

تأثیر مدل وزن دهی و نمره کل سازی سوابق تحصیلی بر کارایی گزینش داوطلبان ورود به دانشگاه‌ها و مراکز آموزش عالی کشور

میثم صادقی*

محمد رضا فلسفی نژاد***

علی دلاور***

نورعلی فرخی****

احسان جمالی*****

چکیده

هدف از پژوهش حاضر شناسایی وزن هریک از دروس سوابق تحصیلی و نمره کل سازی براساس رویکردهای روانسنجی تحلیل عاملی و مدل پیوسته IRT و رویکردهای مبتنی بر نظر متخصصان تاپسیس و AHP، و همچنین مقایسه اثر مدل وزن دهی بر واریانس، پایایی و روایی پیش بین نمرات کل سوابق تحصیلی بود. روش پژوهش حاضر ترکیبی بود. به منظور اجرای پژوهش در بخش اول ۱۱ آیتم (دروس) سوابق تحصیلی از گروه متخصصان مورد نظرخواهی قرار گرفت و با استفاده از تکنیک‌های تاپسیس و AHP وزن دهی و اولویت بندی شد. در بخش دوم ابتدا نمرات دروس نهایی سال سوم متوسطه داوطلبان و پذیرفته شدگان رشته‌های روان‌شناسی و مشاوره در دانشگاه‌های سراسر کشور در گروه علوم انسانی از سازمان سنجش آموزش کشور دریافت شد و جهت برآورد وزن‌های استاندارد در قالب روش تحلیل عاملی و نیز محاسبه نمرات تتا در قالب مدل پیوسته پرسش پاسخ استفاده شد. در بخش سوم و به منظور مقایسه‌ی مدل‌های مختلف وزن دهی و نمره کل سازی، ابتدا گروه نمونه براساس نمرات حاصل از هر چهار رویکرد نمره گذاری شده، و از لحاظ شاخص‌های پراکندگی و میزان پایایی و روایی پیش بین مقایسه شدند. نتایج مقایسه چهار رویکرد نشان داد بیشترین واریانس مربوط به روش نمره کل سازی با مدل پیوسته IRT است. همچنین نتایج تحلیل داده‌ها نشان داد نمرات بدست آمده از مدل پیوسته IRT بیشترین همبستگی را با معدل دانشگاه و نمره روان‌شناسی کنکور دارند و بنابراین نسبت به سه مدل دیگر وزن دهی و نمره کل سازی بالاترین کارایی را در پیش بینی معدل دانشگاه و نمره‌ی روان‌شناسی کنکور دارد. نتایج نشان داد مدل وزن دهی تأثیر معناداری بر

* دانشجوی دکتری سنجش و اندازه گیری، دانشگاه علامه طباطبائی.

** دانشیار گروه سنجش و اندازه گیری، دانشگاه علامه طباطبائی. (نویسنده مسئول)

falsafinejad@yahoo.co.uk

*** استاد گروه سنجش و اندازه گیری، دانشگاه علامه طباطبائی.

**** دانشیار گروه سنجش و اندازه گیری، دانشگاه علامه طباطبائی.

***** استادیار سازمان سنجش آموزش کشور.

پایایی نمرات ندارد. مقایسه توزیع فراوانی نمره - توانایی در چارچوب هر چهار رویکرد نشانگر این امر بود که توزیع فراوانی در IRT دامنه وسیع‌تری از سطوح توانایی را در بر می‌گیرد و واریانس بیشتری ایجاد می‌کند و از روایی پیش بین بالاتری نیز برخوردار است. براین اساس به نظر می‌رسد با اطلاع رسانی کافی می‌توان چارچوب نمره گذاری آزمون‌ها را به سمت نمرات تتا که روایی پیش بین و قدرت تشخیص بالاتری دارند تغییر داد.

واژه‌های کلیدی: گزینش دانشجوی، وزن دهی، نمره کل سازی

مقدمه

هر ساله حدود یک میلیون داوطلب ورود به دانشگاه‌ها و مؤسسات آموزش عالی در کشور با هم رقابت می‌کنند و به علت تأثیری که ورود به دانشگاه و ادامه تحصیل در آینده شغلی و اجتماعی داوطلبان دارد، از اهمیت بالایی برای آنها برخوردار است (حیدری، فقیهی و اقبال، ۲۰۱۴). مطلوب نبودن مکانیسم جذب دانشجو به دانشگاه، علاوه بر ایجاد نگرانی، احتمالاً تضييع حقوق بخشی از داوطلبان را نیز به همراه دارد. از این رو طرح جایگزینی سوابق تحصیلی دانش‌آموزان در ورود به دانشگاه حساسیت و نگرانی بخش عظیمی از آحاد جامعه، به ویژه دانش‌آموزان سالهای پایانی دوره متوسطه را به دنبال داشته است. در صورت موفقیت آمیز نبودن کاهش اثرکنکور سراسری و برآورده نمودن خواست‌های عامه داوطلبان، نگرانی و نارضایتی از طرح افزایش می‌یابد و به تبع آن مشکلات فرهنگی و اجتماعی دیگری بدنبال خواهد داشت.

علی رغم اهمیت گزینش داوطلبان ورود به نظام آموزش عالی کشور، تحقیقات مدون علمی در انتخاب رویکرد بهینه وزن دهی و نمره گذاری نمرات پیوسته‌ی سوابق تحصیلی انجام نشده است. در یکی از معدود پژوهش‌هایی که در این زمینه در کشور صورت گرفته (ثاقبی، ۲۰۱۰). رابطه بین معدل دیپلم، رتبه کنکور سراسری دانشگاه‌ها و پیشرفت تحصیلی دانشجویان را در دانشگاه گیلان بررسی نمود. نتایج این پژوهش نشان داد بین معدل کل دانشجویان در دانشگاه (به عنوان متغیر ملاک) با رتبه کنکور سراسری و معدل دیپلم (به عنوان متغیرهای پیش بین) در سطح ۹۹ درصد اطمینان رابطه مثبت معنی داری وجود دارد. نتایج حاصل از تحلیل رگرسیون در دانشکده‌های دانشگاه گیلان بیانگر این مطلب بود که در دانشکده ادبیات و علوم انسانی و دانشکده منابع طبیعی، رتبه کنکور و در دانشکده‌های تربیت بدنی، علوم پایه و معماری معدل

دیپلم از قدرت پیش بینی بالاتری برخوردار می‌باشند. در دو دانشکده فنی مهندسی و کشاورزی هر دو متغیر معدل دیپلم و رتبه کنکور قادر به پیش بینی پیشرفت تحصیلی در دانشگاه می‌باشند که در دانشکده فنی مهندسی معدل دیپلم و در دانشکده کشاورزی رتبه کنکور از قدرت پیش بینی بالاتری برخوردار بود.

وزن نمراتی که مبنای تصمیم گیری گزینش قرار می‌گیرند و نحوه ترکیب آنها باید به گونه‌ای انجام شود تا به روایی و اعتبار آنها اطمینان وجود داشته باشد. 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 تصمیم گیری را تشکیل می‌دهند، ضرایب آنها می‌تواند به صورت علمی تعیین شود. به طوریکه برای بدست 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 اندازه گیری مجموعه‌ی آزمون را بهتر کنترل کرد و با توجه به هدف‌های اندازه گیری، تفسیرهای روشن‌تری را پدید آورد.

هنگامی که تأثیر سئوالات یا خرده آزمونها را بر نمره کل را به شکل متفاوت بخواهیم از وزن دهی 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 رتبه بندی آزمودنی‌ها لحاظ کند (آدامز و مورفی^۴، ۱۹۸۲؛ فرنچ^۵، ۱۹۸۵). بنابراین مطلوب آن است که مؤلفه‌ها در زمان تشکیل نمرات کل وزن دهی شوند.

یکسانی توزیع آماری روش‌های وزن دهی نشان می‌دهند که استفاده از روش‌های مختلف وزن دهی نمی‌تواند به طور قابل ملاحظه نتایج پایایی را تحت تأثیر قرار دهد و

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

پایایی را به مقدار رضایت بخشی افزایش دهد. وزن دهی می‌تواند بیشترین تأثیر را بر روایی آزمون داشته باشد (فلت^۱، ۲۰۰۴).

یک مقیاس مطمئن و دقیق برای بررسی روایی نمرات ترکیبی بدست آمده از آزمون‌های مختلف وجود ندارد. با این وجود روایی می‌تواند با بررسی کیفیت نمره‌ها در کاربردها و تفسیرهای مورد نظر مورد بررسی قرار گیرد (مسیک^۲، ۱۹۹۴).

اساساً دو نوع رویکرد در وزن دهی و رتبه بندی آیتم‌ها و مقیاس‌ها وجود دارد. اولین رویکرد، به کارگیری روش‌های مبتنی بر نظر متخصصان است که طی آن از افراد متخصص و آگاه نظر سنجی می‌شود و قضاوت در مورد آیتم‌ها براساس معیارهای معین بر حسب نظرات آنها انجام می‌پذیرد. در این رویکرد استفاده از روش‌های تاپسیس^۳ و تحلیل سلسله مراتبی^۴ AHP بیشتر از سایر روش‌ها مورد تأکید بوده‌اند (اونت و سانر^۵، ۲۰۰۸).

در شرایطی که در تصمیم‌گیری با روش‌های چند معیاره صورت می‌گیرد و پژوهشگر با انتخاب، وزن دهی و رتبه بندی تعداد معینی از آیتم‌ها مواجه 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 است. تاپسیس با مفهوم معیار فاصله‌ای آیتم‌ها از راه حل ایده آل که برای اولین بار توسط یانگ و هونگ^۶ (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 آل‌ترین راه حل نزدیک‌ترین باشد.

همچنین در لحاظ کردن نظر متخصصان برای وزن دهی آیتم‌ها تکنیک دیگر تحلیل سلسله مراتبی (AHP) است. AHP یکی از معروفترین فنون تصمیم‌گیری چند منظوره

-
1. Feldt
 2. Messick
 3. Topsis
 4. Analytic Hierarchy Process
 5. Ount & Soner
 6. Multiple Attribute Decision Making
 7. Yoon & Hwang

است که اولین بار توسط ساتی و وارگاس^۱ (۱۹۸۷) 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 وزن دهی مناسب معیارهای تصمیم‌گیری نیستند. استفاده از رویکردهای روانسنجی در وزن دهی و نمره گذاری سؤالات سابقه‌ای طولانی در ادبیات سنجش و اندازه‌گیری دارد (گالیکسن^۲، ۱۹۵۰). همچنین وانگ و استنلی^۳ (۱۹۷۰) و واینر و تیسن^۴ (۱۹۹۳) به چالش وزن دهی و ترکیب نمرات به منظور ساخت یک نمره کل پرداختند. همچنین واینر و تیسن (۱۹۹۳) ویکردهای روانسنجی وزن دهی و نمره کل سازی را به دو دسته روش‌های مستقیم و غیرمستقیم تقسیم کرده‌اند. روش‌های غیرمستقیم وزن دهی و نمره کل سازی شامل: وزن دهی براساس دشواری، براساس اعتبار و براساس تحلیل عاملی است. این روش‌ها از این روی غیرمستقیم نامیده شده‌اند که در آنها وزن آیتم‌ها به طور مستقل محاسبه شده و پس از آن نمرات کل براساس وزن‌ها ساخته می‌شوند. روش‌های مستقیم وزن دهی و نمره کل سازی، مجموع ساده‌ی آیتم‌ها و استفاده از رویکرد نظریه پرسش پاسخ IRT می‌باشند. پژوهشگران زیادی در حوزه‌ی سنجش بر استفاده از تحلیل عاملی در وزن دهی و یافتن ترکیب مناسب نمرات حاصل از عوامل مختلف تأکید کرده‌اند (ولز^۵، ۱۹۹۹؛ حبیبی، ایزانلو و خدایی^۶، ۲۰۱۳). استفاده از تحلیل عاملی قابلیت آنرا دارد که از طریق یافتن m تابع خطی غیر همبسته از ترکیب آیتم‌ها، بهترین ممیزها را از بین آنها انتخاب کرد. در این صورت می‌توانیم توابعی را که واریانس کمتری منعکس می‌کنند، کنار بگذاریم و فقط توابعی را در نظر بگیریم که واریانس بیشتری دارند، به طوری که درصد بیشتری از تغییرات را منعکس می‌کنند. در تحلیل

-
1. Saaty, & Vargas
 2. Gulliksen
 3. Wang & Stanley
 4. Wainer & Thissen
 5. Wells
 6. Habibi, Isanlou & Khoda'i

عاملی، وزن‌های استاندارد شده برای محاسبه نمرات کل به کار گرفته می‌شوند (آدام^۱، ۲۰۱۱).

تحلیل عاملی جزئی از تکنیک‌های کلاسیک وزن دهی و ترکیب نمرات تلقی می‌شود و پژوهش‌های قبلی وابسته بودن پارامتر آیت‌ها به گروه نمونه را از بزرگترین ایرادات این تکنیک‌ها بر شمرده‌اند (هامبلتون^۲ و ساویلسکی^۳، ۲۰۰۰؛ لرد، ۱۹۸۰؛ تامپسون و هسه^۴، ۲۰۰۰). این نقطه ضعف تحلیل کلاسیک می‌تواند ترکیب نمرات را تحت تأثیر 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» معروف‌اند، جستجو کرد. در این نظریه برآورد پارامترها به دلیل ویژگی نامتغیر بودن، نایسته به گروه نمونه است.

مدل‌های IRT با استخراج سطح توانایی (تتا) براساس پارامترهای مدل یک روش مستقیم وزن دهی و نمره کل سازی محسوب می‌شوند. در مدل‌های تک پارامتری تتا فقط براساس سطح دشواری آیت‌ها وزن دهی و استخراج می‌شود. در مدل دوپارامتری تتا براساس سطح دشواری و سطح تمیز آیت‌ها وزن دهی و استخراج می‌شود. در این راه سامجیا^۵ (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 (زاپل اوغلو^۶، ۲۰۱۳).

با وجود اینکه رویکردهای وزن دهی در پژوهش‌های قبلی معرفی شده‌اند ولی تأثیر وزن‌ها در روایی نمرات ساخته شده و گزینش منصفانه به اندازه کافی مورد بررسی قرار نگرفته است. لذا جای یک پژوهش علمی که به طور جامع رویکردهای وزن دهی و نمره کل سازی را به طور همزمان در مطالعه‌ای دقیق مورد هم بینی قرار دهد در ادبیات روانسنجی خالی است. مطالعه‌ای که علاوه بر اطمینان از منصفانه بودن سنجش بتواند

-
1. Odum
 2. Hambleton
 3. Sawilowsky
 4. Thompson & Haase
 5. Samejima
 6. Zopluoglu

کارایی گزینش را تضمین کند. 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، روایی روش‌های مبتنی بر نظر متخصصان (تاپسیس و AHP) و روش‌های مبتنی بر تحلیل عاملی و IRT به منظور وزن دهی بهینه به سوابق تحصیلی و تشکیل نمره کل مورد بررسی قرار می‌گیرند.

روش

به منظور پاسخگویی به سؤالات و ارزیابی جنبه‌های مختلف وزن دهی به نمرات از روش ترکیبی که در آن شیوه‌های کمی و مبتنی بر نظر متخصصان به طور توأمان به کار گرفته می‌شود، مورد استفاده قرار گرفت. روش‌های کمی امکان استفاده از روش‌های آماری پیشرفته را فراهم می‌کند و روش‌های مبتنی بر نظر متخصصان امکان درک عمیق پدیده مورد بررسی و شفاف سازی جنبه‌های نهفته و ناپیدای آن را فراهم می‌آورد. این تحقیق به لحاظ هدف پژوهش، یک پژوهش توسعه‌ای به شمار می‌رود چراکه به دنبال تدوین الگوی بهینه وزن دهی است. به دلیل بلا مصداق نبودن اجرای پژوهش و فراهم آوردن امکان اجرای پژوهش در شرایط واقعی، سوابق تحصیلی داوطلبان رشته‌های روان‌شناسی و مشاوره در گروه علوم انسانی مورد بررسی قرار می‌گیرد. در این صورت می‌توان گفت که جامعه آماری در این بخش کلیه داوطلبان و پذیرفته شدگان رشته‌های روان‌شناسی، مشاوره در سال تحصیلی ۱۳۹۳ می‌باشند که دارای سوابق تحصیلی بودند. در بخش نظرخواهی از متخصصان، از نظرات متخصصان صاحب نظر در رشته‌های روان‌شناسی و مشاوره استفاده شد. بدین ترتیب جامعه‌ی پژوهش در این قسمت کلیه افراد صاحب نظر در این رشته‌ها بود.

در بخش نظرخواهی از متخصصان گروه نمونه از افراد خبره، متخصص و صاحب نظر رشته‌های روان‌شناسی و مشاوره به صورت هدفمند انتخاب شد. در این قسمت اساتیدی انتخاب شدند که حداقل دارای مرتبه علمی دانشیاری بوده و در دانشگاه‌های 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 تحلیل‌های

تصمیم‌گیری 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 رشته‌های روان‌شناسی و علوم تربیتی به عنوان گروه نمونه مورد بررسی قرار گرفت. در بخش رویکردهای روانسنجی تحلیل عاملی 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 وزن دهی 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 درس	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

یافته‌ها

در فرایند تحلیل سلسله مراتبی، پس از ساخت مدل در برنامه Expert Choice و ورود ماتریس مقایسات زوجی، وزن آیتم‌ها 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	جامعه شناسی
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، جامعه شناسی، 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	
۸	۰/۳۶۳	
۱۰	۰/۰۲۹	
۳	۰/۷۹۱	

اولویت	وزن	عوامل (دروس)
۴	۰/۷۵۶	
۵	۰/۷۴۷	
۱	۱	
۹	۰/۱۰۴	
۲	۰/۹۳۶	
۷	۰/۴۹۵	
۱۱	۰	

با توجه به وزن اصلی بدست آمده در روش تاپسیس به ترتیب مؤلفه‌های 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 تحلیل عاملی و 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	
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 داوطلبان و پذیرفته شدگان رشته‌های روان‌شناسی و مشاوره در کنکور ۱۳۹۲ دختران بودند. همچنین بیشترین فراوانی برای معدل کتبی کمتر از ۱۲ بوده است.

ابتدا در مورد قابلیت ۱۱ درس سوابق تحصیلی برای تحلیل عاملی، از آزمون KMO_1 استفاده شد که ضریب به دست آمده رضایت‌بخش بود ($KMO = 0/97$). همچنین نتیجه آزمون کرویت بارتلت^۲ در سطح $P < 0/0001$ معنی‌دار بود ($70.667/58$ = χ^2). برای تحلیل عاملی از روش مؤلفه‌های اصلی، با استفاده شد. براساس نتایج فقط یک عامل برای ۱۱ درس مورد نظر استخراج شد. پژوهش حاضر ۱ عامل را شناسایی کرد که در مجموع $69/43$ درصد واریانس را تبیین می‌کنند. عامل دوم فقط ۴ درصد واریانس را تبیین می‌کرد و دارای ارزش ویژه $0/50$ بود. نتایج حاصل از بارهای عاملی این عامل یا بعد بر روی ۱۱ درس در جدول ۵ ارائه شده است.

جدول ۵. بارهای عاملی برای ۱۱ درس سوابق تحصیلی در کل گروه نمونه

اولویت	بار عاملی	درس
۵	۰/۸۳۹	معارف اسلامی
۶	۰/۸۳۸	آرایه‌های ادبی
۱۱	۰/۷۴۰	ریاضی
۱	۰/۸۹۲	فلسفه و منطق
۲	۰/۸۷۶	عربی
۷	۰/۸۳۳	جامعه‌شناسی
۱۰	۰/۷۷۳	انگلیسی
۹	۰/۸۰۹	تاریخ
۳	۰/۸۷۰	زبان فارسی
۴	۰/۸۵۹	ادبیات فارسی
۸	۰/۸۲۶	جغرافیا

با توجه به جدول ملاحظه می‌گردد با توجه به وزن اصلی بدست آمده در روش تحلیل عاملی، فلسفه منطق، عربی، زبان فارسی، ادبیات فارسی، معارف اسلامی،

1. Kaiser- Meyer- Olkin
2. Bartlett's test of Sphericity

آرایه‌های ادبی، جامعه‌شناسی، جغرافیا، تاریخ، انگلیسی و ریاضی از بیشترین تا کمترین وزن استاندارد شده برخوردار هستند. 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 همانگونه که مشاهده شد دروس سوابق تحصیلی یک سازه‌ی تک عاملی است. 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. 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 هریک از دروس آورده شده است.

جدول ۶. شاخص‌های مدل پیوسته IRT دروس سوابق تحصیلی

دروس	دشواری	تشخیص	آلفا
معارف اسلامی	-۰/۹۰	۱/۳۷	۱/۷۷
آرایه‌های ادبی	-۰/۷۶	۱/۳۲	۰/۹۲
ریاضی	-۰/۳۳	۰/۸۰	۱/۸۱
فلسفه و منطق	-۰/۶۷	۱/۶۸	۱/۳۹
عربی	-۰/۵۷	۱/۴۴	۱/۱۱
جامعه‌شناسی	-۰/۹۶	۱/۲۶	۱/۳۳
انگلیسی	-۰/۰۶	۱/۱۰	۰/۸۸
تاریخ	-۰/۶۲	۱/۱۱	۱/۲۰
زبان فارسی	-۰/۴۶	۱/۵۹	۰/۹۶
ادبیات فارسی	-۰/۶۴	۱/۴۹	۱/۱۴
جغرافیا	-۰/۶۰	۱/۴۳	۱/۲۱

با توجه به جدول فوق، در مدل پیوسته سمیجیما، دشوارترین آیت‌ها به ترتیب انگلیسی و ریاضی و ساده‌ترین آیت‌ها به ترتیب جامعه‌شناسی و معارف اسلامی بودند. همچنین فلسفه و منطق و زبان فارسی دارای بیشترین قدرت تشخیص است.

به منظور مقایسه‌ی مدل‌های مختلف وزن دهی 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 وزن دهی شد. در مورد نمرات تنها هم با یک تبدیل خطی به نمرات تراز شده تبدیل شدند.

جدول ۷. میانگین و انحراف معیار نمرات کل ایجاد شده

شاخص روش وزن دهی	میانگین نمرات تراز شده	انحراف معیار نمرات تراز شده
روش تاپسیس	۵۰۰۳/۱۷	۸۳۶/۴۰
روش AHP	۵۰۰۳/۱۷	۸۲۸/۴۶
روش تحلیل عاملی	۵۰۰۲/۵۳	۸۲۰/۶۸
مدل پیوسته IRT	۴۹۹۷/۴۹	۹۷۹/۳۶

نتایج جدول نشان می‌دهد نمرات بدست آمده از رویکرد IRT پراکندگی بیشتری دارند. به منظور مقایسه‌ی کارایی مدل‌های مختلف وزن دهی و نمره کل سازی، براساس روش مونت کارلو، سوابق داوطلبان به همراه معدل دانشگاه براساس میزان همبستگی گزارش شده در پژوهش‌های قبلی (۰/۵۴ همبستگی بین معدل دیپلم و معدل دانشگاه‌ها) و همچنین استفاده از نمرات روان‌شناسی کنکور پذیرفته شدگان، یک نمونه ۲۰۰۰ نفری شبیه سازی شد و بررسی شد که نمرات بدست آمده از وزن‌های اعمال شده‌ی کدام رویکرد وزن دهی، در پیش بینی معدل دانشگاه و نمره روان‌شناسی کنکور، بیشترین روایی را دارد.

جدول ۸. ضرایب همبستگی بین نمرات کل حاصل از چهار مدل وزن دهی

با معدل دانشگاه و نمرات روان‌شناسی کنکور

روان‌شناسی کنکور	معدل دانشگاه	
۰/۵۴۱	۰/۵۴۵	نمرات بدون وزن
۰/۵۴۵	۰/۵۳۳	نمرات تاپسیس
۰/۵۴۷	۰/۵۳۵	نمرات AHP
۰/۵۳۹	۰/۵۴۶	نمرات تحلیل عاملی
۰/۶۵۹	۰/۶۹۸	نمرات مدل پیوسته IRT

نتایج جدول نشان می‌دهد نمرات بدست آمده از مدل پیوسته IRT بیشترین همبستگی را با معدل دانشگاه و نمره روان‌شناسی کنکور دارند و بنابراین نسبت به سه مدل دیگر وزن دهی و نمره کل سازی بالاترین کارایی را در پیش بینی معدل دانشگاه و نمره‌ی روان‌شناسی کنکور دارد.

به منظور مقایسه‌ی پایایی مدل‌های مختلف وزن دهی و نمره کل سازی ضرایب آلفای کرونباخ نمرات حاصل از هر سه روش مقایسه شد.

جدول ۹. ضریب پایایی آلفای کرونباخ نمرات حاصل از سه روش

شاخص روش وزن دهی	آلفای کرونباخ
روش تاپسیس	۰/۸۹۷
روش AHP	۰/۸۹۹
روش تحلیل عاملی	۰/۹۵۲
مدل پیوسته IRT	۰/۹۴۱

در مجموع به نظر می‌رسد نمرات روش IRT پایایی بالاتری نسبت به روش‌های مبتنی بر نظر خواهی از متخصصان دارد. با این وجود به نظر نمی‌رسد روش وزن دهی بر میزان پایایی نمرات تأثیر معناداری داشته باشد.

بحث و نتیجه‌گیری

نتایج پژوهش بیانگر این مهم بود که مدل وزن دهی در نمره کل سازی و رتبه بندی داوطلبان ورود به دانشگاه تأثیر می‌گذارد. به نظر می‌رسد تفاوت ایجاد شده 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 (واینر و تیسن، ۱۹۹۳).

مدل‌های IRT با استخراج سطح توانایی (تتا) براساس پارامترهای مدل یک روش مستقیم وزن دهی و نمره کل سازی محسوب می‌شوند. در مدل دوپارامتری همچون مدل سمیجیما که در این پژوهش بکار گرفته شد نمرات کل (تتا) براساس سطح دشواری و سطح تمیز آیتم‌ها وزن دهی و استخراج شد.

نتایج تحلیل داده‌ها نشان داد نمرات بدست آمده از مدل پیوسته IRT بیشترین همبستگی را با معدل دانشگاه و نمره روان‌شناسی کنکور دارند و بنابراین نسبت به سه مدل دیگر وزن دهی و نمره کل سازی بالاترین روایی را در پیش بینی معدل دانشگاه و نمره‌ی روان‌شناسی کنکور دارد.

در مدل‌های تک پارامتری تتا فقط براساس سطح دشواری آیتم‌ها وزن دهی 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 آیتم‌ها وزن دهی و استخراج شد و این مزیت باعث شده نمرات تتا روایی بیشتری در پیش بینی معدل دانشگاه و نمرات روان‌شناسی کنکور داشته باشد. استفاده از روش‌های مختلف وزن دهی نمی‌تواند به طور قابل ملاحظه نتایج پایایی را تحت تأثیر قرار دهد و پایایی را به مقدار رضایت بخشی افزایش دهد. وزن دهی می‌تواند بیشترین تأثیر را بر روایی آزمون داشته باشد (فلت، ۲۰۰۴).

آدامز و مورفی (۱۹۸۲) اشاره می‌کنند که وزن‌های اسمی و وزن‌های بدست آمده متفاوت می‌باشند. آن‌ها توضیح می‌دهند 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 آن‌ها برداشت درستی نیست که فرض کرد اگر مؤلفه‌ای دو برابر مؤلفه‌های دیگر وزن داشته باشد در رتبه بندی گزینش نهایی داوطلبان دو برابر سهم دارد. از نظر آن‌ها رتبه بندی داوطلبان بستگی به فاصله‌ی بین نمرات داوطلبان پیش‌تاز و در نهایت پراکندگی (واریانس و کوواریانس) نمرات دارد که به نظر می‌رسد در نمرات تتا این موارد بهتر لحاظ می‌شود.

در مجموع صرف نظر از این که کدام یک از روش‌ها برای به دست آوردن وزن‌ها انتخاب شوند، فلسفه وزن دهی و نمره گذاری بر اساس این حقیقت است که وزن بزرگ‌تر به متغیری تعلق گیرد که ملاک (Y) انتخاب شده را بهتر اندازه بگیرد یعنی

بتواند عملکرد موفقیت آمیز بعدی فرد را پیش بینی کند و وزن کمتر و شاید منفی به متغیری تعلق گیرد که ملاک (Y) انتخاب شده را بدتر اندازه بگیرد (رادنر، ۲۰۰۱؛ غلت، ۲۰۰۴).

در شرایط فعلی کنکور به نظر می‌رسد اکثر وزن‌های دروس به یکدیگر نزدیک بوده و ضرایب دروس در نحوه‌ی مطالعه‌ی و نتایج کنکور تأثیر چندانی ندارد. براین اساس به نظر می‌رسد با اطلاع رسانی کافی می‌توان چارچوب نمره گذاری آزمون‌ها را به سمت نمرات تتا که روایی پیش بین و قدرت تشخیص بالاتری دارند تغییر داد. مقایسه توزیع فراوانی نمره - توانایی در چارچوب هر سه رویکرد نشانگر این امر بود که توزیع فراوانی در IRT نزدیک‌تر به وضعیت توانایی واقعی گروه نمونه بود (بیشترین فراوانی برای معدل کمتر از ۱۲) و دامنه وسیع‌تری از سطوح توانایی را در بر می‌گیرد و واریانس بیشتری ایجاد می‌کند و از روایی پیش بین بالاتری نیز برخوردار است.

منابع

- Adams, R. M. & Murphy, J. L. (1982). The achieved weights of examination components. *Educational Studies*, 8(1), 15-22.
- 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.
- 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.
- Gulliksen, H.O (1950). *Theory of Mental Tests*. New York: John Wiley and Sons, Inc.
- Habibi, M, Isanlou, B, & Khoda'i, I (2013). Application of Factor Analysis in Determining the Weight of Units in Comparison with Weight Measurement with Constant Quantities in Educational Psychometric Measurements. *Journal of Educational Measurement*, (9), 104-82. (in Persian).
- Hambleton, R & Swaminathan, H & Rogers, H.J (1991). *fundamentals of item response theory*. Newbury Park, CA.
- Heydari, M.R; Faghihi, M., & Iqbal, A.A. (2014). The study of the correlation between the degree of diploma and the rank of the national university curriculum with the academic achievement indicators of medical students of Yazd Shahid Sadoughi University of Medical Sciences. *Journal of Medical Education Development Center in Yazd*, 9 (4), 19-. (in Persian).
- Lindquist, E. F. (1940). *Statistical analysis in educational research*. Boston: Houghton Mifflin.
- Lord, M(1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Messick, S. (1994). Interplay of evidence and consequences in the validation of performance assessments, *Kducational Researcher*, 23,13-23.

- Odum, M. (2011). Factor scores, structure and communality coefficients: A primer. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio.
- Ount S, Soner S(2008). Transshipment site selection using the AHP and TOPSIS approaches under fuzzy environment. *Waste Management.*;28(9):1552-59.
- Rudner, L. M. (2001). Informed Test Component Weighting. *Educational Measurement: Issues and Practice.*
- Saaty, T. L. & Vargas., L.U. (1987). Uncertainly and Rank Ordering in the Analytic Hierarchy Process, *European Journal of perational Research*, Vol. 32, 1987.
- Saghebi, K (2010). The study of the relationship between the grade of diploma, the rank of university internship and the academic achievement of students at the University of Guilan. Master's dissertation, Psychology, Guilan University. (in Persian).
- 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.
- Wells, R.D. (1999). Factor scores and factor structure. *In Advances in social science methodology*, (Vol. 5, pp. 123-138). Greenwich, CT: JAI Press.
- 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.