K-1} \right] \left[1 - \frac{\bar{X}(K-\bar{X})}{KV} \right] = 0.7364$$

K=تعداد سؤال‌های آزمون

\bar{X} =میانگین نمره‌های آزمون

V=واریانس

۶- نتیجه‌گیری

هر آزمون خوب و استاندارد باید دست کم دارای ۴ ویژگی روایی (Validity)، پایایی (Reliability)، عملی بودن (Efficiency and Practicality) و ارتباط (Relevance) باشد. مقصود از روایی این است که آزمون، آنچه را که باید بسنجد، می‌سنجد یا نه و تا چه اندازه از عهده این کار برمی‌آید. پایایی، به درجه ثبات نمره‌هایی گفته می‌شود که از اجرای پایایی آزمون به دست می‌آید. عملی بودن، به انجام پذیری آزمون با امکان‌های موجود، آسانی برگزاری و سادگی تفسیر نتایج بدست آمده اشاره دارد. ارتباط، به انطباق آنچه آزمون می‌سنجد و هدف مورد نظر آزمون مربوط می‌شود (جعفرپور، ۱۳۷۵). مهم‌ترین ویژگی از بین این ویژگی‌ها، روایی آزمون است بدین معنا که اگر آزمونی روایی داشته باشد، بی‌گمان پایایی هم دارد، اما عکس آن همیشه صادق نیست. شکی نیست که یک آزمون استاندارد و خوب به عنوان ابزار تشخیصی و کمکی

مجموعه مقالات همایش ملی الزامات آموزش زبان و ادبیات فارسی... دانشگاه فردوسی مشهد، آبان ۱۳۹۹

برای مدرسان اهمیت بسزائی دارد و اگر آزمون‌هایی که در مراکز آزفا برگزار می‌شود، استاندارد باشند، بی‌گمان در رشد و پیشرفت فارسی‌آموزان مؤثر خواهد بود. نتایج به‌دست آمده از پژوهش حاضر در دو حوزه پایایی و روایی آزمون، حاکی از آن بود که در بخش روایی آزمون، نتایج نشان می‌دهد که این آزمون از نظر درجه دشواری و دشواری سؤالات، تقریباً دارای سطح متوسط بود و قابل قبول بود ولی ضریب تشخیص سؤالات از حد متوسط به پایین‌تر بود و بیشتر از نیمی از سؤالات قدرت تشخیص ضعیف یا مرزی داشتند. نتایج به‌دست آمده از سنجش پایایی آزمون نشان داد که آزمون، پایایی متوسطی دارد. بر این اساس عملکرد فارسی‌آموزان در آزمون نیز متوسط بوده‌است. با بررسی این آزمون پی می‌بریم که برای بهبود وضعیت موجود در طراحی آزمون‌های استاندارد، لازم است در وهله اول کارگاه‌های تخصصی آزمون‌سازی برای مدرسان و طراحان آزمون برگزار شود. هر چند مجموعه جامعه المصطفی در تربیت مدرسان آزفا و برگزاری کارگاه‌های آموزشی فعال است، اما به نظر می‌رسد در بخش تولید آزمون‌های استاندارد نیاز به دقت عمل بیش‌تری است.

درنهایت پیشنهاد می‌شود آزمون‌های جامع دیگر مجموعه جامعه المصطفی نیز مورد بررسی قرار گیرند و نتایج آن با این پژوهش مقایسه شوند.

مجموعه مقالات همایش ملی الزامات آموزش زبان و ادبیات فارسی... دانشگاه فردوسی مشهد، آبان ۱۳۹۹

کتابنامه

- بلوم، بنیامین ساموئل، هستینگر، جان توماس و ماداوس، جرج. (۱۹۷۱). «راهنمایی ارزشیابی تکوینی مجموعه‌ای از آموخته‌های دانش آموزان»، ترجمه ابراهیم کظیمی، تهران: انتشارات دانشگاه تربیت معلم.
- جعفر پور عبدالجواد. (۱۳۷۵). «مقدمه‌ای بر آزمون‌شناسی زبان»، شیراز، انتشارات دانشگاه شیراز
- سرافراز یزدی، سعیده. (۱۳۹۷). «اعتبار سنجی و روایی سنجی آزمون تعیین سطح مرکز آرفای دانشگاه فردوسی مشهد»، کارشناسی ارشد، دانشگاه فردوسی، مشهد.
- عباسی، زهرا. (۱۳۹۳). «تهیه و تدوین آزمون بسندگی زبان فارسی»، نخستین همایش زبان فارسی، تهران، دانشگاه تربیت مدرس.
- عباس زادگان، سید محمدرضا؛ فتوت، احمدرضا. (۱۳۸۴). «کاربرد پایایی و روایی در پژوهش»، تهران: دفتر پژوهش‌های فرهنگی .
- قنسولی، بهزاد. (۱۳۸۹). «طراحی و روایی سنجی آزمون بسندگی زبان فارسی»، پژوهش‌های زبان‌های خارجی، شماره ۵۷، بهار.

منبع اینترنتی: سایت جامعه المصطفی <http://miu.ac.ir>

منابع انگلیسی:

- Bachman, L. F. (1995). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Brown, H.D. (2004). *Language assessment: Principles and classroom practice*. New York: Pearson Education.
- Cohen, A. D. (1994). *Assessing language ability in the classroom*. (2nd ed). Boston: Heinle & Heinle.
- Dornyei, Z. (2011). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Farhady, H., Ja'farpur, A., & Birjandi, P. (2004). *Testing language skills: From theory to practice*. Tehran: SAMT.
- Fulcher, G. (2013). *Practical language testing*. New York: Routledge.
- Glen, S. (2016). Kuder-Richardson 20 (KR-20) & 21 (KR-21). Retrieved From statisticsHowTo.com: Elementary Statistics for the rest of us! <https://www.statisticshowto.com/kuder-richardson/> on July 28,2020.
- Hughes, A. (2012). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Traub, R.E. (1994). *Reliability for the Social Sciences: Theory and Applications*. Thousand Oaks CA: Sage.