

## برآورد ویژگی‌های سؤال با استفاده از مدل همگنی مضاعف در رویکرد ناپارامتریک نظریه سؤال پاسخ

علی مقدم‌زاده\*

رضا پیروی\*\*

سیده‌طیبه مطیعی لنگرودی\*\*\*

### چکیده

هدف از پژوهش حاضر، برآورد ویژگی‌های سؤال با استفاده از مدل همگنی مضاعف در رویکرد ناپارامتریک نظریه سؤال پاسخ است. بدین منظور، پاسخنامه‌های داوطلبان رشته ریاضی فیزیک آزمون ورودی دانشگاه‌های کشور در سال ۱۳۸۴، در بخش آزمون اختصاصی درس ریاضی مورد استفاده قرار گرفت. از بین کلیه داوطلبان شرکت‌کننده در این رشته، با توجه به نرم‌افزار مورد استفاده در سازمان سنجش، به روش نمونه‌گیری سیستماتیک یک گروه نمونه ۳۰۰۰ نفری و سپس از میان آنها، دو نمونه ۸۰ و ۸۰۰ نفری به‌طور کاملاً تصادفی انتخاب شدند و دو سؤال این پژوهش، بر روی این دو گروه مورد بررسی قرار گرفت. به‌منظور مقایسه شاخص‌های برآوردشده سؤال‌ها در دو گروه، از آزمون  $t$  وابسته و ضریب همبستگی استفاده شد. نتایج تحقیق نشان داد که بین شاخص‌های محاسبه‌شده سؤال با استفاده از مدل همگنی مضاعف در رویکرد ناپارامتریک نظریه سؤال پاسخ، در دو نمونه تفاوتی وجود ندارد و این مدل در نمونه‌های کم نیز قابل استفاده است.

**واژگان کلیدی:** مدل همگنی مضاعف، رویکرد ناپارامتریک، نظریه سؤال پاسخ، ویژگی‌های سؤال

\*دانشجوی دکترای سنجش و اندازه‌گیری دانشگاه علامه طباطبایی

\*\*کارشناس ارشد دفتر آزمون‌سازی و روان‌سنجی سازمان سنجش آموزش کشور

\*\*\*کارشناس ارشد دفتر آزمون‌سازی و روان‌سنجی سازمان سنجش آموزش کشور

## مقدمه

تلاش‌های نخستین تکوین نظریه کلاسیک اندازه‌گیری<sup>۱</sup> در دهه ۱۸۹۰ آغاز شد. این نظریه یکی از روش‌های دیرینه در ساخت و توسعه آزمون‌ها در حوزه علوم انسانی است که از اوایل دهه ۱۹۰۰ برای ساخت و نمره‌گذاری ابزارهای اندازه‌گیری استفاده شده است. اوج تکامل این نظریه را بعد از اسپیرمن<sup>۲</sup>، می‌توان در دو کتاب: (۱) «مبانی نظری آزمون‌های روانی»<sup>۳</sup> گالیکسن<sup>۴</sup> (۱۹۵۰) و «تئوری‌های آماری نمرات آزمون‌های روانی»<sup>۵</sup> لرد<sup>۶</sup> و ناویک<sup>۷</sup> (۱۹۶۸) مشاهده کرد (همبلتون<sup>۸</sup> و واندر لیندن<sup>۹</sup>، ۱۹۸۲).

اگرچه نظریه کلاسیک اندازه‌گیری مدت زمان طولانی به جامعه روان‌سنجی خدمت کرده است، اما برخی مطالعات، نشان از وجود محدودیت‌هایی در این نظریه (از جمله گالیکسن، ۱۹۵۰؛ لرد و ناویک، ۱۹۶۸؛ همبلتون و سوامیناتان<sup>۱۰</sup> و راجرز<sup>۱۱</sup>، ۱۹۹۱) و آزمون‌های ساخته شده بر اساس آن دارند (به لرد، ۱۹۸۰؛ و همبلتون، ۱۹۸۹ مراجعه شود).

زمینه ارائه نظریه‌های جدید اندازه‌گیری از اوایل نیمه دوم قرن بیستم به وسیله افرادی چون لرد در فاصله (۱۹۵۳ تا ۱۹۵۲)، راش<sup>۱۲</sup> (۱۹۶۸ تا ۱۹۵۸)، رایت<sup>۱۳</sup> (۱۹۷۹)، همبلتون و جونز<sup>۱۴</sup> (۱۹۸۳ و ۱۹۷۹) و غیره فراهم شد. نظریه‌های جدید اندازه‌گیری چه از لحاظ روش‌های آماری و به‌کارگیری توابع و مدل‌های ریاضی و چه از جهت مفروضه‌های نظری و نتایج کاربردی، تفاوت‌های چشمگیری با نظریه کلاسیک اندازه‌گیری دارند. نظریه‌های جدید در دهه ۱۹۵۰ به‌عنوان یک جایگزین برای نظریه کلاسیک اندازه‌گیری معرفی شدند.

نظریه سؤال پاسخ<sup>۱۵</sup>، یک نظریه جامع آماری درباره عملکرد سؤال‌های آزمون، آزمودنی‌ها و چگونگی سنجش توانایی‌هایی است که به وسیله سؤال‌ها در آزمون

1. Classical Test Theory (CTT)
2. Spearman
3. Theory of Mental Tests (IRT)
4. Gulliksen
5. Statistical theories of mental test scores
6. Lord
7. Novick
8. Vander Linden
9. Hambleton
10. Swaminathan
11. Rojers
12. Rasch
13. Wright
14. Jones
15. Item response theory

اندازه‌گیری می‌شود. مقیاس اندازه‌گیری سؤال پاسخ‌ها در این نظریه می‌تواند گسسته یا پیوسته باشد و به صورت دو ارزشی و یا چند ارزشی نمره‌گذاری شود. طبقات نمره سؤال نیز می‌تواند منظم و یا نامنظم باشد؛ می‌تواند یک توانایی و یا چند توانایی در آزمون مستتر باشد؛ و همچنین در ارتباط بین سؤال پاسخ‌ها و توانایی یا توانایی‌هایی که اندازه‌گیری می‌شود، چند روش (یا مدل) وجود دارد (همبلتون و جونز، ۱۹۹۳). نظریه سؤال پاسخ را می‌توان به دو رویکرد پارامتریک<sup>۱</sup> و ناپارامتریک<sup>۲</sup> تقسیم نمود. رویکرد ناپارامتریک سؤال پاسخ، این امکان را فراهم می‌سازد تا پارامترهای هر سؤال مثل ضریب دشواری و ضریب تمیز سؤال را با رتبه‌بندی پاسخ‌دهندگان بر اساس نمره آنان (تعداد پاسخ‌های درست به علاوه خطای تصادفی) که بر اساس توانایی ( $\theta$ ) مرتب شده‌اند، برآورد کنیم. مدل‌های رویکرد ناپارامتریک سؤال پاسخ عبارت‌اند از: (الف) مدل همگنی یکنواخت<sup>۳</sup> و (ب) مدل همگنی مضاعف<sup>۴</sup>. مدل همگنی مضاعف بر مبنای مفروضه‌های تک‌بعدی بودن<sup>۵</sup>، استقلال موضعی<sup>۶</sup> و یکنواختی<sup>۷</sup> سؤال‌ها قرار گرفته است (سیجت‌سما<sup>۸</sup> و مولن‌آیر<sup>۹</sup>، ۲۰۰۲/ به نقل از مقدم‌زاده، ۱۳۸۶):

### الف) تک‌بعدی بودن

اولین مفروضه، تک‌بعدی بودن است؛ یعنی همه سؤال‌های تست، یک صفت مکنون مشابه را اندازه می‌گیرند (سیجت‌سما و مولن‌آیر، ۲۰۰۲/ ترجمه ذوالفقارنسب، ۱۳۸۵)؛ به تعبیر دیگر، تک‌بعدی بودن آزمون بدین معنی است که احتمال عملکرد موفقیت‌آمیز آزمودنی در مجموعه‌ای از سؤال‌ها می‌تواند به صورت یک مدل ریاضی که فقط یک پارامتر توانایی دارد، تعیین شود (درانز<sup>۱۰</sup> و کینگستن<sup>۱۱</sup>، ۱۹۵۰ به نقل از مطیعی‌لنگرودی، ۱۳۸۱). همبلتون (۱۹۸۹) می‌گوید که البته مفروضه تک‌بعدی به صورت کامل صادق نخواهد بود؛ زیرا همیشه یک‌سری عوامل شناختی، شخصیتی، اجرایی و غیره وجود دارند که حداقل تا اندازه‌ای عملکرد آزمودنی در آزمون را تحت تأثیر قرار می‌دهند؛

1. Parametric approach
2. Nonparametric approach
3. Monotone homogeneity model
4. Double homogeneity model
5. Unidimensional
6. Local independence
7. Monotonic
8. Sijistma
9. Molenaar
10. Dorans
11. Kingston

### ب) استقلال موضعی

مفروضه استقلال موضعی بدین معنی است که پاسخ فرد به سؤال  $i$  تحت تأثیر پاسخ‌های او به سؤال‌های دیگر تست نیست. این یک مفروضه قوی است؛ برای مثال، استقلال موضعی ممکن است به وسیله یادگیری که بر اثر تمرین به وجود آمده نقض شود. این امر در طول اجرای آزمون هنگامی که نمره‌گذاری روی صفت مکنون صورت می‌گیرد، رخ می‌دهد (سیجت‌سما و مولن‌آیر، ۲۰۰۲/ ترجمه ذوالفقارنسب، ۱۳۸۵). در صورتی که احتمال پاسخ آزمودنی  $A$  به سؤال معینی  $0/2$  باشد و احتمال پاسخگویی به همین سؤال برای آزمودنی  $B$  مساوی  $0/9$  باشد و چنانچه پاسخ‌های آزمودنی‌ها به صورت موضعی مستقل از یکدیگر باشند، در این صورت احتمال اینکه هر دوی آنها به آن سؤال پاسخ درست بدهند، مساوی است با  $0/18 = (0/2)(0/9)$  (آلن و یین، ۱۹۷۹/ ترجمه دل‌اور، ۱۳۷۴)؛

### ج) یکنواختی توابع سؤال پاسخ<sup>۱</sup>

مفروضه بعدی این است که احتمال شرطی  $P_i(\theta)$  به‌طور یکنواخت روی  $\theta$  بدون کاهش است. این مفروضه در معادله (۱) نشان داده شده است. به‌روشنی می‌توان دید که احتمال  $X_i = 0$  نیز به وسیله تابع سؤال پاسخ قابل توصیف است:

$$p(X_i = 0|\theta) = 1 - p(X_i = 1|\theta) \quad \text{معادله (۱)}$$

این مدل به صورت همگون ایجاد شده است و توابع سؤال پاسخ آنها به‌صورت یکنواخت با صفت مکنون ارتباط دارد. اهمیت کاربردی مدل همگنی یکنواخت این است که می‌توان پاسخ‌دهندگان را بر اساس نمره کل آنها در مقیاس  $\theta$  رتبه‌بندی کرد؛ بنابراین، مدلی است که در آن افراد براساس یک مقیاس ترتیبی مورد اندازه‌گیری قرار می‌گیرند. اگر  $c$  را مقداری ثابت در نظر بگیریم و این مقدار ثابت را با  $s$  و  $t$  جمع کنیم، به شرط برقراری رابطه  $0 \leq s \leq t \leq k$  خواهیم داشت:

$$P(\theta > c | X_+ = s) \leq P(\theta > c | X_+ = t) \quad \text{معادله (۲)}$$

مطابق این معادله، افراد را می‌توان به وسیله  $X_+$  به‌صورت احتمالی روی  $\theta$  رتبه‌بندی کرد. در سؤال‌های دو ارزشی که به‌صورت  $0$  و  $1$  نمره‌گذاری می‌شود، نمره شرطی مورد انتظار هر سؤال برابر است با مقادیر تابع سؤال پاسخ آن که می‌توان به‌صورت معادله زیر نوشت:

$$E(X_i | \theta) = 0 \times P(X_i = 0 | \theta) + 1 \times P(X_i = 1 | \theta) = P_i(\theta) \quad \text{معادله (۳)}$$

به عبارت دیگر، تغییرناپذیری ترتیب قرار گرفتن سؤال‌های بر این دلالت دارد که احتمال پاسخگویی به سؤال  $i$  کوچک‌تر یا برابر احتمال پاسخگویی به سؤال  $j$  است ( $P_i \leq P_j$ ) (سیجت‌سما و مولن‌آیر، ۲۰۰۲ / ترجمه ذوالفقارنسب، ۱۳۸۵).  
برای تحلیل داده‌ها در مدل همگنی مضاعف ناپارامتریک سؤال پاسخ، دو بررسی صورت می‌گیرد:

### الف) بررسی پذیرش یا محبوبیت سؤال $(P_i)$

در نظریه سؤال پاسخ، عموماً و در مدل ناپارامتریک آن خصوصاً، ویژگی‌های سؤال‌ها مستقل از ویژگی‌های آزمودنی‌های نمونه معین و در واقع، بر اساس جامعه آزمودنی‌ها برآورد می‌شود (سیجت‌سما و مولن‌آیر، ۲۰۰۲ / به نقل از مقدم‌زاده، ۱۳۸۶). این پارامتر نشان می‌دهد که منحنی ویژگی سؤال در کجای مقیاس توانایی قرار دارد و معادل دشواری سؤال در نظریه سؤال پاسخ در رابطه با سطح صفت زیربنایی  $\theta$  است (مولن‌آیر، ۱۹۹۷):

### ب) بررسی مقیاس‌پذیری $(H_i)$

این پارامتر که با نماد  $H_i$  نشان داده می‌شود، قدرت سؤال را در تمایزگذاری یا تشخیص آزمودنی‌ها در سطوح مختلف توانایی نشان می‌دهد و بیان می‌کند که هر سؤال آزمون تا چه حد قادر به تفکیک و تمیز بین آزمودنی‌ها در سطوح مختلف توانایی است. در همه مدل‌ها، ضریب  $H_i$  نقش اصلی را بازی می‌کند و برآوردی سریع و کاربردی در اندازه‌گیری‌های کیفی است. به نظر موکن<sup>۳</sup> (۱۹۹۷) در این رابطه می‌توان اشاره نمود که معتقد است، مقیاسی مفید است که ضریب مقیاس‌پذیری آن برای همه سؤال‌ها بزرگ‌تر از  $0/۳۰$  باشد ( $H_i > 0/۳۰$ )؛ در واقع، این شاخص نشان‌دهنده همسویی و هماهنگی سؤال با کل آزمون و یا بیانگر آن است که آیا سؤال همان ویژگی موردنظر آزمون را اندازه‌گیری می‌کند یا نه. این شاخص به نوعی با ضریب قدرت تشخیص سؤال در مدل پارامتریک نظریه سؤال پاسخ همخوانی دارد. یک شیوه برای تعریف ضریب  $H_i$ ، اصطلاح خطای احتمالاتی گاتمن<sup>۴</sup> است. فرض می‌کنیم که احتمال خطای گاتمن برای جفت سؤال‌های  $(i, k)$  بدین صورت تعریف شده است:

$$e_{ik} = \pi_{ik}(1, 0) \quad e_{ik}^{(0)} = \pi_i(1 - \pi_k) \text{ if } i < k$$

$$e_{ik}^{(0)} = (1 - \pi_i)\pi_k \text{ if } i > k$$

1. Acceptance
2. Scalibility
3. Mokken
4. Guttman probability error

که احتمال‌هایی از نمره مشاهده شده و خطای مورد انتظار تحت استقلال کناری را نشان می‌دهد.

$$H_{ik} = 1 - \frac{e_{ik}}{e_{ik}^{(o)}}$$

در مقیاس گاتمن، احتمال اینکه آزمودنی در یک سؤال بسیار آشنا (سؤال آسان)، نمره صفر و در یک سؤال ناآشنا (سؤال بسیار سخت) نمره ۱ بگیرد، خیلی کم است. برعکس شدن مطلب فوق را خطای گاتمن می‌نامند؛ برای مثال، اگر پاسخ‌دهنده‌ای در دو سؤال  $i$  و  $k$ ، به سؤال  $i$  پاسخ صحیح و به سؤال  $k$  که آسان‌تر از سؤال  $i$  است، پاسخ غلط بدهد، مرتکب خطای گاتمن شده است. اگر  $H_{ik} = 1$  باشد، مقیاس‌پذیری کامل است و صفر زمانی به دست می‌آید که استقلال آماری وجود دارد و  $i$  و  $k$  کاملاً مستقل از یکدیگر هستند. از تقسیم دو خطا بر یکدیگر، مقدار مقیاس‌پذیری که به دست می‌آید همواره از یک کم‌تر خواهد بود. ارزش‌های  $H_i$  می‌تواند برای ارزیابی برآزش سؤال  $i$  با توجه به سؤال‌های دیگر به صورت یک ماتریس همبستگی از مجموعه کل سؤال‌ها مورد استفاده قرار گیرد. در نهایت، یک ضریب می‌تواند از مجموعه کل سؤال‌ها به عنوان نسبت خطا و ترکیبی از  $H_i$  یا  $H_k$  تعریف شود. با فرض  $e = \sum_{k < i} k \cdot i e_{ik}$  و  $e^{(o)} = \sum_{k < i} k \cdot i e_{ik}^{(o)}$  ضریب برای مقیاس به صورت زیر بیان می‌شود:

$$H = 1 - \frac{e}{e^{(o)}} = \frac{\sum_i^n \sum_{k < i} k \pi_{ik}^{(o)} H_i}{\sum_i^n \sum_{k < i} k \pi_{ik}^{(o)}} = \frac{\sum_i^n \sum_{k < i} k \pi_{ik}^{(o)} H_{ik}}{\sum_i^n \sum_{k < i} k \pi_{ik}^{(o)}}$$

بنابراین، برای تک‌تک سؤال‌ها، یک مقدار ثابت  $c$  فرض می‌شود؛

$$(H_i; i=1, \dots, n) = c$$

پس  $H \geq c, 0 \leq c \leq 1$  و ضریب مقیاس‌پذیری تک‌تک سؤال‌ها تا اندازه‌ای شباهت به ضریب مقیاس‌پذیری کل آزمون دارد. ضریب مقیاس‌پذیری می‌تواند به عنوان نسبت واریانس با توجه به نمره به حالت نرمال درآید (موکن، ۱۹۹۷). مقیاس همگنی یکنواخت مانند مدل همگنی مضاعف به صورت مجموعه سؤال‌هایی تعریف می‌شود که هر کدام همبستگی مثبت با ویژگی هر سؤال (صفر و یک) دارند که بزرگ‌تر یا برابر مقدار ثابت  $C$  ( $C = 0.3$ ) است. ضریب مقیاس‌پذیری، مجموعه سؤال‌ها را نیز به صورت یک کل مورد آزمون قرار می‌دهد که باید بزرگ‌تر از  $C$  تک‌تک سؤال‌ها باشد. موکن در ادامه طبقه‌بندی زیر را برای پذیرش مقیاس‌ها پیشنهاد می‌کند:

۰/۵۰ < H مقیاس قوی

۰/۵۰ < H < ۰/۴۰ مقیاس متوسط

۰/۴۰ < H < ۰/۳۰ مقیاس ضعیف

## روش پژوهش جامعه آماری

به‌منظور پاسخ به پرسش‌های پژوهشی موردنظر، جامعه‌ای از کلیه داوطلبان ورود به دانشگاه‌های کشور در رشته ریاضی - فیزیک در کنکور سراسری سال ۱۳۸۴ انتخاب شد. جدول (۱)، تعداد داوطلبان شرکت‌کننده در کنکور سراسری ۱۳۸۴ را در رشته ریاضی فیزیک نشان می‌دهد.

جدول (۱) تعداد داوطلبان شرکت‌کننده در کنکور سراسری ۱۳۸۴ در رشته ریاضی فیزیک

رشته	زن	درصد	مرد	درصد	کل
ریاضی فیزیک	۱۳۸۸۶۸	۴۵/۹۱	۱۶۳۶۴۳	۵۴/۰۹	۳۰۲۵۱۱

نمونه‌گیری بر اساس دسترسی به فهرست داوطلبان و با توجه به برنامه کامپیوتری طراحی‌شده در سازمان سنجش آموزش کشور، با استفاده از روش نمونه‌گیری نظام‌انجام شد. از بین کلیه داوطلبان شرکت‌کننده در گروه آزمایشی ریاضی- فیزیک، یک گروه ۳۰۰۰ نفری به‌طور تصادفی انتخاب شد و پس از حذف آزمودنی‌هایی که به‌درستی به سؤال‌ها پاسخ نگفته بودند (افرادی با نمره پایین‌تر از  $\frac{1}{4}$  نمره کل یا حد شانس)، از بین افراد باقیمانده چند نمونه‌گیری صورت گرفت.

مطالعات اندکی وجود دارد که تأثیر حجم نمونه را بر برآوردهایی که برای پایایی پارامترها صورت می‌گیرد، به‌طور منظم بررسی کرده باشند. مطابق پژوهش‌های موجود در زمینه آزمون‌سازی، کاربرد موفق مدل‌های پارامتریک سؤال پاسخ و مدل‌های ناپارامتریک سؤال پاسخ، مستلزم استفاده از سؤال‌ها و آزمودنی‌هایی با حجم بزرگ است تا بتوان به‌طور همزمان صفت مکنون و پارامترهای سؤال را برآورد کرد (لرد، ۱۹۶۸؛ همبلتون و سوآمیناتان، ۱۹۸۵)؛ بنابراین، بر اساس تحقیقات انجام شده در هر دو نظریه و با توجه به نرم‌افزار کامپیوتری به‌کار رفته، از بین داوطلبان گروه آزمایشی ریاضی فیزیک، در نهایت دو نمونه بزرگ و کوچک (۸۰۰ نفری و ۸۰ نفری) انتخاب شد.

## ابزار، روش اجراء و تحلیل داده‌ها

در آزمون ورودی دانشگاه‌ها در ایران از آزمون‌های پیشرفت تحصیلی استفاده می‌شود که در سازمان سنجش به‌وسیله استادان با تجربه کشور طراحی و تهیه می‌شود. ابزار مورد استفاده برای جمع‌آوری داده‌های پژوهش، پاسخنامه‌های داوطلبان در آزمون اختصاصی درس ریاضی در رشته ریاضی فیزیک با ۵۵ سؤال بود. برای تجزیه و تحلیل داده‌ها از نرم‌افزارهای کامپیوتری SPSS و MSP استفاده شد. ابتدا تحلیل‌ها و مشخصه‌های کلاسیک سؤال‌ها و آزمون از طریق SPSS محاسبه و سپس داده‌ها با نرم‌افزار MSP تحلیل شد. به‌منظور تحلیل داده‌ها در مدل ناپارامتریک سؤال پاسخ، از روش‌های آماری مانند آزمون  $t$  وابسته، ضریب همبستگی پیرسون و آزمون‌های معنی‌داری آنها استفاده شد و سپس مقایسه‌های موردنظر انجام گرفت.

## نتایج پژوهش

در پژوهش حاضر، سؤال اول تحقیق، یعنی تأثیر خصوصیات آزمودنی‌ها بر ویژگی‌های سؤال، بر اساس مدل ناپارامتریک سؤال پاسخ این سؤال‌ها تحلیل شد. به دنبال آن، آزمون برازندگی برای هر سؤال و کل آزمون با استفاده از ضریب مقیاس‌پذیری محاسبه شد. سؤال‌هایی که ضریب مقیاس‌پذیری کمتر از  $0/3$  یا منفی داشتند، حذف شدند.

در جدول (۱) (ضمیمه یک)، نتایج تحلیل سؤال‌ها با نرم‌افزار ناپارامتریک سؤال پاسخ یعنی MSP، روی ۵۵ سؤال ریاضی آزمون سراسری سال ۱۳۸۴ نشان داده شده است. همان‌طور که در جدول (۱) ملاحظه می‌شود، تنها یک سؤال با توجه به ضریب منفی، کنار گذاشته شده است. لازم به ذکر است که در هر دو نمونه، همین ضرایب مقیاس‌پذیری و محبوبیت سؤال به دست آمده است.

جدول (۲) مقایسه ضرایب همبستگی شاخص یا محبوبیت سؤال (دشواری سؤال) در دو گروه نمونه به روش همگنی مضاعف در مدل ناپارامتریک

روش مقایسه	تعداد سؤال‌ها	ضریب همبستگی $\Gamma_{pbis}$	سطح معنی‌داری
NIRT	۵۵	$0/984$	$0/0001$

همچنین همان‌طور که جدول (۲) نشان می‌دهد، نتایج مقایسه دو نمونه ۸۰۰ نفری و ۸۰ نفری حاکی از آن است که در مدل ناپارامتریک سؤال پاسخ (NIRT)، همبستگی بسیار بالایی بین مقادیر برآورد شده این دو گروه نمونه وجود دارد. این نکته بیانگر آن است که برآورد این شاخص‌ها در مدل ناپارامتریک یک برآورد عالی است و مقدار همبستگی در حدود  $0/99$  می‌باشد که در سطح  $0/0001$  هم معنی‌دار است. لازم به ذکر است که در نرم‌افزار MSP که برای تحلیل داده‌ها با مدل‌های ناپارامتریک به کار می‌رود، محبوبیت سؤال ( $P_i$ ) معادل ضریب دشواری و مقیاس‌پذیری ( $H_i$ ) معادل ضریب تشخیص است. از آزمون  $t$  وابسته نیز برای نشان دادن تفاوت دو نمونه در برآورد شاخص محبوبیت سؤال استفاده شد. جدول (۳)، نتایج مقایسه این دو نمونه را نشان می‌دهد:

جدول (۳) مقایسه NIRT از نظر شاخص محبوبیت سؤال‌ها با استفاده از آزمون  $t$  وابسته در دو نمونه ۸۰۰ و ۸۰ نفری

روش مورد مقایسه	تعداد سؤال‌ها	میانگین سطح دشواری (محبوبیت سؤال)				خطای استاندارد میانگین سطح دشواری	آزمون $t$	درجه آزادی (d.f)	سطح معنی‌داری
		گروه A (۸۰۰ نفر)	گروه B (۸۰ نفر)	گروه A (۸۰۰ نفر)	گروه B (۸۰ نفر)				
NIRT	۵۵	$0/723$	$0/723$	$0/25$	$0/25$	$0/74$	۵۳	$0/425$	



همان طور که ملاحظه می شود، میانگین سطح دشواری در دو نمونه بزرگ و کوچک با یکدیگر متفاوت نبوده و تقریباً یکسان است. خطای استاندارد میانگین سطح دشواری در مدل ناپارامتریک سؤال پاسخ نیز یکسان است. آزمون  $t$  نشان می دهد که در مدل ناپارامتریک سؤال پاسخ، تفاوت بین دو گروه نمونه از نظر برآورد پارامتر دشواری معنی دار نیست و مقادیر برآورد شده پارامتر دشواری در دو گروه نمونه متفاوت، یکسان یا تقریباً یکسان است. این مسئله بیانگر عدم تأثیرپذیری برآورد محبویت سؤال در مدل ناپارامتریک نظریه سؤال پاسخ از ویژگی های آزمودنی ها می باشد.

جدول (۴) مقایسه میانگین ضرایب همبستگی شاخص مقیاس پذیری (قدرت تشخیص) سؤال ها در دو گروه نمونه با روش همگنی مضاعف در مدل ناپارامتریک

روش مقایسه	تعداد سؤال ها	ضریب همبستگی ( $r_{pbis}$ )	سطح معنی داری
NIRT	۵۵	۰/۹۹۵	۰/۰۰۰۱

در مدل ناپارامتریک سؤال پاسخ، مقیاس پذیری معادل شاخص یا پارامتر قدرت تشخیص یا شیب سؤال ( $a_i$ ) است. جدول (۴)، ضریب همبستگی برآورد شاخص قدرت تشخیص سؤال های آزمون ریاضی را در دو گروه نمونه در مدل ناپارامتریک سؤال پاسخ نشان می دهد. همان طور که ملاحظه می شود، ضریب همبستگی برابر ۰/۹۹۵ شده است که در سطح ۰/۰۰۰۱ معنی دار می باشد که نشان دهنده همبستگی بسیار بالا در مدل ناپارامتریک سؤال پاسخ در برآورد شاخص یا پارامتر قدرت تشخیص است.

جدول (۵) مقایسه شاخص مقیاس پذیری (پارامتر قدرت تشخیص) سؤال در دو گروه نمونه با روش همگنی مضاعف در مدل ناپارامتریک با استفاده از آزمون  $t$  وابسته

روش مورد مقایسه	تعداد سؤال ها	میانگین مقیاس پذیری		خطای استاندارد میانگین سطح دشواری		آزمون $t$	درجه آزادی (d.f)	سطح معناداری
		گروه A (نفر ۸۰۰)	گروه B (نفر ۸۰)	گروه A (نفر ۸۰۰)	گروه B (نفر ۸۰)			
NIRT	۵۵	۰/۸۳۲	۰/۸۳۱	۰/۰۴۱	۰/۰۴۲	-۰/۰۱	۵۴	۰/۸۹۶

همان طور که در جدول (۵) ملاحظه می شود، در مدل همگنی مضاعف ناپارامتریک نظریه سؤال پاسخ، میانگین قدرت تشخیص و انحراف استاندارد (SD) آنها در دو گروه آزمودنی بسیار نزدیک به یکدیگر و خطای استاندارد در مدل همگنی مضاعف ناپارامتریک تقریباً یکسان است. بر اساس نتایج آزمون  $t$  وابسته در مدل همگنی مضاعف ناپارامتریک سؤال پاسخ، تفاوت بین دو گروه در برآورد این شاخص معنی دار

نیست؛ به عبارت دیگر، در مدل ناپارامتریک نظریه سؤال پاسخ، ضریب مقیاس‌پذیری (پارامتر قدرت تشخیص) تحت تأثیر ویژگی‌های گروه آزمودنی‌ها و حجم نمونه قرار نمی‌گیرد و از یک گروه به گروه دیگر، مقادیر آن مشابه است و در نتیجه مقادیر برآوردشده در دو گروه تفاوت معنی‌داری با یکدیگر ندارند.

### نتیجه‌گیری

در بررسی سؤال اول پژوهش، تفاوت شاخص سطح دشواری یا محبوبیت سؤال در دو گروه نمونه ۸۰ و ۸۰۰ نفری در مدل ناپارامتریک سؤال پاسخ مورد مطالعه قرار گرفت و نتایج نشان داد که در مدل ناپارامتریک نظریه سؤال پاسخ، ضریب مقیاس‌پذیری (Hi)، نامتغیر و تقریباً ثابت است. لازم به ذکر است که مزیت عمده این مدل نسبت به مدل‌های پارامتریک این است که در نمونه‌های خیلی کمتر نیز شاخص‌ها تحت تأثیر ویژگی‌های آزمودنی‌ها یا سؤال‌های دیگر تغییر نمی‌کند. در این زمینه، موکن (۱۹۹۷) به نتایج مشابه‌ای دست یافت که مدل‌های ناپارامتریک در آزمون‌های استخدامی و نمونه‌هایی با حجم کمتر به کار می‌رود، ولی در مدل ناپارامتریک نظریه سؤال پاسخ، ضریب مقیاس‌پذیری بدون تأثیرپذیری از میزان توانایی افراد برآورد شده و ویژگی‌های سؤال‌های آزمون، تأثیری بر مقدار برآوردشده ندارد. رمسی<sup>۱</sup> (۱۹۹۷)، تحلیل‌هایی را با استفاده از مدل‌های همگنی یکنواخت و همگنی مضاعف انجام داد و بررسی پژوهشی وی در این رابطه نیز، مطالب فوق را تأیید کرد.

در بررسی سؤال دوم تحقیق، یعنی تفاوت دو گروه از نظر شاخص مقیاس‌پذیری، نتایج زیر به دست آمد: شاخص مقیاس‌پذیری نیز مانند شاخص دشواری سؤال، از نمونه‌های مختلف آزمودنی تأثیر نمی‌پذیرد. در این مدل، ضریب مقیاس‌پذیری، شاخصی است که میزان آگاهی را که یک سؤال درباره سطح توانایی مورد سنجش ارائه می‌دهد، نشان می‌دهد. ضریب مقیاس‌پذیری در مدل همگنی مضاعف ناپارامتریک سؤال پاسخ نامتغیر بوده و به گروه‌های مختلف آزمودنی‌ها قابل تعمیم است. عدم تغییر، یکی از مهم‌ترین ویژگی‌های این مدل است (اونا<sup>۲</sup>، ۲۰۰۴).

1. Ramsey
2. Onna

## منابع

- آلن، مری جی و یین، وندی ام (۱۹۷۹). *مقدمه‌ای بر نظریه‌های اندازه‌گیری (روان‌سنجی)*. ترجمه علی دلاور (۱۳۷۴)، تهران: انتشارات سمت.
- سیجت سیما، کلاس و مولن‌آیر، ایو دبلو (۲۰۰۲). *مقدمه‌ای بر تئوری ناپارامتریک سؤال پاسخ*. ترجمه سلیمان (سالار) ذالفقارنسب (۱۳۸۵). نشر رایانه.
- مطیعی لنگرودی، سیده طیبه (۱۳۸۱). *بررسی مزایا و محدودیت‌های دو روش اندازه‌گیری: کلاسیک و سؤال پاسخ (IRT) از لحاظ نظری و عملی*. پایان‌نامه کارشناسی ارشد، دانشگاه تهران: دانشکده روان‌شناسی و علوم تربیتی.
- مقدم‌زاده، علی (۱۳۸۶). *مقایسه مدل‌های پارامتریک با مدل‌های غیرپارامتریک نظریه سؤال پاسخ*. پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبایی: دانشکده روان‌شناسی و علوم تربیتی.
- Gulliksen, H. (1995). *Theory of mental tests*. Newyork: John Wiley & Sons.
- Hambleton, R.K. (1989). *Principles and selected applications of Item Response Theory*. In R. Linn (Ed), *Educational Measurement (3<sup>rd</sup> end)*. Newyork: Macmillan. 147-200.
- 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.
- Hambleton, R. K., Swaminathan, H. & Rojers, H. J. (1991). *Fundamentals of Item Response Theory*. Newburypark: Sage.
- Hambleton, R. K. & Vander Linden, W. J. (1982). Advance in Item Response Theory and applications: An introduction. *Applied Psychplogical Measurement*. 6(4), 373-387.
- 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.
- Mokken, R.J. (1997). Nonparametric models for dichotomous responses. In: Hambleton, R.K. and Van der Linden, W.J. (Ed's). *Handbook of Modern Item Response Theory*. New York-Berlin: Springer -Verlag, pp.351-367.

- 
- Molenaar, I. W. (1997). Nonparametric models for dichotomous responses. In: Hambleton, R.K. and Van der Linden, W.J. (Ed's). *Handbook of modern Item Response Theory*. New York-Berlin: Springer-Verlag, pp. 369-379.
- Onna, M. J. H. (2004). Ordered latent class models in nonparametric Item Response Theory. Lawrence Erlbaum.
- Ramsey, J. (1997). *Nonparametric models for dichotomous responses*. In: Hambleton, R.K. and Van der Linden, W.J. (Ed's). *Handbook of Modern Item Response Theory*. New York-Berlin: Springer -Verlag, pp. 369-379.

## ضمیمه جدول (۱) نتایج تحلیل داده‌ها با نرم‌افزار ناپارامتریک (MSP) سؤال پاسخ

مقیاس پذیری سؤال ( $H_i$ ) در مرحله انتخاب سؤال						شماره سؤال
گام ۵	گام ۴	گام ۳	گام ۲	گام ۱	ویژگی مقبولیت سؤال ( $p_i$ )	
۰/۹۱	۰/۶۱	۰/۸۵	۰/۸۹	۰/۷۶	۰/۹۷	۱
۰/۹۲	۰/۶۱	۰/۷۸	۰/۸۱	۰/۹۲	۰/۸۵	۲
۰/۶۶	۰/۳۹	۰/۵۴	۰/۸۷	۰/۸۲	۰/۴۵	۳
۰/۸۱	۰/۴۵	۰/۵۶	۰/۸۵	۰/۷۳	۰/۷۴	۴
۰/۸۴	۰/۴۷	۰/۵۲	۰/۸۸	۰/۵۲	۰/۸۹	۵
۰/۸۲	۰/۴۸	۰/۳۶	۰/۸۷	۰/۹۱	۰/۸۱	۶
۰/۷۴	۰/۸۵	۰/۳۵	۰/۷۸	۰/۹۲	۰/۸۷	۷
۰/۸۸	۰/۸۸	۰/۸۸	۰/۸۶	۰/۸۱	۰/۸۵	۸
۰/۶۹	۰/۸۷	۰/۸۷	۰/۸۷	۰/۸۶	۰/۶۵	۹
۰/۶۱	۰/۷۸	۰/۷۸	۰/۷۴		۰/۵۴	۱۰
۰/۶۱	۰/۸۱	۰/۸۲	۰/۸۹		۰/۹۸	۱۱
۰/۳۹	۰/۸۲	۰/۵۲	۰/۸۱		۰/۸۸	۱۲
۰/۴۵	۰/۸۴	۰/۸۱	۰/۸۷		۰/۸۷	۱۳
۰/۴۷	۰/۸۴	۰/۸۲	۰/۸۵		۰/۷۸	۱۴
۰/۴۸	۰/۸۲	۰/۸۴	۰/۷۲		۰/۸۶	۱۵
۰/۴۹	۰/۵۲	۰/۸۴	۰/۷۶		۰/۸۱	۱۶
۰/۹۵	۰/۹۱	۰/۸۶			۰/۸۲	۱۷
۰/۹۳	۰/۹۲	۰/۸۱			۰/۸۴	۱۸
۰/۸۷	۰/۸۲	۰/۸۴			۰/۸۴	۱۹
۰/۸۶	۰/۵۲	۰/۸۲			۰/۸۲	۲۰
۰/۹۱	۰/۸۱	۰/۵۸			۰/۵۲	۲۱

ادامه ضمیمه جدول (۱) نتایج تحلیل داده‌ها با نرم‌افزار ناپارامتریک (MSP) سؤال پاسخ

مقیاس‌پذیری سؤال ( $H_j$ ) در مرحله انتخاب سؤال						شماره سؤال
گام ۵	گام ۴	گام ۳	گام ۲	گام ۱	ویژگی مقبولیت سؤال ( $p_j$ )	
۰/۹۲	۰/۸۲	۰/۵۱			۰/۹۱	۲۲
۰/۶۶	۰/۸۴	۰/۸۷			۰/۹۲	۲۳
۰/۳۷	۰/۸۴	۰/۸۵			۰/۶۶	۲۴
۰/۵۸	۰/۸۲	۰/۶۵			۰/۷۵	۲۵
۰/۸۵	۰/۸۲	۰/۷۶			۰/۵۸	۲۶
۰/۷۸	۰/۷۴	۰/۷۶			۰/۸۵	۲۷
۰/۵۴	۰/۸۸	۰/۷۳			۰/۸۵	۲۸
۰/۵۶	۰/۸۷	۰/۸۷			۰/۵۴	۲۹
۰/۵۲	۰/۷۸				۰/۶۵	۳۰
۰/۳۶	۰/۸۶				۰/۸۱	۳۱
۰/۳۵	۰/۷۲				۰/۸۴	۳۲
۰/۳۷	۰/۸۱				۰/۸۲	۳۳
۰/۴۹	۰/۸۲				۰/۷۴	۳۴
۰/۵۸	۰/۸۴				۰/۷۳	۳۵
۰/۵۷	۰/۸۴				۰/۵۷	۳۶
۰/۵۹	۰/۸۲				۰/۳۵	۳۷
۰/۵۸	۰/۵۲				۰/۶۵	۳۸
۰/۵۹	۰/۹۱				۰/۵۲	۳۹
۰/۳۹	۰/۵۷				۰/۵۴	۴۰
۰/۵۹	۰/۳۵				۰/۷۵	۴۱
۰/۶۱					۰/۷۶	۴۲

ادامه ضمیمه جدول (۱) نتایج تحلیل داده‌ها با نرم‌افزار ناپارامتریک (MSP) سؤال پاسخ

مقیاس‌پذیری سؤال ( $H_j$ ) در مرحله انتخاب سؤال						شماره سؤال
گام ۵	گام ۴	گام ۳	گام ۲	گام ۱	ویژگی مقبولیت سؤال ( $p_j$ )	
۰/۵۷					۰/۶۳	۴۳
۰/۶۱					۰/۶۱	۴۴
۰/۶۱					۰/۷۲	۴۵
۰/۶۴					۰/۵۳	۴۶
۰/۶۱					۰/۵۷	۴۷
۰/۵۳					۰/۵۸	۴۸
۰/۵۷					۰/۵۱	۴۹
۰/۵۸					۰/۶۱	۵۰
۰/۵۱					۰/۶۷	۵۱
۰/۵۷					۰/۷۳	۵۲
۰/۳۵					۰/۶۹	۵۳
۰/۶۵					۰/۶۷	۵۴
۰/۵۶					۰/۷۲	کل