

تأثیر مدل وزن‌دهی بر نمره کل‌سازی سوابق تحصیلی و رتبه‌بندی داوطلبان ورود به دانشگاه‌ها و مراکز آموزش عالی کشور

میثم صادقی^۱، محمدرضا فلسفی نژاد*^۲، علی دل‌آور^۳، نورعلی فرخی^۴، احسان جمالی^۵

تاریخ دریافت: ۹۷/۰۷/۲۵

تاریخ پذیرش: ۹۸/۰۲/۰۸

چکیده

هدف از پژوهش حاضر شناسایی وزن‌دهی هر یک از دروس سوابق تحصیلی و نمره کل‌سازی براساس رویکرد روانسنجی مدل پیوسته IRT و رویکردهای مبتنی بر نظر متخصصان تاپسیس و AHP، و همچنین مقایسه اثر مدل وزن‌دهی بر رتبه‌بندی داوطلبان بود. روش پژوهش حاضر آزمایشی بود. به منظور اجرای پژوهش در بخش اول ۱۱ آیتم (دروس) سوابق تحصیلی گروه علوم انسانی از گروه متخصصان رشته‌های روان‌شناسی و مشاوره مورد نظرخواهی قرار گرفت و با استفاده از تکنیک‌های تاپسیس و AHP وزن‌دهی و اولویت‌بندی شد. در بخش دوم ابتدا نمرات دروس نهایی سال سوم متوسطه پذیرفته شدگان رشته‌های روان‌شناسی و مشاوره در دانشگاه‌های سراسر کشور در گروه علوم انسانی از سازمان سنجش آموزش کشور دریافت شد و جهت برآورد نمرات تتا در قالب مدل پیوسته پرسش پاسخ استفاده شد. در بخش سوم و به منظور مقایسه‌ی مدل‌های مختلف وزن‌دهی و نمره کل‌سازی، ابتدا گروه نمونه براساس نمرات حاصل از هر سه رویکرد رتبه‌بندی شده، و از لحاظ شاخص‌های پراکندگی و میزان تفاوت رتبه‌ها مقایسه شد. نتایج

۱. دانشجوی دکتری سنجش و اندازه‌گیری دانشگاه علامه طباطبائی، تهران، ایران.

۲. *دانشیار سنجش و اندازه‌گیری دانشگاه علامه طباطبائی. تهران، ایران falsafinejad@yahoo.co.uk

۳. استاد ممتاز سنجش و اندازه‌گیری دانشگاه علامه طباطبائی، تهران، ایران.

۴. دانشیار سنجش و اندازه‌گیری دانشگاه علامه طباطبائی، تهران، ایران.

۵. استادیار سازمان سنجش آموزش کشور، تهران، ایران.

مقایسه سه رویکرد نشان داد بیشترین واریانس مربوط به روش نمره کل سازی با مدل پیوسته IRT است. همچنین نتایج تحلیل آماری مبنی بر آزمون ویل کاکسون جهت مقایسه متوسط رتبه‌ها نشان داد در حجم نمونه‌ی کم روش وزندهی به هر سه روش نسبت به روش وزن‌های یکسان تفاوت معنادار در رتبه‌ها ایجاد می‌کند و بنابراین مدل وزندهی در رتبه‌بندی داوطلبان در حجم نمونه کم تاثیر می‌گذارد. اما بین سه مدل وزندهی تفاوت معناداری در رتبه‌بندی داوطلبان مشاهده نمی‌شود. همچنین نتایج نشان داد در حجم نمونه بالا (۲۰۰۰ نفر) روش وزندهی به هر سه روش نسبت به روش وزن‌های یکسان تفاوت معنادار در رتبه‌ها ایجاد می‌کند و وزن دروس در رتبه‌بندی داوطلبان ورود به دانشگاه تاثیر می‌گذارد. همچنین نتایج تحلیل داده‌ها نشان داد نمرات بدست آمده از مدل پیوسته IRT نسبت به دو روش دیگر در رتبه‌بندی داوطلبان تاثیر متفاوتی می‌گذارد ولی دو روش دیگر تفاوت معناداری در رتبه‌بندی داوطلبان ندارند. مقایسه توزیع فراوانی نمره - توانایی در چارچوب هر سه رویکرد نشانگر این امر بود که توزیع فراوانی در IRT نزدیک تر به وضعیت توانایی واقعی گروه نمونه بود و دامنه وسیع‌تری از سطوح توانایی را در بر می‌گیرد و واریانس بیشتری ایجاد می‌کند.

واژگان کلیدی: وزندهی، نمره کل سازی، رتبه‌بندی، مدل پیوسته IRT، تاپسیس

مقدمه

روش‌های جذب و راهیابی داوطلبان به دانشگاه و آموزش عالی در سال‌های گذشته تغییرات و دگرگونی‌هایی داشته است. در سال ۱۳۸۶ طرح حذف کنکور سراسری و جایگزینی سوابق تحصیلی در ورود داوطلبان به دانشگاه مطرح و در مجلس شورای اسلامی به تصویب رسید. این طرح به عنوان چاره‌ای برای رفع مشکلات کنکور سراسری ارائه شده است، در صورتی که نتواند آن مشکلات را حل کند ممکن است مشکلات دیگری را برای دانش‌آموزان و خانواده آنان و نظام آموزشی کشور دامن زند که خود تبعات و آثار سوء و جبران ناپذیری را به دنبال دارد. طرح حذف کنکور که قرار بود از سال ۱۳۹۰ اجرایی شود، به دلایل عملی اجرایی نگردید، و در سال ۱۳۹۲ مورد بازنگری قرار گرفت (یوسف زاده چوسری، محمدگنجی نیک و خرم آبادی، ۱۳۹۲).

در قانون جدید نمرات دروس سال آخر دوره متوسطه به همراه آزمون عمومی و آزمون اختصاصی که توسط سازمان سنجش آموزش کشور برگزار می‌شود، مبنای گزینش دانشجو خواهد بود (مصوبه شورای عالی انقلاب فرهنگی، ۱۳۹۲).

مبنای دقیق و علمی نمره کل سازی که هم اکنون در کنکور سراسری و امتحانات نهایی مورد استفاده قرار می‌گیرد، مشخص نیست. نمراتی که مبنای تصمیم‌گیری گزینش قرار می‌گیرند و نحوه ترکیب آنها باید به گونه‌ای انجام شود که منصفانه باشد و نسبت به گروهی خاص تبعیض آمیز نباشد. برای مثال در کنکور سراسری گروه علوم انسانی برای درس تاریخ و جغرافیا ضریب ۱ اعمال می‌شود در حالی که در امتحانات نهایی سال سوم برای محاسبه معدل کل برای درس جغرافیا ضریب ۳ و برای درس تاریخ ضریب ۲ اعمال می‌شود. زمانی که چندین نمره اساس تصمیم‌گیری را تشکیل می‌دهند، ضرایب آنها می‌تواند به صورت علمی تعیین شود. به طوریکه برای آوردن نمره کل هر فرد، نمره‌های او را در چند آزمون فرعی که هر یک مستقلاً نمره گذاری شده، ترکیب شود. به طور کلی استفاده از مجموعه‌ی آزمون‌هایی^۲ که متشکل از چند آزمون تک بعدی^۳ هستند، توصیه می‌شود چرا که قادرند متغیر مورد اندازه‌گیری مجموعه‌ی آزمون را بهتر کنترل کرد و با توجه به هدف‌های اندازه‌گیری، تفسیرهای روشن‌تری را پدید آورد.

هنگامی که تأثیر سؤالات یا خرده‌آزمونها را بر نمره کل را به شکل متفاوت بخواهیم از وزن‌دهی استفاده می‌شود. ساده‌ترین روش ترکیب نمرات، وزن دادن برابر به آنها و جمع کردن نمرات است. به مرور زمان مشخص شده است که این روش رضایت بخش نیست چرا که نمی‌تواند اهمیت نسبی هر مؤلفه را در رتبه‌بندی آزمودنی‌ها لحاظ کند (آدامز^۴ و مورفی^۵، ۱۹۸۲؛ فرنچ^۶، ۱۹۸۵). بنابراین مطلوب آن است که مؤلفه‌ها در زمان

1. Weights
2. Test batteries
3. Unidimensional
4. Adams
5. Murphy
6. French

تشکیل نمرات کل وزن دهی شوند. آدامز و مورفی (۱۹۸۲) اشاره می کنند که وزن های اسمی و وزن های بدست آمده متفاوت می باشند. آن ها توضیح می دهند که وزن بدست آمده بیانگر درجه نسبی است که یک مؤلفه در واقع در نتایج نهایی یک گزینش سهم دارد. وزن اسمی همان وزن در نظر گرفته شده است. به نظر آن ها برداشت درستی نیست که فرض کرد اگر مؤلفه ای دو برابر مؤلفه های دیگر وزن داشته باشد در رتبه بندی گزینش نهایی داوطلبان دو برابر سهم دارد. از نظر آن ها رتبه بندی داوطلبان بستگی به فاصله ی بین نمرات داوطلبان پیشتاز و در نهایت پراکندگی (واریانس و کوواریانس) نمرات دارد. آدامز و مورفی تاکید می کنند که ارزش نمرات و نمره ی میانگین در وزن های بدست آمده و رتبه بندی داوطلبان سهم چندانی ندارد.

بررسی پیشینه های قدیمی تر مبحث وزن دهی به مؤلفه های ارزیابی نشان می دهد که وزن های بدست آمده برای مؤلفه ها اساسا بستگی به انحراف معیار نمرات دارد (فورست^۱، ۱۹۷۴؛ لیندکوئیت^۲، ۱۹۴۰؛ ورنون^۳، ۱۹۴۰؛ تین^۴، ۱۹۶۶). با این وجود، مقدار انحراف معیار نمرات ترکیبی ایجاد شده هم به واریانس و هم به کوواریانس نمرات مؤلفه ها بستگی دارد (آدامز و مورفی، ۱۹۸۲). اساسا دو نوع رویکرد در وزن دهی و رتبه بندی آیتم ها و مقیاس ها وجود دارد. اولین رویکرد، به کارگیری روش های مبتنی بر نظر متخصصان است که طی آن از افراد متخصص و آگاه نظر سنجی می شود و قضاوت در مورد آیتم ها براساس معیارهای معین بر حسب نظرات آنها انجام می پذیرد. در این رویکرد استفاده از روش های تاپسیس^۵ و تحلیل سلسله مراتبی^۶ AHP بیشتر از سایر روش ها مورد تاکید بوده اند (اونات^۷ و سونر^۸، ۲۰۰۸). در شرایطی که در تصمیم گیری با روش های چند معیاره صورت می گیرد و پژوهشگر با انتخاب، وزن دهی و رتبه بندی تعداد معینی از آیتم ها مواجه

1. Forrest
2. Lindquist
3. Vernon
4. Thyne
5. Topsis
6. Analytic Hierarchy Process
7. Ount
8. Soner

می‌شود، بدیهی است که در این شرایط استفاده از تکنیک‌های تصمیم‌گیری تک معیاره کارساز نبوده و از تکنیک‌های چندمعیاره^۱ (MADM) برای استخراج وزن‌ها استفاده می‌شود. یکی از تکنیک‌هایی که قابلیت اجرا و استفاده بالایی در این شرایط دارد تکنیک تاپسیس می‌باشد. تاپسیس با مفهوم معیار فاصله ای آیتم‌ها از راه حل ایده آل که برای اولین بار توسط یون^۲ و هوانگ^۳ (۱۹۹۵) ارائه شده است. در این تکنیک فرض می‌شود که k متخصص به ارزیابی m آیتم یا گویه می‌پردازند که براساس n معیار مورد ارزیابی قرار می‌گیرند و در آن آیتم‌هایی به عنوان راه حل ایده آل ارائه می‌شود که نسبت به ایده آل ترین راه حل نزدیکترین باشد.

همچنین در لحاظ کردن نظر متخصصان برای وزن‌دهی آیتم‌ها تکنیک دیگر تحلیل سلسله مراتبی (AHP) می‌باشد. AHP یکی از معروفترین فنون تصمیم‌گیری چند منظوره است که اولین بار توسط توماس (۱۹۸۸) ابداع گردید. این تکنیک، مسائل پیچیده را بر اساس آثار متقابل آنها مورد بررسی قرار می‌دهد و بر مقایسه‌های زوجی استوار است. فرآیند تحلیل سلسله مراتبی به گونه‌ای ماتریسهای حاصل از مقایسه‌های زوجی را با یکدیگر تلفیق می‌سازد که وزن بهینه برای آیتم‌ها حاصل آید.

روش‌های مبتنی بر نظر متخصصان تنها روش‌های موجود در وزن‌دهی مناسب مؤلفه‌ها نیستند. استفاده از روش‌های روانسنجی در وزن‌دهی و نمره‌گذاری آیتم‌ها سابقه ای طولانی در ادبیات سنجش و اندازه‌گیری دارد (گالکسن^۴، ۱۹۵۰). همچنین وانگ^۵ و استلی^۶ (۱۹۷۰)، داوز^۷ (۱۹۷۶)، و واینر^۸ و تیسن^۹ (۱۹۹۳) به چالش وزن‌دهی و ترکیب نمرات به منظور ساخت یک نمره کل پرداختند. واینر و تیسن (۱۹۹۳) رویکردهای روانسنجی

1. Multiple Attribute Decision Making
2. Yoon
3. Hwang
4. Gulliksen
5. Wang
6. Stanley
7. Dawes
8. Wainer
9. Thissen

وزن‌دهی و نمره کل سازی را به دو دسته روش‌های مستقیم و غیرمستقیم تقسیم کرده‌اند. روش‌های غیرمستقیم وزن‌دهی و نمره کل سازی شامل: وزن‌دهی براساس دشواری، پایایی و تحلیل عاملی می‌باشد. این روش‌ها از این روی غیرمستقیم نامیده شده‌اند که در آن‌ها وزن آیت‌ها به طور مستقل محاسبه شده و پس از آن نمرات کل براساس وزن‌ها ساخته می‌شوند. روش‌های مستقیم وزن‌دهی و نمره کل سازی، مجموع ساده‌ی آیت‌ها و استفاده از رویکرد نظریه پرسش پاسخ IRT می‌باشند.

وابسته بودن پارامتر آیت‌ها به گروه نمونه را از بزرگترین ایرادات تکنیک‌های کلاسیک وزن‌دهی و ترکیب نمرات بر شمرده اند (تامسون^۱ و هاسه^۲، ۲۰۰۰؛ ساویلوفسکی^۳، ۲۰۰۰). این نقطه ضعف تحلیل کلاسیک می‌تواند ترکیب نمرات را تحت تاثیر قرار دهد؛ لذا بهینه سازی فرآیند گزینش مستلزم استفاده از مدل‌ها و تکنیک‌هایی است این نقاط ضعف را بیشتر پوشش دهد. چنین روش‌هایی را می‌توان در مدل‌های جدید اندازه گیری که به مدل‌های پرسش-پاسخ «IRT» معروف اند، جستجو کرد. در این نظریه برآورد پارامترها به دلیل ویژگی نامتغیر بودن، ناپسته به گروه نمونه است.

مدل‌های IRT با استخراج سطح توانایی (تتا) براساس پارامترهای مدل یک روش مستقیم وزن‌دهی و نمره کل سازی محسوب می‌شوند. در مدل‌های تک پارامتری تتا فقط براساس سطح دشواری آیت‌ها وزن‌دهی و استخراج می‌شود. در مدل دوپارامتری تتا براساس سطح دشواری و سطح تمیز آیت‌ها وزن‌دهی و استخراج می‌شود. در این راه سمیجیما^۴ (۱۹۷۴) مدل پیوسته پرسش پاسخ را معرفی کرده است. زمانی که نمرات مؤلفه‌ها ارزش‌های نامحدودی را می‌گیرند، از مدل‌های پیوسته پرسش پاسخ جهت برآورد پارامترها استفاده می‌شود (زویلوگو، ۲۰۱۳).

با وجود اینکه رویکردهای وزن‌دهی نمرات پیوسته در پژوهش‌های قبلی معرفی شده اند ولی تاثیر وزن‌ها در رتبه‌بندی داوطلبان و گزینش منصفانه به اندازه کافی مورد بررسی

-
1. Thompson
 2. Haase
 3. Sawilowsky
 4. Samejima

قرار نگرفته است. لذا جای یک پژوهش علمی که به طور جامع رویکردهای وزن دهی و نمره کل سازی نمرات پیوسته را به طور همزمان در مطالعه ای مورد هم بینی قرار دهد در ادبیات روانسنجی خالی است. برای بررسی میزان اثربخش بودن این رویکردها در شرایط واقعی گزینش، نیاز به یک مصداق بیرونی است، تا وزن دهی و نمره کل سازی براساس آن صورت پذیرد. بنابراین با استفاده از شرایط واقعی داوطلبین گروه علوم انسانی در ایران، تاثیر روش های مبتنی بر نظر متخصصان (تاپسیس و AHP) و روش IRT به منظور نمره کل سازی سوابق تحصیلی و رتبه بندی داوطلبان مورد بررسی قرار می گیرند.

روش

این تحقیق از لحاظ روش یک پژوهش آزمایشی و به لحاظ هدف پژوهش، یک پژوهش کاربردی به شمار می رود. به دلیل بلا مصداق نبودن اجرای پژوهش و فراهم آوردن امکان اجرای پژوهش در شرایط واقعی، سوابق تحصیلی داوطلبان رشته های روان شناسی و مشاوره در گروه علوم انسانی مورد بررسی قرار می گیرد. در این صورت می توان گفت که جامعه آماری در این بخش کلیه پذیرفته شدگان رشته های روان شناسی، مشاوره در سال تحصیلی ۱۳۹۳ می باشند که دارای سوابق تحصیلی بودند.

در بخش نظرخواهی از متخصصان، از نظرات متخصصان صاحب نظر در رشته های روان شناسی و مشاوره استفاده شد. بدین ترتیب جامعه ی پژوهش در این قسمت کلیه افراد صاحب نظر در این رشته ها بود. در بخش نظرخواهی از متخصصان گروه نمونه از افراد خبره، متخصص و صاحب نظر رشته های روان شناسی و مشاوره به صورت هدفمند انتخاب شد. در این قسمت اساتیدی انتخاب شدند که حداقل دارای مرتبه علمی دانشجویی بوده و در دانشگاه های تهران، شهید بهشتی و علامه طباطبائی مشغول به تدریس بودند و یا سابقه تدریس داشتند. حجم نمونه در این بخش با استفاده از قاعده اشباع تعیین شد. در تحلیل های تصمیم گیری چندمعیاره، در صورت وجود معیارها متعدد حداقل حجم نمونه یک نفر به ازای هر معیار توصیه شده است. در این تحقیق به منظور حصول اطمینان کافی به نتایج ۳۰ نفر از اساتید صاحب نظر در رشته های روان شناسی و علوم تربیتی به عنوان

گروه نمونه مورد بررسی قرار گرفت. در بخش رویکرد روانسنجی IRT با توجه به محدود بودن حجم آیت‌های مورد بررسی در این پژوهش «۱۱ درس امتحان نهایی رشته علوم انسانی در سال سوم متوسطه» و لزوم وزن‌دهی هم‌ی خرد آزمون‌ها، ریز نمرات ۷۵۶۷ نفر از داوطلبان و پذیرفته شدگان رشته‌های روان‌شناسی، مشاوره در دانشگاه‌های سراسر کشور (دولتی، غیرانتفاعی و پیام نور) انتخاب و مورد تجزیه و تحلیل قرار گرفت.

مهمترین ابزار مورد استفاده در این پژوهش دروس امتحانات نهایی گروه علوم انسانی است که مورد نظر خواهی متخصصان قرار گرفتند و داده‌های آن از سازمان سنجش آموزش کشور برای گروه نمونه تهیه شد و این دروس عبارتند از:

جدول ۱. دروس امتحانات نهایی رشته علوم انسانی

نام دروس	ضریب
دین و زندگی	۳
آرایه‌های ادبی	۲
عربی ویژه انسانی	۴
فلسفه و منطق	۳
ریاضی ویژه انسانی	۲
جامعه‌شناسی	۲
زبان خارجی	۲
تاریخ ایران و جهان	۲
زبان فارسی تخصصی	۲
ادبیات فارسی تخصصی	۲
جغرافیا	۳

یافته‌های پژوهش

در فرایند تحلیل سلسله‌مراتبی، پس از ساخت مدل در برنامه Expert Choice و ورود ماتریس مقایسات زوجی، وزن آیت‌ها به گونه‌ای که در زیر نشان داده شده است، بدست آمد.

جدول ۲. اولویت بندی آیتمها

اولویت	وزن	آیتم
۶	۰/۰۷۲	معارف اسلامی
۸	۰/۰۵۳	آرایه‌های ادبی
۱۰	۰/۰۲۵	عربی
۵	۰/۱۰۹	فلسفه و منطق
۳	۰/۱۳۴	ریاضی
۴	۰/۱۱۲	جامعه‌شناسی
۱	۰/۱۹۶	انگلیسی
۹	۰/۰۳۱	تاریخ
۲	۰/۱۷۷	زبان فارسی
۷	۰/۰۷۰	ادبیات فارسی
۱۱	۰/۰۲۲	جغرافیا

با توجه به جدول فوق ملاحظه می‌گردد با توجه به وزن اصلی بدست آمده در روش AHP به ترتیب دروس انگلیسی، زبان فارسی، ریاضی، جامعه‌شناسی، فلسفه و منطق، معارف اسلامی، ادبیات فارسی، آرایه‌های ادبی، تاریخ، عربی و جغرافیا از بیشترین تا کمترین اهمیت در گروه برخوردار هستند. جدول ۳ مقادیر وزن مربوط به هر عامل (درس) را در روش تاپسیس نشان می‌دهد.

جدول ۳. مقدار وزن هر عامل

اولویت	وزن	عوامل (دروس)
۶	۰/۶۵۵	معارف اسلامی
۸	۰/۳۶۳	آرایه‌های ادبی
۱۰	۰/۰۲۹	عربی

۳	۰/۷۹۱	فلسفه و منطق
۴	۰/۷۵۶	ریاضی
۵	۰/۷۴۷	جامعه شناسی
۱	۱	انگلیسی
۹	۰/۱۰۴	تاریخ
۲	۰/۹۳۶	زبان فارسی
۷	۰/۴۹۵	ادبیات فارسی
۱۱	۰	جغرافیا

با توجه به وزن اصلی بدست آمده در روش تاپسیس به ترتیب مولفه‌های انگلیسی، زبان فارسی، فلسفه منطق، ریاضی، جامعه شناسی، معارف اسلامی، ادبیات فارسی، آرایه‌های ادبی، تاریخ، عربی و جغرافیا از بیشترین تا کمترین اهمیت در گروه برخوردار هستند. در بخش مربوط به رویکرد IRT ابتدا ویژگی‌های جمعیت شناختی نمونه‌ی پژوهش ($N=7567$) بررسی می‌شود.

جدول ۴. ویژگی‌های جمعیت شناختی داوطلبان و پذیرفته شدگان رشته‌های روان‌شناسی و مشاوره در کنکور ۱۳۹۳

درصد	فراوانی		
۱۹/۲	۱۴۵۳	پسر	جنسیت
۸۰/۸	۶۱۱۴	دختر	
۶۸/۶	۵۱۹۱	۱۳۹۲	سال اخذ دیپلم
۱۹/۱	۲۳۷۶	پیش از سال ۱۳۹۲	
۶۵/۲	۴۹۳۶	پیام نور	دوره‌ی پذیرش شده
۱۴/۴	۱۰۹۰	غیرانتفاعی	
۵/۹	۴۴۹	شبانه	
۱۴/۴	۱۰۹۲	روزانه	
۴۲	۳۱۷۷	کمتر از ۱۲	معدل کتبی دیپلم
۲۰/۳	۱۵۳۳	بین ۱۲ تا ۱۴	
۱۷/۸	۱۳۵۰	بین ۱۴ تا ۱۶	

تأثیر مدل وزن دهی بر نمره کل سازی سوابق تحصیلی و رتبه بندی داوطلبان ... / ۱۱

۱۲/۷	۹۵۸	بین ۱۶ تا ۱۸
۴/۸	۳۶۰	بین ۱۸ تا ۱۹
۲/۵	۱۸۹	بین ۱۹ تا ۲۰

نتایج جدول ۴ نشان می‌دهد بیشتر پذیرفته شدگان رشته‌های روان‌شناسی و مشاوره در کنکور ۱۳۹۲ دختران بودند. همچنین بیشترین فراوانی برای معدل کتبی کمتر از ۱۲ بوده است.

به منظور تحلیل آیت‌ها با نظریه پرسش پاسخ، ابتدا مفروضه‌های اساسی و اولیه نظریه پرسش پاسخ یعنی تک بعدی بودن و استقلال موضعی بررسی گردید و مشاهده شد دروس سوابق تحصیلی یک سازه‌ی تک عاملی می‌باشد. همبلتون معتقد است جهت تعیین تعداد ابعاد آزمون‌ها می‌توانیم از نسبت حاصل ارزش ویژه عامل اول به ارزش ویژه عامل دوم استفاده کنیم. او اظهار می‌دارد چنانچه عامل اول ۲ تا ۳ برابر عامل دوم باشد می‌توان نتیجه گرفت که آزمون مورد نظر تک بعدی است، یا به عبارت دیگر یک عامل غالب، عملکرد افراد در آزمون را تبیین می‌کند. در آیت‌های پژوهش ارزش ویژه عامل اول ۷/۶۷ بود و ۶۹/۴۳ درصد واریانس را تبیین می‌کرد. ارزش ویژه عامل دوم ۰/۴۷ بود.

همبلتون (۱۹۸۹) ذکر می‌کند که چنانچه فرض تک بعدی بودن برقرار باشد، مفروضه استقلال موضعی نیز برقرار است. با توجه به این مطلب در آزمون ترکیبی مذکور مفروضه استقلال موضعی نیز برقرار است. در جداول بعدی پارامترهای دشواری، تشخیص و حدس مربوط به هر یک از دروس آورده شده است.

جدول ۵. شاخص‌های مدل پیوسته IRT دروس سوابق تحصیلی

دروس	دشواری	تشخیص	آلفا
معارف اسلامی	-۰/۹۰	۱/۳۷	۱/۷۷
آزایه‌های ادبی	-۰/۷۶	۱/۳۲	۰/۹۲
ریاضی	-۰/۳۳	۰/۸۰	۱/۸۱

۱/۳۹	۱/۶۸	-۰/۶۷	فلسفه و منطق
۱/۱۱	۱/۴۴	-۰/۵۷	عربی
۱/۳۳	۱/۲۶	-۰/۹۶	جامعه شناسی
۰/۸۸	۱/۱۰	-۰/۰۶	انگلیسی
۱/۲۰	۱/۱۱	-۰/۶۲	تاریخ
۰/۹۶	۱/۵۹	-۰/۴۶	زبان فارسی
۱/۱۴	۱/۴۹	-۰/۶۴	ادبیات فارسی
۱/۲۱	۱/۴۳	-۰/۶۰	جغرافیا

با توجه به جدول فوق، در مدل پیوسته سمیجیما، دشوارترین آیتم‌ها به ترتیب انگلیسی و ریاضی و ساده ترین آیتم‌ها به ترتیب جامعه شناسی و معارف اسلامی بودند. همچنین فلسفه و منطق و زبان فارسی دارای بیشترین قدرت تشخیص است. به منظور مقایسه‌ی مدل‌های مختلف وزندهی و نمره کل سازی، ابتدا نمرات سوابق تحصیلی این افراد به نمرات تراز شده تبدیل شده و سپس نمرات هر درس براساس وزن‌های بدست آمده در هر رویکرد وزندهی شد. در مورد نمرات تنها هم با یک تبدیل خطی به نمرات تراز شده تبدیل شدند. در ادامه نمرات بدست آمده در هر چهار رویکرد لحاظ شاخص‌های پراکندگی، و میزان تفاوت رتبه‌ها مقایسه می‌شوند و براین اساس میزان تاثیرگذاری مدل وزندهی در رتبه‌بندی و گزینش مورد آزمون قرار می‌گیرد.

جدول ۶. میانگین و انحراف معیار نمرات کل ایجاد شده

شاخص	
میانگین نمرات تراز شده	انحراف معیار نمرات تراز شده
روش وزندهی	
۵۰۰۳/۱۷	۸۳۶/۴۰
روش تاپسیس	

تأثیر مدل وزن دهی بر نمره کل سازی سوابق تحصیلی و رتبه بندی داوطلبان ... / ۱۳

۸۲۸/۴۶	۵۰۰۳/۱۷	روش AHP
۹۷۹/۳۶	۴۹۹۷/۴۹	مدل پیوسته IRT

نتایج جدول نشان می دهد نمرات بدست آمده از رویکرد IRT پراکندگی بیشتری دارند. همچنین به منظور بررسی تأثیر مدل وزن دهی در رتبه بندی دو نمونه ی ۵۰۰ نفری و ۲۰۰۰ نفری از داوطلبان به طور تصادفی انتخاب شد و براساس چهار روش وزن دهی و نمره گذاری انجام شد.

جدول ۷. آزمون ویل کاکسون برای مقایسه رتبه ها (n=200)

Sig	مقدار Z	روش وزن دهی	روش وزن دهی
۰/۰۰۱	۱۲/۲۶	روش تاپسیس	
۰/۰۰۱	۱۲/۲۸	روش AHP	روش وزن های یکسان
۰/۰۰۱	۱۲/۳۴	مدل پیوسته IRT	
۰/۴۸	۰/۷۵	روش AHP	روش تاپسیس
۰/۲۹	۱/۰۴	مدل پیوسته IRT	
۰/۳۹	۰/۸۵	مدل پیوسته IRT	روش AHP

نتایج جدول نشان می دهد در حجم نمونه ی کم روش وزن دهی به هر سه روش نسبت به روش وزن های یکسان تفاوت معنادار در رتبه ها ایجاد می کند و بنابراین مدل وزن دهی در رتبه بندی داوطلبان در حجم نمونه کم تأثیر می گذارد. اما بین سه مدل وزن دهی تفاوت معناداری در رتبه بندی داوطلبان مشاهده نمی شود.

جدول ۸. آزمون ویل کاکسون برای مقایسه رتبه ها (n=5000)

Sig	مقدار Z	روش وزن دهی	روش وزن دهی
۰/۰۰۱	۳۳/۵۸	روش تاپسیس	
۰/۰۰۱	۳۳/۵۵	روش AHP	روش وزن های یکسان
۰/۰۰۱	۳۳/۵۷	مدل پیوسته IRT	
۰/۴۱	۰/۸۱	روش AHP	روش تاپسیس
۰/۰۱۳	۲/۴۸	مدل پیوسته IRT	

۰/۰۲	۲/۲۹	مدل پیوسته IRT	روش AHP
<p>نتایج جدول نشان می‌دهد روش وزن‌دهی به هر سه روش نسبت به روش وزن‌های یکسان تفاوت معنادار در رتبه‌ها ایجاد می‌کند و بنابراین مدل وزن‌دهی در رتبه‌بندی داوطلبان ورود به دانشگاه تاثیر می‌گذارد. به علاوه نتایج جدول نشان می‌دهد از بین سه مدل استفاده شده، مدل پیوسته IRT نسبت به دو روش دیگر در رتبه‌بندی داوطلبان تاثیر متفاوتی می‌گذارد ولی دو روش دیگر تفاوت معناداری در رتبه‌بندی داوطلبان ندارند. نتایج پژوهش بیانگر این مهم بود که مدل وزن‌دهی در رتبه‌بندی داوطلبان ورود به دانشگاه تاثیر می‌گذارد. به علاوه نتایج نشان داد از بین سه مدل استفاده شده، مدل پیوسته IRT نسبت به دو روش دیگر در رتبه‌بندی داوطلبان تاثیر متفاوتی می‌گذارد ولی دو روش دیگر تفاوت معناداری در رتبه‌بندی داوطلبان ندارند. با وجود این وزن‌های بدست آمده از روش تاپسیس واریانس بیشتری نسبت به وزن‌های بدست آمده از روش AHP دارند. روش AHP تکنیکی است که در پرسشنامه‌ی مقایسه زوجی ترکیب دو تایی همه‌ی آیت‌ها می‌بایست در نظر گرفته شود. بنابراین اگر تعداد آیت‌ها و معیارها زیاد باشد باعث می‌شود که تعداد مقایسات زوجی نیز افزایش پیدا کند که این امر باعث طولانی شدن پرسشنامه می‌شود و پاسخ‌دهندگان ممکن است در مقایسات دچار اشتباه شوند، در این مدل وزن‌دهی تعداد آیت‌ها و معیارها باید به حدی باشد که تعداد مقایسات زوجی داخل پرسشنامه در حد معقول و منطقی به دست آید. با در نظر گرفتن این نکته که در پژوهش حاضر ۱۱ آیت برای مقایسه موجود بود در روش AHP از در نظر گرفتن معیار صرف نظر کرد که این موضوع یکی از محدودیت‌های روش AHP در این پژوهش نسبت به روش تاپسیس بود.</p>			

بحث و نتیجه‌گیری

برای انجام روش تاپسیس حتماً بایستی معیار و وزن معیارها وجود داشته باشد. در پژوهش حاضر دو معیار مرتبط بودن و پیش نیاز بودن با وزن برابر در نظر گرفته شد. در روش تاپسیس میزان امتیاز هر گزینه نسبت به هر معیار به صورت طیف لیکرت به دست آمده و با توجه به اینکه داده‌های ماتریس تصمیم‌گیری قضاوتی هستند اجرای آن در تعیین وزن زمانی که آیت‌ها زیاد باشد (مانند دروس کنکور) راحت‌تر است و نتایج دقیق‌تری می‌دهد. به علاوه بایستی به این نکته نیز اشاره کرد که وزن‌های بدست آمده از روش تاپسیس واریانس بیشتری نسبت به وزن‌های بدست آمده از روش قضاوت‌های زوجی مبتنی بر تکنیک AHP داشتند و این نکته برای ایجاد نمرات با واریانس بیشتر یک نکته‌ی مطلوب است. تفاوت ایجاد شده بین روش‌های مبتنی بر نظر متخصصان و روش IRT را می‌توان چنین تبیین کرد که یکی از ویژگی‌های مهم برآورد توانایی در مدل‌های IRT این است که در این مدل‌ها برآوردهای توانایی وابسته به الگوی پاسخ آیت‌ها است (به جز مدل راش). این بدین معناست که آزمودنی‌هایی که دارای نمرات کل یکسانی هستند ممکن است برآورد θ های متفاوتی داشته باشند در صورتی که برخی از آیت‌هایی که نمره بالاتری گرفتند متفاوت باشد. در مدل‌های IRT استفاده از الگوی نمره گذاری موجب افزایش دقت برآورد توانایی می‌شود تا اینکه از فقط از تعداد پاسخ صحیحی یا نمره کل استفاده شود. از آن جهت که θ تبدیل خطی از تعداد پاسخ‌های صحیح نمی‌باشد، θ دقیقاً نمی‌تواند همان نمره Z باشد اما نحوه‌ی تفسیر آنها یکسان خواهد بود. هنگامی که یک آزمودنی ۱ انحراف معیار بالاتر از میانگین، جامعه مرجع می‌باشد، θ آزمودنی برابر ۱ می‌باشد، با این وجود هیچ فرضیه‌ای مبنی بر نرمال بودن توزیع θ وجود ندارد، به همین دلیل θ نمی‌تواند به نمرات درصدی، توزیع نرمال تبدیل شود. نکته‌ی دیگر که باید در مورد توزیع آماری نمره‌ها در نظریه‌ی کلاسیک مد نظر قرار گیرد این است که هیچ

مفروضه ای درباره‌ی توزیع نمره‌ها آنچنان که در نظریه‌ی سوال-پاسخ (IRT) وجود دارد ساخته نشده است. در نظریه‌ی کلاسیک نمره X یک آزمون دهنده به محتوای آزمون وابسته است. به عبارت دیگر نمره آزمون دهنده اندازه ای مطلق از خصیصه مورد اندازه گیری (مثل توانایی ریاضی یا فیزیک) او نیست. نمره هر فرد مجموع نمره‌های به دست آمده بر اساس نمرات خرده آزمون‌ها است و ضرایب سوال‌های و آزمون تحت تاثیر گروه نمونه آزمون دهنده‌ها قرار می‌گیرد و پارامتر توانایی واقعی جامعه آزمون دهندگان همچنان نامشخص است. از این رو باید گفت که آزمون سازی بر اساس نمرات معیار نظریه‌ی کلاسیک رویکرد توزیع آزاد می‌باشد. به عبارت دیگر در این نظریه هیچ گاه مقادیر نمره مشاهده شده X با مقادیر نظری T که در این تئوری به عنوان نمره واقعی به آن اشاره شده همگرا نخواهد شد و مقدار خطای اندازه گیری E تحت تاثیر اجزای مجدد یک آزمون تبدیل به چند نوع خطای سیستماتیک می‌شود و ماهیت و تصادفی بودن خود را از دست می‌دهد و پیش بینی یا برآورد مقادیر میانگین و انحراف معیار خطاهای سیستماتیک غیر ممکن و یا برای محاسبه به روش‌های پیچیده تر آماری مثل تئوری تعمیم پذیری دارند (واینر و تیسن، ۱۹۹۳). مدل‌های IRT با استخراج سطح توانایی (تتا) براساس پارامترهای مدل یک روش مستقیم وزن‌دهی و نمره کل سازی محسوب می‌شوند. در مدل‌های تک پارامتری تتا فقط براساس سطح دشواری آیت‌ها وزن‌دهی و استخراج می‌شود. در مدل دوپارامتری همچون مدل سمیجیما که در این پژوهش بکار گرفته شد نمرات کل (تتا) براساس سطح دشواری و سطح تمیز آیت‌ها وزن‌دهی و استخراج شد و این مزیت باعث شده نمرات تتا تاثیر متفاوتی بر رتبه‌بندی داوطلبان داشته باشد. در مجموع و در شرایط فعلی گزینش دانشجو به نظر می‌رسد اکثر وزن‌های دروس به یکدیگر نزدیک بوده و ضرایب دروس در نحوه‌ی مطالعه‌ی و نتایج تاثیر چندانی ندارد. براین اساس به نظر می‌رسد

با اطلاع رسانی کافی می توان چارچوب نمره گذاری آزمون ها را به سمت نمرات تنا که قدرت تشخیص بالاتری دارند تغییر داد. نتایج پژوهش نشان داد در حجم نمونه ی پایین ، رتبه های افراد در سه رویکرد، نزدیک بود و تفاوت بین رتبه ها اندک بود اما در حجم نمونه ی بزرگ که کنکور نیز این ویژگی را دارد، مقایسه توزیع فراوانی نمره - توانایی در چارچوب هر سه رویکرد نشانگر این امر بود که توزیع فراوانی در IRT نزدیک تر به وضعیت توانایی واقعی گروه نمونه بود (بیشترین فراوانی برای معدل کمتر از ۱۲) و دامنه وسیع تری از سطوح توانایی را در بر می گیرد و واریانس بیشتری ایجاد می کنند.

منابع

- یوسف زاده چوسری، محمدرضا؛ محمد گنجی نیک، فرشته و خرم آبادی، یدالله. (۱۳۹۲). پیامدهای طرح حذف کنکور از دیدگاه دبیران و دانش آموزان دوره ی پیش دانشگاهی شهر همدان در سال تحصیلی ۹۰-۸۹. مجله پژوهش علوم انسانی، ۱۲(۳۰)، ۱۰۳-۱۱۷.
- Adams, R. M. & Murphy, J. L. (1982). The achieved weights of examination components. *Educational Studies*, 8(1), 15-22.
- Dawes, R. (1976). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7), 571-582.
- Forrest, G. M. (1974). The presentation of results. In H. G. Macintosh (Ed.), *Techniques and problems of assessment*. London: Edward Arnold.
- French, S. (1985). The weighting of examination components. *Statistician*, 34. 265-280.
- Feldt, L. S. (2004). Estimating the reliability of a test battery composite or a test score based on weighted item scoring. *Measurement and Evaluation in Counseling and Development*, 37, 184-190.
- Gulliksen, H.O (1950). *Theory of Mental Tests*. New York: John Wiley and Sons, Inc.
- Hambleton, R & Swaminathan, H & Rogers, H.J (1991). *fundamentals of item response theory*. Newbury Park, CA
- Lindquist, E. F. (1940). *Statistical analysis in educational research*. Boston: Houghton Mifflin.
- Lord, F. M (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Ount S, Soner S(2008). Transshipment site selection using the AHP and TOPSIS approaches under fuzzy environment. *Waste Management*.;28(9):1552-59.
- Sawilowsky,S. (2000). Psychometrics versus Datametrics: Comment on Vacha-Haase's "Reliability Generalization" Method and Some Epm Editorial Policies. *Educational and Psychological Measurement*, vol. 60(2):157-173.
- Thompson,B , Haase, T. (2000). Psychometrics is Datametrics: the Test is not Reliable. *Educational and Psychological Measurement*, vol. 60, no. 2, pp. 174-195.
- Thyne, J. M. (1966). What is the weight of an examination paper? *British Journal of Educational Psychology*. 36. 321-326.
- Vernon, P. E. (1940). *The measurement of abilities*. London: University of London Press.
- Wainer, H. & Thissen D. (1993). Combining multiple-choice and constructed response test scores: Toward a marxist theory of test construction, *Applied Measurement in Education*, 6(2), 103-118.
- Wang, M.D. & Stanley J.C. (1970). Differential weighting: A review of methods and empirical studies. *Review of Educational Research*, 40, 663-705.
- Yoon, K.P.; Hwang, C. (1995). *Multiple Attribute Decision Making: An Introduction*. California: SAGE publications.
- Zopluoglu C. (2013). A comparison of two estimation algorithms for Samejima's continuous IRT model. *Behav Res Methods*, 45(1):54-64.