نامهٔ آموزش عالی
دوره جدید، سال اول، شماره اول،
بهار ۱۳۸۷، ص ۴۷-۲۹

مقایسهٔ دو نظریهٔ «کلاسیک آزمون» (CTT) و «پاسخ سؤال» (IRT) از نظر اندازه‌گیری شاخص‌های دقت آزمون

سیده طیبه مطیعی لنگرودی*
بلال ایرانلو**
رضا پیروی***

چکیده

پژوهش حاضر، به منظور مقایسهٔ دو نظریهٔ کلاسیک آزمون (CTT) و پاسخ سؤال (IRT) از نظر اندازه‌گیری شاخص‌های دقت آزمون انجام شد. روش تحقیق، از نوع کاربردی ـ توصیفی است و برای بررسی سؤال پژوهشی، پاسخنامه‌های داوطلبان ورود به دانشگاه‌های کشور در سال ۱۳۸٤، در رشتهٔ علوم تجربی در آزمون اختصاصی درس زیست‌شناسی مورد استفاده قرار گرفت. از بین کلّیهٔ داوطلبان شرکت‌کننده نظام سالی- واحدی در این گروه آزمایشی، یک گروه نمونهٔ ۱۰۰۰ نفری به طور تصادفی انتخاب شد و سؤال تحقیقی بر روی پاسخنامه‌های گروه نمونه بررسی شد. نتایج تحقیق نشان داد که نظریهٔ IRT از نظر اندازه‌گیری شاخص‌های دقت آزمون بر نظریهٔ CTT ارجحیت دارد. با توجه به نتایج تحقیق، در نظریهٔ پاسخ سؤال، برای هر سطح توانایی، آگاهی و خطای استاندارد جداگانه‌ای به دست می‌آید که بر اساس آن می‌توان فهمید آزمون در کدام سطح توانایی، دارای بیشترین دقت و کمترین خطاست؛ مزیتی که نظریهٔ کلاسیک اندازه‌گیری فاقد آن است.

واژگان کلیدی: نظریهٔ کلاسیک اندازه‌گیری، نظریهٔ پاسخ سؤال، تابع آگاهی شاخص‌های دقت آزمون، پایایی

---

* کارشناس ارشد دفتر آزمون سازی سازمان سنجش آموزش کشور
** کارشناس ارشد مرکز مطالعات سازمان سنجش آموزش کشور
*** کارشناس ارشد دفتر آزمون سازی سازمان سنجش آموزش کشور

مقدمه

مانند سایر رشته‌های علمی، در اندازه‌گیری نیز نظریه‌ها و مدل‌ها، معمولاً یک چارچوب کلی جهت بررسی فرایند تهیه آزمون‌ها فراهم می‌آورند. به‌طور کلی، نظریه‌های اندازه‌گیری به دو دسته عمده:۱. نظریهٔ ضعیف نمره واقعی یا نظریهٔ کلاسیک اندازه‌گیری[1]، ۲. نظریهٔ جدید اندازه‌گیری یا نظریهٔ پاسخ سؤال[2] تقسیم می‌شوند (همبلتون[3] و واندرلیندن[4]، ۱۹۸۲).

کوشش‌های اسپیرمن[5] (۱۹۰٤) و دیگر صاحب‌نظران بعد از او به رشد و گسترش نظریه‌ای منجر شد که توانست بیشتر از نیم قرن بر نهضت آزمون‌سازی حکومت کند. به رغم خدمات طولانی‌مدت این نظریه (که نظریهٔ کلاسیک اندازه‌گیری نامیده شد) به جامعهٔ روان سنجی، به‌تدریج و با گذشت زمان، محدودیت‌های آن آشکار گردید و از سوی محققان مختلف اندازه‌گیری (از جمله: گالیکسن، ۱۹۵۰؛ لرد[6] و ناویک[7]، ۱۹٦۸؛ همبلتون[8] و سوآمیناتان[9] و راجرز[10]، ۱۹۹۱) مورد توجه قرارگرفت. با مطالعات برخی از پژوهشگران، مانند لرد (۱۹۸۰)، همبلتون (۱۹۸۹) و ...، شواهد زیادی در رابطه با محدودیت آزمون‌هایی که براساس نظریهٔ کلاسیک اندازه‌گیری ساخته شده‌اند، به دست آمده است. این نظریه، موجب دشواری‌های عملی و مسائل تفسیری و تحلیلی متعددی می‌شود که قادر به پاسخگویی و حل آنها نیست، اما گسترش پرشتاب مبانی نظری نظریهٔ پاسخ سؤال از دههٔ ۱۹۸۰ به بعد و انتشار مقالات و کتب متعدد در این زمینه، چشم‌انداز بسیار روشنی را پیش‌روی سازندگان آزمون قرار داده است (از جمله: همبلتون، ۱۹۸۹؛ همبلتون و سوامیناتان و راجرز، ۱۹۹۱؛ همبلتون و جونز، ۱۹۹۳). نظریهٔ جدید اندازه‌گیری چه از لحاظ روش‌های آماری و به‌کارگیری توابع و مدل‌های ریاضی و چه از جهت مفروضات نظری و نتایج کاربردی، تفاوت‌های چشمگیری با نظریهٔ کلاسیک داشته و چارچوب مفیدی برای حل مسائل گستردهٔ اندازه گیری خصوصیات روانی ـ تربیتی فراهم آورده است. دو نظریهٔ کلاسیک اندازه‌گیری و پاسخ

1. Classical Test Theory
2. Item Response Theory
3 . Hambleton
4. Vanderlinden
5. Spearman, Charles
6. Lord
7. Novick
8. Hambleton, R . K
9. Swaminathan, H
10. Rogers, H. J

سؤال، از بسیاری جهات با یکدیگر تفاوت دارند که در این مقاله به مقایسهٔ عملی‌این دونظریه از نظر اندازه‌گیری شاخص‌های دقت[1] آزمون با استفاده از داده‌های موجود در سازمان سنجش آموزش کشور پرداخته می‌شود.

در نظریهٔ کلاسیک، همان‌طور که آناستازی (۱۹۸۸، ترجمهٔ براهنی، ۱۳۷۹، ص:۱۱۱) بیان کرده است، منظور از دقت اندازه‌گیری یا پایایی[2] درجهٔ ثبات نمره‌هایی است که:

«با بازآزمایی افراد واحد با آزمون واحد در موقعیت‌های مختلف، یا با آزمایش افراد واحد با سلسله ماده‌های همتا[3]، و یا تحت سایر شرایط متغیر به دست می‌آیند. مفهوم پایایی بر اساس محاسبهٔ خطای اندازه‌گیری است که با استفاده از آن می‌توان مشخص کرد که تحت‌تأثیر عوامل نامربوط و تصادفی، نمرهٔ فرد واحدی تا چه اندازه نوسان خواهد داشت. مفهوم پایایی آزمون به چند جنبهٔ مختلف از ثبات نمره‌ها اشاره دارد. پایایی آزمون به معنای وسیع آن شاخص این است که تفاوت‌های مشهود در نمره‌های آزمون تا چه اندازه ناشی از تفاوت‌های واقعی در ویژگی‌های مورد مطالعه است و تا چه اندازه می‌توان آنها را به خطای تصادفی نسبت داد. به بیان فنی‌تر، شاخص‌های پایایی آزمون امکان می‌دهند، برآورد کنیم که چه نسبتی از پراش کل آزمون را پراش خطا[4] تشکیل می‌دهد».

در مطلوب‌ترین شرایط نیز هیچ آزمونی یک وسیلهٔ کاملاً پایا و دقیق نخواهد بود. در نتیجه، همراه هر آزمون، پایایی آن گزارش می‌شود و این شاخص پایایی، مخصوص وضعیتی است که در آن آزمون در شرایط معیار و روی آزمودنی‌هایی همانند نمونهٔ هنجاری به‌کار می‌رود؛ بنابراین، همراه با نوع پایایی محاسبه شده، ویژگی‌های گروه نمونه نیز گزارش می‌شود. نظر به اینکه تمام پایایی مربوط به درجهٔ ثبات یا توافق بین دو سلسله نمره است که مستقلاً به‌دست آمده‌اند، همهٔ آنها را می‌توان با ضریب همبستگی نمایش داد. همان‌طورکه می‌دانیم ضریب همبستگی نشان‌دهندهٔ درجهٔ همخوانی یا رابطهٔ بین دو سلسله نمره است؛ بنابراین، پایایی به‌عنوان ضریب همبستگی، میزان ارتباط بین نمره‌های واقعی[5] و نمره‌های مشاهده شده[6] را نشان می‌دهد.

---

1. Accuracy
2. Reliability
3. Parallel Item
٤. منظور از پراش خطا (Error Variance)، هر نوع شرایطی است که از لحاظ هدف آزمون نامربوط شناخته می‌شود.
5. True Scores
6. Observed scores

فرمولی که به منظور محاسبهٔ ضریب پایایی ارائه شده است، بر مدل ریاضی رابطهٔ بین نمره‌های واقعی، نمره‌های مشاهده شده و نمره‌های خطا استوار است که بر اساس آن می‌توان پایایی را به‌عنوان نسبت واریانس نمره‌های واقعی به واریانس نمره‌های مشاهده شده بیان کرد.

$$\rho_{XX'} = \frac{\delta_T^2}{\delta_X^2}$$ (آلن[1] و ین[2]، ترجمهٔ دلاور، ۱۳۷٤)

که در آن $\rho_{XX'}$ ضریب پایایی، δ_T^2 واریانس نمره‌های واقعی و δ_X^2 واریانس نمره‌های مشاهده شده است. این فرمول یک مفهوم نظری دارد و درعمل، محاسبهٔ این مقدار به جهت غیرقابل مشاهده بودن نمره‌های واقعی و همچنین عدم امکان به دست آوردن تمام نمره‌های مشاهده شده جامعه غیرممکن است؛ بنابراین، به‌جای این فرمول می‌توان گروهی از آزمودنی‌ها را در دو موقعیت متفاوت و با آزمونی یکسان یا با دو فرم همتا از یک آزمون مورد آزمایش قرار داد.

در نظریهٔ کلاسیک اندازه‌گیری، برای تمام توانایی‌ها (آزمودنی‌ها با توانایی‌های مختلف)، مقدار ثابتی از خطای معیار اندازه‌گیری به دست می‌آید که در نتیجه شاخص پایایی نیز برای همهٔ سطوح توانایی ثابت خواهد بود. مسلماً دقت اندازه‌گیری هر آزمون در سطوح مختلف توانایی متفاوت است، که این مسئله در نظریهٔ کلاسیک به‌هیچ‌وجه قابل مشاهده نیست، اما در نظریهٔ پاسخ سؤال، شاخص دقت برآورد پارامترها، تابع آگاهی[3] نامیده می‌شود و میزان خطای آن در حد خطای استاندارد برآورد پارامترهاست. تابع آگاهی که معکوس واریانس خطاست و دقت برآورد پارامترهای سؤال و توانایی را مشخص می‌کند، نشان‌دهندهٔ آن است که سؤال و آزمون در برآورد توانایی در طیف نمرات توانایی چگونه عمل می‌کنند. تابع آگاهی برای سطوح مختلف توانایی مقادیر متفاوتی به‌دست می‌دهد و یک ویژگی بسیار مفید در نظریهٔ پاسخ سؤال است.

مفهوم ضمنی واژهٔ «آگاهی[4]» این است که فرد چیزی دربارهٔ یک موضوع یا عنوان خاصی می‌داند، اما در آمار و روان‌سنجی، اصطلاح آگاهی بیان‌کنندهٔ چیزی مشابه اما فنّی‌تر است. معنی آماری آگاهی ازسوی فیشر[5] مطرح شد که آگاهی را به صورت معکوس میزان دقتی که پارامتری باید برآورد شود، تعریف کرد؛ بنابراین، اگر بتوان یک

1. Allen
2. Yen
3. Information Function
4. Information
5. Fisher

پارامتر را با دقت برآورد کرد، دربارهٔ مقدار پارامتر، نسبت به زمانی‌که با دقت کمتری برآورد شده است، آگاهی وجود دارد. از نظر آماری، دقتی که با آن یک پارامتر برآورد می‌شود با تغییرپذیری برآوردها در اطراف مقدار پارامتر، اندازه‌گیری می‌شود؛ از این‌رو، اندازهٔ دقت برابر است با واریانس برآوردکننده‌ها که با δ^2 نشان داده می‌شود. مقدار آگاهی که با I بیان می‌شود از فرمول زیر به دست می‌آید.

$$I = \frac{1}{\delta^2}$$ (بیکر، ۲۰۰۲)

در نظریهٔ پاسخ سؤال، برآورد مقدار پارامتر توانایی برای یک آزمودنی مورد توجه است. پارامتر توانایی با نماد θ نشان داده می‌شود. $\hat{\theta}$ یک برآورد از θ است. انحراف استاندارد برآورد توانایی، برای پارامتر توانایی آزمودنی محاسبه می‌شود؛ که اگر مجذور شود، واریانس به‌دست می‌آید که از طریق آن، سطح توانایی معیّنی با میزانی از دقت می‌تواند برآورد شود. با توجه به معادلهٔ فوق، می‌توان گفت که مقدار آگاهی در سطح توانایی معیّن، معکوس واریانس است. اگر مقدار آگاهی بزرگ باشد، به این معنی خواهد بود که توانایی یک آزمودنی که میزان توانایی واقعی او در آن سطح است، می‌تواند با دقت برآورد شود؛ یعنی اینکه همهٔ برآوردها به‌صورت قابل قبولی به مقدار واقعی نزدیک خواهد شد. اگر مقدار آگاهی کوچک باشد، به این معنی است که توانایی را نمی‌توان با دقت برآورد نمود و برآوردها به‌طور وسیعی نسبت به مقدار واقعی پارامتر پراکنده هستند. با به‌کارگیری فرمولی مناسب می‌توان مقدار آگاهی را برای هر سطحی از توانایی، در مقیاسی از منفی بی‌نهایت تا مثبت بی‌نهایت، محاسبه کرد (بیکر، ۲۰۰۲).

در نظریه IRT، هم برای سؤال و هم برای آزمون، تابع آگاهی رسم می‌شود. بنا به‌نظر همبلتون (۱۹۸۹)، می‌توان تابع آگاهی یک سؤال برای سطح معیّنی از توانایی را به‌عنوان نسبت مجذور مشتق تابع ویژگی سؤال به واریانس آن سؤال تعریف کرد.

$$I_i(\theta) = \frac{(p_i'(\theta))^2}{p_i(\theta)q_i(\theta)}$$ (همبلتون، ۱۹۸۹)

که در آن: $i = 1,2,...,n$ است. به جای این فرمول که صورت اجایو نرمال تابع آگاهی است، می‌توان از تابع منطقی استفاده نمود که مقادیر حاصل از آن اندکی با مقادیر حاصل از فرمول قبلی تفاوت دارد. لرد (۱۹۸۰) برای محاسبهٔ مقدار آگاهی یک سؤال در مدل سه‌پارامتری، فرمول ساده زیر را پیشنهاد می‌کند.

$$I_i(\theta) = \frac{D^2 a_i^2 (1 - c_i)}{(c_i + e^{DL_i})(1 + e^{-DL_i})^2}$$ (لرد، ۱۹۸۰)

که: $L_i = (\theta - b_i)$ است و a_i، b_i و c_i به‌ترتیب پارامترهای شیب سؤال، دشواری و سطح مجانب منحنی ویژه سؤال هستند. پارامتر دشواری، مکان بیشینهٔ تابع آگاهی سؤال را در پیوستار توانایی نمایش می‌دهد. سؤالاتی که منحنی ویژه آنها دارای شیب بیشتری است، در مقایسه با سؤالاتی که منحنی آنها مسطح‌تر است، اطلاعات بیشتری را برای تصمیم‌گیری به‌دست می‌دهد. در حقیقت، هرگاه نسبت دو پارامتر a_i، ۲ به ۱ باشد، نسبت بیشینهٔ آگاهی حاصل ٤ به ۱ خواهد بود. هنگامی که احتمال حدس کاذب در سؤال وجود داشته باشد، تابع آگاهی برای این نوع سؤال‌ها نامتقارن خواهد بود؛ زیرا آزمودنی‌هایی که توانایی کمتری دارند، بیشتر متوسل به حدس زدن می‌شوند و این امر مقدار آگاهی حاصل از سؤال را در بخش انتهایی پایین مقیاس توانایی کاهش خواهد داد.

از آنجا که نظریهٔ پاسخ سؤال بر تک‌تک سؤالات تشکیل‌دهندهٔ آزمون مبتنی است، به‌عنوان یک نظریهٔ سؤال[1] شناخته می‌شود. تحت این نظریه، هر سؤال آزمون یک ویژگی نهفته را اندازه‌گیری می‌کند. در نتیجه، مقدار آگاهی براساس فقط یک سؤال می‌تواند در هر سطح توانایی محاسبه شود و به‌صورت $I_i(\theta)$ نشان داده می‌شود که i نشان‌دهندهٔ شمارهٔ سؤال است. از آنجا که فقط یک سؤال بررسی می‌شود، مقدار آگاهی در هر نقطه مقیاس توانایی مقدار خیلی کوچکی خواهد بود. یک سؤال، توانایی را با بیشترین دقت در سطحی از توانایی که برابر با پارامتر دشواری سؤال است، اندازه‌گیری می‌کند. مقدار آگاهی سؤال، همچنان‌که توانایی از دشواری سؤال فاصله می‌گیرد، کاهش می‌یابد و در کرانه‌های مقیاس توانایی به صفر نزدیک می‌شود.

از آنجا که یک آزمون برای برآورد توانایی یک آزمودنی به‌کار می‌رود، مقدار آگاهی به‌دست آمده به‌وسیلهٔ آزمون در هر سطح از توانایی نیز می‌تواند به دست آید. یک آزمون، مجموعه‌ای از سؤالات است؛ بنابراین، آگاهی آزمون در سطح توانایی معیّن، به سادگی از جمع آگاهی سؤال‌ها در آن سطح به‌دست می‌آید. در نتیجه، تابع آگاهی آزمون به صورت زیر تعریف می‌شود.

$$I(\theta) = \sum_{i=1}^{N} I_i(\theta) \qquad \text{(بیکر، ۲۰۰۲)}$$

که: $I(\theta)$ مقدار آگاهی آزمون در یک سطح از توانایی θ، $I_i(\theta)$ برابر مقدار آگاهی برای سؤال i در سطح توانایی θ و N برابر تعداد سؤالات آزمون است.

کل سطح تابع آگاهی آزمون خیلی بزرگ‌تر از تابع آگاهی سؤال خواهد بود و در نتیجه، یک آزمون، توانایی را دقیق‌تر و جامع‌تر از یک سؤال تنها اندازه‌گیری می‌کند. هر چه تعداد سؤالات آزمون بیشتر باشد، توانایی یک آزمودنی با دقت بیشتری برآورد

1. Itemized

می‌شود. با ترسیم مقدار آگاهی آزمون در برابر توانایی، نمودار تابع آگاهی آزمون به‌دست می‌آید (بیکر، ۲۰۰۲). در یک آزمون با هدف کلی، تابع آگاهی ایده‌آل، به‌صورت خطی افقی با مقادیر بزرگ I خواهد بود و همهٔ سطوح توانایی با دقت یکسان برآورد خواهد شد. بنا به‌نظر بیکر (۲۰۰۲)، متأسفانه دستیابی به چنین تابع آگاهی دشوار است. اینکه سطوح متفاوت توانایی با درجات متفاوتی از دقت برآورد می‌شود، برای سازندگان و مصرف‌کنندگان آزمون اهمیت قابل ملاحظه‌ای دارد؛ چون بدین معنی است که دقتی که توانایی آزمودنی برآورد می‌شود، بستگی به جایگاه توانایی آزمودنی در مقیاس توانایی دارد.

از آنجایی که هدف از به دست آوردن نمره‌های آزمون در واقع این است که این نمره‌ها ما را در تصمیم‌گیری‌های مناسب یاری نماید، بنابراین، مؤثر بودن آزمون از این طریق تعریف می‌شود که آزمون موردنظر تا چه میزان ما را قادر به تصمیم‌گیری‌های صحیح می‌سازد (ثرندایک، ۱۹۸۲، ترجمهٔ هومن، ۱۳۷۵). زمانی‌که در نظریهٔ پاسخ سؤال مدل مناسب تعیین شد، براساس نوع مدل، می‌توان میزان دقت برآورد توانایی آزمودنی را تعیین نمود و از آنجایی که هر آزمون از تعدادی سؤال تشکیل شده است و هر سؤال می‌تواند نقشی اساسی در تعیین برآورد توانایی آزمودنی داشته باشد، لذا با استفاده از تابع آگاهی سؤال می‌توان میزان آگاهی‌دهندگی هر سؤال را در سطح معّینی ازتوانایی و در نتیجه، میزان آگاهی‌دهندگی کل آزمون را تعیین کرد (همبلتون و کوک[1]، ۱۹۷۷).

با توجه به مطالب نظری فوق، هدف این تحقیق آن است که با استفاده از داده‌های تجربی، در عمل، این دو نظریه را از نظر شاخص‌های دقت آزمون مقایسه کند؛ بنابراین، در مقالهٔ حاضر، پاسخگویی به پرسش زیر موردتوجه قرار گرفته است:

تفاوت دو روش اندازه‌گیری کلاسیک و پاسخ سؤال از نظر اندازه‌گیری شاخص‌های دقت آزمون چیست؟

پژوهشگاه علوم انسانی و مطالعات فرهنگی

جامعه و نمونه

به منظور پاسخگویی به پرسش پژوهشی مورد نظر، از روش کاربردی ـ توصیفی استفاده شد. جامعهٔ این تحقیق شامل کلّیهٔ داوطلبان نظام سالی ــ واحدی ورود به دانشگاه‌های کشور در گروه آزمایشی علوم تجربی سال ۱۳۸٤ هستند. در سال ۱۳۸٤ در رشتهٔ علوم تجربی تعداد ٤۲۱۹٥٦ نفر (۲۹۰۸۷۷ نفر زن و ۱۳۱۰۷۹ نفر مرد) داوطلب شرکت در آزمون سراسری شده بودند که ۳۳٥۱۳۸ نفر داوطلب نظام سالی ـ واحدی و ۸٦۸۱۸ نفر داوطلب نظام قدیم و نظام ترمی ـ واحدی بودند. از کل داوطلبان، تعداد

1. Cook

۳۸٤۲۰۲ نفر (۲٦۸٦۷۲ نفر زن و ۱۱۵۵۳۰ نفر مرد) در این گروه آزمایشی در آزمون شرکت کردند. در نهایت، ۱٤٤۵۸٤ نفر (۱۰۲۱۵۹ نفر زن و ٤۲٤۲۵ نفر مرد) مجاز به انتخاب رشته شدند که ۱۲٦۷٦۷ نفر از نظام سالی ـ واحدی و ۱۷۸۱۷ نفر از نظام قدیم و نظام ترمی ـ واحدی بودند (دفتر بررسی‌های فنی و آماری سازمان سنجش آموزش کشور، ۱۳۸۵). به‌منظور انجام این تحقیق، فقط داده‌های مربوط به داوطلبان نظام سالی ــ واحدی که اکثریت افراد را تشکیل می‌دادند، مورد تحلیل و بررسی قرار گرفت.

با توجه به در دسترس بودن فهرست تصادفی داوطلبان، نمونه‌گیری با استفاده از روش نمونه‌گیری تصادفی انجام شد. ابتدا نمره کلّیهٔ داوطلبان شرکت‌کننده نظام سالی ـ واحدی گروه آزمایشی علوم تجربی محاسبه شد و سپس آزمودنی‌هایی که به درستی به سؤالات پاسخ نگفته بودند (کسانی که پایین‌تر از $\frac{۱}{۴}$ نمره کل یا حد شانس، نمره کسب کرده بودند)، حذف شدند و نمونه‌گیری از بین افراد باقیمانده صورت گرفت.

بنا به‌نظر نانالی (۱۹٦۷، به‌نقل از تروسکوسکی، ۱۹۹۹)، حداقل حجم نمونهٔ پیشنهادی برای به‌کارگیری مؤثر نظریهٔ کلاسیک اندازه‌گیری در دامنه‌ای از حدود ۳۰۰ تا ۵۰۰ متغیر است. البته ترجیح داده می‌شود که حجم نمونه ۱۰۰۰ باشد؛ همچنین، بر اساس نوع مدل انتخاب شده در نظریهٔ پاسخ سؤال، به محقق توصیه می‌شود که حجم نمونه را برای مدل‌های تک‌پارامتری حداقل ۲۰۰ (رایت و استون، ۱۹۷۹)، برای مدل‌های دوپارامتری تا ۵۰۰ (گلدمن و راجو، ۱۹۸٦) و برای مدل‌های سه‌پارامتری بالای ۱۰۰۰ (تیسن و راینر، ۱۹۸۲) درنظر بگیرد.

بر اساس تحقیقات انجام شده در هر دو نظریهٔ کلاسیک و پاسخ سؤال و با توجه به نرم‌افزار کامپیوتری به‌کار رفته برای تحلیل سؤال‌ها، حجم نمونه در هر دو نظریه ۱۰۰۰ نفر انتخاب شد.

پژوهشگاه علوم انسانی و مطالعات فرهنگی

ابزار و روش اجرا

ابزار مورد استفاده برای جمع‌آوری داده‌های پژوهش، پاسخنامه‌های داوطلبان گروه آزمایشی علوم تجربی در سال ۱۳۸٤ در آزمون درس اختصاصی زیست‌شناسی با ۵۰ سؤال بود که در سازمان سنجش آموزش کشور ازسوی استادان زبده و با تجربهٔ کشور طراحی و تهیه شده بود. برای تجزیه و تحلیل داده‌ها از نرم‌افزارهای کامپیوتری SPSS و BILOG.MG استفاده شد. ابتدا تحلیل‌ها و برخی از مشخصه‌های کلاسیک سؤال‌ها و آزمون‌ها از طریق SPSS محاسبه شد و سپس داده‌ها با نرم‌افزار BILOG.MG تحلیل شد.

نتایج

در مدل کلاسیک، دقت اندازه‌گیری به‌وسیلهٔ برآورد پایایی نمرات آزمون سنجیده می‌شود که برای برآورد درجهٔ دقت آزمون‌ها بر اساس این مدل، پایایی به سه روش بررسی گردید. این سه روش عبارت‌اند از: روش آلفای کرونباخ[1]، روش دونیمه‌ای اسپیرمن ـ براون[2] و روش گاتمن[3]. مقادیر پایایی به سه روش مذکور در جدول (۱) گزارش شده است.

جدول(۱) مقادیر پایایی در روش کلاسیک اندازه‌گیری به سه روش آلفای کرونباخ، دونیمه‌ای اسپیرمن ـ براون وگاتمن

آلفای کرونباخ	۰/۸۸۸۲
دونیمه‌ای اسپیرمن ـ براون	۰/۸٤۸۷
گاتمن	۰/۸٤٤۸

علاوه بر آن، برای بررسی تأثیر تک‌تک سؤال‌ها بر پایایی، آزمون زیست‌شناسی رشتهٔ تجربی سال ۱۳۸٤ که شامل ۵۰ سؤال بود، به دو نیمه تقسیم شد. سؤال‌های زوج، یک نیمه و سؤال‌های فرد، نیمهٔ دیگر آزمون را تشکیل دادند. در ضمیمهٔ ۱، مشخصه‌های آماری سؤال‌های نیمهٔ اول آزمون، و در ضمیمهٔ ۲، مشخصه‌های آماری سؤال‌های نیمهٔ دوم آزمون ارائه شده است. بررسی ستون مقدار آلفا در صورت حذف هر سؤال نشان می‌دهد که سؤال ٤ در نیمهٔ اول آزمون و سؤال ۲۷ در نیمهٔ دوم آزمون اگر حذف شوند، مقدار آلفا افزایش می‌یابد. حذف بقیهٔ سؤال‌ها تغییر معناداری در مقدار پایایی ایجاد نمی‌کند؛ همچنین، بررسی ستون همبستگی سؤال با کل آزمون نشان می‌دهد که سؤال‌های ٤، ۲۲ و ۳٦ (در نیمهٔ اول آزمون) و سؤال‌های ۱۹، ۲۷ و ٤۹ (در نیمهٔ دوم آزمون) رابطهٔ کمی با نمرهٔ کل آزمون دارند، که باید مورد بررسی قرار گیرند.

در نظریهٔ IRT ابتدا با استفاده از رابطهٔ $r_{xx'} = \frac{s_t^2}{s_t^2 + s_e^2}$ (ثرندایک، ۱۹۸۲، ترجمهٔ هومن، ۱۳۷۵) برای تعیین پایایی که یکی از شاخص‌های مهم و شناخته شده در نظریهٔ کلاسیک است، اقدام به برآورد ضریب پایایی در نقاط مختلفی از سطوح توانایی

1. Cronbach Alpha
2. Split-half Spearman-Brown
3. Guttman

شـــد و ضـــریب پایـــایی آزمـــون در ســـطوح مختلـــف توانـــایی از طریـــق فرمول $r_{xx'}(\theta)=\frac{s_\theta^2}{s_\theta^2+s_e^2}$ برآورد گردید که نتایج آن در جدول (۲) آمده است.

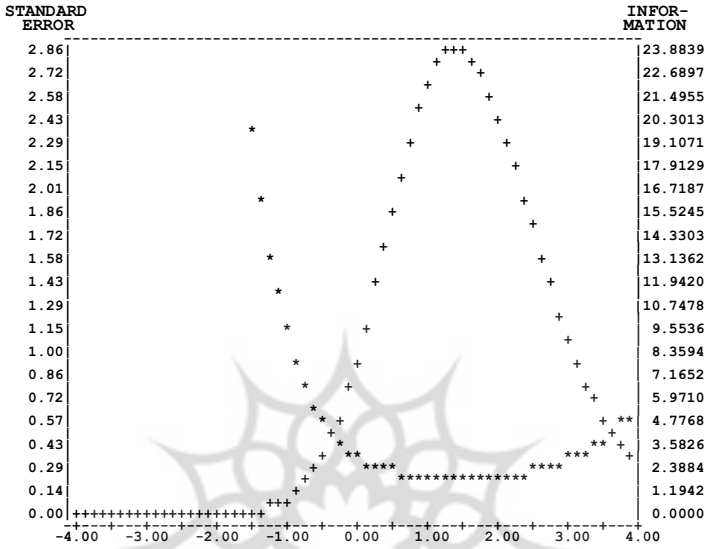
جدول(۲) برآورد ضریب پایایی در سطوح مختلف توانایی

سطوح مختلف توانایی (θ)	$\theta=+3$	$\theta=+2$	$\theta=+1$	$\theta=0$	$\theta=-1$	$\theta=-2$	$\theta=-3$
برآورد ضریب پایایی	۰/۹۷۸	۰/۹۹۷	۰/۹۹۹	۰/۹۷٦	۰/۸٦٦	۰/۵۲۹	۰/۱٦٦
میزان خطای استاندارد اندازه‌گیری	۰/۲۹	۰/۰۰۲	۰/۰۰۱	۰/۰۲۲	۰/۱۳۹	۰/۸۰۳	٤/۵۱

همان‌طور که در جدول (۲) ملاحظه می‌شود، بیشترین دقت اندازه‌گیـری در سـطح توانایی(θ=+۱) به دست آمده است که دارای کمترین خطـای اسـتاندارد انـدازه‌گیـری می‌باشد. میزان برآورد ضریب پایایی برای θ = +۳ نیـز برابـر ۰/۹۷۸ بـه دسـت آمـد. جدول (۲) به‌خوبی نشان می‌دهد که مقادیر پایایی در توانایی‌های مختلف متفاوت است و این یک ویژگی مهم در IRT است؛ در حالی‌که در CTT برای کل آزمون فقط یـک مقدار برای ضریب پایایی به دست می‌آید.

در نظریهٔ پاسخ سؤال، پس از بـرآورد مشخـصه‌هـای سـؤال، مقـادیر آگـاهی بـرای تک‌تک سؤال‌ها و برای کل آزمون محاسبه مـی‌گـردد. در ضـمیمهٔ(۳)، مقـادیر آگـاهی سؤال‌های آزمون ارائه شده است. همان‌طور که گفتـیم، در نظریـهٔ IRT دقـت بـرآورد پارامترها، از جمله پارامتر توانایی را می‌توان با استفاده از تابع ریاضی $\frac{P'(\theta)^2}{P(\theta)\cdot Q(\theta)}$ که به تابع آگاهی معروف است، به‌دست آورد. نمودار(۱)، تابع آگاهی آزمون و خطـای استاندارد برآورد آزمون زیست‌شناسی سال ۱۳۸٤ را نشان می‌دهد.

نمودار(۱) تابع آگاهی و خطای استاندارد برآورد آزمون



پس از محاسبهٔ میزان پایایی، میزان آگاهی برای تک‌تک سؤال‌ها و کل آزمون محاسبه گردید. مقادیر آگاهی در سطوح متفاوت توانایی(θ =۳- تا θ =+۳) در جدول (۳) نشان داده شده است. بررسی جدول (۳) نشان می‌دهد که در فاصله ۰/۵ تا ۲، توانایی افراد با بیشترین دقت برآورد شده است.

جدول (۳) مقادیر آگاهی در سطوح مختلف توانایی

	θ =+۳	θ =+۲	θ =+۱	θ =0	θ = -۱	θ = -۲	θ = -۳
مقادیر آگاهی برآوردشده	۸/۹۵	۲۰/۵٤	۲۲/۰۸	۸/۰۳	۰/۸۱	۰/۰۳	۰/۰۰۱
میزان خطای استاندارد اندازه‌گیری	۰/۲۹	۰/۰۰۲	۰/۰۰۱	۰/۰۲۲	۰/۱۳۹	۰/۸۰۳	٤/۵۱

همان‌طور که در جدول (۳) ملاحظه می‌شود، مقدار آگاهی برآوردشده در توانایی (۱+=θ) بیشتر از سطوح دیگر است (لازم به‌ذکر است که حداکثر آگاهی آزمون دقیقاً در ۱/۳۷۵۰+=θ برابر با ۲۳/۸۸ است)؛ بنابراین، آزمون مذکور برای آزمودنی‌ها با این سطح توانایی بیشترین کارایی را دارد. از بررسی تابع آگاهی این آزمون مشاهده می‌شود که میزان آگاهی در توانایی ۳+= θ در حدود ۸/۹۵ است؛ متوسط آگاهی نیز ۹/۹ به‌دست آمد.

بحث و نتیجه‌گیری

در بررسی تفاوت دو نظریهٔ کلاسیک اندازه‌گیری و پاسخ سؤال از نظر شاخص‌های دقت اندازه‌گیری آزمون، باید گفت، که ضریب پایایی به‌عنوان یک شاخص دقت اندازه‌گیری در نظریهٔ کلاسیک، یک مقدار ثابت برای تمام توانایی‌ها به دست می‌دهد. با توجه به اینکه دقت اندازه‌گیری هر آزمون در سطوح مختلف توانایی متفاوت است، یکسان فرض کردن دقت اندازه‌گیری در نظریهٔ کلاسیک برای تمام سطوح توانایی قابل قبول به‌نظر نمی‌رسد، اما در نظریهٔ پاسخ سؤال، شاخص دقت برآورد پارامترها و توانایی، تابع آگاهی است که با خطای استاندارد برآورد پارامترها همراه بوده و برای سطوح مختلف توانایی، مقادیر متفاوتی را نشان می‌دهد (همبلتون، ۱۹۸۹ و بیکر، ۲۰۰۲؛ و ...). تابع آگاهی آزمون یک خصیصهٔ بسیار مفید در نظریهٔ پاسخ سؤال است. اساساً، تابع آگاهی بیان می‌کند که آزمون در برآورد توانایی در سراسر طیف چگونه عمل می‌کند.

در بررسی عملی سؤال تحقیق، مشخص شد که تابع آگاهی در IRT دقیقاً نشان می‌دهد که در هر سطح از توانایی میزان دقت برآورد θ به‌چه میزان است و آزمون مذکور در کدام سطح از توانایی بیشترین مقدار آگاهی را دارد که این امر در نظریهٔ کلاسیک قابل مشاهده نیست. همان‌طور که در جدول (۲) مشاهده می‌شود، مقادیر پایایی برای سطوح مختلف توانایی یکسان به‌دست نیامده است، (مطابق با نظرهای همبلتون، ۱۹۸۹، بیکر، ۲۰۰۲ و ...). یکسان نبودن مقادیر پایایی در سطوح مختلف توانایی، امری بسیار روشن و محسوس و ناشی از آن است که دقت اندازه‌گیری آزمون برای توانایی‌های مختلف نمی‌تواند یکسان باشد. در حالی‌که، در روش CTT فقط یک مقدار ثابت پایایی برای تمام توانایی‌ها حاصل می‌شود. این امر یکی از نارسایی‌های نظریهٔ کلاسیک اندازه‌گیری نسبت به IRT است. متفاوت بودن دقت اندازه‌گیری هر آزمون در سطوح مختلف توانایی، در نظریهٔ کلاسیک اندازه‌گیری قابل بررسی نیست؛ بدین ترتیب، می‌توان نتیجه گرفت که در IRT، سازندهٔ آزمون با استفاده از نمودار تابع آگاهی قادر است، مشخص نماید، در سطوح مختلف توانایی تا چه حد اطلاعات مؤثر

به‌دست می‌آید و همچنین می‌تواند با توجه به آگاهی حاصل از سؤال‌ها، کارایی آزمون در سطوح مختلف توانایی را ارزیابی نماید؛ بنابراین، نظریهٔ IRT از نظر اندازه‌گیری شاخص‌های دقت آزمون بر نظریهٔ CTT ارجحیت دارد. از این ویژگی نظریهٔ پاسخ سؤال می‌توان در ساخت آزمون‌ها استفاده کرد؛ بدین‌صورت که، بسته به هدف آزمون می‌توان سؤال‌هایی را انتخاب کرد که بیشترین دقت اندازه‌گیری را در سطح خاصی از توانایی داشته باشند؛ بنابراین، فقط افرادی قادر به پاسخ‌گویی به سؤال‌های آزمون خواهند بود که توانایی آنها در سطح موردنظر است. البته بسته به هدف سازندهٔ آزمون، ملاک انتخاب سؤال‌ها متفاوت خواهد بود؛ به‌عنوان مثال، اگر هدف سازندهٔ آزمون این است که تمام افراد را با دقت یکسانی اندازه‌گیری کند، باید سؤال‌هایی را در فرم نهایی آزمون بگنجاند که دامنهٔ دشواری آنها کل پیوستار خصیصه را بپوشاند و در عین‌حال، همهٔ این سؤال‌ها نیز باید قدرت تشخیص زیادی داشته باشند.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

منابع

آلن، مری‌جی و ین، وندی ام (۱۹۷۹). *مقدمه‌ای بر نظریه‌های اندازه‌گیری (روان‌سنجی)*. ترجمه علی دلاور(۱۳۷٤)، تهران: انتشارات سمت.

آناستازی، آ. (۱۹۸۸). *روان‌آزمایی*. ترجمه محمدنقی براهنی(۱۳۷۹)، تهران: انتشارات دانشگاه تهران.

ثرندایک، رابرت (۱۹۸۲). *روان‌سنجی کاربردی*. ترجمه حیدرعلی هومن(۱۳۷۵)، تهران: انتشارات دانشگاه تهران.

دفتر بررسی‌های فنی و آماری سازمان سنجش آموزش کشور- حوزه معاونت فنی و آماری (۱۳۸۵). *کارنامه آزمون سراسری سال ۱۳۸٤*. تهران: مرکز انتشارات سازمان سنجش آموزش کشور.

Baker, F. (2002). *The basic of item response theory*. ERIC Clearinghouse on Assessment and Evaluation (1-886047-03-0).

Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley & Sons.

Hambleton, R.K. (1989). Principles and selected applications of item response theory. In: Linn, R. (ed), *Educational measurement* (3rd ed). Newyork: Macmillan, 147-200.

Hambleton, R.K. & Cook, L.L. (1977). Latent trait models and their use in the analysis of educational test data. *Journal of ducational Measurement*, 14(2), 75-94.

Hambleton, R.K. & VanderLinden, W.J. (1982). Advance in item response theory and applications: an introduction. *Applied Psychological Measurement*, 6(4), 373-378.

Hambleton, R.K., Swaminathan, H. & Rojers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: sage.

Hambleton, R.K.; Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.

Golman, S. H. & Raju, N. S. (1986). Recovery of one- and two – parameter logistic Item parameters: an empirical study. *Educational and Psychological Measuerment*. 46, 11-21.

Lord, F. M. (1980). *Applications of item response theory to practice testing problems*. Hillsdale, N.J: Lawrence Erlbaum.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores reading*. MA: Addison-Wesley.

Truskosky, D. M. (1999). *An empirical examination of classical test theory and item response theory parameters: implications for research and practice in small-and large-sample assessment.* Department of Psychology in the graduate school Southern Illinois university at Carbondale.

Thissen, D. & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika*, 47, 397-412.

Wright, B.D. & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

ضمیمهٔ (۱) جدول مشخصه‌های آماری نیمهٔ اول آزمون
(سؤالات زوج) درCTT

شمارهٔ سؤال در آزمون اولیه	میانگین مقیاس در صورت حذف سؤال	واریانس مقیاس درصورت حذف سؤال	همبستگی سؤال با کل آزمون	مقدار آلفا در هر سؤال
۲	۱۳/۷۱۲٤	۷۰/٦۱٤۸	۰/۲۱۱۵	۰/۸۸۸۳
٤	۱۳/۵۵٤۹	۷۱/٤۵٦۲	۰/۱۰۲۵	۰/۸۹۰۱
٦	۱۳/۸۷٦۳	۷۰/٤۸۷۰	۰/۲۷۹۸	۰/۸۸۷۱
۸	۱۳/۹۲۳۳	۷۰/۳۷۲۷	۰/۳۳۱۷	۰/۸۸٦۵
۱۰	۱۳/۷۵۸۱	٦۹/۵۲۲۰	۰/۳٦۰۸	۰/۸۸٦۱
۱۲	۱۳/۸٤۳۳	٦۹/۱٤۹۳	۰/٤۵۳۱	۰/۸۸٤۸
۱٤	۱۳/۵۰۰۲	۷۰/۵۵٤۳	۰/۲۱٤٤	۰/۸۸۸۳
۱٦	۱۳/۸٦۵۳	۷۰/۰۹٦۲	۰/۳۳۰۷	۰/۸۸٦۵
۱۸	۱۳/۹۷۵۱	۷۰/٦۳۳۰	۰/۳٤٦٤	۰/۸۸٦٤
۲۰	۱۳/۷٦۰۵	٦۸/٦۹۸٦	۰/٤٦۹۷	۰/۸۸٤٤
۲۲	۱۳/۹۹۱۲	۷۱/۹۳۸۱	۰/۱۰۹۱	۰/۸۸۸۷
۲٤	۱۳/٦۲۸٤	۷۰/۵۳۹۸	۰/۲۱۲۹	۰/۸۸۸٤
۲٦	۱۳/۷۸۷۵	٦۸/۸۸٤۳	۰/٤۵۷۱	۰/۸۸٤۷
۲۸	۱۳/۸۲٦۱	٦۹/٤۸۳۱	۰/۳۹۵۲	۰/۸۸۵٦
۳۰	۱۳/۸۸۷۳	٦۸/۷۵۳۸	۰/۵۵۱٤	۰/۸۸۳۷
۳۲	۱۳/۹۰۰۷	۷۰/۳۷۰۵	۰/۳۱۳٤	۰/۸۸٦۷
۳٤	۱۳/۸۳٦۲	۷۰/۰۱۳۷	۰/۳۲٦٤	۰/۸۸٦۵
۳٦	۱۳/۷۲۳٤	۷۱/۵۰۳۲	۰/۱۰۲۷	۰/۸۸۹۹
۳۸	۱۳/۸۵۸٤	٦۹/۸۳۷۹	۰/٤٦۸۰	۰/۸۸۵۱
٤۰	۱۳/۸۸٤٤	٦۹/۱۹۸۸	۰/٤۷۹۸	۰/۸۸٤٦
٤۲	۱۳/۸۷۱۰	٦۹/۱۷۰۷	۰/٤۷۱۷	۰/۸۸٤٦
٤٤	۱۳/۹۱۹۹	٦۹/٦٦۷۰	۰/٤٤۳٦	۰/۸۸۵۱
٤٦	۱۳/۷۹۵۸	٦۸/۱٦۸۱	۰/۵۵۹۱	۰/۸۸۳۲
٤۸	۱۳/۷٤۳٦	٦۸/۱۹۸۸	۰/۵۲۸۵	۰/۸۸۳۵
۵۰	۱۳/۵۹۲۲	۷۰/۱٤۵۷	۰/۲۵۹٦	۰/۸۸۷۷

ضمیمهٔ (۲) جدول مشخصه‌های آماری نیمهٔ دوم آزمون (سؤالات فرد) در CTT

شمارهٔ سؤال در آزمون اولیه	میانگین مقیاس در صورت حذف سؤال	واریانس مقیاس درصورت حذف سؤال	همبستگی سؤال با کل آزمون	مقدار آلفا در هر سؤال
۱	۱۳/۵۸۳۹	۷۰/٦٦۷۹	۰/۱۹٦۵	۰/۸۸۸۷
۳	۱۳/٦۵۲۲	٦۸/۳۹۰۳	۰/٤۷۹۷	۰/۸۸٤۲
۵	۱۳/۸۱۹۳	٦۸/۵۳۰۹	۰/۵۲٤۱	۰/۸۸۳۷
۷	۱۳/۸۰٤۳	۷۰/۵۷٤۹	۰/۲۳٦۸	۰/۸۸۷۸
۹	۱۳/۸۱۵٦	٦۹/۰٦۸۹	۰/٤٤٦۸	۰/۸۸٤۸
۱۱	۱۳/۸۱۳۵	٦۸/۳۷۲۱	۰/۵٤۲٤	۰/۸۸۳۵
۱۳	۱۳/۹۱۳۲	۷۰/٦۷۸۹	۰/۲۷٤۱	۰/۸۸۷۲
۱۵	۱۳/۷٦۰۲	٦۸/۸۵۰٤	۰/٤٤۹۵	۰/۸۸٤۷
۱۷	۱۳/۸۵٦۲	٦۹/۹٤۸۸	۰/۳٤٦٦	۰/۸۸٦۳
۱۹	۱۳/٦۲۰۸	۷۱/۲٤٦۱	۰/۱۲۷٦	۰/۸۸۹۷
۲۱	۱۳/۹۵۸٦	۷۱/۲۷۵٦	۰/۲۰۹۲	۰/۸۸۷۸
۲۳	۱۳/۸۳۰۳	٦۹/٤۹۰۰	۰/۳۹٦٦	۰/۸۸۵٦
۲۵	۱۳/۵۷۲۷	۷۰/٦۵۱۹	۰/۱۹۸۵	۰/۸۸۸٦
۲۷	۱۳/۸۰۹۷	۷۲/۰۸۲۳	۰/۰۳۷۷	۰/۸۹۰٦
۲۹	۱۳/۸۵۰۱	٦۸/۷۷۳٤	۰/۵۱۲۵	۰/۸۸٤۰
۳۱	۱۳/۹۱۳۸	٦۹/۷۲۹٤	۰/٤۲٦٦	۰/۸۸۵۳
۳۳	۱۳/۸۳۵٤	٦۸/۷٤٤۰	۰/۵۰۵٤	۰/۸۸٤۱
۳۵	۱۳/۹٤۱۰	۷۰/۱۸۷۷	۰/۳۸۱٤	۰/۸۸۵۹
۳۷	۱۳/۹٦۹۹	۷۱/۲۳٤٤	۰/۲۲٦۸	۰/۸۸۷٦
۳۹	۱۳/۷۳٤۷	٦۸/۹۷٦٤	۰/٤۲٤۱	۰/۸۸۵۱
٤۱	۱۳/۹۷۵۳	۷۰/۸۱۸۹	۰/۳۱۱٤	۰/۸۸٦۸
٤۳	۱۳/۸٦۳۰	٦۹/٦۰۳٤	۰/٤۰۱۳	۰/۸۸۵۵
٤۵	۱۳/۷٦۰٦	٦۸/٤۳٦۹	۰/۵۰٤۳	۰/۸۸۳۹
٤۷	۱۳/۷۰٤٤	٦۷/۷۰٦۹	۰/۵۷۸٤	۰/۸۸۲۷
٤۹	۱۳/۷۹۷۰	۷۱/۲۵٤۰	۰/۱٤٤۹	۰/۸۸۹۱

ضمیمهٔ(۳) جدول مقادیر آگاهی سؤالات آزمون زیست‌شناسی رشتهٔ علوم تجربی ۱۳۸٤

شمارهٔ سؤال	ماکزیمم آگاهی	خطای استاندارد	نقطهٔماکزیمم آگاهی	خطای استاندارد
۱	۰/۲۹۹۷	۰/۰٤۰۲	۱/۷۳۸۱	۰/۱٦۱۹
۲	۰/۳۵۱٤	۰/۰۵۰٦	۲/۲۷۳۸	۰/۱۸٤٤
۳	۰/۸۱۱۸	۰/۱۱۳۷	۰/٤۷۹٤	۰/۰۷۲۰
٤	۰/۱۹۵٦	۰/۰۲۳۸	۲/۲۷۱۹	۰/۲۱۰٤
۵	۰/۸۸۷۳	۰/۱۲٤۲	۰/۹۲۱۱	۰/۰۵۸۷
٦	۰/٤۹۷۷	۰/۰۷۵۱	۲/۰۲۷۸	۰/۱۳۱٦
۷	۰/٤۱۱۳	۰/۰٦۰۵	۲/۰۵۸۵	۰/۱۵۰۳
۸	۰/۵۱۵۱	۰/۰۷۳۰	۱/٦۸٤۳	۰/۱۰۲٤
۹	۰/٦٦٦٦	۰/۰۹۸۱	۱/۳٦۷۵	۰/۰۸۹٤
۱۰	۰/٤۷۰۸	۰/۰٦۳۳	۱/۲۵۵۰	۰/۰۹۸۲
۱۱	۱/۰۳۵۱	۰/۱٤٤۹	۰/۸۱٤۲	۰/۰۵۲۸
۱۲	۰/۷۷۳۰	۰/۱۱۷۱	۱/۲۵۵٦	۰/۰۷٦٤
۱۳	۰/۵۲۳۳	۰/۰۸۰٦	۲/۱۳۷۱	۰/۱۳۹٦
۱٤	۰/۲٦۸۲	۰/۰۳٤۱	۱/۵۵۲۵	۰/۱۵۷۱
۱۵	۰/۵۲٦۷	۰/۰٦۵۰	۰/۹٦۳٤	۰/۰۷٦۷
۱٦	۰/۵۵۲۰	۰/۰۷۹۸	۱/۵۵۲۰	۰/۰۹۷۸
۱۷	۰/۵۳۵۳	۰/۰۸۱۰	۱/۹۱۵۸	۰/۱۲۳۸
۱۸	۰/٦۷۵۸	۰/۱۰۸٦	۲/۱۱۳۹	۰/۱۲۵۲
۱۹	۰/۳۲۷۲	۰/۰٤٦۹	۲/۱۵۰۱	۰/۲۱۱۱
۲۰	۰/٦۵٦۰	۰/۰۸۹۷	۱/۰۱۹۰	۰/۰۸۰۰
۲۱	۰/۵۳۱۳	۰/۰۸٤۹	۲/۵۰٦۹	۰/۱٦۹۳
۲۲	۰/۵۰۲۵	۰/۰۸۱۰	۲/۸۲٦٤	۰/۲۰۱٦
۲۳	۰/٦۰٤۱	۰/۰۹٦۳	۱/۲٤۸۳	۰/۰۸۷٤
۲٤	۰/۳۳۳۲	۰/۰٤۵٤	۱/۷۰۸۹	۰/۱٤۹۹
۲۵	۰/۳٤۸۳	۰/۰٤۹۵	۱/۷٤۰۳	۰/۱٦۵۱
۲٦	۰/٦٤۹٤	۰/۰۹۳۸	۰/۹٤۳۸	۰/۰٦٤۷
۲۷	۰/۳۸۰۳	۰/۰۵۸٤	۲/۹٦٦۲	۰/۲٦۵۸
۲۸	۰/٦۰۱۷	۰/۰۸۷۳	۱/٤۲۳۹	۰/۰۹٤۳
۲۹	۰/۸۸۷۳	۰/۱۲۵۹	۱/۰۷٤۷	۰/۰٦۰۸
۳۰	۰/۰٤۰٦	۰/۱۵۹٤	۱/۲۲٤۲	۰/۰٦۲۸
۳۱	۰/۷۳۵۰	۰/۱۱۰۲	۱/۵۵۲۲	۰/۰۸۵۳
۳۲	۰/۷۲۰۱	۰/۱۲۱۹	۱/۸۸۳٦	۰/۱۱۸۲
۳۳	/۷۷۹۲	۰/۱۱٤٤	۱/۱۹۷٤	۰/۰۷٤۵

ادامهٔ ضمیمهٔ (۳) جدول مقادیر آگاهی سؤالات آزمون زیست‌شناسی
رشتهٔ علوم تجربی ۱۳۸٤

شمارهٔ سؤال	ماکزیمم آگاهی	خطای استاندارد	نقطهٔ ماکزیمم آگاهی	خطای استاندارد
۳٤	۰/۵۱۸۲	۰/۰۷۷٦	۱/۷۵۸٤	۰/۱۱۵۵
۳۵	۰/۵۳۱۹	۰/۰۷۳٤	۱/۷۳۷٤	۰/۰۹٦٦
۳٦	۰/۳٦٤٦	۰/۰۵۳۹	۲/۵۷۰۸	۰/۲۲٤۳
۳۷	۰/٦٦٦۲	۰/۱۱۲۲	۲/۳۲۲۲	۰/۱٤۹٤
۳۸	۰/۸۹۰٦	۰/۱٤۸۰	۱/۷۱٤٦	۰/۰۹٤۰
۳۹	۰/٦۳۸۷	۰/۰۹۲۱	۰/۹۹۳۲	۰/۰۸۵۲
٤۰	۰/٦٤۰٦	۰/۰۹۱٦	۱/٤۵٦۹	۰/۰۸٤۷
٤۱	۰/٦۷۹۹	۰/۱۱۱۹	۲/۲۱٤٦	۰/۱۳۳۹
٤۲	۰/۷۵٦۳	۰/۱۰٦۹	۱/۲٤۷۳	۰/۰۷۲۳
٤۳	۰/٤۸۸۱	۰/۰٦٦٦	۱/۵۱۸٤	۰/۰۹٦۷
٤٤	۰/۸۱۷۰	۰/۱۲٦٤	۱/۵۱۸٦	۰/۰۸۲٤
٤۵	۰/۷٤۳۸	۰/۰۹۹۰	۰/۸۵۵۰	۰/۰٦۹۷
٤٦	۰/۸۹٤۱	۰/۱۱٤۱	۰/۷۸۱۵	۰/۰۵٤۰
٤۷	۰/۲٦۰۰	۰/۱۷۲٤	۰/٤۵۳٤	۰/۰٤۹٦
٤۸	۰/۹٤۵٤	۰/۱۳۷۸	۰/۷۵۱۰	۰/۰٦۲۹
٤۹	۰/۳۹۷۰	۰/۰۵۹٤	۲/۳٦۸۹	۰/۱۸۲۱
۵۰	۰/۳٤۹٤	۰/۰٤۹۳	۱/۷۰۰٦	۰/۱۵۵۰

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی