

## اعتباریابی و رواسازی آزمون چندگزینه‌ای - چند پاسخ پیشرفت تحصیلی در بین دانشجویان زن دانشکده علوم تربیتی دانشگاه آزاد اسلامی تبریز

شیرین داوری فیض پور آذر<sup>1</sup>

دکتر علی نقی اقدسی<sup>2\*</sup>

دکتر غلامرضا گل محمد نژاد<sup>3</sup>

### چکیده

هدف از تحقیق حاضر اعتباریابی و رواسازی آزمون چندگزینه‌ای - چند پاسخ پیشرفت تحصیلی با استفاده از روش روانسنجی بود. جامعه آماری شامل کلیه دروس ارائه شده در نیمسال دوم سال تحصیلی 95-94 در مقطع کارشناسی ارشد رشته علوم تربیتی دانشگاه آزاد اسلامی تبریز، به تعداد 65 کلاس درس و نمونه 4 درس می‌باشد که به روش تصادفی خوشه‌ای انتخاب شده‌اند. ابزار اندازه‌گیری، شامل آزمون‌های پیشرفت تحصیلی محقق ساخته برای دروس انتخابی بود که روایی محتوایی آنها به تأیید 5 نفر از اساتید گروه علوم تربیتی رسید و پایایی آنها از طریق آلفای کرونباخ محاسبه گردیده است. روایی آزمون‌های چندگزینه‌ای - چند پاسخ نیز از طریق بررسی همبستگی آن با سایر انواع آزمون‌ها و با استفاده از آزمون همبستگی پیرسون برآورد گردیده است. نتایج نشان داد پایایی آزمون‌های چندگزینه‌ای - چند پاسخ طراحی شده مناسب بوده و همچنین همبستگی بین نمرات آزمون چندگزینه‌ای - چند پاسخ با نمرات آزمون‌های تشریحی گسترده پاسخ، صحیح-غلط، کوتاه پاسخ و جور کردنی از نظر آماری معنی‌دار می‌باشد و ضرایب همبستگی به ترتیب 0/865، 0/628 و 0/771 می‌باشد، لذا می‌توان ادعا نمود آزمون‌های چندگزینه‌ای - چند پاسخ طراحی شده دارای اعتبار و روایی می‌باشند.

**واژه‌های کلیدی:** آزمون چندگزینه‌ای - چند پاسخ، آزمون جور کردنی، آزمون کوتاه پاسخ، آزمون صحیح-غلط، آزمون تشریحی گسترده پاسخ

1. دانش آموخته کارشناسی ارشد دانشگاه آزاد اسلامی واحد تبریز، تبریز، ایران. [shirindavary@yahoo.com](mailto:shirindavary@yahoo.com)

2. استادیار گروه مشاوره، واحد تبریز، دانشگاه آزاد اسلامی، تبریز، ایران (نویسنده مسؤول). [Alinaghi.aghdasai@yahoo.com](mailto:Alinaghi.aghdasai@yahoo.com)

3. استادیار گروه روانشناسی، واحد تبریز، دانشگاه آزاد اسلامی، تبریز، ایران. [dr.golmohammad@yahoo.com](mailto:dr.golmohammad@yahoo.com)

## مقدمه

فرایند آموزش و میزان دستیابی فراگیران به اهداف آموزشی و موفقیت آنان همواره از دغدغه‌های اصلی هر نظام آموزشی می‌باشد. هدف از ساخت آزمون‌ها نیز تدوین ابزاری برای اندازه‌گیری میزان تحقق این اهداف است. از میان انواع آزمون‌ها، آزمون‌های پیشرفت تحصیلی دارای اهمیت بیشتری می‌باشند، چراکه اطلاعات حاصل از این آزمون‌ها پایه و اساس قضاوت درباره دانش و توانایی‌های فراگیران و تصمیم‌گیری درباره مطلوبیت نتایج فعالیت‌های آموزشی معلم و کوشش‌های دانش‌آموزان است (سیف، 1392) و از سویی ممکن است بر موقعیت اجتماعی، عزت‌نفس، انگیزه‌های تحصیلی، انتخاب‌ها و نهایتاً بر آینده و سرنوشت فراگیران تأثیرگذار باشند. آزمون‌های بسیار متنوعی برای اندازه‌گیری میزان پیشرفت تحصیلی فراگیران وجود دارند که هر یک محاسن و معایب خود را دارا می‌باشند و بسته به اهداف آموزشی، موضوع، ساختار محتوا، امکانات و بسیاری عوامل دیگر می‌توان هر یک از آن‌ها را به کار برد. اله دادی و یعقوبی (1390) انواع آزمون‌های مداد و کاغذی را در سه دسته تشریحی (گسترده پاسخ و محدود پاسخ)، کوتاه پاسخ (پرسشی، کامل کردنی و تشخیصی) و عینی (صحیح-غلط، جور کردنی و چندگزینه‌ای) دسته‌بندی نموده‌اند.

برخی صاحب‌نظران چون بوکارد<sup>1</sup> معتقدند در جهان واقعی کسی فاقد دانش نیست و به همین علت استفاده از سؤالات انشایی و باز پاسخ را امری فاجعه‌بار و بسیار پرهزینه می‌داند (به نقل از کریمی، فلسفی نژاد و درتاج 1390)، از نظر دلاور و زهراکار (1387) نیز جهت اندازه‌گیری‌های خوب و ثمربخش، عینیت در نمره‌گذاری ضرورت دارد. از سویی برخی محققان چون ویلکوکس<sup>2</sup> و پالاک<sup>3</sup> (2014) به مزایای آزمون‌های چندگزینه‌ای از جمله، سهولت و عینیت نمره‌گذاری، آن‌ها را برای مقیاس‌های گسترده ایده آل شاره نموده و بر این باورند که این مزیت آشکار باید موجب ایجاد انگیزه برای بهبود استراتژی‌های ساخت این آزمون‌ها گردد، لذا می‌توان لزوم تلاش برای بهبود شیوه‌های ارزشیابی عینی را توجیه نمود. بدین منظور جهت ارزشیابی پیشرفت تحصیلی دانشجویان نوع خاصی از آزمون‌های چندگزینه‌ای استفاده گردیده است که با عنوان آزمون چندگزینه‌ای-چند پاسخ نامبرده خواهد شد. شکل خاصی از این آزمون، برای

1. Bukard
2. Wilcox
3. Pollock

نخستین بار توسط کروناخ در سال 1979 با نام چند پاسخ معرفی گردیده است (به نقل از تاراسوا<sup>1</sup> و آیور<sup>2</sup>، 2013). وربیک<sup>3</sup> (2012) نیز با اشاره به این سوالات آنها را به عنوان سوالاتی با پتانسیل بالا معرفی نموده که دهه‌هاست این سوالات مورد غفلت واقع شده است. مطالعه این سوالات زمینه معرفی سوالاتی بود که با تأکید بر نقاط قوت آزمون‌های چندگزینه‌ای مرسوم، از نقاط ضعف این آزمون‌ها از جمله مشکل حدس زدن پاسخ‌ها توسط آزمون‌دهندگان و مشکل طراحی سوالات با استفاده از گزینه‌های انحرافی معیوب بکاهد. نحوه طراحی سوالات آزمون‌های چندگزینه‌ای - چند پاسخ به نحوی می‌باشد که هر سؤال آن می‌تواند تعداد گزینه‌های مختلفی داشته باشد، در صورتی که یافتن گزینه مناسب دشوار باشد لزومی به استفاده از گزینه‌های معیوب که روایی و پایایی آزمون را تهدید می‌کنند نیست و می‌توان به حداقل گزینه‌ها اکتفا کرد و در صورتی که امکان استفاده از گزینه‌های بیشتری وجود داشته باشد، می‌توان جهت کاهش احتمال درستی پاسخ حدسی از آنها استفاده نمود. همچنین هر سؤال لزوماً یک گزینه صحیح نیست و این تعداد می‌تواند متغیر باشد که خود عاملی برای کاهش احتمال درستی یک پاسخ انتخاب شده بر اساس حدس می‌باشد، چرا که یک آزمون چهار گزینه‌ای با تنها یک گزینه صحیح را به چهار شکل می‌توان پاسخ داد و لذا احتمال درستی پاسخ حدسی 0/25 می‌باشد، در حالی که یک آزمون چندگزینه‌ای - چند پاسخ با چهار گزینه را می‌توان به 16 طریق مختلف پاسخ داد که احتمال درستی هر پاسخ حدسی در این حالت به 0/06 کاهش می‌یابد.

سوالات چندگزینه‌ای را اولین بار در سال 1914 فردریک در دانشگاه کانزاس ابداع نموده است (به نقل از ملاحادی، 1388). هر سؤال از سه قسمت اصلی تشکیل می‌شود: (1) ساقه سؤال، (2) گزینه درست یا گزینه کلید، (3) گزینه‌های انحرافی. از انواع مختلف آزمون‌های چندگزینه‌ای می‌توان به نوع پرسشی، نوع ناتمام، نوع تنها گزینه درست، نوع بهترین گزینه درست، نوع منفی، نوع جایگزینی، نوع ناتمام پیشنهادی و نوع پاسخ ترکیبی اشاره نمود (سیف، 1393). ان جی<sup>4</sup> و چان<sup>5</sup> (2009) نیز انواع مختلف آزمون‌های چندگزینه‌ای اشاره نموده‌اند: چندگزینه‌ای مرسوم، چندگزینه‌ای لیبرال، چندگزینه‌ای

1. Tarasowa
2. Auer
3. Verbic
4. Ng
5. Chan

حذفی، ارزیابی اطمینان، تست احتمال و اولویت‌بندی که تمامی آن‌ها برای مقابله با مشکل حدس زدن معرفی گردیده‌اند.

از نظر ثرندایک (1386) در سؤالات چندگزینه‌ای مهم است که گزینه‌های گمراه‌ساز خوش‌ظاهر و موجه باشند و بتوانند دانش‌آموزان ناآگاه را جلب کنند و به نظر پاشا شریفی (1390)، گزینه‌های انحرافی باید کارایی لازم را داشته باشند یعنی توجه افراد بی اطلاع از موضوع مورد اندازه‌گیری را به اندازه گزینه درست به خود جلب کنند و برای آزمودنی‌هایی که در مورد موضوع مورد اندازه‌گیری توانایی و مهارت کافی دارند، گول زنده نباشند، چرا که کاپلان و ساکوزو (1386) معتقدند، گزینه‌های انحرافی نامناسب می‌توانند به‌طور نامطلوبی کیفیت آزمون را تحت تأثیر قرار دهند. پاشا شریفی (1390) همچنین در زمینه تعداد گزینه‌های سؤالات چندگزینه‌ای اشاره می‌کند که در آزمون‌هایی که با سؤال‌های 5 یا 6 گزینه‌ای ساخته می‌شوند، احتمال یافتن تصادفی پاسخ‌های درست کمتر از سؤال‌های 3 یا 4 گزینه‌ای است، اما یک پرسش سه گزینه‌ای که همه گزینه‌های آن با رعایت اصول تهیه شوند به یک پرسش چهار یا پنج گزینه‌ای که بعضی از گزینه‌های آن نادرست باشد، ارجح‌تر می‌باشد. لرد (1391) نیز عنوان می‌کند که چنین به نظر می‌رسد که هر چه تعداد گزینه‌ها بیشتر باشد، بهتر است، اما باید در نظر داشت که افزایش تعداد گزینه‌ها منجر به افزایش زمان سنجش می‌گردد. از نظر کاپلان و ساکوزو (1386) نظریه‌های روان‌سنجی پیشنهاد می‌دهند که اضافه کردن گزینه‌های انحرافی بیشتر باعث افزایش پایایی سؤالات می‌شوند، اما در عمل اضافه کردن گزینه‌های انحرافی ممکن است واقعاً باعث افزایش پایایی آزمون نشود، زیرا پیدا کردن گزینه‌های خوب مشکل است.

یکی از نگرانی‌هایی که همواره در مورد سؤالات چندگزینه‌ای وجود داشته است، پاسخ حدسی آزمون‌شوندگان به این سؤالات است. به‌طور کلی دو الگو در حدس زدن مشهود است. الگوی اول که حدس ناآگاهانه یا تصادفی نامیده می‌شود و الگوی دوم که حدس آگاهانه یا آموخته‌شده نام دارد، در هر حالت حدس زدن پایایی آزمون را کاهش می‌دهد. برخی از راه‌کارهای ارائه شده جهت مقابله با حدس زدن عبارت‌اند از: افزایش تعداد گزینه‌های هر سؤال، استفاده از گزینه نمی‌دانم، نمره‌دهی منفی، مدل ارزیابی اطمینان، حذف گزینه و انتخاب لیبرال و معمول‌ترین راه‌حل، نمره دهی منفی می‌باشد،

ایراد نمره دهی منفی این است که در این حالت نمرات دانشجویان علاوه بر دانش آن‌ها متأثر از شخصیت آن‌ها نیز می‌باشد (مرتاض هجری، خباز و جلیلی، 1393).

دامنیچ و همکاران (2015) از مزایای آزمون‌های چندگزینه‌ای به عینیت و سرعت در سنجش موضوعات متعدد و امکان کامپیوتری بودن نمره‌دهی و از نقاط ضعف این آزمون‌ها به کاهش روایی و پایایی بر اثر حدس زدن و عدم توانایی این سؤالات در تمایز بین دانش نسبی و نبود دانش اشاره نموده‌اند و یکی از روش‌های رفع این نواقص را استفاده از سؤالات چند پاسخ به‌جای تک پاسخ ذکر کرده‌اند که انواع آن‌ها با نام‌ها و روش‌های نمره‌گذاری مختلف مورد استفاده قرار گرفته‌اند از جمله: چند صحیح-غلط<sup>1</sup>، چند علامتی<sup>2</sup>، چند پاسخ<sup>3</sup> و چند چندگزینه‌ای<sup>4</sup>.

از نظر مورگان<sup>5</sup> (1979)، در حالت عمومی یک آزمون چندگزینه‌ای 4 یا 5 گزینه دارد که یکی از آن‌ها صحیح می‌باشد و اغلب دانشجویان در دوران تحصیل با آن آشنا هستند، اما نوع دیگری از آزمون‌های چندگزینه‌ای وجود دارد که هر سؤال آن ممکن است چند پاسخ درست داشته باشد. اسکالیس<sup>6</sup> و گیفورد<sup>7</sup> (2006) نیز به 28 دسته از نوآوری‌ها در زمینه طراحی های آزمون‌های مبتنی بر کامپیوتر، از جمله سؤالات چند پاسخ اشاره نموده‌اند. این محققان اشاره می‌کنند که هالادینا<sup>8</sup> معتقد بود که اگرچه آزمون‌های چند پاسخ دهه‌هاست که فراموش شده‌اند، اما پیش‌بینی می‌کرد که به‌زودی علاقه به این تست‌ها باز خواهد گشت. وربیک (2012) نیز سؤالات چند پاسخ را نوعی سؤال پاسخ‌گزين شبیه به سؤالات چندگزینه‌ای می‌داند با این تفاوت که هر سؤال ممکن است بیش از یک گزینه صحیح داشته باشند و این آزمون‌ها را دارای پتانسیل بالا جهت اندازه‌گیری‌های آموزشی می‌داند. باینکه مدت‌هاست از زمان معرفی این آزمون‌ها می‌گذارد اما مطالعات بسیار کمی در این زمینه صورت گرفته است که با جستجو در پایگاه‌های مختلف<sup>9</sup> موارد یافت شده و مرتبط مورد مطالعه قرار گرفته است.

1. multiple True-False
2. multiple-mark
3. multiple-answer
4. Multiple-multiple-choice
5. Morgan
6. Scalise
7. Gifford
8. Haladyna
9. ERIC, Science Direct, Springer, IEEE...

یکی از مسائل اصلی در مورد سؤالات چند پاسخ، انتخاب قوانین درست نمره‌دهی می‌باشد.

دامنیچ و همکاران (2015)، وربیک (2012) و تاراسوا و آیور (2013) به چند الگوریتم نمره‌دهی برای سؤالات چند پاسخ اشاره نموده اند، از جمله الگوریتم مورگان که مورد استفاده در تحقیق فوق است.

دامنیچ و همکاران (2015)، در تحقیق خود به تأثیر الگوریتم‌های مختلف نمره‌دهی به سؤالات چندگزینه‌ای- چند پاسخ، پرداخته‌اند. جامعه آماری در این تحقیق دانش‌آموزان 7 مدرسه متوسطه در منطقه شهری جنوا در سال تحصیلی 2012-2013 بودند و نمونه‌گیری به‌صورت داوطلبانه انجام شده بود. ابزار تحقیق شامل 14 سؤال بود که بعد از تجزیه و تحلیل داده‌ها دو سؤال به دلیل نقص حذف و در مجموع 12 سؤال در تجزیه و تحلیل نهایی استفاده شد که دو سؤال آن باز پاسخ و 10 سؤال آن چندگزینه‌ای- چند پاسخ بود. تعداد گزینه‌ها در سؤالات چندگزینه‌ای- چند پاسخ از 4 تا 8 متغیر و تعداد گزینه‌های درست بین 2 تا 5 و تعداد گزینه‌های انحرافی بین 1 تا 5 بود. پاسخ‌ها با 8 الگوریتم مختلف نمره‌دهی، نمره‌گذاری و سپس تحلیل شدند. در تمام 286 پرسش‌نامه که تحلیل گردیدند، الگوریتم‌های نمره‌دهی نسبی ویژگی‌های روان‌سنجی بهتری نشان دادند. همچنین نتایج نشان داد که الگوریتم‌های نمره‌دهی مختلف سبب تفاوت در پایایی، ضریب دشواری، قوه تمیز و حساسیت سؤالات در برابر حدس می‌گردند و میانگین نمرات در هر هفت الگوریتم نسبی بیشتر از الگوریتم دو ارزشی بوده و روش نمره‌دهی دو ارزشی نسبت به بقیه روش‌ها ضریب آلفای کرونباخ کمتری داشته است. وربیک (2012) نیز در تحقیقی با عنوان ارزش اطلاعاتی سؤالات چند پاسخ، 32 سؤال که 29 سؤال آن چندگزینه‌ای و 3 سؤال آن چند پاسخ با 5 گزینه بود، طراحی نمود و بر روی 926 دانش‌آموز از 50 مدرسه ابتدایی که به‌طور داوطلبانه شرکت کرده بودند، اجرا نمود. از شش شیوه نمره‌دهی برای نمره دادن به این سه سؤال استفاده کرد. نتایج نشان داد تأثیر روش نمره‌دهی بر قوه تمیز و پایایی هر سه سؤال معنی‌دار است و پایایی روش‌های نمره‌دهی نسبی بالاتر از روش‌های دو ارزشی می‌باشد.

ویلکوکس و پولاک (2014) در تحقیقی به مقایسه آزمون چندپاسخ ترکیبی<sup>1</sup> و آزمون باز پاسخ در درس فیزیک الکترواستاتیک پرداخته‌اند. آزمون بر روی دانشجویان دو کلاس الکترواستاتیک از دانشگاه کلرادو اجرا شد. در هر یک از این دو کلاس دانشجویان بر اساس نمرات میان ترمشان به دو گروه همتراز تقسیم شدند و یک بخش از دانشجویان به سؤالات باز پاسخ و بخش دیگر به سؤالات چند پاسخ ترکیبی پاسخ دادند. نتایج نشان داد میانگین نمرات، ضریب دشواری، روایی و پایایی سؤالات در دو نوع آزمون تفاوت معنی‌داری با هم ندارند. همچنین توزیع نمرات در هر دو حالت نرمال بود و واریانس مشابه داشت. برای بررسی همبستگی سؤالات از ضریب همبستگی پیرسون استفاده گردیده که این میزان بین 0/24 تا 0/71 برای دوبه‌دوی سؤالات هر نسخه بود.

علی<sup>2</sup>، کار<sup>3</sup> و روت<sup>4</sup> (2016)، در تحقیقی با عنوان روایی و پایایی نمرات به دست‌آمده از سؤالات چندگزینه‌ای، دو نسخه از سؤالات باز پاسخ و چندگزینه‌ای را با سؤالات یکسان به‌صورت تصادفی به شش گروه از دانشجویان سال اول پزشکی دانشگاه نورث داکوتا ارائه دادند. تحلیل داده‌ها نشان داد در سؤالات چندگزینه‌ای شاخص دشواری کمتر و ضریب آلفای کرونباخ بیشتر از سؤالات باز پاسخ است. همچنین نتایج نشان داد که عملکرد دانشجویان در سؤالات چندگزینه‌ای بهتر از سؤالات باز پاسخ می‌باشد و دلیل آن به ماهیت سؤالات برمی‌گردد، چرا که در آزمون‌های چندگزینه‌ای جواب درست انتخاب ولی در آزمون‌های باز پاسخ از حافظه فراخوانی می‌شود به همین علت ضریب دشواری سؤالات در آزمون‌های چندگزینه‌ای پایین‌تر از حد انتظار بود.

در تحقیقی تجربی که توسط زونتی<sup>5</sup>، بیوس<sup>6</sup>، گریفین<sup>7</sup> و والاس<sup>8</sup> (2013) با عنوان ثابت یا متغیر: مقایسه سؤالات چندگزینه‌ای با 3 گزینه، 4 گزینه‌ها و گزینه‌های مختلف انجام گرفته است، 60 سؤال چندگزینه‌ای در کالج زنان و مامایی رویال استرالیا و نیوزلند

1. این سؤالات به شکل چند پاسخ دومرحله‌ای می‌باشند، هر سؤال دو بخش دارد که آزمون‌دهندگان بعد از انتخاب پاسخ‌های بخش اول با خواندن بخش دوم سؤال به انتخاب پاسخ‌های بخش دوم می‌پردازند.

2. Ali
3. Carr
4. Ruit
5. Zoanetti
6. Beaves
7. Griffin
8. Wallace

به سه شکل اجرا گردید. در سری اول، سؤالات 4 گزینه‌ای اجرا گردید، در سری دوم، یکی از ناکارترین گزینه‌ها حذف و سؤالات سه گزینه‌ای اجرا و در سری سوم بهترین تعداد گزینه برای هر سؤال انتخاب گردید، یعنی برخی سؤالات سه گزینه و برخی از سؤالات 4 گزینه داشتند. نتایج تحلیل سؤالات نشان داد که در کل، آزمون سؤالات چهارگزینه‌ای از نظر اعتبار و خطای اندازه‌گیری بهتر از سؤالات سه گزینه‌ای می‌باشند، اما تحلیل تک‌تک سؤالات نشان داد با حذف برخی گزینه‌ها، ویژگی‌های روان‌سنجی برخی سؤالات بهتر شده است و همچنین سؤالات با تعداد گزینه‌های مختلف، بهترین ویژگی‌های روان‌سنجی از لحاظ پایایی و قوه تمیز را نشان دادند.

تتسوهیتو<sup>1</sup>، اوسامو<sup>2</sup>، توموکو<sup>3</sup> و کی‌یومی<sup>4</sup> (2006) در تحقیقی، آثار کاهش تعداد گزینه‌ها بر روی ویژگی‌های روان‌سنجی آزمون ورودی دانشگاه کانسای را مورد بررسی قرار داده‌اند. این آزمون که به‌صورت یک سری سؤالات چهارگزینه‌ای بود، با حذف یک گزینه به یک آزمون سه گزینه‌ای تبدیل شد و بر روی گروه دیگری اجرا گردید. مقایسه نتایج دو آزمون نشان داد که شاخص‌های مرکزی در دو گروه تفاوت معناداری ندارند. نتایج این تحقیق پیشنهاد می‌کند، از آنجا که سؤالات 3 گزینه‌ای همان کارایی سؤالات 4 گزینه‌ای را دارند، می‌توان به‌جای 4 گزینه از 3 گزینه استفاده کرد.

در تحقیقی که توسط رساییان، نخعی و صادقی قندهاری (1381) جهت مقایسه روش‌های آزمون‌های چندگزینه‌ای، صحیح-غلط و کوتاه پاسخ انجام گرفته است. نتایج نشان داده است که معتبرترین روش مورد بررسی، سؤالات کوتاه پاسخ است. جاوید (2014) نیز در تحقیقی به مقایسه سؤالات چندگزینه‌ای و صحیح-غلط پرداخته است. نتایج نشان داد، تفاوت عملکرد دانش‌آموزان در دو نسخه از لحاظ آماری معنی‌دار است. همچنین پایایی سؤالات صحیح-غلط بیشتر از سؤالات چند گزینه به‌دست آمده است.

### روش‌شناسی تحقیق

روش تحقیق حاضر روان‌سنجی بوده و جامعه آماری شامل کلیه دروس ارائه شده در دوره کارشناسی ارشد رشته علوم تربیتی دانشگاه آزاد اسلامی تبریز در نیمسال دوم سال تحصیلی 94-95 به تعداد 65 درس می‌باشد. نمونه انتخابی 4 درس از بین 65 درس

1. Tetsuhito
2. Osamu
3. Tomoko
4. Kiyomi



ارائه شده و در مجموع شامل 12 کلاس می‌باشد که به روش تصادفی خوشه‌ای از بین گرایش‌های مختلف علوم تربیتی، انتخاب گردیده‌اند و به چهار گروه با سه کلاس جهت اجرای هم زمان آزمون چندگزینه‌ای-چندپاسخ با یکی از انواع آزمون‌های مذکور دسته‌بندی شدند، در تحقیق حاضر داده‌های مربوط به دانشجویمان زن ارائه گردیده است. ابزارهای اندازه‌گیری شامل آزمون‌های پیشرفت تحصیلی محقق ساخته بر اساس منابع درسی، مطابق با سرفصل‌های وزارت علوم: آزمون چندگزینه‌ای-چندپاسخ به همراه آزمون‌های تشریحی گسترده پاسخ، صحیح-غلط، کوتاه پاسخ و جور کردنی. روایی محتوایی تمامی آزمون‌های محقق ساخته به تأیید 5 نفر از اساتید گروه علوم تربیتی رسیده و پایایی تمامی آزمون‌ها به جز آزمون تشریحی گسترده پاسخ، به روش آلفای کرونباخ محاسبه گردیده است. مقدار آلفا برای تمامی آزمون‌ها در جدول 1 آورده شده است.

جدول شماره 1: پایایی آزمون‌ها

مقدار آلفای کرونباخ	نوع آزمون
0/808	آزمون چندگزینه‌ای-چند پاسخ گروه 1
0/815	آزمون چندگزینه‌ای-چند پاسخ گروه 2
0/840	آزمون چندگزینه‌ای-چند پاسخ گروه 3
0/828	آزمون چندگزینه‌ای-چند پاسخ گروه 4
0/772	آزمون صحیح-غلط گروه 1
0/703	آزمون جور کردنی گروه 2
0/817	آزمون کوتاه پاسخ گروه 3

چنانچه ملاحظه می‌شود پایایی تمامی آزمون‌ها در سطح قابل قبولی می‌باشند. روش نمره‌گذاری: آزمون‌های کوتاه پاسخ، جور کردنی و صحیح-غلط به صورت دو ارزشی 0 و 1 نمره‌گذاری گردیدند. آزمون تشریحی گسترده پاسخ توسط استاد مربوطه نمره‌گذاری و نتایج در اختیار محقق قرار گرفته است. آزمون چندگزینه‌ای-چند پاسخ با استفاده از الگوریتم مورگان نمره‌گذاری گردید، الگوریتم مورگان یک روش دهی به صورت سه ارزشی 1، 0، 1- می‌باشد:

1: زمانی که گزینه درست از طرف آزمون‌دهنده انتخاب گردد.

1: زمانی که گزینه نادرست از طرف آزمون‌دهنده انتخاب گردد.

0: زمانی که گزینه، از طرف آزمون‌دهنده انتخاب نگردد.

دلایل محقق برای انتخاب این روش از بین روش‌های موجود برای نمره‌گذاری آزمون‌های چندگزینه‌ای- چند پاسخ:

- 1) منصفانه‌تر بودن این روش نمره‌دهی نسبت به روش نمره‌دهی دو ارزشی.
- 2) در نظر گرفتن نمره منفی، چون در غیر این صورت و در صورت انتخاب همه گزینه‌ها آزمون‌دهنده می‌تواند نمره کامل را اخذ نماید.
- 3) عدم تعلق نمره به گزینه‌های علامت زده نشده، چرا که اگر برای علامت نزدن گزینه‌ای که نباید علامت زده می‌شد نمره در نظر بگیریم، آزمون‌دهنده‌ای که هیچ گزینه‌ای را انتخاب نکند نمره خواهد گرفت.

### یافته‌ها

در این بخش ابتدا آماره‌های توصیفی از قبیل میانگین و انحراف معیار نمرات آزمون‌ها ارائه گردیده است.

جدول شماره 2: فراوانی، میانگین، انحراف معیار نمرات و تعداد سؤالات هر کلاس

نوع آزمون	تعداد	میانگین	انحراف معیار	کمترین	بیشترین	چولگی	کشیدگی
چندگزینه ای - چند پاسخ	94	5/16	2/16	1/92	8/85	0/38	-0/90
صحیح-غلط	94	13/10	3/69	6/67	17/33	-0/31	-1/59
چندگزینه ای - چند پاسخ	98	9/02	3/71	3	18	0/86	0/61
کوتاه پاسخ	98	11/87	3/81	6/67	17/33	0/14	-1/70
چندگزینه ای - چند پاسخ	102	3/70	3/28	0	12	0/62	-0/22
چور کردنی	102	9/91	5/00	0	20	-0/03	-0/60
چندگزینه ای - چند پاسخ	98	10/36	4/71	3	18	0/10	-1/31
تشریحی	98	14/94	4/26	7	20	-0/48	-1/08

نتایج آزمون همبستگی برای دوبه‌دو آزمون‌ها در جدول 3 آورده شده است.

جدول شماره 3: نتایج آزمون همبستگی پیرسون

مقدار sig	مقدار ضریب همبستگی	آزمون‌ها
0/000	0/628	آزمون چندگزینه‌ای-چند پاسخ با آزمون صحیح-غلط
0/000	0/771	آزمون چندگزینه‌ای-چند پاسخ با آزمون کوتاه پاسخ
0/000	0/669	آزمون چندگزینه‌ای-چند پاسخ با آزمون جور کردنی
0/000	0/865	آزمون چندگزینه‌ای-چند پاسخ با آزمون تشریحی

با توجه به جدول در تمامی موارد  $\text{sig} < 0/05$ ، لذا با اطمینان 95 درصد می‌توان ادعا داشت همبستگی بین آزمون چندگزینه‌ای-چند پاسخ و سایر انواع آزمون‌ها، معنادار می‌باشد.

آزمون سؤالات:

سؤال (1) آیا آزمون چندگزینه‌ای-چند پاسخ از طریق محاسبه همسانی درونی آن بر اساس فرمول کرونباخ دارای اعتبار است؟  
با توجه به نتایج حاصل از محاسبه آلفای کرونباخ برای سؤالات چندگزینه‌ای-چند پاسخ طبق جدول 1، پایایی تمام آزمون‌ها در سطح قابل قبول بود.

سؤال (2) آیا آزمون چندگزینه‌ای-چندپاسخ پیشرفت تحصیلی از طریق بررسی همبستگی آن با سایر انواع آزمون‌ها روا است؟

نتایج آزمون همبستگی پیرسون طبق جدول 2 نشان داد که در سطح اطمینان 95 درصد بین نمرات آزمون چندگزینه‌ای-چندپاسخ با آزمون صحیح-غلط، آزمون کوتاه پاسخ، آزمون جور کردنی و آزمون تشریحی همبستگی وجود دارد و می‌توان ادعا کرد آزمون چندگزینه‌ای-چندپاسخ از طریق برآورد همبستگی آن با سایر انواع آزمون‌های ذکر شده، دارای روایی است.

### بحث و نتیجه‌گیری

یافته‌های پژوهش نشان دادند که آزمون‌های چندگزینه‌ای - چندپاسخ طراحی شده، دارای روایی و پایایی بودند.

با توجه به اینکه از نظر کاپلان و ساکوزو (1386) ضریب آلفای کرونباخ بین 0/7 و 0/8 برای بیشتر هدف‌ها به اندازه‌ی کافی خوب است، لذا میزان پایایی آزمون‌های محقق ساخته، مناسب برای ارزشیابی پیشرفت تحصیلی می‌باشند. نتایج تحقیق فوق با

تحقیق ویلکوکس و پالاک (2014) که نشان داده‌اند سؤالات چندپاسخ و بازپاسخ تفاوت چندانی از لحاظ ویژگی‌های روانسنجی با یکدیگر ندارند، همسو می‌باشد. همچنین می‌توان به نتایج تحقیق جاوید (2014) که نشان داده پایایی سؤالات صحیح-غلط بیشتر از چندگزینه‌ای می‌باشد و تحقیق رساییان، نخعی و صادقی قندهاری (1381) که نشان داده‌اند سؤالات کوتاه‌پاسخ از اعتبار بیشتری در مقایسه با آزمون‌های چندگزینه‌ای برخوردارند، اشاره نمود. در تبیین یافته فوق می‌توان اشاره کرد که آزمون چندگزینه‌ای- چندپاسخ نوعی آزمون چندگزینه‌ای است لذا با رعایت اصول ساخت چنین آزمون‌هایی، می‌توان سؤالاتی پایا طراحی نمود.

همچنین با توجه به نتایج، همبستگی بین نمرات آزمون چندگزینه‌ای-چند پاسخ و نمرات سایر انواع آزمون‌ها از نظر آماری معنی‌دار و مقدار آن به لحاظ عددی مثبت می‌باشد که نشان‌دهنده رابطه مستقیم بین نمرات دو آزمون می‌باشد و بیانگر روایی این آزمون به عنوان ابزاری جهت ارزشیابی پیشرفت تحصیلی می‌باشد. نتایج تحقیق فوق با تحقیق ویلکوکس و پالاک (2014) که نشان داده‌اند سؤالات چندپاسخ و بازپاسخ تفاوت چندانی از لحاظ ویژگی‌های روان‌سنجی با یکدیگر ندارند، همسو می‌باشد.

در تبیین یافته‌های فوق می‌توان اظهار داشت، آزمونی که با سایر آزمون‌هایی که همان سازه را می‌سنجند همبستگی بالایی داشته باشد دارای روایی همگرایی خوب است (شکوهی یکتا و پرند، 1388)، از طرفی با توجه به اینکه سؤالات تمامی آزمون‌ها برای سنجش پیشرفت تحصیلی دانشجویان و از یک محتوا طراحی شده بودند؛ لذا هر دو از سازه‌ای یکسان، عمق یادگیری فراگیران را می‌سنجیدند، به همین دلیل انتظار می‌رفت که بین نمرات حاصل از دو آزمون همبستگی برقرار باشد که میزان این همبستگی شاهدهی بر میزان روایی همگرایی ابزار می‌باشد.

هرچند به سادگی نمی‌توان ادعا کرد که نوع خاصی از سؤال بهتر از انواع دیگر سؤال‌ها است و قبل از تعیین نوع سؤال ضروری است که به محاسن و معایب آن سؤال و کارایی آن در سنجش هدف‌های موردنظر توجه شود (کرمی، 1393) و کاملاً آشکار است که اندازه‌گیری دقیق مستلزم متناسب بودن سطح دشواری آزمون با سطح توانایی آزمودنی است (لرد، 1391)، با این وجود با توجه به اینکه آزمون چندگزینه‌ای-چند پاسخ نوعی آزمون عینی است لذا محاسن این آزمون‌ها را که سیف (1393) سرعت، سهولت تصحیح، عینیت نمره‌گذاری و پوشش بخش گسترده‌تری از محتوا ذکر می‌کند را دارا

می‌باشد و علاوه بر این محاسن به دلیل امکان استفاده از تعداد گزینه‌های متفاوت برای هر سؤال، استفاده از گزینه‌های معیوب هیچ ضرورتی نداشته و از طرفی نیز تفاوت در تعداد گزینه‌های صحیح عاملی است که موجب می‌گردد، احتمال درستی پاسخ حدسی بسیار کاهش یابد. لذا بخشی از معایب آزمون‌های چندگزینه‌ای را کاسته و می‌تواند جایگزینی مناسب برای انواع آزمون‌ها بنا به اهداف، ضرورت و شرایط باشد. در تحقیق فوق محقق با محدودیت‌هایی مواجه بوده از جمله اینکه، تحقیق فوق در دانشکده علوم انسانی در رشته علوم تربیتی و در مقطع کارشناسی ارشد و در شهر تبریز انجام گرفته است، لذا تعمیم نتایج آن به سایر دانشگاه‌ها، دانشکده‌ها، رشته‌ها، مقاطع و آموزش و پرورش باید با احتیاط صورت گیرد و با توجه به اینکه تحقیق فوق در سال تحصیلی 94-95 صورت گرفته، نتایج آن شاید در سال‌های آتی متفاوت باشد. با توجه به یافته‌های تحقیق پیشنهاد می‌شود معلمان، اساتید و طراحان آزمون‌های پیشرفت تحصیلی که در آزمون‌های خود از سؤالات صحیح-غلط استفاده می‌کنند، با توجه به وجود همبستگی بین نمرات آزمون چندگزینه‌ای-چند پاسخ و آزمون صحیح-غلط، از آزمون چندگزینه‌ای-چند پاسخ و به دلیل اینکه احتمال درستی پاسخ حدسی در آن کمتر می‌باشد و قادر به سنجش هدف‌های سطح بالاتری نسبت به آزمون صحیح-غلط می‌باشد، استفاده نمایند. در صورتی که در آزمون‌های خود از سؤالات کوتاه پاسخ استفاده می‌کنند، با توجه به وجود همبستگی بین نمرات آزمون چندگزینه‌ای-چند پاسخ و آزمون کوتاه پاسخ، از آزمون چندگزینه‌ای-چند پاسخ به دلیل که اینک می‌تواند اهداف سطوح بالاتر را بسنجد و نمره‌گذاری آن آسان‌تر است، استفاده نمایند، در صورتی که در آزمون‌های خود از سؤالات جور کردنی استفاده می‌کنند، با توجه به وجود همبستگی بین نمرات آزمون چندگزینه‌ای-چند پاسخ و آزمون جور کردنی، از آزمون چندگزینه‌ای-چند پاسخ به دلیل که اینک می‌تواند اهداف سطوح بالاتر را بسنجد و نمره‌گذاری آن آسان‌تر است، استفاده نمایند و در صورتی که در آزمون‌های خود از سؤالات تشریحی گسترده پاسخ استفاده می‌کنند، با توجه به وجود همبستگی بین نمرات آزمون چندگزینه‌ای-چند پاسخ و آزمون تشریحی گسترده پاسخ، از آزمون چندگزینه‌ای-چند پاسخ به دلیل که اینک نمره‌گذاری آن عینی بوده و به سهولت و در زمان کمتری انجام پذیر است و محتوای گسترده‌تری را تحت پوشش قرار می‌دهد، استفاده نمایند.

## References

- Ali, S.H., Carr, P.A., & Ruit, K.G. (2016), Validity and reliability of scores obtained on multiple-choice questions: Why functioning distractors matter. *Journal of the Scholarship of Teaching and Learning*, 16 (1), 1-14.
- Allah Dady, A., & Yaghubi, N. (2011), *Evaluating and Measuring in Education*. Tehran:Sobhan, (In Persian).
- Delavar, A., & Zaharakar, K. (2008), *Assessment and Measuring in Psychology, Conseling and Educational Science*. Tehran: Arasbaran Publications, (In Persian).
- Domnich, A., Pantatto, D., Arata, L., Bevilacqua, I., Apprato, L., Gasparini, R., & Amicizia, D. (2015), Impact of different scoring algorithms applied to multiple-mark survey items on outcome assessment: An in-field study on health related knowledge. *Journal of Preventive Medicine and Hygiene*, 56, 162-171.
- Javid, L. (2014), The comparison between multiple-choice (MC) and multiple true- false (MTF) test formats in Iranian intermediate EFL learners' vocabulary learning. *Procedia-Social and Behavioral Sciences*, 98, 784-788.
- Huang, Y.M., Trevisan, M., & Storfer, A. (2007), "The Impact of the "all-of-the-above" option and Student Ability on Multiple Choice Tests." *International Journal for the Scholarship of Teaching and Learning*, 1(2), 2-5.
- Kaplan, R.M., & Saccuzzo, D.P. (2004). *Psychological Assessment*. Translation: Delavar, A., Dortaj, F., & Farrokhi, N.A. (2007). Tehran: Arasbaran Publications, (In Persian).
- Karami, A. (2014), *Introduction to Testing and Psychological Tests*. Tehran: Psychometric Publishing Center, (In Persian).
- Karimi, B., Falsafi Nezhad, M.R., & Dortaj, F. (2011), The effect of the number of question options on the psychometric properties of the test and the ability satisfied in the classic question-and-answer

- measurement models. *Educational Measurement Quarterly*, 6, (In Persian).
- Lord, F.M. (1980), *Applications of Question-Answer Theory*. Translation: Delavar, A., & Younesi, J. (2012). Tehran: Ariyan Publications, (In Persian).
- Mollahadi, M. (2009), Evaluating by multichoice tests. *Educational Strategies in Medical Science*, 2(4), 177-182, (In Persian).
- Morgan, M.R.J. (1979), An interactive computer program for multiple-choice self-testing. *Biochemical Education*, 7(3), 67-69.
- Mortaz Hejri, S., Khabaz Mafinezhad, M., & Jalili, M. (2014), Guessing in multiple choice questions: challenges and strategies. *Iranian Journal of Medical Education*, 14, 594-604, (In Persian).
- Ng, A.W.Y., & Chan, A.H.S. (2009), Different methods of multiple-choice test: Implications and design for further research. *Proceedings of the International Multi-Conference of Engineers and Computer Scientists: Hong Kong, Vol: II*.
- Pasha Sharifi, H. (2011), *Psychometric Principles and Psychological Assessment*. Tehran Roshd, (In Persian).
- Rassaian, N., Nakhaei, S., & Sadeghi Ghandehari, N.(2002), Comparison of three exam techniques in medical students: multiple choice, true-false and short answer questions. *Hakim Research Journal*, 5(4), 271-278, (In Persian).
- Rodriguez, C.M. (2005), Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24(2),3-13.
- Saif., A.A. (2013), *Modern Educational Psychology*. Tehran: Doran Publications, (In Persian).
- Saif., A.A. (2014), *Assessing Process and Product of Learning*. Tehran: Doran Publications, (In Persian).

- Shukuhi Yekta, M., & Parand, M. (2009), Psychological and Educational Tests (Theoretical Foundations of Assessment and Application of Psychological and Educational Tests Include Intelligence, Academic Achievement and Personality Tests. Tehran: Tabib Publications, (In Persian).
- Tetsuhito, Sh., Osamu, T., Tomoko, Y., & Kiyomi, Y. (2006), Comparison of three and four option English test for university entrance selection, purposes in Japan. *Language Testing*, 23, 35-57.
- Thorndike, B.L. (1910), *Applied Psychometrics*. Translation: Human, H.A. (2007), Tehran: Tehran University, (In Persian).
- Scalise, K., & Gifford, B. (2006), Computer-based assessment in E-Learning: A framework for constructing "Intermediate constraint" questions and tasks for technology platforms. *The Journal of Technology, Learning and Assessment*, 4(6).
- Tarasowa, G., & Auer, S. (2013), Balanced scoring method for Multiple-mark questions Conference: Computer-Supported Education (CSEDU), 411-416.
- Verbic, S. (2012), Information value of Multiple-response questions. *Psihologija*, 45 (4), 467- 485.
- Wilcox, B.R., & Pollock, S.J. (2014), Coupled multiple-response versus free-response conceptual assessment: An example from upper-division physics. *Physical Review Special Topics\_Physics Education Research*, 10(020124), 1\_11.