» تحلیل و اعتبارسنجی آزمون مهارت درک­ شنیداری زبان فارسی

 سطح فرامیانی«

لیلا گل‌پور: استادیارگروه آزفا، دانشگاه بین‌المللی امام خمینی (ره)

محمدباقر میرزایی حصاریان:‌ استادیارگروه آزفا، دانشگاه بین‌المللی امام خمینی (ره)

امیررضا وکیلی‌فرد: دانشیارگروه آزفا، دانشگاه بین‌المللی امام خمینی (ره)

**چکیده**

مهارت شنیداری یکی از مهارت­های اصلی و ضروری در یادگیری زبان فارسی است، چراکه آغازگر فرآیند یادگیری زبان دوّم/ خارجی است و امکان بازگشت به آنچه گفته‌شده و بازنگری آن به‌گونه‌ای که در درک متن خواندن میّسر است، ‌مقدور نیست. از این رو،‌ با توجه به اهمیت مهارت شنیداری، در این پژوهش تلاش شده است که به­منظور دسترسی به یک آزمون استاندارد و تهیه بانک سؤال­های استاندارد، یکی از آزمون­های طراحی شده و برگزارشده­ی مهارت شنیداری سطح فرامیانی مورد تحلیل و بررسی قرار گرفته است. نمونه­ی آماری این پژوهش ۵۹ نفر از فارسی­آموزان مرکز آموزش زبان فارسی دانشگاه بین­المللی امام خمینی(ره) بوده که از لحاظ زبانی در سطح فرامیانی و در رده‌ی سنّی 18 تا 30 قرار داشتند و به‌طور تصادفی انتخاب شده بودند. ابزار این پژوهش، یک آزمون سه­گزینه­ای درک شنیداری بود که شامل۲۰ پرسش از نوع درک موضوع اصلی متن،‌ درک موضوعات بیان شده در یک مصاحبه،‌درک توالی یک رویداد، درک مقایسه کردن، درک طبقه‌بندی، تشخیص درست یا نادرست بودن جمله، دگرگفت و بندش بوده است که زبان­آموزان با توجه به فایل صوتی شنیده شده به پرسش­ها پاسخ داده بودند. متون فایل صوتی شامل موضوعات آموزشی،‌ اجتماعی، پزشکی،‌ فرهنگی،‌ محیط زیست و تغذیه بدون نیاز به دانش و تخصصی خاص بود. متغیرهای این پژوهش شاخص ضریب دشواری، ضریب تمیز و پایایی پرسش­های آزمون درک شنیداری بود. پس از واکاوی داده­ها، مشخص شد که ۷۵٪ پرسش­ها ( ۱۵ پرسش)، دارای ضریب دشواری بین ۱- ۷/۰ ،‌ ۱۵٪ ( ۳ پرسش)‌ دارای ضریب دشواری بین ۷/۰- ۳/۰ و ۱۰ ٪ باقی­مانده­( ۲ پرسش) در محدوده­ی ضریب دشواری بین ۳/۰-۰ قرار داشتند. در بررسی ضریب تمیز این آزمون نیز مشخص شد که ۱۵٪ پرسش­ها ( ۳ پرسش) دارای ضریب تمیز 3/0 <D ، ۶۵٪ پرسش­ها (‌۱۳ پرسش) 3/0 >D ، ‌۱۵ ٪ پرسش­ها ( ۳ پرسش) ۰ =D و ۵٪ (۱ پرسش) ‌دارای ضریب تمیز منفی است. دستاوردی که از ضریب دشواری این پژوهش به دست آمد نشان­دهنده­ی سهولت و آسانی آزمون و نتایج به دست آمده از ضریب تمیز نیز گویای این امر بود که این آزمون به­خوبی نتوانسته است زبان­آموزان قوی را از ضعیف جدا کند و سرانجام، ‌۱۵٪ پرسش­ها ( ۳پرسش ) با ضریب دشواری بین ۷/۰- ۳/۰ و ضریب تمیز 3/0 <D به­عنوان سؤالات استاندارد انتخاب شدند و در بانک سؤالات استاندارد قرار گرفتند و بعد از آسیب­شناسی پرسش­های باقی­مانده، مشخص شد که عملکرد نادرست گزینه­های انحرافی، نشاندار بودن گزینه­ی کلید، تشخیص گزینه­ی صحیح حتی بدون نیاز به فایل صوتی،‌ مبهم بودن گزینه­های انحرافی،‌ وابسته بودن گزینه­های انحرافی به اطلاعات عمومی نه متن شنیداری، از علت­های اصلی عملکرد نامناسب بقیه سؤالات و در نتیجه نامناسب بودن ضریب دشواری و ضریب تمیز بوده است که در طراحی آزمون درک شنیداری باید به آن توجه شود.

**واژگان کلیدی: ارزشیابی، مهارت درک شنیداری،‌ ضریب دشواری،‌ ضریب تمیز،‌ پایایی**

**۱-مقدمه**

 هر سیستمی دارای سه عنصر ورودی، تجزیه­-تحلیل و خروجی است. مهارت شنیدن و خواندن به عنوان دو مهارت ورودی، آغازگر فرآیند یادگیری زبان دوّم می­باشند و مغز، کار تجزیه و تحلیل را بر عهده دارد و در نهایت صحبت کردن و نوشتن، خروجیِ فرآیند یادگیری زبان دوّم است. صحبت کردن در نتیجه­ی بلوغ شنیداری رخ می­دهد، بنابراین مهارت شنیداری در میان مهارت­های زبانی از اهمیّت ویژه­ای در آموزش برخوردار است. در گذشته، گوش دادن یک مهارت منفعل و غیر پویا به شمار می­رفت زیرا قابل مشاهده نبود و توجّه زیادی را در برنامه­های درسی به خود جلب نمی­کرد، دلیل دیگری که برای نادیده گرفتن گوش دادن در گذشته وجود داشت، تأکید زیادی بود که بر روی مهارت­های تولیدی و صحبت کردن وجود داشت امّا امروزه در انگاره­های اخیر مهارت شنیداری، یک فرآیند فعّال و پویا به­شمار می­رود که می­تواند تاثیر به­سزایی در یادگیری زبان داشته باشد، از این رو، توجّه خاصی را از سوی مدرسان زبان دوّم می­طلبد (کومب و همکاران، 2007 : 90). در صنعت آزمون­سازیِ استاندارد شده، چندین آزمون بیان شفاهی جداگانه قابل دسترس هستند امّا به ندرت می­توان یک آزمون شنیداری جداگانه یافت. یکی از دلایل این موضوع این است که شنیدن اغلب بخشی از صحبت کردن در نظر گرفته می­شود. چگونه می­توان بدون داشتن مهارت در شنیدن، به زبانی دیگر صحبت کرد ؟ علاوه­براین، ماهیّت کاملاً قابل مشاهده­ی صحبت کردن، سبب می­شود که در مقایسه با شنیدن، آزمونی عینی­تر باشد. امّا شاید جهت­گیری­های جهانی در قبال صحبت کردن، دلیل مؤثّرتری باشد. ارزش یک سخنگوی خوب متأسفانه از یک شنونده­ی خوب بسیار بیشتر است. مردم معمولاً برای اینکه بپرسند آیا فردی در یک زبان دیگر مهارت دارد، می­گویندآیا می­توانید انگلیسی صحبت کنید؟ آنها به­ندرت می­پرسند آیا زبان انگلیسی را درک می­کنید؟ هر آموزگار زبانی می­داند که توانایی بیان شفاهی افراد به­جز در تک­گویی­ها، سخنرانی­ها، قرائت کردن­ها و امثال آن، دقیقاً به سودمندیِ توانایی درک شنیداری آن­ها است. در زندگیِ روزمرّه، چه در محیط کار و چه در بافت­های آموزشی و محیط خانه، درک شنیداری از لحاظ زمان، تعداد، واژگان ، دانش و توجّه – که همگی قابل اندازه­گیری هستند- به­مراتب بیشتر از صحبت کردن است، بنابراین، لازم است که توجّه دقیقی به به­منظور ارزشیابی مهارت شنیدن داشت (براون،2004: 119). در مباحث مرتبط با آموزش زبان دوم / خارجی،‌ آموزش و آزمون ارتباط تنگاتنگی با یکدیگر دارند. در امر آموزش زبان فارسی به غیرفارسی­زبانان به عنوان زبان دوم/ خارجی،‌ اگرچه در بحث آموزش با توجه به توسعه‌ی دوره‌های آموزش زبان فارسی به غیر فارسی زبانان و تدوین کتاب‌های مختلف آموزشی مرتبط در این زمینه و ...،‌ تحوّلات ارزشمندی صورت گرفته اما با وجود آثار مثبت آن، کاستی‌هایی را نیز به همراه دارد. مهمترینِ این نقص‌ها، گستره‌ی آزمون‌سازی زبان فارسی و اعتبار سنجی و اعتبار بخشی آزمون زبان فارسی و تهیه بانک سؤالات استاندارد در ۴ مهارت زبان فارسی است. با توجّه به اهمیّت مهارت شنیداری در امر آموزش زبان دوم / خارجی، در این پژوهش به­منظور ارزیابیِ مهارت درک شنیداری،‌ یکی از آزمون­های پایان دوره­ی سطح فرامیانی مرکز آموزش زبان فارسی دانشگاه بین­المللی امام خمینی (ره)،‌ برطبق استانداردهای آزمون­سازی مورد تجزیه و تحلیل قرارگرفته است.

**۲-چارچوب نظری پژوهش**

مهارت شنیداری در میان مهارت‌های زبانی از اهمیّت ویژه‌ای در آموزش زبان دوم / خارجی برخوردار است. کودکان هم مدّت‌ها فقط گوش می‌کنند و مقداری از دانش زبانی اولیّه‌ی آن‌‌‌‌‌ها صرفاً با شنیدن کسب می‌شود. کودک پانزده‌ماهه می‌تواند در مقابل گفتار‌هایی از قبیل "دست نزن" "به‌ من نگاه کن" "بیا اینجا" و بسیاری دیگر عکس‌العمل مناسب بروز دهد که نشانه‌ی درک درست وی از مفهوم این گفتارها و نشانه‌ی پردازش این گفتار‌ها برپایه‌ی دانشِ زبانی است که بوسیله‌ی شنیدن کسب کرده است. شاید به همین سبب باید آموزش زبان دوّم / خارجی را نیز با مهارت شنیدن آغاز کرد (ضیاءحسینی،1385 :110).

در گذشته، گوش دادن یک مهارت منفعل و غیر پویا به شمار می‌رفت، زیرا قابل مشاهده نبود و توجّه زیادی را در برنامه‌های درسی به خود جلب نمی‌کرد، دلیل دیگری که برای نادیده گرفتن گوش دادن در گذشته وجود داشت، تاکید زیادی بود که بر روی مهارت‌های تولیدی و صحبت کردن وجود داشت، امّا امروزه در انگاره‌های اخیر، گوش دادن، یک فرآیند فعّال و پویا به‌شمار می‌رود که می‌تواند تاثیر به‌سزایی در یادگیری زبان داشته باشد،‌ از این رو، توجّه خاصی را از سوی استادان زبان دوّم می‌طلبد (کومب و همکاران، 2007(

**۲-1 . سنجش و آزمودن مهارت شنیداری**

در این بخش موضوعاتی از قبیل ساختار آزمون مهارت شنیداری، زمان در آزمون مهارت شنیداری و عوامل مؤثر در سهولت و دشواری آزمون مهارت شنیداری مورد بررسی قرار می‌گیرد که باید برای طرّاحی و تهیّه‌ی آزمون مهارت شنیداری به آن توجّه نمود.

**۲-۱-۱ ساختار آزمون مهارت شنیداری**

بهتر است آزمون با یک سؤال آسان شروع شود تا باعث کاهش اضطراب آزمون‌دهندگان در آغاز آزمون شود. آزمودن طیف وسیعی از مهارت‌ها در خلال آزمون، اهمیّت بسیاری دارد. ترتیب چینش سؤال‌ها و ترتیب متن شنیداری باید یکسان باشد. همچنین سؤال‌ها باید در طول متن شنیداری پخش شود و توزیع آن تقریباً یکسان باشد. توصیه می‌شود که از 15 تا 20 ثانیه اوّل متن شنیداری سؤالی مطرح نشود تا آزمون‌دهندگان بتوانند خود را با فضای متن تطبیق دهند. نباید تنها به آزمودن درک جزئیّاتی مانند تاریخ و اعداد اکتفا کرد بلکه باید مهارت‌های شنیداری مختلف مانند درک موضوع اصلی، استنتاج ،‌نتیجه‌گیری و توانایی پاسخ به سؤالات را هم مورد ارزیابی قرار داد ( کومب و همکاران، 2007 : 96 ).

غالباً در آزمون‌های گوش دادن، از سؤال‌های عینی مانند سؤال‌های چند‌گزینه‌ای و درست و نادرست استفاده می‌شود زیرا هم مطمئن‌تر است و هم تصحیح و تحلیل کردن آن راحت‌تر است، در استفاده از این نوع سؤال‌ها به جای 4 گزینه پاسخ بهتر است که 3 گزینه ارائه شود چراکه حافظه‌ی کوتاه مدّت نقش مهمّی در درک شنیداری دارد و از آنجاییکه آزمون‌دهندگان دسترسی ِ متعدّدی به متن شنیداری ندارند، ‌تعداد گزینه‌های زیاد در آزمون گوش‌دادن، به بار حافظه‌ی آن‌ها می‌افزاید و پاسخ دادن به سؤال‌های آزمون را دشوارتر می‌کند. تصاویر هم معمولاً به عنوان بخشی از آزمون گوش‌دادن به‌کار می‌روند. در صورت استفاده از تصاویر، باید تصاویر واضح و مشخّص باشند ( همان:96).

برای دانشجویانی که می‌خواهند در فضایی ادامه‌ی تحصیل بدهند که تدریس به زبان مورد نظر آزمون انجام می‌گیرد، درک و فهم صحبت‌ها و سخنرانی‌های رسمی و غیر رسمی، یک مهارت ضروری برای آنان است، بنابراین در آزمون‌های گوش دادن باید از این متون نیز استفاده شود (فرهادی و همکاران، 2007: 220 ).

**۲-۱-۲ زمان در آزمون مهارت شنیداری**

طول یک آزمون گوش‌دادن به دو عامل بستگی دارد : طول نوار ضبط شده و تعداد تکرار شدن‌ متن شنیداری. آزمون‌های مهارتی تافل و آیلتس، به آزمون‌دهندگان اجازه می‌دهند که تنها یک بار متن شنیداری را بشنوند،‌ در حالیکه در آزمون‌های پیشرفت متن مورد نظر دو بار پخش می‌شود. باک (2001 ) پیشنهاد می‌کند که اگر هدف شما ارزیابی ِ ایده‌ی اصلی متن است بهتر است یک‌بار پخش شود، امّا اگر هدف شما ارزیابی جزئیّات است بهتر است متن شنیداری دو بار پخش شود (کومب و همکاران، 2007 : 97 ). طبق گفته‌ی کرول (1972) آزمون‌های شنیدن نباید بیشتر از 30 دقیقه طول بکشند. همچنین باید به آزمون‌دهندگان، زمانی را برای پیش‌خوانیِ سؤال‌های آزمون داده شود و آن‌ها در خلال شنیدن متن شنیداری، به سؤال‌ها پاسخ دهند. اگر قرار است که آنان پاسخ سؤالات را از برگه‌ی سؤال به پاسخ‌نامه منتقل نمایند باید مقداری زمان اضافی نیز در نظر گرفته شود (همانجا). در آزمون‌های مهارتیِ تافل و آیلتس زمان آزمون گوش‌دادن 30 تا 35 دقیقه است و 10 دقیقه‌ی اضافی نیز برای انتقال پاسخ‌ها به پاسخنامه در نظر گرفته می‌شود.

**رعایت توصیه‌های زیر در ارتقای کیفیِ آزمون مهارت شنیداری تاثیر به‌سزایی دارد:**

* انتخاب تصاویر در صورت استفاده، باید با دقّت بسیاری انجام گیرد. تصاویر باید ساده و قابل فهم باشند و برای تمام آزمون‌دهندگان تفسیر یکسانی داشته‌ باشند(فرهادی و همکاران ،‌2007 : 223).
* صحبت‌های موجود در متن شنیداری باید نمونه‌ای از فعّالیّت‌های گفتاری در زندگی واقعی باشد (همانجا).
* متون شنیداری نباید به‌نحوی باشد که فهمیدن آن‌ها مستلزم داشتن دانشی خاص (مانند داشتن تحصیلات در یک رشته‌ی دانشگاهی) باشد،‌ به عبارت دیگر متون آزمون باید عمومی باشند .
* در بسیاری از آزمون‌های گوش دادن به‌ویژه در آزمون گوش دادن دانشگاهی،‌آزمون‌دهندگان می‌خواهند در حین گوش دادن به یادداشت برداری بپردازند، ‌اینکه آنان مجاز به این کار باشند یا نه، بستگی به ساختار و اهداف آزمون دارد،‌ امّا هیل و کورتنی (1994) دریافتند که اجازه‌ی یادداشت‌برداری دادن به آزمون‌دهندگان نه تنها عملکرد آنان را در آزمون گوش دادن تافل بهبود نمی‌بخشد ،‌بلکه عملکردشان را هم تضعیف می‌نماید

(باک، 2001: 138).

* بریندلی (1998- نقل از باک ،2001:138) اشاره می‌کند که فاصله‌ی بین سؤال‌ها در متن شنیداری نباید به‌نحوی باشد که همزمان آزمون‌دهندگان در حال پاسخگویی به یک سؤال باشند و نیاز به گوش دادن برای پاسخ به سؤال بعدی را داشته باشند.

**۲-۱-۳ عوامل مؤثر در سهولت و دشواری آزمون مهارت شنیداری**

 **سرعت صحبت کردن،** یکی از عواملی است که باعث می‌شود مطلب شنیداری برای شنونده سخت یا آسان باشد. سرعت صحبت کردن شرکت‌کنندگان در یک متن شنیداری باید طبیعی باشد. سرعت طبیعی متن شنیداری 165- 180 واژه در دقیقه است. هلگسن ( 2003 : 33 ) علاوه بر متغیّر سرعت که باعث سختی یا آسانی ِمتن شنیداری می‌شود مسأله بار شناختی را هم مطرح می‌کند.

براون شش عامل که در افزایش یا کاهش راحتی درک مطلب شنیداری مؤثر هستند را به شرح زیر معرّفی می‌کند:

**1)‌ تعداد صداهای افراد یا اشیاء در یک متن :** مثلاً تعداد صداهای بیشتر دشواریِ متن را افزایش می‌دهد.

**2) میزان تمایز صداهای افراد یا اشیاء از یکدیگر :** مثلاً وجود صدای یک زن و یک مرد، درک متن را آسانتر می‌کند تا اینکه صدای دو مرد یا دو زن باشد.

**3) ارتباطات فضایی ساده و خاص، آسانتر از ارتباطات پیچیده فهمیده می‌شوند** : مثلاً در یک جمله‌ی امری مانند" مستقیم برو جلو" آسان‌تر از جمله‌ی " همینطور به راهت ادامه بده " درک خواهد شد.

**4) ترتیب وقایع :** ‌وقتی اطّلاعات دریافتی به‌گونه‌ای تنظیم شده باشد که وقایع در آن به ترتیب روی داده‌اند مطلب راحت­تر فهمیده می‌شود.

**5 ) تعداد استنباط‌های مورد نیاز** شنونده هر چقدر کمتر باشد او راحت‌تر متن را می‌فهمد .

**6 ) اطّلاعات و تجربیّات پیشین شنونده** در زمینه‌ی مطلبی که متن شنیداری درباره‌ی آن است باعث آسان‌ شدن درک متن خواهد شد.

**۲-۲ انواع ارزشیابی ِکنش‌های شنیداری**

**شنیدن فشرده**[[1]](#footnote-1) : آزمون شنیدن فشرده، ارزشیابیِ شناسایی عناصر واجی و ساختواژی زبان است .در این نوع ارزیابی،‌ محرکی گفتاری ارائه می‌گردد و از آزمون‌دهندگان خواسته می‌شود که از میان دو یا چند گزینه، ‌این محرّک را تشخیص دهند. این نوع فعّالیّت‌ها به دلیل نبود بافت،‌ واقعی به نظر نمی‌رسد.

**تشخیص دگرگفت**[[2]](#footnote-2):‌ در سنجش مهارت‌های خرد درک شنیدن، اغلب از طریق ارائه‌ی یک جمله‌ی محرّک و فراخواندن آزمون‌دهندگان به انتخاب دگرگفت درست از میان چند‌گزینه‌ به سنجش واژگان، عبارت و جملاتی می‌پردازد.

**شنیدن واکنشی**[[3]](#footnote-3) : شنیدن رشته کلامی نسبتاً کوتاه (‌احوال‌پرسی، پرسش، دستور، اطمینان از درک و غیره) به منظور بیان پاسخی کوتاه. یک قالب پرسش و پاسخ می‌تواند تعاملی را در این فعالیّت‌های شنیداری به وجود آورد واکنش آزمون‌دهنده، پاسخی مناسب به یک پرسش است.

**شنیدن گزینشی[[4]](#footnote-4)** : این نوع شنیدن، تحلیل رشته گفتمان‌هایی مانند تک‌گویی‌های کوتاه به مدّت چند دقیقه و به منظور دستیابی به اطلاعاتی خاص می‌باشد. هدف چنین کنشی الزاماً جستجوی معناهای کلّی و جامع نیست بلکه کسب توانایی در درک اطّلاعاتی به خصوص در بافتی متشکّل از رشته کلامی طولانی‌تر (مانند توضیحات یک آموزگار در کلاس، ‌خبرهای تلویزیونی یا رادیویی یا داستان) است. فعّالیّت‌های ارزشیابی در شنیدن گزینشی، ‌از زبان‌آموزان می‌خواهند که به عنوان مثال به اسامی، اعداد، توضیحات و یا واقعیّت‌ها و حوادثی خاص دقّت کنند.

**تکمیل کردن جملات** : نوعی از شنیدن گزینشی است که نیمی از جمله با تأکید ویژه‌ای شنیده می‌شود و آزمون‌دهنده باید از بین گزینه‌ها ادامه‌ی منطقی جمله را انتخاب کند .

**بندش شنیدن**[[5]](#footnote-5) :‌ فعالیّت‌های بندش شنیدن که گاهی دیکته‌های بندش و یا دیکته‌های ناقص نیز نامیده می‌شوند نوعی ارزیابی شنیدن گزینشی است که از آزمون‌دهنده می‌خواهند که به داستان، تک‌گویی یا مکالمه‌ای گوش کنند و به‌طور همزمان متنی نوشتاری را بخوانند که واژگان یا عبارت‌هایی گزینش از آن حذف شده‌اند. این آزمون شامل متنی است که با فاصله­های مشخّص ( معمولاً هفت واژه در میان) واژگانی از آن حذف شده‌اند و از آزمون‌دهندگان خواسته می‌شود که با واژه‌هایی مناسب جاهای خالی را پر کنند. یکی از نقاط ضعف روش بندش شنیدن این است که به راحتی تبدیل به فعالیّت‌های درک مطلب شوند به عبارت دیگر آزمون‌دهنده بدون گوش کردن موفّق به کامل کردن واژگان یا عبارات مناسب می‌شود. برای جلوگیری از چنین مشکلی باید واژه‌هایی برای جاهای خالی در نظر گرفته شود که بار اطّلاعاتی زیادی دارند چون در این شیوه ،‌هدف سنجش درک شنیداری است نه حدس زدن گرامر یا واژگان .

**انتقال اطّلاعات**[[6]](#footnote-6) : می‌توان شنیدن گزینشی را نیز به کمک تکنیک انتقال اطّلاعات ارزشیابی نمود. در این شیوه، ‌اطّلاعات شنیداری پردازش شده باید در قالب ارائه تصویر ،‌نامگذاری اجزای یک نمودار ،‌تشخیص یک شی‌ء در یک عکس، پر کردن فرم و یا نشان دادن مسیرها در یک نقشه منتقل شوند. فعّالیّت‌های انتقال اطّلاعات می‌توانند با استفاده از نمودار ،‌نقشه ،‌جدول، جدول زمانی و دیگر اطّلاعات زندگی روزمرّه واقعی‌تر به‌نظر برسند مانند : شنیدن برنامه‌ی هفتگی و پر کردن جدول .

**تکرار جملات**[[7]](#footnote-7) : فعّالیّت ساده‌ی تکرار یک جمله یا شبه جمله یا تکرار جملات نیز به عنوان نوعی از ارزشیابی درک شنیداری به کار گرفته می‌شود. مانند آزمون دیکته .آزمون‌دهندگان باید بتوانند یک رشته‌ی زبانی را برای مدّتی در ذهن خود نگاه دارند تا قادر باشند که آن را بیان کنند و سپس با تکرار شفاهی محرّک، واکنش نشان دهند. درک شنیداری نادرست ،چه در سطح واجی و چه در سطح گفتمان را می‌توان در تکرار نادرست مشاهده کرد. نادرستی در تکرار مشابه، نادرستی در درک شنیداری است. به‌عنوان مثال ،‌آزمون " فون پاس" [[8]](#footnote-8)یک نوع آزمون تجاری است که در ارزشیابیِ بیان شفاهی و درک شنیداری به شدّت بر تکرار جملات تاکید می‌ورزد. بدیهی است که ارزشیابی با روش تکرار جملات خالی از اشکال نیست (باک،2001 :79 )، چرا که این‌گونه فعّالیّت‌ها، آزمون‌ مهارت‌های شفاهی هستند و نه صرفاً آزمون‌های شنیداری. علاوه بر این ممکن است که این آزمودن فقط ارزشیابی تشخیص صداها باشد و عدم تکیه بر حافظه‌ی کوتاه مدّت می‌تواند از ارزش آن بکاهد و اعتبار آن را به عنوان شیوه‌ای برای ارزشیابی که فقط درک شنیداری را می‌سنجد را زیر سؤال ببرد، همچنین ممکن است که معلّم نتواند خطای درک شنیداری را از خطای بیان شفاهی تشخیص دهد .

**شنیدن گسترده** [[9]](#footnote-9) :در این نوع شنیدن هدفِ درک بالا به پایین، مورد نظر است. دامنه‌ی کنش گسترده، شنیدن سخنرانی‌های طولانی در یک سو و شنیدن یک مکالمه و استنباط و هدفی جامع در سوی دیگر را در بر می‌گیرد. شنیدن با هدف پی بردن به محتوای کلام پیام اصلی یا استنباط کردن همگی از انواع شنیدن گسترده هستند.

**دیکته[[10]](#footnote-10)** : یکی از ابزار‌های قدیمی ارزشیابیِ زبان خصوصاً درک شنیداری است. در دیکته، آزمون‌دهندگان به متنی که عموماً از 50 تا 100 واژه تشکیل شده است و سه بار تکرار می‌شود، گوش می‌دهند : نخست با سرعت معمولی، سپس با مکث‌هایی بلند بین عبارت‌ها و واژگان که در آن آزمون‌دهندگان آنچه را که لحظاتی قبل شنیدند را می‌نویسند و سرانجام بار دیگر با سرعت طبیعی تا نوشته‌هایشان را بررسی و ویرایش کنند. هیوز (1989)، کوهن (1994 )، بیلی (1998 ) و باک (2001 ) همگی از منطق بودن دیکته به عنوان آزمونی که لازمه‌ی آن سطحی از مهارت زبانی به منظور پردازش و نگارش درست تمامی اجزاست،‌ جانبداری می‌کنند. سطح دشواری فعّالیّت دیکته را می‌توان به کمک طول گروه‌های واژگانی، طول مکث‌ها، سرعت خوانده شدن متن ،‌پیچیدگی‌ِ گفتمان، دستور و واژگان کنترل کرد. البته قوه‌ی حافظه و پردازش معنا شناختی هم در این فعّالیّت مورد تاکید قرار می‌گیرند (باک، 2001 : 78 ). با وجود این، کاربرد‌پذیری برگزاری آزمون‌های دیکته، میزان مناسب از پایایی در یک سیستم نمره‌گذاری منسجم و ارتباط قوی با دیگر توانایی‌های زبانی سبب می‌شوند که آزمون دیکته در ارزشیابی‌های گسترده برای سنجش درک شنیداری به کار رود.

**بازگویی[[11]](#footnote-11)** : نوع دیگری از فعّالیّت های شنیدن گسترده، بازگویی است. دراین فعّالیّت آزمون‌دهندگان به یک داستان یا رویداد خبری گوش می‌کنند و فقط آن را بازگو و یا خلاصه می‌نمایند چه به‌طور شفاهی و چه به‌طور کتبی.آزمون‌دهندگان حین پاسخ دادن باید پیام، ‌مفهوم اصلی، هدف، دلایل و یا نتیجه‌گیری را تشخیص دهند تا نشان دهند که به‌طور کامل متن را درک کرده‌اند. نمره‌گذاری تا حدودی طبق معیار از پیش تعیین شده صورت می‌پذیرد زیرا می‌توان عناصر ضروری در بازگوکردن داستان را مشخّص کرد. در این نوع فعّالیّت نیز ممکن است که پایایی چندانی وجود نداشته باشد و زمان و زحمتی که صرف خواندن و ارزیابی پاسخ می‌شود ،‌از میزان کاربرد‌پذیری می‌کاهد. روایی، پردازش شناختی، توانایی ارتباطی و واقعی بودن به خوبی در این فعّالیّت مشاهده می‌شود (براون،2004 : 122 – 138).

به­منظور طراحی آزمون درک شنیداری از موارد زیر نیز می­توان بهره برد: حدس زدن، درک اعداد، درک الفبا، درک توصیفات، درک زبان، درک توالی یک رویداد، درک هم معنی‌ها، درک احساسات، درک توضیحات، درک طبقه­بندی، درک مقایسه­کردن، درک ترتیب زمانی اتفاقات (جلیلی،1390).

**2-3 معیارهای اعتبار**

**۲-۳-۱ ضریب دشواری سؤال[[12]](#footnote-12):**

درصد کل آزمون­شوندگانی که به یک سوال جواب صحیح داده­اند را ضریب دشواری گویند. که با حرف P نمایش داده می­شود.

**تعداد جوابهای صحیح گروه پایین + تعداد جوابهای صحیح گروه بالا**

 = ضریب دشواری سوال= P

**تعداد افراد گروه بالا + تعداد افراد گروه پایین**

هر اندازه ضریب دشواری یک سوال بزرگتر (به یک نزدیکتر) باشد آن سوال آسان­تر و هر اندازه این ضریب کوچک­تر (به صفر نزدیکتر) باشد سوال دشوارتر است. شاید بهتر باشد به جای ضریب دشواری واژه ضریب سهولت یا آسانی[[13]](#footnote-13) را به­کار برد.

**۲-۳-۲ ضریب تمیزسؤال[[14]](#footnote-14)**

برخلاف ضریب دشواری که میزان آسان یا سخت بودن یک سؤال را برای آزمودنی­ها نشان می­دهد، ضریب تمیز، قدرت سؤال را در تمایز و تشخیص گروه­های قوی و ضعیف آزمودنی­ها را نشان می­دهد که با حرف D مشخص می­شود. ضریب تمیز مشخص می­کند که سؤال تاچه اندازه توانسته گروه قوی را از ضعیف جدا سازد. ضریب تمیز زیر 3/0 قابل قبول نیست و برای محاسبه آن از این فرمول استفاده می شود:

 تعداد جواب­های صحیح گروه پایین - تعداد جواب­های صحیح گروه بالا

 = ضریب تمیز سوال=DI(d)

تعداد افراد یک گروه

D < 1 > 1- **ضریب تمیز بزگتر** = قوه تمیز بیشتر ، **ضریب کوچکتر**= قوه تمیز کمتر

**ضریب تمیزصفر**= عدم توانایی سوال در ایجاد تمایز بین گروه قوی وضعیف

**ضریب تمیز منفی** = گروه ضعیف به سوال پاسخ داده ولی گروه قوی خیر

اگر معلم بخواهد با استناد به یک نمرة واحد، چگونگی تمرکز نمرات یک کلاس را مشخص کند، از اندازه‌های گرایش مرکزی[[15]](#footnote-15) استفاده می‌کند. این اندازه‌ها عبارتند از نما[[16]](#footnote-16)، میانگین[[17]](#footnote-17) و میانه[[18]](#footnote-18).

**۳-۳-۳ نما**

 نمای هر دسته از نمرات عبارت است از نمره‌ای که بیشترین فراوانی را دارد. مورد استفادة نما زمانی است که معلم بخواهد با یک نگاه سطحی، نمره‌ای را به عنوان نمایندة نمرات گروه نشان دهد.

**۲-۳-۴ میانگین**

 میانگین یا معدل دسته‌ای از نمرات، بنابه تعریف، عبارت است از حاصل جمع همة آن نمرات بخش بر تعداد آن‌ها. فرمول محاسبة میانگین اعداد طبقه‌بندی نشده به قرار زیر است:

$$\overbar{X}=\frac{\sum\_{}^{}X}{N}$$

$$\sum\_{}^{}X=نمرات مجموع$$

***۲-۳ -۵ میانه***

*میانه نقطه‌ای است در روی مقیاس نمرات یا توزیع نمرات که نصف بالا و نصف پایین گروه نمرات را از یکدیگر جدا می‌کند. این نقطه‌ای است که 50 درصد نمرات در بالای آن و 50 درصد دیگر نمرات در پایین آن قرار دارند. در مواردی که نمرات اعداد یک توزیع فرد است، نمرة وسطی توزیع به عنوان میانه انتخاب می‌شود، اما در توزیعاتی که تعداد نمرات آن‌ها زوج است، دو نمرة میانی را باید با هم جمع و حاصل را بر 2 تقسیم کنیم تا میانة آن اعداد به دست آید (سیف 1390: 354-358)، یعنی:*

$$میانه=Md=\frac{میانی نمره دو مجموع}{2}$$

**۳- پیشینه­ی پژوهش**

موسوی (1379) در پژوهشی با عنوان«بررسی راهکار‌های نظری آزمون‌های بسندگی و تهیۀ یک آزمون بسندگی استاندارد زبان فارسی برای غیر­فارسی ‌زبانان»به مطالعۀ مبانیِ تهیۀ یک آزمون چهاربخشی پرداخته است که قصددارد مهارت زبانیِ فارسی­آموزان خارجی را بسنجد. چهار بخش آزمون مورد نظر، گوش دادن، ‌قواعد ساختاری، ‌واژگان و درک ‌خواندن است. در این آزمون دو مهارت تولیدی صحبت کردن و نگارش لحاظ نشده است و برای بخش دستور همانند تافل بخش ‌جداگانه‌ای در نظر گرفته شده است.

قنسولی (1389) در مقاله‌ای با عنوان «طراحی و روایی سنجی آزمون بسندگی زبان فارسی» از میان مدل‌های مختلف اجزای بسندگی زبان، مدل چند بعدی زبان را که با واقعیت شرایط آزمون تناسب دارد، ‌انتخاب کرده است. در این مدل به دو مهارت اصلی زبان مانند خواندن، ‌شنیدن و مقوله‌های دستور و واژگان توجه شده است. در طراحی این آزمون تنها به دو مهارت اصلی خواندن و شنیدن توجه شده و مهارت‌های گفتن و نوشتن به­دلیل محدودیت‌های تحقیق مانند پژوهش موسوی، لحاظ نشده است. در این پژوهش ضمن ساخت یک نمونه آزمون بسندگی زبان فارسی با تأکید بر آزمون‌های شنیدن و خواندن و دستور، ‌روایی و پایایی آزمون نیز سنجیده شده است. برای روایی‌سنجی آزمون،‌ از فرایند دو مرحله‌ای شامل تحلیل بازنگری سه دانشجوی خارجی دورۀ دکتری زبان و ادبیات فارسی و بهره‌گیری از مدل فولچر و دیویدسون بر مبنای نظریۀ بحث در سنجش روایی بهره گرفته شده است. این مدل چارچوبی مفهومیاست و اساساً روایی را فرایند بحث دربارۀ معنای نمرات و کاربرد نمرات می­داند. پایایی آزمون نیز به شیوۀ آلفای کرونباخ محاسبه شده است. نتایج این آزمون نشان داد که دانشجویان خارجی در بخش خواندن و درک مفاهیم بهترین عملکرد را داشتند و بخش دستور دشوارترین بخش برای آنان بود و پایایی آزمون نیز 70/0 تعیین شد.

 جلیلی (1390) در پژوهشی با عنوان "آزمون مهارتی فارسی برپایه‌ی چهار مهارت اصلیِ زبانی" تلاش کرده است یک آزمون مهارتی برای زبان فارسی طرّاحی نماید. تفاوت موسوی با جلیلی این است که موسوی قواعد ساختاری (دستور زبان) و واژگان (دایره لغات) را به‌عنوان دوبخش جداگانه در آزمون خود گنجانده است و دارای فرمت آزمون تافل است امّا جلیلی آزمون مهارتی زبان فارسی را بر پایه‌ی آیلتس طراحی نموده و آن را آمفا نامیده است. آزمون آمفا این دو بخش را به‌طور ضمنی در بخش‌های دیگر می‌آزماید و نیازی به جداسازی آن‌‌‌‌ها نمی‌بیند. به علاوه ‌دو بخش نوشتن و صحبت کردن که از مهارت‌های زبانیِ ضروری برای دانشجویان خارجی هستند در پایان‌‌نامه‌ی موسوی نادیده گرفته شده است.

 شقاقی و همکاران (1390) در مقاله‌ای با عنوان " هنجاریابیِ آزمون بین­المللی زبان فارسی" به هنجاریابی آزمون استاندارد کانون زبان پرداخته است. این آزمون به ترتیب شامل بخش‌های شنیدن (30 سؤال)، دستور (30 سؤال )،واژگان و خواندن (هر یک 20 سؤال)‌ است و مهارت صحبت کردن در این آزمون لحاظ نگردیده و همانند آزمون تافل برای دستور بخش جداگانه‌ای در نظر گرفته شده است.

ممقانی (1390) در پژوهشی با عنوان "ارزیابی روایی و پایایی آزمون وزارت علوم و تحقیقات با هدف دستیابی به یک آزمون مهارت برای زبان فارسی" به ارزیابی پایایی و روایی آزمون زبان انگلیسی وزارت علوم و تحقیقات و فناوری (ام اِس آر تی )پرداخته است. هدف از این پژوهش این بوده است که آیا می‌توان از ارزیابی نتایج روایی و پایایی آزمون (ام اِس آر تی) به یک الگوی مناسب برای طراحی یک آزمون مهارت زبان فارسی رسید. در همین راستا از دو آزمون دیگر یعنی آمفا (آزمون مهارتی فارسی) ‌و آزمون کانون زبان ایران نیز استفاده شد و روایی و پایایی این دو آزمون نیز در این پژوهش مورد بررسی قرار گرفتند. از داده‌های این پژوهش این نتایج به دست آمد که از آزمون(ام اِس آر تی ) نمی‌توان برای طرّاحی یک آزمون مهارت فارسی استفاده کرد چون مبنای آن تافل است و آزمون مهارتی آیلتس، ‌آزمون مناسب‌تری برای سنجش توانایی‌هایِ زبانیِ فارسی آموزان است چون دارای بخش صحبت کردن است و همچنین دو بخش دستور و واژگان را جداگانه مورد سنجش قرار نداده است،‌ در صورتی که آزمون تافل فاقد چنین ویژگی‌هایی است. بنابراین نتیجه‌گیری شد که از میان دو آزمون آمفا و کانون زبان ایران،‌ آزمون آمفا مبنای مناسب‌تری برای ارزیابیِ مهارت‌های زبانیِ غیر فارسی‌زبانان است.

در گذشته پژوهش‌هایی درمورد نوع سؤالات آزمون شنیدن مانند چند‌گزینه‌ای[[19]](#footnote-19)، پرکردن جای خالی[[20]](#footnote-20)، سؤالات کوته‌پاسخ[[21]](#footnote-21) و سؤالات باز پاسخ[[22]](#footnote-22) انجام شده است به عنوان مثال ایکین[[23]](#footnote-23) (1992) به‌منظور ارزیابیِ مهارت شنیدن زبان‌آموزان فرانسه، سؤالات چند‌گزینه‌ای را با دیگر سؤالات مورد بررسی قرار داد و در نهایت مشخّص شد که آزمون‌دهندگان از سؤالات چهار گزینه‌ای نمره‌ی بهتری گرفتند.

تِنگ[[24]](#footnote-24) (1998)، سه نوع سؤالات چند‌گزینه‌ای، بندش و سؤالات کوته پاسخ را مورد بررسی قرار داد و نتیجه گرفت که دانشجویان دانشگاهی در پاسخ به سؤالات چندگزینه‌ای موفّق‌تر بودند.

مطالعه‌ی دیگری که در سال (2004) توسط چِنگ[[25]](#footnote-25) انجام گرفت، مشخّص کرد که آزمون‌دهندگان در سؤالات چند‌گزینه‌ای و بندش[[26]](#footnote-26) در مقایسه با سؤالات کوته‌پاسخ، عملکرد بهتری داشتند.

چَنگ (2005)[[27]](#footnote-27) در بررسی‌های خود نشان داد که آزمون‌دهندگانی که سطح مهارت زبانیِ آن‌ها در سطح مقدّماتی تا متوسّط است در سؤالات چهارگزینه‌ای نمرات بالاتری کسب کرده‌اند. همچنین وی پیشنهاد نمود که آزمون‌هایی که مستلزم نوشتن است با ساختار توانایی ِشنیدن همسو نیست چرا که تقاضایی به جز درک شنیداری دارد.

چَنگ (2005) و وو [[28]](#footnote-28)(1998) خاطر نشان کردند که اگرچه سؤالات چند‌گزینه‌ای نسبت به سؤال‌های بازپاسخ، آسان‌تر است امّا شواهدی از تحقیقاتی دیگر به‌دست آمده که این نوع آزمون برای زبان‌آموزانی که سطح مهارت زبانیِ‌آن‌ها پایین است، مفید نیست چون نیازمند مهارت خواندن است. "وو" بیان کرد که سؤالات چند‌گزینه‌ای اگر برای زبان‌آموزان سطح پایین استفاده شود پایایی ِآزمون شنیدن را تحت‌تاثیر قرار می‌دهد .

پژوهش دیگری که توسّط فریدل[[29]](#footnote-29)و کوستین [[30]](#footnote-30)(2004) انجام گرفت حاکی از این امر است که علاوه بر نوع سؤالات (چندگزینه‌ای یا بازپاسخ)‌، عوامل مهم دیگری در سهولت و دشواریِ آزمون شنیدن مؤثر است .به عنوان مثال، ‌اگر اطّلاعات مورد نیاز برای پاسخ‌دهی به سوال آزمون در ابتدای متن شنیده شده، باشد آسان‌تر از زمانی است که در وسط یا انتهای متن باشد و اگر واژه‌هایی کi در سؤال استفاده می‌شود با واژه‌های متن همپوشانی داشته باشد نیز ضریب سهولت آزمون نیز بیشتر خواهد شد.

یاناگاو[[31]](#footnote-31)ا و گرین[[32]](#footnote-32) (2008) هم در پژوهشی نیز به این نتیجه رسیدند که سؤالات استنباطی[[33]](#footnote-33) نسبت به سؤالات عینی[[34]](#footnote-34) در مقوله‌ی دشواریِ‌ آزمون تاثیرگذار است.

 آنّا[[35]](#footnote-35) و چنگ[[36]](#footnote-36) و رید[[37]](#footnote-37) (2013) هم در پژوهشی بیان کردند که نوع سؤالات آزمون بر عملکرد زبان‌آموزان در آزمون تاثیر بسیاری دارد آن‌ها با مقایسه سؤالات چندگزینه‌ای شفاهی و کتبی در آزمون درک شنیداری به این نتیجه رسیدند که زبان‌آموزانی که در سطح پایین‌تری قرار دارند در سوالات چند‌گزینه‌ای کتبی نمره بالاتری گرفتند ولی دانشجویان پیشرفته در هر دو حالت نمره‌ی یکسانی کسب کردند و این در حالی بود که در بررسیِ نظرسنجی زبان‌آموزان 75/0 آنان سؤالات چندگزینه‌ای کتبی را ترجیح داده بودند.

**۴- روش­شناسی پژوهش**

ابتدا به منظور ارائه مبانی و مفاهیم نظری جهت مطالعه‌ی‌ پژوهش‌های قبلی و چارچوب نظری از روش توصیفی- تحلیلی و به منظور تعیین اعتبارسنجی آزمون موردنظر، از روش میدانی استفاده شده است . در این پژوهش، به­منظور دسترسی به یک آزمون استاندارد و تهیه بانک سؤال­های استاندارد، یکی از آزمون­های طراحی شده و برگزارشده­ی مهارت شنیداری سطح فرامیانی مرکز آموزش زبان فارسی دانشگاه بین­المللی امام خمینی(ره)  مورد تحلیل و بررسی قرار گرفته است. که ابزار این پژوهش است. این آزمون درک­شنیداری سه­گزینه­ای و شامل ۲۰ پرسش از نوع درک موضوع اصلی متن،‌ درک موضوعات بیان شده در یک مصاحبه،‌درک توالی یک رویداد، درک مقایسه کردن، درک طبقه‌بندی، تشخیص درست یا نادرست بودن جمله، دگرگفت و بندش بوده است که زبان­آموزان با توجه به فایل صوتی شنیده شده به پرسش­ها پاسخ داده بودند. متون فایل صوتی نیز شامل موضوعات آموزشی،‌ اجتماعی، پزشکی،‌ فرهنگی،‌ محیط زیست و تغذیه بدون نیاز به دانش و تخصصی خاص بوده است. نمونه­ی آماری این پژوهش ۵۹ نفر از فارسی­آموزان مرکز آموزش زبان فارسی دانشگاه بین­المللی امام خمینی(ره) بوده که از لحاظ زبانی در سطح فرامیانی و در رده‌ی سنّی 18 تا 30 قرار داشتند و به‌طور تصادفی انتخاب شده بودند. متغیرهای این پژوهش شاخص ضریب دشواری، ضریب تمیز و پایایی پرسش­های آزمون درک شنیداری بود.

**۵- ارائه و واکاوی داده­ها**

هدف از تحلیل سؤال‌های یک آزمون وارسی ِتک‌تک سؤالات و تعیین میزان دقّت و نارسایی آن‌هاست. در تحلیل سؤال‌های آزمون،‌ نقاط ضعف و قوّت یک آزمون و کیفیّت سؤالات آن تعیین می‌شود. بنابراین لازم است که آزمونگر ‌پس از اجرای هر آزمون، ‌سؤال‌های آن را مورد تحلیل قرار دهد و با استفاده از نتایج حاصل از آن به تجدید نظر در آزمون و بهبود کیفیّت سؤال‌ها برای استفاده‌های بعدی اقدام نماید.(سیف، 1390: 421 ). در فرآیند تحلیل سؤالات یک آزمون، میزان ضریب دشواری،‌ ضریب تمیز،‌ همبستگی سؤال با کل آزمون و پایایی و روایی آن سنجیده می‌شود که در این پژوهش به میزان ضریب دشواری،‌ ضریب تمیز و پایایی پرداخته شده است. جد.ل زیر نتایج این واکاوی را نشان می­دهد:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| پایایی | ضریب تمیز  | ضریب دشواری | گزینهپ | گزینهب | گزینه الف | شماره پرسش |
| **704/0** | 0.06 | 86% | 51 | 7 | 1 | 1 |
| **709/0** | 0.06 | 10% | 50 | 2 | 6 | 2 |
| **724/0** | 0.13 | 94% |  | 56 | 3 | 3 |
| **714/0** | 0.06 | 84% | 50 | 4 | 5 | 4 |
| **714/0** | 0.03 | 84% | 7 | 50 | ۲ | 5 |
| **697/0** | 0.31 | 77% | 7 | 46 | 6 | 6 |
| **۶۰۳/0** | 0 | 1 | - | - | 59 | 7 |
| **698/0** | 0 | 89% | 2 | 53 | 4 | 8 |
| **708/0** | 0.03 | 94% | 56 | 1 | 2 | 9 |
| **728/0** | 0.17 | 74% | 4 | 11 | 44 | 10 |
| **714/0** | 0.37 | 67% | 11 | 40 | 8 | 11 |
| **713/0** | 0.20 | 89% | 4 | 53 | 2 | 12 |
| **699/0** | 0.10 | 83% | 3 | 49 | 7 | 13 |
| **701/0** | 0.10 | 15% | 39 | 11 | 9 | 14 |
| **721/0** | 0 | 1 | 59 | - | - | 15 |
| **697/0** | 0.06 | 96% | - | 1 | 57 | 16 |
| **711/0** | 0.13 | 93% | 55 | 1 | 3 | 17 |
| **700/0** | 0.34 | 49% | 18 | 29 | 12 | 18 |
| **702/0** | 0.41 | 64% | 38 | 9 | 12 | 19 |
| **705/0** | 0.06 | 96% | 1 | 1 | 57 | 20 |

**نمره میانگین، میانه و نما آزمون شنیدن**

10+10+11+12+13+13+13+14+14+14+14+14+14+15+15+15+15+15+15+15+15+15+ 15+15+16+16+16+16+16+16+16+16+16+16+16+16+16+16+16+16+16+17+17+17+17+17+17+17+17+17+17+17+17+18+18+18+18+18+19=899

59=15.23 ÷ 899 = میانگین

16= نما

16= میانه

**۵-۱ واکاوی ضریب دشواری و ضریب تمیز و گزینه­های انحرافی**

استفاده از پرسش­های سه­گزینه­ای برای آزمون­های درک شنیداری بهترین انتخاب هستند، چرا که از روایی بیشتری برخوردار است چون اگر چهارگزینه باشد مهارت خواندن نیز مورد ارزیابی قرار می­گیرد و ضعف در درک خوانداری ممکن است بر روی ارزیابی درک شنیداری و نیز روایی و پایایی آزمون تأثیرگذارباشد. در ادامه تک­تک سؤالات آزمون مورد بحث قرار می­گیرد:

**سوال اول : موضوع اصلی متن شنیداری پخش شده چیست؟**

الف) راه­های جلوگیری از مسمومیت غذایی (۱) نفر

ب) عوارض مسمومیت غذایی (۷) نفر

پ) عوامل ایجاد مسمومیت غذایی (۵۱) نفر

طبق داده­های آماری گزینه­ی (الف) 1.7% و گزینه (ب) 11.9% و گزینه (پ) 86.5% را بخود اختصاص داده­اند. یعنی یک نفر گزینه الف، هفت نفر گزینه ب و 51 نفر گزینه پ را انتخاب کرده­ است. با در نظر گرفتن آمار نتیجه می­گیریم که گزینه­های انحرافی مناسب نیستند و نتوانسته­اند عملکرد خوبی داشته باشند. با بدست آمدن ضریب دشواری86% نتیجه می­گیریم که این سوال ساده است چراکه هرچه ضریب دشواری (سهولت) بزرگ­تر باشد سوال ساده­تر خواهد بود. ضریب­تمیز این پرسش نیز 0.06 است که با توجه به توضیحات پیشین قابل قبول نیست زیرا کمتر از ۳/۰ است.

**سوال دوم: موضوع اصلی متن شنیداری پخش شده چیست؟**

الف) اهمیت تحصیل زنان (۶) نفر

ب) علت مرگ و میر نوزادان (۲) نفر

پ) مراقبت­های دوران بارداری (۵۰) نفر

از بین سه گزینه،‌ موضوع گزینه­­ی (الف) و (ب) تقریبا دریک حوزه است اما گزینه (پ) کاملا متفاوت است و همین امر ممکن است درانتخاب آن به­عنوان گزینه­ی درست تأثیرگذار باشد. طبق داده­های آماری 6 نفر گزینه (الف) یعنی حدود 10% و 2 نفرگزینه (ب) حدود 3.4% و 50 نفر گزینه (پ) حدود 85% را انتخاب کرده­اند. با توجه به اینکه گزینه­ی درست (پ) بوده است گزینه­های انحرافی عملکرد مناسبی نداشتند.ضریب دشواری این پرسش سوال 85% محاسبه شده است که باتوجه به اینکه ضریب دشواری مطلوب بین ۷/۰ p< >3/0 است،‌ بنابراین، این پرسش درگروه پرسش­های ساده قرار خواهد گرفت.ضریب تمیز سوال نیز ۰۶/۰- است و با توجه به این­که ضریب تمیز 3/۰ D< قابل قبول نیست. با توجه به این که ضریب تمیز این پرسش منفی است این بدان معنی است که گروه ضعیف از گروه قوی بهتر عمل کرده­اند.

**سوال سوم: با توجه به متن شنیداری پخش شده، رشته­ی تحصیلی و تخصص خانم دکتر دلیرانی چیست؟**

الف) کارشناس زیست شناسی (۳ نفر)

ب) متخصص محیط زیست (۵۶)

پ) کارشناس اقتصادی (۰)

در این پرسش،‌گزینه (الف) را سه نفر یعنی حدود 5% ، گزینه (ب) را 56 نفر یعنی 95% و همانطور که داده­های این پرسش نشان می­دهد گزینه­ی انحرافی (پ)‌ اصلا انتخاب نشده است چون متن در مورد موضوعات اقتصادی نبوده است بنابراین،‌ این گزینه­ی انحرافی نیاز به بازنگری و اصلاح دارد. ضریب دشواری(سادگی) سوال 94% می باشد که باتوجه به اینکه هرچه ضریب دشواری بزرگ­­ترباشد سوال ساده­تر خواهد بود و ضریب دشواری مطلوب۷/۰p<>3/0 می باشد لذا این سوال درگروه سوال­های بسیار ساده خواهد بود. ضریب تمیز سوال ۱۳/۰ است که با توجه به اینکه هرچه ضریب تمیز سوال بزرگ­تر باشد قدرت جداسازی سوال بیشتر است، ضریب تمیز 3/0 D< قابل قبول نیست.

**سوال چهارم: مهمترین علت خشک شدن دریاچه ارومیه، چه می باشد؟**

الف) شوری بیش از اندازه ی دریاچه (۵ نفر)

ب) ساختن سد بر روی این دریاچه (۴ نفر)

پ) کاهش میزان بارندگی به طور مداوم ( ۵۰ نفر)

در این پرسش بدون نیاز به شنیدن فایل صوتی می­توان گزینه­ی موردنظر را انتخاب کرد چرا که گزینه­ی (الف) که نمی­تواند علت خشک شدن هر دریاچه­ای باشد و از انتخاب ها کنار می رود. گزینه­ی (ب) نیز به نظر اشکال دارد چرا که سد را برروی رودخانه احداث می­کنند. این گزینه مورد مناسبی برای انتخاب نمی­تواند باشد بنابراین انتخاب گزینه (پ) حتی بدون گوش کردن به فایل صوتی محتمل­ترین انتخاب می باشد. با توجه به داده­ها، گزینه (الف) را 5 نفر یعنی 8.5% گزینه (ب) را 4 نفر یعنی حدود 8.5% و گزینه (پ) را 50 نفر، حدود 85% انتخاب کرده­اند. این خود نشان دهنده نامناسب بودن گزینه­های انحرافی و سادگی سوال می باشد. ضریب دشواری این سوال 84% می باشد باتوجه به اینکه هرچه ضریب­ دشواری بزرگ­­ترباشد سوال ساده­تر خواهد بود و ضریب دشواری مطلوب۷/۰p<>3/0 می باشد لذا این سوال درگروه سوال­های بسیار ساده خواهد بود. ضریب تمیز سوال 0.06 است با توجه به اینکه هرچه ضریب تمیز سوال بزرگ­تر باشد قدرت جداسازی سوال بیشتر است، ضریب تمیز 3/0 D< قابل قبول نیست.

**سوال پنجم: آرتمیا .......................................................................................................................................................**

الف) درحال حاضر ازدریاچه ارومیه صید می شود ودرصنعت پرورش ماهی از آن استفاده می گردد. ( ۲ نفر)

ب) به دلیل داشتن ارزش غذایی بالا، منبع خوبی برای پرورش ماهی است. (۵۰ نفر)

پ) در آب های شیرین زندگی می کند وغذای اصلی ماهی های آب شیرین است. (۷ نفر)

این پرسش از نوع جای خالی بوده و هر سه گزینه در صورت نداشتن اطلاعات در مورد متن می­توانند گزینه­ی انتخابی باشند. در این پرسش ۲ آزمودنی گزینه الف را انتخاب کرده­اند که حدود 3.4% را بخود اختصاص می­دهد و 50 آزمودنی ( 85% ) گزینه ب گزینه پ را نیز 7 آزمودنی یعنی14% انتخاب کرده­اند. با توجه با آمار ارائه شده ضریب دشواری این پرسش 84% بوده است که با توجه به توضیحات پیشین و موارد ذکر شده در پرسش­های بالا این پرسش نیز بسیار ساده است. ضریب تمیز این پرسش 0.03 بوده و با توجه به توضیحات پیشین این ضریب تمیز نیز قابل قبول نیست.

**سوال 6: با توجه به متن، کدام یک از جمله های زیر نادرست است؟**

الف) در حال حاضر، تعداد پرندگانی که به دریاچه ارومیه مهاجرت می کنند، کمتر شده است. (۶ نفر)

ب) امروزه، دریاچه ی ارومیه یکی از مهمترین منابع تولید آرتمیا در جهان به شمار می رود. (۴۶ نفر)

پ) خشک شدن دریاچه ارومیه، باعث کاهش محصولات کشاورزی نیز می شود. ( ۷ نفر)‌

از آنجایی که گزینه­ی (ب) یکی ازمهمترین منابع تولید درجهان،‌ دارای بارمعنایی مثبت و گزینه­های (الف) و (ب) – کمتر شدن مهاجرت پرندگان و کاهش محصولات کشاورزی، دارای بار منفی است همین تفاوت معنایی گزینه­ی صحیح را نشاندار کرده است. گزینه (الف) را 6 نفر یعنی کمی بیش از 10%، گزینه ی (ب) را 46 نفر، حدود 78% و گزینه ی (پ) توسط 7 نفر که حدود 12% نفر ازپاسخ دهندگان انتخاب کرده­اند. . ضریب دشواری این پرسش ۷۷/۰ و ضریب تمیز این پرسش 0.31 است که ضریب تمیز قابل قبولی می باشد با در نظر گرفتن ضریب دشواری این پرسش می توان گفت که به نسبت پرسش های پیشین این پرسش بهتر عمل کرده و با ضریب تمیز 0.31 نسبت به پرسش­های پیشین طراحی بهتری داشته است.

**سوال 7: مهمترین دلیل ترس از امتحان چیست؟**

الف) نتیجه ی امتحان ( نفر ۵۹)

ب) سؤالات سخت (‌۰ )

پ) مرور نکردن (۰ )

سؤال به گونه­ای طراحی شده است که درابتدا هر شخص بدون شنیدن متن شنیداری و دیدن گزینه های پاسخ، درذهن خود پاسخی برای آن دارد. درادامه با دیدن گزینه­ها و بعد با شنیدن متن به تطبیق پاسخ خود با آنها می پردازد. به نظر گزینه­های خوبی برای انحراف انتخاب نشده­اند چراکه اولاً نتیجه­ی امتحان مهمترین عامل ترس می­باشد و اینکه سؤال سخت و مرور نکردن درنهایت درنتیجه تأثیر دارد. با بررسی داده­ها مشخص شد که تمامی داوطلبان گزینه ی الف را انتخاب کرده­اند پس پرسش بسیار ساده بوده که همه توانسته­اند پاسخ صحیح را انتخاب کنند و گزینه­های انحرافی اصلا نتوانسته­اند خوب عمل کنند. ضریب دشواری این پرسش 1 بوده که باتوجه به اینکه ضریب دشواری مطلوب p<./7 >3/0 است به این نتیجه می رسیم که این پرسش بسیار ساده است. ضریب تمیز نیز 0 است پس نتیجه می­گیریم که ضریب تمیز این پرسش نیز غیرقابل قبول است.

**سوال 8: با توجه به متن کدام یک از گزینه های زیر درایجاد اضطراب ناشی از امتحان نقش مهمی ندارد؟**

الف) انتظار بیش از حد والدین (۴ نفر)‌

ب) آماده نبودن برای امتحان ( ۵۳ نفر)‌

پ) نا مشخص بودن سؤالات ( ۲ نفر)‌

هر سه گزینه را می­توان از عوامل ایجاد اضطراب ناشی از امتحان دانست. بنابراین برای پیدا کردن پاسخ صحیح احتیاج به شنیدن متن شنیداری می باشد. هرچند که موضوع گزینه­ی (الف) کمی با دیگر گزینه­ها متفاوت است. گزینه ی (الف) توسط 4 نفر، حدود 6.8%، گزینه (ب) را 53 نفر، حدود 90% ، و گزینه (پ) را 2 نفر یعنی چیزی حدود 3.4% انتخاب کرده­اند. ضریب دشواری این پرسش 89% که با توجه به توضیحات پیشین، این پرسش در میان سوال­های بسیار ساده قرار می گیرد. ضریب تمیز این پرسش نیز 0 است که با توجه به اینکه ضریب تمیز 3/0 D< قابل قبول نیست، نتیجه می­گیریم که ضریب تمیز این پرسش نیز غیرقابل قبول می باشد.

**سوال 9: براساس متن برای کاهش اضطراب چه روشی مفید است؟**

الف) رفع اشکالات از معلم (۲ نفر)

ب) مرور مطالب به صورت گروهی (۱نفر)

پ) شرکت در آخرین جلسات کلاس ( ۵۶)

در این پرسش به نظر می رسد که گزینه­های انحرافی بخوبی ارائه نشده­اند چراکه گزینه (الف) دارای ابهام معنایی است و مشخص نیست که واژه اشکالات به چه چیزی و چه کسی برمی­گردد. آیا معلم اشکالاتی دارد که باید رفع شود؟ یا مشکلات درسی؟ به هر حال این گزینه را 2 نفر انتخاب کرده­اند که 3.4% از کل آمار را به خود اختصاص داده است. 1نفر گزینه (ب) که 1.7% بوده و 56 نفر گزینه (پ) یعنی 95% را انتخاب کرده­اند. ضریب دشواری این پرسش %94 است که بازهم درگروه پرسش­های بسیار ساده قرار می گیرد. ضریب تمیز این پرسش نیز 03/0 است که غیرقابل قبول می باشد.

**سوال 10: با توجه به متن، ویژگی پشه ماده کدام است؟**

الف) برای تولید مثل از خون انسان تغذیه می­کند. (۴۴ نفر)‌

ب) از عصاره ی گیاهان تغذیه می­کند. (۱۱ نفر)

پ) این نوع پشه وز وز نمی­کند. ( ۴ نفر)

سوال آزمون شنیداری باید تا حد امکان، وابسته به متن شنیداری باشد نه اطلاعات عمومی آزمون­دهنده. در این پرسش اگر آزمون­دهنده بداندکه پشه­ی ماده برای تولید مثل به خون انسان احتیاج دارد، این سؤال و متن مربوطه بطورکلی بی فایده است چون آزمون شنیداری می باشد و آزمون اطلاعات عمومی نیست. بنابراین، از روایی آن کاسته می­شود. گزینه (الف) را 44 نفر، حدود 75% و گزینه­ (ب) را 11 نفر حدود 18.6% و گزینه (پ) را 4نفر حدود 6.8% به­عنوان پاسخ درست انتخاب کرده­اند. که این داده­ها نشان­دهنده­ی موفقیت این سؤال و گزینه­های انحرافی آن می باشد. ضریب دشواری سوال74% می باشد. و ضریب تمیز آن 17/0 می­باشد. هرچند از محدوده قابل قبول فاصله دارد اما با توجه به گزینه­های انحرافی مناسب­تر،‌ پرسش قابل قبول­تری نسبت به سؤالات قبلی است.

**۱۱- با توجه به متن کدام گزینه نادرست است ؟**

الف - پشه نر شاخک دارد و وزوز نمی کند. (۸نفر )

ب - تعداد کمی از افراد به نیش پشه حساسیت دارند. ( ۴۰ نفر )

پ - پشه نر، درشت­تر از پشه ماده است. (۱۱ نفر)

در این پرسش،‌ گزینه (الف) را ۸ نفر، گزینه (ب) را 4۰ نفر، و گزینه (پ) توسط ۱۱ نفر ازآزمودنی­ها انتخاب کرده­اند. . ضریب دشواری این پرسش **۶۷٪**و ضریب تمیز این پرسش **۳۷/ ۰** است که ضریب تمیز قابل قبولی می­باشد با در نظر گرفتن ضریب دشواری و ضریب تمیز این پرسش و همچنین به­علت این که ۴۹ نفر گزینه­های انحرافی را انتخاب کرده­اند می­توان نتیجه گرفت که این پرسش سؤال مناسبی است.

**۱۲ - با توجه به متن، کدام گزینه در مورد خرافات درست نیست؟**

الف - بخشی از آداب و رسوم ، گذشتگان ما هستند. ( ۲ نفر)

ب – می­توان باورهای خرافی را ترک کرد. ( ۵۳ نفر )

پ - باوری است که بر اساس اصول علمی نمی باشد. ( ۴نفر )

پرسش درست / نادرست،‌ یکی از انواع پرسش­های پربسامد آزمون­های درکی زبان است. در این پرسش گزینه (الف) توسط ۲ نفر، گزینه (ب) را 53 نفر و گزینه (پ) را ۴ نفر انتخاب کرده­اند. ضریب دشواری این پرسش 89% که با توجه به توضیحات پیشین، این پرسش در میان سوال­های بسیار ساده قرار می گیرد. ضریب تمیز این پرسش نیز**۲۰/ ۰** است که با توجه به اینکه ضریب تمیز 3/0 D< قابل قبول نیست، نتیجه می­گیریم که ضریب تمیز این پرسش نیز غیرقابل قبول می باشد.

**13- چرا ما تحت تاثیر خرافات قرار می گیریم؟**

الف - چون از گذشتگان به ما رسیده (۷ نفر)

ب - به دلیل ترس و احساس درونی ( ۴۹ نفر)

پ - چون بدیمن هستند ( ۳ نفر)

در این پرسش گزینه (الف) توسط ۷ نفر، گزینه (ب) را ۴۹ نفر و گزینه (پ) را ۳ نفر انتخاب کرده­اند. ضریب دشواری این پرسش 8۳% که با توجه به توضیحات پیشین، این پرسش در میان سوال­های بسیار ساده قرار می گیرد. ضریب تمیز این پرسش نیز**۱۰/ ۰** است که با توجه به اینکه ضریب تمیز 3/0 D< قابل قبول نیست، نتیجه می­گیریم که ضریب تمیز این پرسش نیز غیرقابل قبول می باشد.

**14 - کدام گزینه، از دلایل گرایش به خرافات است ؟**

الف - تحصیلات پایین (۹نفر)

ب - اطلاع از نتایج رواج خرافی ( ۱۱نفر )

پ - نداشتن اعتماد به نفس ( ۳۹نفر)

در این پرسش گزینه (الف) توسط ۹ نفر، گزینه (ب) را ۱۱ نفر و گزینه (پ) را ۳۹ نفر انتخاب کرده­اند. ضریب دشواری این پرسش ۱۵% که با توجه به توضیحات پیشین، این پرسش در میان سوال­های بسیار سخت قرار می گیرد. ضریب تمیز این پرسش نیز**۱۰/ ۰** است که با توجه به اینکه ضریب تمیز 3/0 D< قابل قبول نیست، نتیجه می­گیریم اگر چه گزینه­های انحرافی نقش خود را به­خوبی ایفا کرده است اما ضریب دشواری و ضریب تمیز این پرسش قابل قبول نمی­باشد.

**15- کشور ژاپن :**

الف - از نظر رتبه سلامتی جهان در جایگاه دوم می باشد ( ۰)

ب - از لحاظ سلامتی غذایی در رتبه اول جهان قرار دارد ( ۰ )

پ - به دلیل افزایش طول عمر رتبه اول جهان را کسب کرده است ( ۵۹ )

با بررسی داده­ها مشخص شد که تمامی داوطلبان گزینه ی(پ)‌ را انتخاب کرده­اند پس پرسش بسیار ساده بوده که همه توانسته­اند پاسخ صحیح را انتخاب کنند و گزینه­های انحرافی اصلا نتوانسته­اند خوب عمل کنند. ضریب دشواری این پرسش 1 بوده که باتوجه به اینکه ضریب دشواری مطلوب p<./7 >3/0 است به این نتیجه می­رسیم که این پرسش بسیار ساده است. ضریب تمیز نیز 0 است پس نتیجه می­گیریم که ضریب تمیز این پرسش نیز غیرقابل قبول است.

**16- علت معروف بودن غذاهای فرانسوی چیست؟**

الف- مزه کیفیت غذا برای آنان از اهمیت زیادی برخوردار است ( ۵۷)

ب - غذاهای فرانسوی رنگ و تنوع زیاد دارد ( ۲ )

پ - آنها از ادویه و چاشنی های مخصوص استفاده می کنند ( ۰ )

گزینه (الف) توسط ۵۷ نفر، گزینه (ب) را ۲ نفر، و گزینه (پ) را کسی انتخاب نکرده است. ضریب دشواری این پرسش ۹۶% که با توجه به توضیحات پیشین، این پرسش در میان سوال­های بسیار ساده قرار می گیرد. ضریب تمیز این پرسش نیز ۰۶/۰ است که با توجه به اینکه ضریب تمیز 3/0 D< قابل قبول نیست.

**17- با توجه به متن کدام یک از گزینه های زیر از عادات غذایی کشور هلند به شمار می آید؟**

الف - در برنامه غذایی خود همیشه به طور روزانه از پنیر استفاده می کنند. ( ۱ )

ب - غذا را به صورت تکه های کوچک در ظرف می گذارند. ( ۳ )

پ - زود هنگام غذا میخورند و تنها یک وعده غذایی داغ در طول روز مصرف می کنند. ( ۵۵ )

در این پرسش نیز گزینه (الف) توسط ۱ نفر، گزینه (ب) را ۳ نفر، و گزینه (پ) را ۵۵ نفر انتخاب کرده­اند. ضریب دشواری این پرسش ۹۳% که با توجه به توضیحات پیشین، این پرسش در میان سوال­های بسیار ساده قرار می گیرد. ضریب تمیز این پرسش نیز ۱۳/۰ است که با توجه به اینکه ضریب تمیز 3/0 D< قابل قبول نیست. همچنین گزینه­های انحرافی به­خوبی عمل نکرده­اند.

**18- موضوع اصلی متن درباره چیست ؟**

الف - حالت های چهره ( ۱۲ )

ب - زبان بدن ( ۲۹ )

پ - انتقال اطلاعات ( ۱۸ )

درک موضوع اصلی متن در آزمون­های درکی زبان یکی از پرسش­های استنتاجی است که در آزمون درک شنیداری نیز کاربرد زیادی دارد. در این پرسش،‌ گزینه (الف) را ۱۲ نفر، گزینه (ب) را ۲۹ نفر، و گزینه (پ) توسط ۱۸ نفر ازپاسخ دهندگان انتخاب کرده­اند. . بنابراین گزینه­های انحرافی نقش خود را با موفقیت ایفا نموده­اند. ضریب دشواری این پرسش **۴۹٪**و ضریب تمیز این پرسش **۳۴/ ۰** است که ضریب تمیز قابل قبولی می­باشد با در نظر گرفتن ضریب دشواری و ضریب تمیز این پرسش و همچنین به­علت این که ۳۰ نفر گزینه­های انحرافی را انتخاب کرده­اند می­توان نتیجه گرفت که این پرسش سؤال مناسبی است.

 **19 - با توجه به متن کدام گزینه درست نیست؟**

الف -لحن صدا و بلند بودن یا آهسته بودن صدا در زمان گفتگو بر درک مخاطب تاثیر گذار است ( ۱۲ )

ب - احساساتی که از طریق صورت نشان داده میشود تغییر مشابهی در همه جهان دارند ( ۹ )

پ - مهمترین بخش ارتباطات انسان از طریق زبان برقرار می شود ( ۳۸)

. در این پرسش،‌ گزینه (الف) را ۱۲ نفر، گزینه (ب) را ۹ نفر، و گزینه (پ) توسط ۳۸ نفر ازپاسخ دهندگان انتخاب کرده­اند. . بنابراین گزینه­های انحرافی نقش خود را با موفقیت ایفا نموده­اند. ضریب دشواری این پرسش **۶۴٪**و ضریب تمیز این پرسش **۴۱/ ۰** است که ضریب تمیز قابل قبولی می­باشد با در نظر گرفتن ضریب دشواری و ضریب تمیز این پرسش و همچنین به­علت این که ۲۱ نفر گزینه­های انحرافی را انتخاب کرده­اند می­توان نتیجه گرفت که این پرسش سؤال مناسبی است.

**20- چند درصد از پیام ما از طریق حرکات بدن به مخاطبان انتقال می یابد؟**

الف - ۶۰٪ ( ۵۷ )

ب - ۳۳ ٪ ( ۱ )

پ - ۷ ٪ ( ۱ )

در این پرسش گزینه (الف) توسط ۵۷ نفر، گزینه (ب) و گزینه (پ) را نیز ۱ نفر انتخاب کرده­ است. ضریب دشواری این پرسش ۹۶% که با توجه به توضیحات پیشین، این پرسش در میان سوال­های بسیار آسان قرار می­گیرد. ضریب تمیز این پرسش نیز**۶ ۰/ ۰**  است که با توجه به اینکه ضریب تمیز 3/0 D< قابل قبول نیست.

**۵- ۲ واکاوی پایایی**

پایایی به دقّت، ‌اعتمادپذیری، ‌ثبات یا تکرارپذیری ِنتایج آزمون اشاره می‌کند .رابطه‌ی بین پایایی و روایی بدین‌گونه است که یک آزمون باید پایا باشد تا بتواند روا باشد یعنی پایایی شرط روایی است امّا روایی برای پایایی ضروری نیست. ‌از آنجایی­که این آزمون دارای گزینه‌های چند ارزشی است به عبارت دیگر بیش از دو گزینه است، برای محاسبه‌ی پایاییِ این آزمون‌،‌ از فرمول آلفای کرونباخ استفاده شده است. در روش ضریب آلفا کرونباخ هر سؤال آزمون با تک‌تک سؤال‌‌های دیگر مقایسه می‌شود. در محاسبه‌ی آلفای کرونباخ گفته شده است که اگر ضریب آلفا بیشتر از 70/0 باشد آزمون از پایاییِ قابل قبولی برخوردار است (مؤمنی و قیومی، 1389: 212 ). بنابراین با توجه به جدول داده‌های بخش آماری، پایاییِ درونی این آزمون، مطلوب و قابل قبول می‌باشد.

**۶- نتیجه­گیری و پیشنهادات**

تهیه بانک سؤال یکی از اهداف ضروری مراکزآموزش زبان فارسی است که تحلیل و وارسی سؤالات موجود در هر مرکز،‌ می­تواند گامی برای رسیدن به این هدف باشد. هدف از تحلیل سؤال‌های یک آزمون وارسی ِتک‌تک سؤالات و تعیین میزان دقّت و نارسایی آن‌هاست. در تحلیل سؤال‌های آزمون،‌ نقاط ضعف و قوّت یک آزمون و کیفیّت سؤالات آن تعیین می‌شود. بنابراین لازم است که آزمونگر ‌پس از اجرای هر آزمون، ‌سؤال‌های آن را مورد تحلیل قرار دهد و با استفاده از نتایج حاصل از آن به تجدید نظر در آزمون و بهبود کیفیّت سؤال‌ها برای استفاده‌های بعدی اقدام نماید. در این پژوهش گامی هر چند کوچک برای رسیدن به این هدف برداشته شده است. پس از واکاوی داده­ها، مشخص شد که ۷۵٪ پرسش­ها ( ۱۵ پرسش)، دارای ضریب دشواری بین ۱- ۷/۰ سؤالات آسان ،‌ ۱۵٪ ( ۳ پرسش)‌ دارای ضریب دشواری بین ۷/۰- ۳/۰ سؤالات مناسب و ۱۰ ٪ باقی­مانده­( ۲ پرسش) در محدوده­ی ضریب دشواری بین ۳/۰-۰ سؤالات سخت قرار داشتند. در بررسی ضریب تمیز این آزمون نیز مشخص شد که ۱۵٪ پرسش­ها ( ۳ پرسش) دارای ضریب تمیز 3/0 <D ، ۶۵٪ پرسش­ها (‌۱۳ پرسش) 3/0 >D ،‌۱۵ ٪ پرسش­ها ( ۳ پرسش) ۰ =D و ۵٪ (۱ پرسش) ‌دارای ضریب تمیز منفی است. دستاوردی که از ضریب دشواری این پژوهش به دست آمد نشان­دهنده­ی سهولت و آسانی آزمون و نتایج به دست آمده از ضریب تمیز نیز گویای این امر بود که این آزمون به­خوبی نتوانسته است زبان­آموزان قوی را از ضعیف جدا کند و سرانجام، ‌۱۵٪ پرسش­ها ( ۳پرسش: پرسش 11، 18، 19 ) با ضریب دشواری بین ۷/۰- ۳/۰ و ضریب تمیز 3/0 <D به­عنوان سؤالات استاندارد انتخاب شدند و در بانک سؤالات استاندارد قرارگرفتند و بعد از آسیب­شناسی پرسش­های باقی­مانده، مشخص شد که عملکرد نادرست گزینه­های انحرافی، نشاندار بودن گزینه­ی کلید، تشخیص گزینه­ی صحیح حتی بدون نیاز به فایل صوتی،‌ مبهم بودن گزینه­های انحرافی،‌ وابسته بودن گزینه­های انحرافی به اطلاعات عمومی نه متن شنیداری، از علت­های اصلی عملکرد نامناسب بقیه سؤالات و در نتیجه نامناسب بودن ضریب دشواری و ضریب تمیز بوده است که در طراحی آزمون درک شنیداری باید به آن توجه شود. همچنین با توجه به این موضوع که در هر آزمونی برحسب اهداف آزمون به سؤالاتی با ضریب دشواری متفاوت ( آسان،‌ مناسب، سخت)‌ نیاز است از داده­های این پژوهش می­توان در بانک سؤالات طبقه­بندی شده استفاده کرد. خصوصا در آزمون­های کامپیوتر – بنیانی که آزمون­ها بر اساس توانش زبانی زبان­آموزان طراحی شده­اند و زبان­آموزان در صورت شکست در پاسخ یک سؤال با سؤال ساده­تری روبه­رو خواهد شد. در چنین آزمون­هایی نیز از پرسش­هایی که ضریب دشواری آن مشخص شده است مانند داده­های این پژوهش می­توان استفاده کرد.

پایان سخن، به علاقه­مندان حوزه­ی آزمون­سازی، پیشنهاد می­شود که به­منظور اعتبارسنجی و اعتباربخشی دقیق­تر آزمون حاضر،‌ انواع روایی( روایی محتوایی،‌ روایی صوری،‌ روایی سازه­ای،‌ روایی ملاکی) این آزمون مورد بررسی قرار گیرد چرا که روایی­سنجی آزمون پژوهش دیگری را می­طلبد.

**منابع**

1. جلیلی،اکبر.(1390). آزمون مهارتی فارسی (آمفا) بر پایه چهار مهارت اصلی زبان. پایان نامه‌ی کارشناسی ارشد رشته‌ی آموزش زبان فارسی به غیر فارسی زبانان. دانشگاه علامه طباطبایی
2. سیف،علی اکبر.(1387).اندازه گیری ،سنجش و ارزشیابی آموزشی(ویرایش پنجم).تهران: نشر دوران
3. شقاقی و همکاران. (1390).هنجاریابیِ آزمون بین‌المللی زبان فارسی .فصلنامه‌ی پازند :91.
4. ضیاءحسینی،سیدمحمّد. .(1385 ). اصول و نظریه‌های آموزش زبان فارسی به غیر فارسی زبانان .تهران : انتشارات سخن .
5. قنسولی،بهزاد.(1389 ).طراحی و روایی‌سنجیِ آزمون بسندگی زبان فارسی. مجلّه پژوهش زبان‌های خارجی ،ش: 57 .115-129 .
6. ممقانی،هانیه.(1390). ارزیابی روایی و پایایی آزمون وزارت علوم و تحقیقات و فناوری (MSRT) با هدف دستیابی به یک آزمون مهارت برای زبان فارسی .پایان نامه‌ی کارشناسی ارشد رشته ی آموزش زبان فارسی به غیر فارسی زبانان.دانشگاه علامه طباطبایی
7. موسوی،ثریا.(1379). بررسی راهکارهای نظری آزمون های بسندگی و تهیه یک آزمون بسندگی استاندارد زبان فارسی. پایان نامه‌ی کارشناسی ارشد رشته ی آموزش زبان فارسی به غیر فارسی زبانان.دانشگاه شیراز
8. مومنی،م.و قیومی،ع.(1389).تحلیل های آماری با استفاده از SPSS. تهران:انتشارات مولف

**Reference**

1. Brindley, G. (2003). "Classroom-based Assessment" in D. Nunan (ed.) *Practical English Language Teaching*. US: McGraw-Hill.
2. Broun, H .Douglas.(2004).”Language assessment principles and classroom practice”. Pearson Education Inc. Longman.
3. Buck, G. (2001). *Assessing Listening*. Cambridge University Press.
4. Chang, A.C.-S.,( 2005). The perceived effectiveness of question preview in listening comprehension tests by EFL learners. New Zealand Stud. Appl.Linguistics 11, 75-99.
5. Cheng, H.-F.,(2004). A comparison of multiple-choice and open-ended response formats for the assessment of listening proficiency in English. Foreign Lang. Ann. 37, 544-555
6. Coombe, C., K. Folse and N. Hubley (2007). *A Practical Guide to Assessing English Language Learners*. US: The University of Michigan Press.
7. Eykyn , L.B.,( 1992). The Effects of Listening Guides on the Comprehension of Authentic Texts by Novice Learners of French as a Second Foreign Lang. Ann. 37, 544e555.
8. Farhadi, H. , A .Jafarpour and P .Birjandi.(2007). "testing language skills from theory to practice" .tehran:SAMT
9. Freedle, R., Kostin, I., (1999). Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL’s
10. Helgesen, M. (2003). "Listening" in D. Nunan (ed.) *Practical English Language Teaching*. US: McGraw-Hill
11. Wu, Y., (1998). What do tests of listening comprehension test? A retrospection study of EFL test-takers performing a multiple-choice task. Lang Test. 15, 21-44

**Analyzing and validating upper-intermediate Persian listening comprehension skill test**

**Leila Golpour**

Corresponding author, Assistant Professor of Teaching Persian Language to speakers of other languages, Imam Khomeini International University, Qazvin, Iran.

**Mohammad Bagher Mirzaei Hesarian**

 Assistant Professor of Teaching Persian Language to speakers of other languages Imam Khomeini International University, Qazvin, Iran.

**Amirreza Vakilifard:**

 Associate Professor of Teaching Persian Language to speakers of other languages Imam Khomeini International University , Qazvin, Iran.,

**Abstract**

Listening skill is one of the essential skills in learning Persian, as it is the beginning of the learning process of a second/foreign language. Moreover, it is not possible to go back to what was discussed as well as to modify it in a way that is possible for understanding the reading text. As such, concerning the importance of listening skills, this paper investigates one of the designed and conducted upper-intermediate level listening skill tests to access a standard test as well as to prepare a bank of standard questions. To do so, the statistical sample includes 59 Persian learner of the Persian Language Education Center of Imam Khomeini International University, who are at the upper-intermediate level in terms of language in the age group of 18 to 30 years, which are randomly chosen. The research tool is a three-alternative listening comprehension test including 20 questions such as understanding the main issue of the text, the issue expressed in an interview, the sequence of an event, understanding comparison and classification, recognizing whether a sentence is true or false, paraphrase, and cloze. Afterwards, the students answered the questions based on the heard audio file.

The texts of the audio file encompass educational, social, medical, cultural, environmental, and nutritional issues without the need for special knowledge and expertise. The research variables are the difficulty coefficient index, discrimination coefficient and reliability of listening comprehension test questions. After performing data mining, it can be observed that 75% of the questions (15 questions) have a difficulty coefficient between 0.7-1, 15% (3 questions) have a difficulty coefficient between 0.3-0.7, and 10% (2 questions) have in the range of difficulty coefficient between 0-0.3. Besides, in terms of the discrimination coefficient (D), 15% of the questions (3 questions) have D>0, while D< 0.3 for 65% of the questions (13 questions), D = 0 for 15% of the questions (3 questions), and 5% (1 question) has a negative discrimination coefficient. The obtained difficulty coefficient indicates the ease and simplicity of the test. Meanwhile, the results of the discrimination coefficient reveal that this test cannot separate appropriately strong language learners from weak ones. Ultimately, 15% of the questions (3 questions) via a difficulty coefficient between 0.3-0.7 and a discrimination coefficient greater than 0.3, are chosen as the standard questions and then placed in the bank of standard questions.

After the analyzing of the remaining questions, it can be concluded that the incorrect performance of deviant alternatives, the marked key alternative, the recognition of the correct alternative even without the audio file, the vagueness of deviant alternatives, and the dependence of deviant alternatives on public information, not the listening text, are the main reasons for the inappropriate performance of the rest of the questions as well as inappropriateness of the difficulty coefficient and discrimination coefficient, which should be taken into account to design the listening comprehension test.

**Keywords:** Evaluation, Listening Comprehension Skill, difficulty coefficient, discrimination coefficient.

1. - intensive listening [↑](#footnote-ref-1)
2. - Paraphrase recognition [↑](#footnote-ref-2)
3. -responsive listening [↑](#footnote-ref-3)
4. - selective listening [↑](#footnote-ref-4)
5. - listening cloze [↑](#footnote-ref-5)
6. - information transfer [↑](#footnote-ref-6)
7. - sentence repetition [↑](#footnote-ref-7)
8. - phone pass [↑](#footnote-ref-8)
9. - extensive listening [↑](#footnote-ref-9)
10. - dictation [↑](#footnote-ref-10)
11. - Retelling [↑](#footnote-ref-11)
12. - Difficulty Index [↑](#footnote-ref-12)
13. - Facility Index [↑](#footnote-ref-13)
14. - Discriminative Index [↑](#footnote-ref-14)
15. . measures of central tendency [↑](#footnote-ref-15)
16. . mood [↑](#footnote-ref-16)
17. . mean [↑](#footnote-ref-17)
18. median [↑](#footnote-ref-18)
19. 1- multiple choice [↑](#footnote-ref-19)
20. 2 - filling the blank [↑](#footnote-ref-20)
21. - short answer question [↑](#footnote-ref-21)
22. open-ended [↑](#footnote-ref-22)
23. - Eykyn [↑](#footnote-ref-23)
24. - Teng [↑](#footnote-ref-24)
25. - Cheng [↑](#footnote-ref-25)
26. - cloze [↑](#footnote-ref-26)
27. - Chang [↑](#footnote-ref-27)
28. - Wu [↑](#footnote-ref-28)
29. - Freedle [↑](#footnote-ref-29)
30. - Kostin [↑](#footnote-ref-30)
31. - Yanagava [↑](#footnote-ref-31)
32. - Green [↑](#footnote-ref-32)
33. - inference [↑](#footnote-ref-33)
34. - explicit [↑](#footnote-ref-34)
35. - Anna [↑](#footnote-ref-35)
36. - chang [↑](#footnote-ref-36)
37. - Read [↑](#footnote-ref-37)