

کاربرد آزمون سازی انطباقی

در تعیین سطح زبان انگلیسی دانشجویان



نگار نوروززاده، کارشناس ارشد آموزش زبان انگلیسی
عضو هیئت علمی دانشگاه آزاد اسلامی، واحد استهبان
email: abbas.mehrpooya@yahoo.com



Knowledge Improvement

چکیده

«آزمون‌های انطباقی» را می‌توان پیشرفتی در فناوری سنجش به حساب آورد. این آزمون‌ها برای برطرف کردن نقاط ضعف آزمون‌های سنتی ارائه شده‌اند. از آن‌جا که در آزمون‌های سنتی ممکن است آزموده‌ای که از سطح و دانش زبانی بالایی برخوردار است، با سؤالاتی ساده سنجیده شود و یا برعکس، آزموده‌ای که از سطح و دانش پایینی برخوردار است، با سؤالاتی دشوار سنجیده شود، طبیعی است که به نتیجه این گونه آزمون‌ها نمی‌توان اعتماد کرد و سطح دانش زبانی یک فرد را نمی‌توان با این گونه آزمون‌ها تشخیص داد. در واقع این گونه می‌توان بیان کرد که سؤال‌های بسیار ساده و یا بسیار دشوار نمی‌توانند سطح دانش زبانی افراد را مشخص کنند. در آزمون‌های انطباقی، با در نظر گرفتن این نکته و با استفاده از رایانه، سؤال‌ها برای تعیین سطح دانش زبانی یک فرد دقیقاً با توجه به سطح دانش او انتخاب می‌شوند. در واقع رایانه به شکلی برنامه‌نویسی می‌شود که به صورت کاملاً هوشمند سؤال‌های یک آزمون را انتخاب کند.

روش برگزاری آزمون نیز متفاوت است. برگزاری آزمون توسط رایانه صورت می‌گیرد و آزمون‌ها دارای سؤال‌های تفکیک شده و به صورت تک به تک هستند. بدین شکل که ابتدا سؤالی از بین سؤال‌هایی با سطح دشواری متوسط انتخاب و به شخص مورد نظر نشان داده می‌شود. پس از پاسخ‌گویی شخص به سؤال، رایانه جواب را بررسی می‌کند و چنانچه پاسخ صحیح باشد، سؤال بعدی از بین سؤال‌هایی انتخاب می‌شود که سطح دشواری آن‌ها از سطح دشواری سؤال قبلی بالاتر است. در واقع این سؤال دشوارتر از سؤال قبل است. اما اگر پاسخ آزموده صحیح نباشد، سؤال بعدی یکی از سؤال‌های ساده‌تر خواهد بود. بدین صورت با تغییر سطح دشواری سؤالات تا زمانی که آزموده به سطحی از سؤالات برسد که به تمامی آن‌ها پاسخ صحیح و یا به تمامی آن‌ها پاسخ غیر صحیح دهد، سطح و دانش زبانی آزموده تعیین می‌شود.

کلیدواژه‌ها: آزمون‌های سنتی، آزمون‌های انطباقی، آزموده، سطح دشواری، برآورد دانش زبانی

Abstract

The emergence of Computer Adaptive Tests (CATs) can be considered as an improvement in evaluation technology. These tests are presented in order to remove the weak points of traditional tests. In traditional tests, one cannot determine the English level of the testees based on the results of these tests, as too simple or too difficult items which are not adapted to the level of testees' language ability cannot determine their level of language proficiency. In CATs, the items will be adapted to the level of testees' language ability to estimate their level of ability. In fact, there will be a software programming which chooses the items adjusted to the level of testees' language ability.

The method of administering tests is also different in CATs. The administration of the test is done by the computerised tests which will be performed for each individual and by single items i.e., first an item will be chosen from a set of items with an average level of difficulty to be answered by each individual testee. The computer will check the answer. If the answer appears to be correct, the subsequent item will be chosen from among items of a difficulty level higher than that of the previous one. In fact, the more difficult item will be chosen according to the difficulty level of the previously administered one. If the testee's answer isn't correct, the movement will be downward toward the simpler items; therefore, the estimated level of the testee's language ability will be determined by changing the difficulty level of the items answered. Therefore, it is clear that CATs are preferable to traditional tests as they are adapted to the level of the testees's language ability. Hence they can be more accurate in determining the testee's level of language ability.

Key Words: Traditional Tests, Computer Adaptive Tests (CATs), Testee, Level of Difficulty, Language Ability Estimate

مقدمه

در عرصه تهیه و تدارک آزمون‌های سنجش توانایی زبانی، دیده‌ایم که به کارگیری آزمون‌هایی که برای آزمودنی بسیار ساده بوده‌اند، نتیجه‌ای در پی نداشته‌اند، مگر اتلاف وقت. چراکه تنها به برانگیختن رفتارها و واکنش‌های ناخواسته، هم‌چون اشتباهات ناشی از بی‌دقتی و یا گزینش عمومی پاسخ‌های نادرست می‌انجامند. از سوی دیگر، پرسش‌هایی که بسیار دشوار هستند، عموماً نتایجی بی‌هوده و فاقد ارزش اطلاعاتی بالا در پی داشته‌اند. چراکه آزمودنی‌ها از تلاش جدی برای پاسخ‌گویی به پرسش‌ها دست کشیده و به رفتارهایی چون به کارگیری حدس و گمان در رسیدن به پاسخ، پاسخ‌گویی گروهی پرسش‌ها و... روی آورده‌اند [Linacer, 2000].

در این راه، همان‌گونه که در «تاریخ آزمون‌های بهره‌ هوشی»

(۲۰۱۱) آمده و لیناگر (۲۰۰۰) نیز یادآوری می‌کند، پیدایش پدیده آزمون‌های هوشمند نخستین گامی بود که به سال ۱۹۰۵ به ابتکار آلفرد بینه برداشته شد. آزمون‌های هوشمند در خلال دهه‌های بعد در فرایند بهینه‌سازی قرار گرفتند که نتیجه آن از پی پیدایش روش «Flexilevel» به ابتکار لرد به سال ۱۹۷۰، پیدایش روش گام‌های^۱ به سال ۱۹۸۷ با نوآوری هنینگ، و نیز «آزمونک»‌ها^۲ به ابتکار لوییس و شیهان به سال ۱۹۹۰، این بود که همه آن‌ها به سادگی از سوی فرد یا با بهره‌گیری از رایانه قابل اجرا هستند. روند تکاملی به کارگیری رایانه‌های هوشمند در آزمون‌های تطبیقی، پیشرفت‌های دیگر را در این عرصه میسر ساخت که همان سادگی اجرا و گزینش یک‌به‌یک پرسش‌هاست [پیشین].

در این روش‌ها، پرسش‌ها بر پایه سطح دشواری گروه‌بندی می‌شوند و در هر سطح چندین زیرگروه از پرسش‌ها تشکیل

پاتریشیا ای. دانکل (۱۹۹۹)
ویژگی‌هایی چون قابلیت هماهنگی
آزمون با سرعت پاسخ‌گویی
آزمودنی، برانگیختن حس چالش
در او بدون ناامید ساختنش،
دریافت سریع بازخورد، بهبود
ضریب امنیت آزمون، و قابلیت
چند رسانه‌ای یافته‌ها را برای
آزمون‌های انطباقی برمی‌شمرد

و هوسر (۱۹۹۳) این‌گونه بیان می‌کنند: «در واقع ارزش آزمون‌های انطباقی، به عنوان جانشین مناسبی برای آزمون‌های سنتی، روز به روز بیشتر می‌شود.» و یا براون (۱۹۹۷) این‌گونه بیان می‌کند: «آزمون‌های انطباقی به دلیل برگزاری به صورت فردی، قدمی مثبت و مؤثر در امر سنجش هستند.» کارلسون (۱۹۹۴) می‌نویسد: «آزمون‌های انطباقی نه تنها دارای دقت بیشتری نسبت به آزمون‌های سنتی هستند، بلکه قادرند به سنجش سطح دانش زبانی افراد در سطوح گسترده و متفاوت بپردازند.» وینر (۱۹۹۰) نیز می‌گوید: «از آنجا که بانک سؤالات آزمون‌های انطباقی دربرگیرنده شمار زیادی از پرسش‌هاست، لذا در این آزمون‌ها آزمودنی نمی‌تواند همانند آزمون‌های سنتی که دارای تعداد محدودی سؤال هستند، با به‌خاطر سپردن تعدادی از سؤالات نمرهٔ بهتری کسب کند.»

از این‌رو با توجه به کاربردهای آزمون انطباقی، تصمیم گرفته شد که طرحی پژوهشی در این زمینه انجام پذیرد. این طرح پژوهشی را می‌توان به دو قسمت تقسیم کرد: «بانک سؤالات» و «نرم‌افزار طرح». در ادامه به بررسی جزئیات این طرح پژوهشی می‌پردازیم.

آزمون‌های انطباقی با استفاده از رایانه

«آزمون‌های انطباقی با استفاده از رایانه»^۴ بر مبنای توانایی آزمودنی برگزار می‌شوند و از این‌رو به «آزمون مناسب»^۵ نیز معروفاند.

این آزمون‌ها به‌طور متوالی سؤالاتی را براساس پاسخ‌های آزمودنی انتخاب می‌کنند. از منظر آزمودنی، آزمون با توجه به توانایی فرد، سطح دشواری خود را تعیین می‌کند. برای

می‌گردد. سپس آزمون با اجرای زیرگروه‌ها آغاز می‌شود و هماهنگ با میزان موفقیت در هر زیرگروه، با جهش به سطوح بالا و پایین ادامه می‌یابد. پس از اجرای چند زیرگروه، برآورد پایانی توانایی آزمودنی به دست می‌آید. اگرچه این رویکرد، رویکردی خام است، ولی همین روش‌های ابتدایی نیز توانسته‌اند هم‌چون تکنیک‌های پیشرفتهٔ «CAT»، نتایجی سودمند در پی داشته‌باشند [Yao, 1991].

پژوهشگران و کارشناسان ویژگی‌های خاصی را در مقایسه با این روش‌ها برای آزمون‌های انطباقی CAT برشمرده‌اند که آن‌ها را نسبت به روش‌های دیگر ممتاز می‌گردانند. پاتریشیا ای. دانکل (۱۹۹۹) ویژگی‌هایی چون قابلیت هماهنگی آزمون با سرعت پاسخ‌گویی آزمودنی، برانگیختن حس چالش در او بدون ناامید ساختنش، دریافت سریع بازخورد، بهبود ضریب امنیت آزمون، و قابلیت چند رسانه‌ای یافته‌ها را برای آزمون‌های انطباقی برمی‌شمرد. گرین و همکارانش (۱۹۹۵) این نکته را یادآور می‌شوند که شماری از سازمان‌ها و گروه‌های گوناگون می‌توانند از توان آزمون‌های رایانه‌ای بهره‌گیرند که از آن جمله می‌توان به مربیان آموزشی در برنامه‌های زبان دوم و خارجی، انجمن‌های گزینشی و استخدامی، سازمان‌های آموزشی - پرورشی، ادارات دولتی آموزشی، و بخش‌های زبان خارجی که در پی برگزاری CAT برای ورود و خروج دانشجویان از دوره‌های آموزشی زبان هستند، اشاره کرد.

پیش از این، در دهه ۱۹۶۰ بود که وزارت دفاع ایالات متحده به نقاط قوت آزمون‌های انطباقی پی برد و از تحقیقات نظری در زمینهٔ آزمون‌های انطباقی و «نظریهٔ تحلیل پاسخ‌ها»^۳ حمایت کرد [Wainer, 1990]. نظریهٔ تحلیل پاسخ‌ها براساس نظریهٔ احتمالات است که میزان احتمال این که آزمودنی پاسخ صحیح به یک پرسش بدهد را محاسبه می‌کند [Alderson, Clapham & Wall, 1995].

طرفداران آزمون‌های انطباقی، افزایش روزافزون استفاده از رایانه را، انقلابی در امر سنجش می‌دانند. در آمریکا، کاربرد آزمون انطباقی رایانه‌ای به‌طور فزاینده‌ای در ارزش‌یابی آموزشی، به‌ویژه در آزمون‌های سراسری، از جمله آزمون فراغت از تحصیل، آزمون انتخاب مدیران و آزمون‌های توانایی حرفه‌ای ارتش رواج یافته است [Jian-quan et al. 2007].

هم‌چنین در چین، آزمون انطباقی رایانه‌ای توانش زبانی، برای دانش‌آموزان دورهٔ میانی و نیز آزمون انطباقی رایانه‌ای برای توانمندی ریاضی دانش‌آموزان دبستانی ساخته و به کار گرفته شده‌اند [Q I Shu-qing 2003].

در ارتباط با آزمون‌های انطباقی، برای مثال، کینگزبری

اشاره شده است:

- I hope you know that you --- come with me if you don't want to.

- a) can't b) don't have to
c) don't need d) mustn't

Maximum likelihood estimation (correct answer): 100%

- I'm glad to hear that at least today's --- cheerful.

- a) news are b) the news are
c) the news is d) news is

Maximum likelihood estimation (correct answer): 94%

- They have an expression in Australia when a famous person mixes with the crowd, they say they're going on a walk --- .

- a) across b) along
c) out d) about

Maximum likelihood estimation (correct answer): 64%

هر چند که آزمون‌های انطباقی دارای الگوریتمی هستند که از چند سؤال به دفعات زیاد استفاده نشود، ولی از آن‌جا که ممکن است سطح توانایی افرادی یکسان باشد، لذا بدیهی است برای گروهی از افراد با سطح توانایی یکسان، احتمال تکرار یک یا چند سؤال وجود دارد. در آزمون‌های انطباقی، مرور کردن سؤال‌های برگزارد شده، امکان‌پذیر نیست. در واقع آزمون‌های انطباقی باید سؤالی ساده‌تر را در صورت پاسخ‌گویی اشتباه به یک سؤال توسط آزمودنی به او نشان دهند. زیرا شخص باهوش و زیرک می‌تواند از این روش برای یافتن جواب‌های صحیح استفاده کند. در واقع، با مقایسه سؤال بعدی با سؤال قبلی، می‌تواند متوجه شود که جواب او به یک سؤال صحیح بوده است یا نه.

پنج جزء اصلی در ساخت آزمون‌های انطباقی وجود دارد که عبارت‌اند از:

۱. سطح‌بندی کردن بانک سؤالات

بانکی از پرسش‌ها باید با آزمون انطباقی همراه شود. این بانک باید با یک مدل هوش‌سنجی درجه‌بندی شود. به این منظور معمولاً از نظریه پاسخ به سؤال (IRT) استفاده می‌شود. دلیل استفاده از نظریه مذکور آن است که در این روش، آزمودنی و سؤالات با یک معیار سنجیده می‌شوند. که این امر کمک شایانی به انتخاب سؤال می‌کند.

مثال، اگر آزمودنی به سؤالی با سطح دشواری متوسط پاسخ صحیح بدهد، سؤال بعدی سؤالی با سطح دشواری مشکل‌تری خواهد بود. یا اگر پاسخ آزمودنی به سؤالی صحیح نباشد، سؤالی ساده‌تر برای او انتخاب می‌شود. در مقایسه با «آزمون‌های چهار گزینه‌ای ثابت»، آزمون‌های انطباقی با تعداد سؤالات کمتری و با دقت بیشتری آزمون‌شوندگان را می‌سنجند.

می‌توان گفت که آزمون‌های انطباقی، «الگوریتم تکرار»^۷ هستند که دارای مراحل زیرند:

۱. براساس تقریبی از توانایی آزمودنی، سؤالی از بانک سؤالات موجود انتخاب می‌شود.
۲. سؤال انتخاب شده به آزمودنی نشان داده می‌شود و آزمودنی پاسخ صحیح و یا غلط را انتخاب خواهد کرد.
۳. براساس جواب داده شده توسط آزمودنی، سطح توانایی او تصحیح می‌شود.
۴. مراحل ۱ تا ۳ تا زمانی که به ضابطه تعریف شده‌ای برای پایان دادن به حلقه سؤالات دسترسی حاصل شود.

از آن‌جا که در آغاز آزمون، سطح توانایی آزمودنی مشخص نیست، لذا برای نخستین پرسش، الگوریتم سؤالی را با سطح دشواری متوسط و یا ساده انتخاب می‌کند. براساس الگوریتم انتخاب سؤال، طبیعی است که آزمون‌های متفاوت با سؤالات متفاوت‌رو به‌رو شوند.

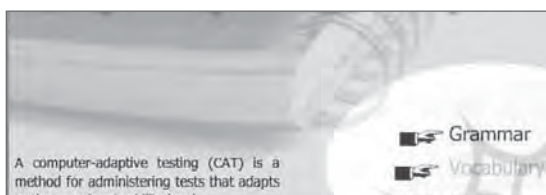
«فناوری هوش‌سنجی»^۸ که سؤالات را به درستی از بین سؤالات متفاوت با سطح دشواری‌های متنوع متناسب با توانایی آزمودنی انتخاب می‌کند، «نظریه پاسخ به سؤال»^۹ نام دارد. این نظریه اساس انتخاب بهترین سؤال برای آزمودنی است و سبب می‌شود براساس اطلاعات به‌دست آمده از آزمون سؤالات انتخاب شوند.

در آزمون‌های انطباقی با استفاده از رایانه، اولین اقدام تهیه بانک سؤالات سطح‌بندی شده است. برای این که بتوان مشخصات سؤالات را تعیین نمود (تا بتوان بهترین سؤالات را برگزید) تمام سؤالات باید مورد تحلیل قرار گیرند. برای دستیابی به این هدف، سؤالات جدید باید به سؤالات مؤثر یک آزمون افزوده شوند. برای مثال، نمی‌توان آزمون انطباقی را با استفاده از سؤالات کاملاً جدید تهیه کرد. تمام سؤالات باید با یک نمونه آماری مناسب سنجیده شوند تا بتوان مشخصات سؤال را به‌دست آورد. این نمونه آماری می‌تواند شامل هزار نفر آزمودنی باشد. در واقع تعداد آزمودنی‌ها با توجه به اهداف مورد نظر تعیین می‌شود.

در زیر به چند نمونه از نتایج به‌دست آمده از تحلیل سؤالات

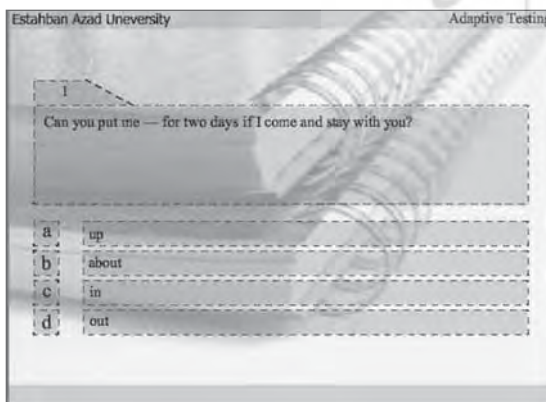


شکل ۱. صفحه اصلی نرم افزار آزمون سازی انطباقی



شکل ۲. انتخاب نوع آزمون

کاربر می تواند یکی از گزینه های «Grammar» یا «Vocabulary» را انتخاب کند. با انتخاب نوع آزمون توسط کاربر، نرم افزار آزمون را در همان زمینه انتخاب شده توسط کاربر برگزار و سؤالات را از بانک داده های متناسب انتخاب می کند. آزمون به صورت تک سؤال برگزار می شود. بدین شکل که ابتدا یک سؤال با سطح دشواری متوسط یا آسان به کاربر نشان داده می شود و پس از پاسخ گویی کاربر به سؤال، نرم افزار سؤالات بعدی را براساس جواب های داده شده توسط کاربر و بر مبنای IRT، به او نشان خواهد داد. پنجره مرحله برگزاری آزمون در شکل ۳ نشان داده شده است.



شکل ۳. برگزاری آزمون توسط نرم افزار

۲. انتخاب سطح دشواری سؤالات آغازین آزمون

در آزمون های انطباقی انتخاب سؤال براساس سطح توانایی آزمودنی انجام می گیرد. از آن جا که آزمون های انطباقی در آغاز نمی توانند توانایی آزمودنی را تخمین بزنند، معمولاً در آزمون های انطباقی، آزمودنی دارای سطح متوسط توانایی در نظر گرفته می شود و آزمون با سؤالی با سطح دشواری متوسط آغاز می شود.

۳. الگوریتم انتخاب سؤال

همان طور که اشاره شد، IRT آزمودنی و سؤالات را با یک معیار می سنجد. به این دلیل که اگر آزمون انطباقی تقریبی از توانایی آزمودنی داشته باشد، قادر خواهد بود که سؤالی متناسب با آن انتخاب کند. در این روش سؤالی با سطح دشواری بیشتر از سطح توانایی شخص انتخاب می شود.

۴. روش امتیازدهی

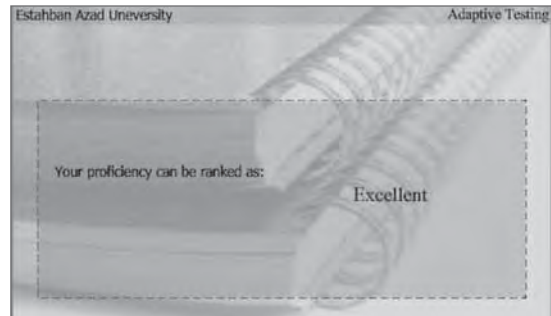
پس از برگزار شدن یک سؤال، آزمون انطباقی تقریب خود را از توانایی آزمودنی تصحیح می کند. اگر آزموده به سؤال پاسخ صحیح داده باشد، آزمون انطباقی درجه توانایی آزمودنی را افزایش می دهد و برعکس. این عمل براساس «تابع پاسخ گویی به سؤال»^{۱۱} از IRT بر اساس «تابع احتمال»^{۱۲} توانایی آزمودنی است. دو روشی که به این منظور به کار می روند عبارتند از: ۱. تخمین بیشترین احتمال ممکن، و: ۲. تقریب بایسیان^{۱۳}. تخمین بیشترین احتمال ممکن، تأثیرپذیر نیست و در مواردی که پاسخ به سؤالات یکنواخت باشد (به همه سؤالات پاسخ صحیح و یا اشتباه داده شود)، کارایی خود را از دست می دهد. در این موارد بهتر است از روش بایسیان استفاده شود.

۵. معیار پایان آزمون

الگوریتم آزمون های انطباقی به گونه ای طراحی شده است که در یک حلقه تکرار، سؤالی را انتخاب می کند و پس از آن سطح توانایی آزمودنی را تصحیح می کند. این رویه تا زمانی که از تمام سؤالات بانک داده ها استفاده شود، ادامه پیدا می کند. از این رو باید معیاری برای پایان دادن به آزمون در نظر گرفت. به طور معمول، زمانی که «خطای استاندارد» آزمودنی از میزان تعیین شده کمتر شود، آزمون متوقف می شود.

استفاده از نرم افزار طرح

در طراحی نرم افزار سعی شده است که استفاده از آن بسیار آسان باشد تا بدین وسیله عموم بتوانند از آن استفاده کنند. با اجرای نرم افزار، پنجره اصلی آن نمایان می شود (شکل ۱).



شکل ۴. نشان دادن نتیجهٔ آزمون برگزار شده به کاربر توسط نرم‌افزار

با پاسخ‌گویی کاربر به پرسش‌ها و تصحیح سطح دشواری پرسش‌ها براساس پاسخ‌های او، مرحله‌ای فرا می‌رسد که در آن سطح دشواری پرسش نشان داده شده به کاربر، منطبق بر توانایی اوست. در این مرحله نرم‌افزار برگزاری آزمون را به پایان می‌رساند و نتیجه را به کاربر نشان می‌دهد. همان‌طور که مشاهده می‌شود، توسط این نرم‌افزار، در سه مرحله کاربر می‌تواند سطح توانایی خود را در پاسخ‌گویی به سؤالات زبان انگلیسی مشخص کند.

نتیجه‌گیری و پیشنهادها

در آزمون‌های سنتی، آزمودنی ممکن است با سؤال‌های بسیار دشوار مواجه شود که این امر می‌تواند بر اعتماد به نفس آزمودنی در پاسخ‌گویی به سؤالات تأثیر منفی بگذارد و سنجش سطح دانش آزمودنی را غیرممکن سازد. این اشکال یکی از اشکالات آزمون‌های سنتی است. با توجه به اشکالات موجود در آزمون‌های سنتی، نیاز به روشی جدید در برگزاری آزمون احساس می‌شود. از این‌رو در مقاله حاضر سعی شد تا با بررسی آزمون‌سازی انطباقی، نحوه بکارگیری

توسط آزمون‌های انطباقی می‌توان افراد را با هر سطحی از دانش زبانی آزمود و برخلاف آزمون‌های سنتی محدود به یک گروه خاص نیست. به علاوه می‌توان از آن در همه جا (مدرسه، دانشگاه و...) استفاده کرد

آن برای تعیین سطح زبان انگلیسی دانشجویان نشان داده شود. در یک مقایسهٔ کلی آزمون‌های انطباقی با آزمون‌های سنتی می‌توان به موارد زیر اشاره کرد:

۱. در آزمون‌های انطباقی، برخلاف آزمون‌های سنتی که دارای تعداد محدودی سؤال هستند، از بانک سؤالات زیاد و متنوع استفاده می‌شود و آزمودنی نمی‌تواند با به‌خاطر سپردن تعدادی از سؤالات و یا حتی به‌خاطر سپردن نوع سؤالات نمره بهتری کسب کند (ویتر ۱۹۹۰).

۲. توسط آزمون‌های انطباقی می‌توان افراد را با هر سطحی از دانش زبانی آزمود و برخلاف آزمون‌های سنتی محدود به یک گروه خاص نیست. به علاوه می‌توان از آن در همه جا (مدرسه، دانشگاه و...) استفاده کرد.

۳. آزمون‌های انطباقی به راحتی پس از برگزاری آزمون می‌توانند مشخص کنند که آزمودنی در چه مهارتی ضعف دارد.

۴. از آن‌جا که آزمون‌های انطباقی به‌صورت تمام رایانه‌ای برگزار می‌شوند، امکان اشتباهات ناشی از خطای شخصی، مانند خطا کردن در تصحیح آزمون‌ها، از بین می‌رود.

۵. تمام آزمون‌های انطباقی با سطح دانش زبانی آزمودنی متناسب هستند و این امر در تعیین سطح و دانش آزمودنی حائز اهمیت است.

۶. برخی از آزمودنی‌ها ممکن است براساس حدس و گمان به سؤال‌ها پاسخ دهند. در آزمون‌های سنتی این امر قابل تشخیص نیست، درحالی که در آزمون‌های انطباقی این امر قابل شناسایی است و رایانه به راحتی می‌تواند با بررسی الگوی پاسخ‌های آزمودنی به سؤال‌ها و تجزیه و تحلیل آن‌ها، تشخیص دهد که آزمودنی به یک سؤال خاص به‌صورت تصادفی پاسخ صحیح داده است یا خیر، و یا در آخر اگر آزمودنی به تمام سؤالات بر اساس حدس و گمان پاسخ داده است، این موضوع را تشخیص دهد و نمره‌ای برای این آزمون در نظر نگیرد، تا برای تعیین سطح دانش زبانی آزمودنی، آزمون دیگری برگزار کند.

۷. امروزه در بسیاری از دانشگاه‌های معتبر دنیا از آزمون‌های انطباقی برای سنجش دانش زبانی افراد در زمینه‌های متفاوت، از جمله لغت، دستور زبان، درک مطلب و... استفاده می‌شود. در حال حاضر تمامی برنامه‌های سنجش دانش زبانی در اروپا و امریکا به‌صورت انطباقی و یا در حال تبدیل شدن به آزمون‌های انطباقی است. در ایران در زمینهٔ ساخت این‌گونه تست‌ها تاکنون تلاشی جدی صورت نگرفته است و نیاز مبرمی به دستیابی به این فناوری احساس می‌شود.

در واقع با پیشرفت علم و کاربرد رایانه در تمام امور، می‌توان آزمون‌های انطباقی را انقلابی در زمینهٔ سنجش افراد به‌شمار آورد.

Present Research and some Future Directions. *Language Learning & technology*, 1(1), 44-59. Retrieved November 1, 1998 from the World Wide Web: <http://polyglot.cal.msu.edu/llt/vollnum1/brown/default.html>.

Carlson, R. (1994). Computer-Adaptive Testing: A Shift in the Evaluation Paradigm. *Journal of Educational Technology Systems*. 22, 213-224.

Green, B., Kingsbury, G., Lloyd, B., Mills, C., Plake, B., Skaggs, G., Stevenson, J., Zara, T., & Schwartz, J., (1995). *Guidelines for Computerized-Adaptive Test Development and Use in Education*. Washington, DC: American Council on Education Credit by Examination Program.

Henning, G. (1987), *A guide to language testing*, Cambridge, Mass.: Newbury House.

History of IQ Tests, (2011), IQ TestExperts. com, <http://www.iqtestexperts.com/iq-history.php>

Jian-quan, TIAN, Dan-min, MIAO, Xia, ZHU & Jing-jing, GONG, (2007). An Introduction to the Computerized Adaptive Testing, *US-China Education Review*, Jan. 2007, Volume 4, No. 1 (Serial No.26), 72-81.

Kingsbury, G., & Houser, R. (1993). Assessing the Utility of Item Response Models: Computer-Adaptive Testing. *Educational Measurement: Issues and Practice*. 12,21-27.

Linacre, John M., (2000), Computer-Adaptive Testing CAT: A Methodology Whose Time Has Come., *MESA, Memorandum, No.69*.

Lord, Frederic M. (1970), *The Self-Scoring Flexilevel Test [microform]*, [Washington, D.C.]: Distributed by ERIC Clearinghouse, 1970. 13 p. Online version: <http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?acno=ED042813>

Lewis, C. & Sheehan, K., (1990), Using Bayesian Decision Theory to Design a Computerized Mastery Test, *Applied Psychological Measurement*, 14,367-386.

Dunkel, Patricia A., (1999). Considerations in Developing or Using Second/Foreign Language Proficiency Computer-Adaptive Testing. *Journal of Language Learning & Technology, Volume 2, Number 2*, pp. 77-93.

Qi Shu-qing, (2003). *The Application of Modern Measurement Theory in Examination*. Wuhan: Huazhong Normal University Press, 350-369.

Wainer, H. (1990). *Computer Adaptive Testing: A Primer*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wainer, H., & Mislevy, R.J. (2000). Item-Response Theory, Calibration, and Estimation, Book Chapter in Schank, Roger C., (2000), *Explanation Patterns: Understanding Mechanically and Creatively*, Lawrence Erlbaum Associates, Inc, Publishers,

Yao, T. (1991) CAT with a poorly calibrated item bank. *Rasch Measurement Transactions* 5:2, P. 141.

در این مقاله سعی شد با بررسی آزمون‌های انطباقی و چگونگی استفاده از نتایج آن برای به‌دست آوردن سطح دانش زبانی دانشجویان، قدمی هرچند کوچک، در گسترش این آزمون‌ها در ایران برداشته شود.

پیشنهادات

در پایان، شایسته است با یادآوری ویژگی‌های آزمون‌های انطباقی، از جمله قابلیت این آزمون‌ها در هماهنگی با سرعت پاسخ‌گویی آزمودنی، برانگیختن حس رویارویی با چالش آزمون بدون ناامید ساختن آزمودنی، دست‌یابی زودتر به بازخورد، بهبود ضریب امنیت آزمون، و قابلیت چند رسانه‌ای یافته‌ها از راه رایانه‌ای کردن داده‌ها، دست‌یابی آسان به نمودار و گراف، و سرانجام جلوگیری از اتلاف وقت، توجه دبیران و پژوهشگران را به کوشش برای برداشتن گام‌های بعدی در این زمینه جلب نمود.

از میان پیشنهادهای قابل طرح در این زمینه، می‌توان به بهره‌گیری از آزمون‌های انطباقی برای گروه‌بندی سریع و آسان دانش‌آموزان و دانشجویان، و همگون کردن کلاس‌های درس اشاره کرد. از این راه با قرار دادن دانش‌آموزان و دانشجویان در گروه‌های گوناگون برخوردار از توانایی زبانی یکسان، می‌توان مواد آموزشی مناسب‌تر و هماهنگ‌تر با توانایی زبانی دانش‌آموزان و دانشجویان در اختیار آن‌ها قرار داد، و به اهداف آموزشی مورد نظر برای هر گروه در کمترین زمان و با کمترین هزینه و انرژی دست یافت.

پی‌نوشت

1. Step Procedure
2. Testlets
3. Item Response Theory
4. Computer-Adaptive Testing (CAT)
5. tailored testing
6. Static Multiple Choice Tests
7. iterative algorithm
8. Psychometric Technology
9. Item Response Theory (IRT)
10. Item Response Function
11. likelihood function
12. Maximum Likelihood Estimation
13. Bayesian Estimation

منابع

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. New York: Cambridge University Press.

Binet, A., Simon, and Th. (1905). Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *Ann'ee psychol.*, 191-244.

Brown, J. D. (1997). Computer in Language Testing: