



دانشگاه علامه طباطبائی
دانشکده‌ی روانشناسی و علوم تربیتی

پایان نامه جهت اخذ مدرک کارشناسی ارشد رشته سنجش و اندازه گیری

ارزشیابی آزمون کارشناسی ارشد خردانمای ریاسی با استفاده از مدلهای

جدید اندازه گیری و تعیین روکیری جدیدتی در آن

استاد راهنما

جناب دکتر علی دلاور

استاد مشاور

جناب دکتر محمد رضا فلسفی نژاد

پژوهشگر

مینا آسیابی

تابستان 91

6.....	فصل اول: کلیات پژوهش
7.....	مقدمه
9.....	بیان مساله
12.....	اهمیت و ضرورت پژوهش
13.....	اهداف پژوهش
14.....	سوالات پژوهش
14.....	تعاریف مفهومی
Error! Bookmark not defined.	فصل دوم: ادبیات پژوهش
18.....	انواع آزمونها
20.....	نظریه های آزمون سازی
21.....	تئوری کلاسیک آزمون (CTT)
21.....	نظریه سوال - پاسخ (IRT)
26.....	روشهای برآورد همزمان توانایی و پارامتر های سوال
29.....	روشهای مختلف کشف DIF
30.....	روشهای آماری برای آشکار سازی DIF
32.....	چارچوب کلی DIF
34.....	روش خی دو مانتل - هنسزل (MH)
35.....	روش های مجذور کای
36.....	روش رگرسیون لجستیک
41.....	روش نظریه سوال - پاسخ:
42.....	آزمونهای نسبت درستمایی برای بررسی کنش افتراقی سوال
44.....	جغرافیای سیاسی
45.....	روند تکوین جغرافیای سیاسی در ایران
46.....	خلاصه پژوهش های انجام گرفته
50.....	پیشینه روش رگرسیون لجستیک
Error! Bookmark not defined.	فصل سوم: روش پژوهش

53	روش شناسی تحقیق
53	جامعه و نمونه و روش نمونه گیری پژوهش
54	روش گرد آوری اطلاعات و داده ها
54	ابزارهای پژوهش
54	روش تجزیه و تحلیل داده ها
Error! Bookmark not defined. فصل چهارم: تجزیه و تحلیل اطلاعات	
Error! Bookmark not defined. مقدمه	
Error! Bookmark not defined. تجزیه و تحلیل آزمون زبان	
Error! Bookmark not defined. تجزیه و تحلیل آزمون جغرافیای سیاسی	
Error! Bookmark not defined. تجزیه و تحلیل آزمون مبانی جغرافیای انسانی	
Error! Bookmark not defined. تجزیه و تحلیل آزمون ژئو مورفولوژی	
Error! Bookmark not defined. فصل پنجم: بحث و نتیجه گیری	
114	مقدمه
114	خلاصه تحقیق
119	بحث و نتیجه گیری
126	تفسیر DIF
128	پیشنهادات تحقیق
129	محدودیت تحقیق
130	منابع فارسی
132	English reference
138	خروجی نرم افزار FACTOR برای تحلیل عاملی اکتشافی آزمون زبان انگلیسی تخصصی
139	دستور فرمان تحلیل رگرسیون لجستیک برای یک سوال آزمون در نرم افزار SPSS

- جدول 4-1: برآورد شاخص های کلاسیک آزمون زبان 61
- جدول 4-2: میزان پایایی بقیه سوالات بعد از حذف هر سوال (روش لوپ) درس زبان 62
- جدول 4-3: تحلیل گزینه های انحرافی در آزمون زبان (گزینه صحیح با علامت * مشخص شده است) 63
- جدول 4-4: نتایج تحلیل عاملی اکتشافی سوالات آزمون زبان 64
- نمودار 4-1: نمودار اسکری آزمون زبان 65
- جدول 4-5: برآورد پارامترهای سوالات آزمون زبان در مدل سه پارامتری 66
- جدول 4-6: برآورد پارامترهای سوالات آزمون زبان در مدل دو پارامتری 67
- جدول 4-7: برآورد پارامترهای سوالات آزمون زبان در مدل یک پارامتری 68
- جدول 4-8: مقدار 2-LOG LIKELIHOOD- برای هر یک از مدل های 3pl, 2pl, 1pl در آزمون زبان 69
- نمودار 4-2: منحنی های ویژگی همه سوالات آزمون زبان در مدل سه پارامتری 70
- نمودار 4-3: منحنی های ویژگی همه سوالات آزمون زبان در مدل دو پارامتری 70
- نمودار 4-4: منحنی های ویژگی همه سوالات آزمون زبان در مدل یک پارامتری 71
- نمودار 4-5: منحنی آگاهی آزمون و خطای استاندارد آزمون زبان 72
- نمودار 4-6: منحنی هیستوگرام توانایی آزمودنیها در درس زبان 72
- جدول 4-12: نتایج به دست آمده براساس روش رگرسیون لجستیک جهت بررسی سوگیری سوالات 73
- جدول 4-13: برآورد شاخص های کلاسیک آزمون جغرافیای سیاسی 74
- جدول 4-14: میزان پایایی بقیه سوالات بعد از حذف هر سوال (روش لوپ) درس جغرافیای سیاسی 75
- جدول 4-15: تحلیل گزینه های انحرافی در آزمون جغرافیای سیاسی (گزینه صحیح با علامت * مشخص شده است) 76
- جدول 4-16: نتایج تحلیل عاملی اکتشافی سوالات آزمون جغرافیای سیاسی 77
- نمودار 4-7: نمودار اسکری آزمون جغرافیای سیاسی 78
- جدول 4-17: برآورد پارامترهای سوالات آزمون جغرافیای سیاسی در مدل سه پارامتری 79
- جدول 4-18: برآورد پارامترهای سوالات آزمون جغرافیای سیاسی در مدل دو پارامتری 80
- جدول 4-19: برآورد پارامترهای سوالات آزمون جغرافیای سیاسی در مدل یک پارامتری 81
- جدول 4-20: مقدار 2-LOG LIKELIHOOD- برای هر یک از مدل های 3pl, 2pl, 1pl در آزمون جغرافیای سیاسی 82

چکیده

هدف پژوهش حاضر ارزشیابی آزمون کارشناسی ارشد رشته جغرافیای سیاسی با استفاده از مدل‌های جدید اندازه‌گیری و تعیین سوگیری جنسیتی در آن بوده است. جهت تجزیه و تحلیل آزمون و سوالات آن بر طبق نظریه کلاسیک اندازه‌گیری، ضریب دشواری، ضریب تمیز، روش لوپ، ضریب پایایی آزمون، واریانس سوالات و فراوانی گزینه‌های انحرافی محاسبه شد. جهت تجزیه و تحلیل آزمون و سوالات آن بر طبق نظریه IRT ابتدا مفروضه تک بعدی بودن خرده‌آزمون‌ها کلیه خرده‌آزمون‌های زبان، جغرافیای سیاسی، جغرافیای انسانی و ژئومورفولوژی با برنامه FACTOR مورد تحلیل عاملی قرار گرفتند. نتایج تجزیه و تحلیل آزمون‌های مذکور نشان داد که کلیه خرده‌آزمون‌ها تک بعدی هستند و در نتیجه در تمامی آزمون‌های مذکور مفروضه استقلال موضعی نیز برقرار است. سپس برازش مدل -داده‌ها از طریق نرم افزار BILOG_MG مورد بررسی قرار گرفت و در نهایت پارامترهای برآورد شده سوال (دشواری، تمیز و حدس) و توانایی آزمودنی‌ها استخراج گردید. نتایج مربوط به پارامترهای سوالات حاکی از دشوار بودن اغلب سوالات در تمام خرده‌آزمون‌ها بود و از طرفی ترتیب قرارگیری سوالات در دفترچه در 4 خرده‌آزمون نامناسب بود. برای محاسبه آماره شیوه رگرسیون لوجستیک سه گام پایایی با استفاده از نرم افزار SPSS انجام گرفت. نتایج محاسبات مربوط به سوگیری در میان سوالات نشان داد تنها در آزمون زبان یک سوال و در آزمون جغرافیای انسانی 5 سوال و در خرده‌آزمون ژئومورفولوژی 2 سوال دارای DIF یک‌نواخت معنادار شناخته شدند. تحلیل گزینه‌های انحرافی نشان می‌دهد که گزینه‌های انحرافی تمامی سوالات در همه آزمون‌ها هم احتمال نیستند و عملکرد معیوبی داشته‌اند. نتایج تجزیه و تحلیل حاکی از این است که آزمون زبان بهترین برازش را با مدل‌های 2 و 3 پارامتری دارد، خرده‌آزمون‌های جغرافیای سیاسی و جغرافیای انسانی با مدل دو پارامتری بهترین برازش را داشتند و خرده‌آزمون ژئومورفولوژی با مدل یک پارامتری بهترین برازش را داشت.

کلید واژه‌ها: سوگیری، رگرسیون لوجستیک، BILOG-MG, FACTOR, آزمون کارشناسی ارشد رشته جغرافیای دانشگاه آزاد اسلامی.

فصل اول

کلیات پژوهش

مقدمه

مرور تحولات اقتصادی کشورها نشان می‌دهد که در طول تاریخ عوامل مختلفی اهرم قدرت اقتصادی جوامع به شمار می‌رفته است. سالها زمین به عنوان مهم‌ترین عامل در اقتصاد کشاورزی جوامع و نیروی انسانی که باعث می‌شده خانواده‌های پر جمعیت اقتصاد قوی تری داشته باشند، بوده است. اما بدون شک وقوع انقلاب صنعتی و تحولات پیرامون آن نقطه عطف مهمی در این حوزه بوده است. به طوری که بعد از این انقلاب اگرچه همچنان نیروی کار عامل مهمی در اقتصاد کشورها به شمار می‌رفت. اما کم کم عامل نیروی تحصیل کرده و با صلاحیت به عنوان اهرم قدرت جای خود را در فعالیت‌های اقتصادی باز می‌یافت.

رشد شتابان توسعه علم و فناوری در دهه‌های اخیر باعث شده است که نقش دانش و فناوری در اقتصاد بیش از پیش پررنگ شود. و منافع و قدرت حاصل از تسلط کشورها بر حوزه دانش باعث شده که کشورهای با دانش و فناوری بالاتر به عنوان قدرت‌های اقتصادی دنیا شناخته شده و توانایی جذب سایر عوامل اقتصادی مانند سرمایه، منابع اولیه، نیروی کار ارزان و غیره را نیز داشته باشند.

از این رو یکی از دغدغه‌های جامعه امروز برای پیشرفت همه جانبه پرورش نیروی انسانی توانمند و با صلاحیت است. پرورش نیروی انسانی تحصیل کرده کاری دراز مدت و دنباله دار است. پذیرش دانشجو در سطح تحصیلات تکمیلی و تربیت این گروه فرایندی در این جهت است. در کشور ما این فرایند با گزینش افراد برای گذراندن دوره های تحصیلات تکمیلی آغاز می‌شود. از آنجایی که در ایران تاکنون تنها از طریق یک آزمون سراسری برای هر یک از دانشگاهها این گزینش تحقق یافته است، آزمونهای ورودی دانشگاههای کشور یکی از مهم ترین عناصر در گزینش و بعد از آن پرورش نیروی انسانی تحصیل کرده به حساب می‌آید.

بررسی و ارزشیابی این آزمون‌ها امری ضروری جهت هرچه بهتر شدن فرایند گزینش افراد جهت گذراندن دوره‌های تحصیلات عالی است که این مهم یکی از وظایف اصلی افرادی است که در زمینه سنجش و اندازه‌گیری فعال هستند .

آزمون‌ها می‌بایست از لحاظ گوناگون بررسی شوند یکی از جنبه‌های اصلی بررسی آزمون‌ها بحث سوداری آنها علیه زیر گروه‌های شرکت کننده در آزمون است . تحقیق در مورد آنچه امروزه سوگیری سوال نامیده می‌شود، به طور جدی از اوایل دهه اول قرن بیست یعنی دوران جنبش مدنی در آمریکا آغاز شده و نیروی محرکه خود را از جنبش حقوق زنان که پیش از آن پا گرفت دریافت کرد. وجه مشخصه این دوره اهمیت دادن به برابری فرصت‌ها و حمله به نظام‌هایی بود که پنداشته می‌شد میان انسان‌ها فرق می‌گذارند از نظام‌های تربیتی گرفته تا نظام‌های اشتغال و خدمات اجتماعی .

هدف از تحقیقات ابداع روش‌هایی برای مطالعه تفاوت‌های فرهنگی بود. و نیز بررسی این ادعا که دلیل اصلی (و شاید تنها دلیل) تفاوت بارز عملکرد دانش‌آموزان سیاه پوست و دانش‌آموزان اسپانیایی تبار و دانش‌آموزان سفید پوست در تست‌های توانش شناختی به خاطر وجود سؤال‌هایی است که خارج از حوزه فرهنگی گروه‌های اقلیت است. پیش فرض این بود که محتوای این قبیل سؤال‌ها چنان است که دانش‌آموزان اقلیت بخت و فرصت اندکی برای یادگیری دارند. بنابراین هدف شناسایی و خارج ساختن این قبیل سؤال‌ها از تست‌ها بود.

همزمان و موازی با چنین تحقیقاتی که تحقیق در مورد تست‌ها (به عنوان یک کل) آغاز شد. به خصوص تست‌هایی که برای گزینش افراد به کار می‌روند. تست‌هایی که بر علیه گروه‌های اقلیت سوگیری داشتند و این گروه‌ها نمره‌های پایین‌تری از این تست‌ها می‌آوردند.

آنچه این دو خط تحقیقاتی (گذشته از تفاوت هدفشان) و مدل‌ها و روش‌هایی را که برای انجام آنها اشباع شده بود از هم سوا می‌کرد. این بود که در تحقیقات نوع دوم ملاکی طبیعی و بیرونی و آشکار برای بررسی و برآورد سوگیری‌گزینش وجود داشت. در حالی که چنین ملاکی برای برآورد سوگیری در سؤال موجود نبود. به همین دلیل محققانی که سوگیری سؤال را بررسی می‌کردند به علت استفاده از کل تست به عنوان ملاک هم‌تاسازی گروه‌ها بر اساس توانایی متمایل گشتند. به زودی در هر دوی این رویکردها دانسته شد که مطالعه سوگیری در صورتیکه ملاک به هر دلیلی نابسند یا نامناسب باشد به بیراهه کشیده خواهد شد. بالاخص اگر ملاک، خود سوگیر باشد. مسلماً در چنین شرایطی امکان این وجود دارد که برخی از سؤال‌های سوگیر بدون سوگیری جلوه کنند و برخی از سؤال‌های بدون سوگیری نیز سوگیر شناخته شوند.

بیان مساله

در جهان امروز بسیاری از تصمیم‌گیری‌ها براساس نتایج تست‌ها صورت می‌پذیرد. در واقع از آزمون‌ها برای به دست آوردن نمونه‌ای از رفتار فرد برای نیل به اهداف طبقه‌بندی افراد، گزینش داوطلبان، ارزشیابی و تشخیص‌های بالینی و... استفاده می‌شود. در عین حال یکی از عمده‌نگرانی‌های فردی و اجتماعی در باب تاثیر تست‌های روانشناسی، استفاده ناعادلانه از آنها است. همواره این نگرانی وجود داشته است که مبادا تست‌ها فرصت‌های ناعادلانه و جانبدارانه‌ای به نفع گروه‌های قومی، نژادی، مذهبی یا طبقاتی خاص ایجاد کند. در بررسی‌های به عمل آمده، جنبه‌ها و صورتهای مختلفی از عادلانه یا ناعادلانه بودن تست‌ها گزارش شده است که از جمله آنها می‌توان به عادلانه بودن در مرحله استاندارد سازی، نمره گذاری و تصمیم‌گیری براساس نمرات تست اشاره کرد.

بدین ترتیب امکان وجود سوگیری در کاربرد تست‌ها مورد توجه قرار گرفته است. سوگیری می‌تواند منجر به اشتباهاتی در تصمیم‌گیری براساس نمرات تست، طبقه‌بندی و انتخاب دانش‌آموختگان شود

(زومبو¹، 1999). مهمترین هدف مطالعات سوالات سودار، ترویج و ارتقای سنجش عادلانه و در نظر گرفتن فرصتهای برابر برای همه آزمودنی ها و بالا بردن روایی آزمونها است. در این گونه مطالعات معمولاً سه هدف دنبال میشود:

- ارائه روشهایی برای حذف سوالات سودار آزمون سازان به صورتیکه در مرحله تست سازی قابل استفاده باشد.
- شناسایی روشهای آماری تعیین سوالات سودار علیه گروههای خاص.
- شناسایی متغیرها و عوامل بالقوه موثر در سوگیری سوالات علیه گروههای خاص جامعه.

میان روایی آزمون و سودار بودن سوالات آن رابطه تنگاتنگی وجود دارد به طوری که سوگیری یکی از جنبه های روایی آزمونها به حساب می آید. (زومبو، 1999).

در کشور ما نیز آزمونها به طور گسترده ای استفاده می شوند، که از جمله موارد استفاده آنها برگزاری امتحانات ورودی دانشگاهها در مقطع کارشناسی ارشد برای گزینش دانشجو می باشد. نتایج این آزمون نیز مانند سایر آزمونهای روانی و آموزشی می تواند تحت تاثیر قومیت یا جنسیت و یا سایر عواملی قرار گیرد که روی عادلانه بودن تست اثر می گذارند.

کارکرد افتراقی سوال، عملکرد متفاوت آزمودنی ها در سوالات، علیرغم توانایی های یکسان بین آنها را نشان می دهد. (جانسون 1999) احتمال وجود عملکرد متفاوت در سوالات آزمون ورودی کارشناسی ارشد نیز به دور از انتظار نیست و با کارکرد افتراقی سوال قابل تبیین است.

آنچه که در استفاده از هر تست آموزشی یا روانشناسی برای گزینش متقاضیان - که آزمون کارشناسی ارشد جز مهمترین این تستها در کشور ماست - بایستی مد نظر باشد، عادلانه بودن تست برای تمامی متقاضیان و

¹ zumbo

سودار نبودن آن علیه دسته ای از شرکت کنندگان است (زومبو، 1999). و روشی که تحلیل کارکرد افتراقی سوال خوانده می شود جزء روشهای استاندارد جدید تحلیل سوگیری سوال است. وظیفه خطیر محققان و برنامه ریزان آموزشی و کادر های گزینش (به خصوص مسولین سازمان سنجش) به عنوان استفاده کنندگان تست های آموزشی (مخصوصا کنکور های سراسری) آگاه بودن از جریان تفکر و تحلیل سوگیری تست می باشد.

باید توجه کرده باشید که عمده هدف استفاده از آزمونها در کنکور های ورودی دانشگاهها گزینش و انتخاب افراد برای استفاده از امکانات و موقعیت های آموزشی است. آنچه که در استفاده از هر تست آموزشی یا روانشناسی برای گزینش متقاضیان - که آزمون کارشناسی ارشد جز مهمترین این تست ها در کشور ماست - بایستی مد نظر قرار گیرد، معتبر بودن و مناسب بودن و همچنین عادلانه بودن تست برای تمامی متقاضیان و سودار نبودن آن علیه دسته ای از جامعه شرکت کنندگان است (زومبو، 1999).

وظیفه خطیر محققان و برنامه ریزان آموزشی و کادرهای گزینش به عنوان استفاده کننده تست های آموزشی آگاه بودن از ویژگی های آزمون های به کار برده شده و شیوه های سنجش این ویژگی ها و همچنین جریان تفکر و تحلیل سوگیری تست می باشد.

برای ارزیابی مساله سوگیری سوال میان گروههای مختلف آزمودنی می توان از رویکرد های متفاوت استفاده کرد. دو رویکرد عمده آماری و قضاوتی وجود دارد. از این میان شیوه های متنوع آماری درسنجش سوگیری پذیرفته تر و پر کاربردتر هستند. زیرا که به شکلی عینی به بررسی سوداری سوالات می پردازند. از میان شیوه آماری سنجش سوگیری برخی از این رویکردها بر نظریه کلاسیک آزمون و برخی دیگر بر نظریه سوال - پاسخ مبتنی هستند (آیرونسون¹، 1982). روش های مبتنی بر نظریه سوال - پاسخ در مقایسه با دیگر

¹ Ironson

روشهای دیگر در ارزیابی سوگیری سوال جدیدتر هستند البته یک شیوه آماری نوین نیز روش رگرسیون لوجستیک می باشد (زومبو، 1999).

در این پژوهش محقق سعی دارد با بررسی ویژگی های روانسنجی آزمون کارشناسی ارشد رشته جغرافیای سیاسی و نیز بررسی سوداری سوالات این آزمون به شیوه ی رگرسیون لوجستیک به طراحان آن در جهت رفع مشکلات آزمون و همچنین سایر طراحان و مجریان آزمونهای بزرگ و سرنوشت ساز یاری دهد.

اهمیت و ضرورت پژوهش

یکی از حوزه های چالش برانگیز آموزش و پرورش، سنجش و اندازه گیری توانایی ها و صفات افراد متعلق به گروههای متفاوت است که مورد سنجش قرار میگیرند . وقتی که براساس نتایج آزمون ها تصمیمات مهمی چون جایگزینی افراد آزمودنی در پایه های مختلف تحصیلی و ارتقای آنها به پایه ها و مقاطع بالاتر، استخدام در سازمان ها و موسسات، ارائه خدمات مشاوره ای، تجویز دارو گرفته می شود، ضرورت دارد این آزمون ها مطابق قواعد و اصول روانسنجی تدوین شوند .

امروزه با توجه به اهمیتی که به آموزش و نظام آموزشی به ویژه در مقاطع بالاتر تحصیلی داده می شود، میزان کمتر ترک تحصیل افراد پس از دوره دیپلم، حضور زنان در رقابت های ملی، افزایش درصد قبولی زنان در کنکور، می توان به این نتیجه رسید که انگیزه اصلی جوانان در مسیرهای تحصیلی و آموزشی است . بنابراین عادلانه بودن آزمون های سراسری برای گروه های وسیعی از جامعه حایز اهمیت روزافزون است.

از طرف دیگر از آنجایی که آزمون این رشته در سال 89 با یک تعویق یک ساله و به دلیل بازنگری در برنامه و اصلاح رو به رو شد و سپس از طرف سازمان سنجش اعلام شد که در سال 90 با یک سری اصلاحات در کنکور سراسری دانشگاه آزاد برگزار می شود و تا به حال تحقیق مبسوطی راجع به سوالات طرح شده و اجرا شده در این حیطة صورت نگرفته است . به نظر می رسد ارزشیابی دقیق سوالات این مجموعه امری بسیار

مهم باشد. مسلماً نتایج حاصل از پژوهش حاضر برای دست اندر کاران طراحی و ساخت سوالات این رشته و همچنین مجریان و داوطلبان شرکت کننده در کنکور رشته جغرافیای سیاسی بسیار با اهمیت خواهد بود.

از طرفی دیگر نتایج حاصل از پژوهش های انجام گرفته در موضوع سوگیری آزمونهای سراسری در کشور مورد استفاده دست اندر کاران تهیه این گونه آزمونها و افراد ذی نفع در این آزمونها و همچنین گامی جهت گسترش دانش محققین در این زمینه خواهد بود. بعلاوه از انجایی که در این پژوهش سعی شده تا شیوه رگرسیون لوجستیک در سنجش سوگیری به نحوی مبسوط شرح داده شده و به کار گرفته شود.

اهداف پژوهش

هدف اصلی پژوهش عبارت است از: ارزشیابی و همچنین تجزیه و تحلیل سوالات کنکور دانشگاه آزاد در رشته جغرافیای سیاسی و در مقطع کارشناسی ارشد از تمامی جهات و بی نقص کردن هرچه بیشتر سوالات آن به منظور شناسایی سوالاتی که علیه گروههای جنسیتی سوگیر هستند. به علاوه معرفی هرچه بیشتر روش رگرسیون لوجستیک و نظریه سوال - پاسخ و کاربرد های آن در حوزه سنجش و اندازه گیری و ارائه پیشنهادات به سازمان ها و مراکزی که مسولیت تهیه و اجرای آزمون ها را برعهده دارند، از اهداف دیگر پژوهش است.

اهداف فرعی

1. بررسی ویژگی های سوالات کنکور کارشناسی ارشد رشته جغرافیای سیاسی سال 90 در دانشگاه آزاد اسلامی بر طبق نظریه کلاسیک
2. بررسی ویژگی های سوالات کنکور کارشناسی ارشد رشته جغرافیای سیاسی سال 90 در دانشگاه آزاد اسلامی بر طبق نظریه IRT
3. بررسی برازش مدل های IRT با مجموعه داده های آزمون

4. شناسایی سوالاتی که برای دختران دارای کارکرد افتراقی هستند با استفاده از روش رگرسیون

لوجستیک

سوالات پژوهشی

1. ویژگی های سوالات کنکور کارشناسی ارشد رشته جغرافیای سیاسی سال 90 در دانشگاه آزاد

اسلامی برطبق نظریه کلاسیک چگونه است؟

2. ویژگی های سوالات کنکور کارشناسی ارشد رشته جغرافیای سیاسی سال 90 در دانشگاه آزاد

اسلامی برطبق نظریه IRT چگونه است؟

3. کدامیک از مدل های IRT بهترین برازش را با این مجموعه از داده ها دارد؟

4. کدامیک از سوالات کنکور کارشناسی ارشد رشته جغرافیای سیاسی بر طبق شیوه رگرسیون

لوجستیک دچار DIF است؟

تعاریف مفهومی

نظریه کلاسیک آزمون: این نظریه، نمره آزمون را به عنوان نمره ای در نظر می گیرد که دارای دو مولفه یا

دو جزء اصلی جمع پذیر نمره حقیقی و خطا یا اشتباه تصادفی است (ثرنداک، 1982 ترجمه هومن، 1369)

نظریه سوال - پاسخ: این نظریه به کاربرد نظریه یا الگوی صفت مکنون گفته میشود. در این نظریه به جای

تاکید بر نمرات کل آزمون، برپاسخ های آزمودنی ها به تک تک سوالات آزمون تکیه می شود (سیف، 1381).

پارامتر دشواری سوال: مطابق نظریه سوال - پاسخ پارامتر دشواری سوال نقطه ای است روی مقیاس توانایی

که در آن نقطه احتمال پاسخ دهی درست به سوال برای آزمودنی برابر 0.50 باشد. (بریس¹، 2000)

¹ Bryce

پارامتر تمییز سوال: عبارت است از قدرت سوال در جداسازی آزمودنی هایی که توانایشان پایین تر از دشواری سوال است از آزمودنی هایی که توانایشان بالاتر از دشواری سوال است، این ویژگی اساساً بیانگر شیب منحنی ویژگی سوال در وسط منحنی است هر چه شیب منحنی بیشتر باشد قدرت تمییز سوال بالاتر است (بیکر¹، 2001).

پارامتر حدس سوال: عبارت است از پاسخ دهی درست به سوال برای آزمودنی هایی که از لحاظ توانایی مورد سنجش آزمون در نقطه صفر قرار دارند اما به سوال پاسخ درست می دهند (شیپارد²، کامیلی³ و ویلیامز⁴، 1985).

توانایی: ویژگی یا صفتی است که توسط آزمون اندازه گرفته می شود (بریس، 2000).

منحنی ویژگی سوال: خطی است که رابطه بین سطح توانایی آزمودنی و احتمال پاسخ دهی درست به سوال را برای او توصیف می کند (بریس، 2000).

تابع آگاهی سوال: شاخصی است که سهم هر سوال را در اندازه گیری صفت مورد نظر آزمون نشان می دهد (بیکر، 2001).

تابع آگاهی آزمون: شاخصی است که نشان می دهد آزمون تا چه اندازه توانسته است توانایی مورد سنجش آزمودنی را اندازه بگیرد (بیکر، 2001).

خرده آزمون: آزمون ورودی دانشگاهها در هر یک از گروههای آزمایشی از چندین درس مختلف مانند ادبیات فارسی، ریاضی، شیمی.... تشکیل می شود که به هر یک از این درس ها در اصطلاح خرده آزمون گفته

¹ Baker
² Shepard
³ Camilli
⁴ Williams

می شود. در این پژوهش منظور از خرده آزمون، سوالات مربوط به درس های زبان انگلیسی تخصصی، جغرافیای سیاسی، مبانی جغرافیای انسانی و ژئومورفولوژی است.

سوگیری سوال: سؤالی دارای سوگیری است که افراد با توانایی های مساوی اما از گروه های مختلف، شانس یکسانی برای پاسخ صحیح به آن سؤال نداشته باشند. (آیرونسون¹، 1982)

¹ Iranson

فصل دوم

ادبیات پژوهش

آزمون

آزمون یک روش نظام‌دار برای سنجش نمونه‌ای از رفتار و یک وسیله اندازه‌گیری در روانشناسی و تعلیم و تربیت است (لطف آبادی، 1375). در تعریف آزمون یا تست آن را ابزاری عینی و استاندارد شده برای اندازه‌گیری نمونه‌ای از رفتار یا ویژگی‌های انسان در نظر می‌گیرند. منظور از عینی بودن این است که در تفسیر و نمره دهی نتایج آزمون قضاوت و نظر شخصی آزماینده یا ممتحن دخالت ندارد.

انواع آزمونها

الف) آزمونهای وابسته به هنجار

آزمونهای هنجاری حوزه وسیعی از دانش را اندازه می‌گیرند و وضعیت پیشرفت عمومی یک فرد معین را در رابطه با گروههای مختلف یا برای مقیاسه عملکرد یک فرد با یک گروه معین به کار گرفته می‌شود (لطف آبادی، 1375). در آزمونهای گروه مرجع یا نرم مرجع عملکرد هر آزمودنی با متوسط عملکرد افراد گروه نرم، مثلا گروه همسالان آزمودنی مقایسه و تفسیر می‌شود (شریفی، 1383).

ارزشیابی وابسته به هنجار به جای یک ملاک مطلق از پیش تعیین شده به نوعی ملاک نسبی وابسته است. علت نام گذاری این نوع ارزشیابی به ارزشیابی هنجاری این است که در آن ملاک یا معیار ارزشیابی نوعی هنجار گروهی است. (سیف، 1384). در واقع آزمون نرم مرجع آزمونی است که ملاک موفقیت در آزمون پس از اجرای آزمون مشخص می‌شود. آزمون کنکور نمونه روشنی از این نوع آزمون است (یونسی، 1381). از آزمونهای هنجاری در طبقه بندی کردن و تصمیمات مشاوره‌ای و ارزیابی گروههای دانش آموزان، برنامه‌ها و فعالیت‌های پژوهشی، اندازه‌گیری تفاوت‌های فردی در استعداد‌های شناختی و خصایص عاطفی و شخصیتی استفاده می‌شود (لطف آبادی، 1375؛ شریفی، 1383).

در تفسیر آزمونهای هنجاری، گروهی که نمره فرد با آن مقایسه می‌شود گروه هنجار نامیده می‌شود و آمار جمع‌آوری شده درباره این گروه را که معمولا شامل اندازه‌گیری‌های متوسط عملکرد آن است هنجار می‌نامند. در هنجارهای مورد استفاده باید معیارهایی بکار گرفته شود که مقبولیت علمی آنها را تضمین کند.

مهمترین این معیارها عبارتند از معرف بودن¹، مناسب بودن² و جدید بودن³ (لطف آبادی، 1375).

تجزیه و تحلیل آزمونهای هنجاری

ویژگی های سوالات آزمونهای هنجاری از طریق محاسبات آماری تعیین می شود. برای تعیین وجود هماهنگی بین عملکرد در یک سوال با عملکرد در کل آزمون از محاسبه همبستگی استفاده می شود. تحلیل آزمون به وسیله دو شاخص (دشواری و تمیز) انجام می شود.

معمولا در اکثر آزمونهای پیشرفت، سوالاتی که شاخص دشواری آنها 0.5 یعنی در حدود پنجاه درصد باشد، مطلوب به حساب می آیند. شاخص تمیز را با محاسبه همبستگی بین پاسخ صحیح به یک سوال و نمره کل آزمون (معمولا ضریب همبستگی دو رشته ای نقطه ای) به دست می آورند (لطف آبادی، 1375).

ب) آزمونهای ملاکی

اصطلاح اندازه گیری ملاک مرجع برای اولین بار توسط کلاوس⁴ (1962) در توصیف ارزیابی عملکرد انسان در سیستم های ماشینی به کار برده شد. اگرچه از لحاظ مفهومی در کارهای ایبل⁵ (1962)، فلانگال⁶ (1951) و ثرنداک⁷ (1913) مشخص شده بود.

بعدها نیز توسط گلاسر⁸ و کلاوس (1962) مفهوم اندازه گیری ملاک مرجع در اندازه گیری پیشرفت بکار برده شد. تعاریف زیادی از آزمونهای ملاک مرجع شده است اما شکل اصلاح شده تعریف پوفام، توسط همبلتون⁹ (1981) ارائه شده است که به صورت گسترده مورد استفاده قرار می گیرد. بر اساس این تعریف، آزمونهای ملاک مرجع آزمونهایی هستند که ارزیابی از سطوح عملکرد امتحانات در رابطه با مجموعه ای از اهداف روشن و از قبل تعیین شده ارائه می دهند (همبلتون، 1980).

¹ Representativeness

² Relevance

³ Recency

⁴ Klaus

⁵ Ebel, r.l

⁶ Flanagan

⁷ Thorndike

⁸ Glaser

⁹ Hambleton

نظریه های آزمون سازی

آن دسته از نظریه هایی که در روانشناسی برای ساخت و تحلیل آزمونهای روانی و به طور کلی اندازه گیری پدیده های روان شناختی مورد استفاده قرار می گیرند، نظریه های آزمون سازی نامیده می شوند. نظریه آزمون را می توان، نمایش نمادی عوامل موثر بر نمره های مشاهده شده دانست که با به کار گیری مدل های ریاضی و با استفاده از مفروضات آن نظریه توصیف می شود.

نظریه های آزمون سازی در روانسنجی، در واقع در خدمت فراهم کردن مبانی و اصولی هستند که از طریق آنها بتوان به پیش بینی، برآورد یا استنباط راجع به توانایی یا صفت مورد سنجش آزمون پرداخت (همبلتون، 1989) بدیهی است زمانی این پیش بینی و برآورد صفت مورد اندازه گیری از دقت مطلوب برخوردار خواهد بود که منابع ایجاد کننده تغییر در آن صفت، تا حد قابل قبولی شناخته شده باشند.

یکی از مهم ترین موضوعات مورد بررسی در زمینه نظریه های اندازه گیری و مدل های مربوط به آنها، سنجش خطا های اندازه گیری و یا به عبارت دیگر بررسی دقت اندازه گیری است. بدیهی است که از یک نظریه و مدل مناسب، انتظار می رود که بتواند دقت قابل قبولی در اندازه گیری پدیده مورد نظر داشته باشد. یک نظریه یا مدل اندازه گیری خوب در یک آزمون سنجش توانایی با مشخص کردن رابطه بین سوالات آزمون و سطح توانایی، چهارچوبی مرجع جهت طراحی آزمون مناسب و حل مشکلات عملی مربوط به آن فراهم می سازد.

همبلتون و ون در لیدن¹ (1982)، تئوری آزمون را به دو دسته عمده تئوری کلاسیک اندازه گیری و

تئوری سوال - پاسخ تقسیم نمودند که به اختصار، به معرفی هر یک خواهیم پرداخت:

¹ Van der Linden