



دانشگاه علامه طباطبائی

دانشکده روانشناسی و علوم تربیتی

رساله جهت دریافت دکترای رشته سنجش و اندازه گیری

موضوع:

مقایسه قدرت آزمون نسبت در ستمنمایی مبتنی بر مدل پرسش-پاسخ (IRT) با روشهای تحلیل

عاملی تأییدی و رگرسیون لوجستیک در شناسایی کنش افتراقی سوالات (DIF) به منظور

اطمینان از عادلانه بودن سنجش در آزمون های سرنوشت ساز

استاد راهنما:

دکتر محمد رضا فلسفی نژاد

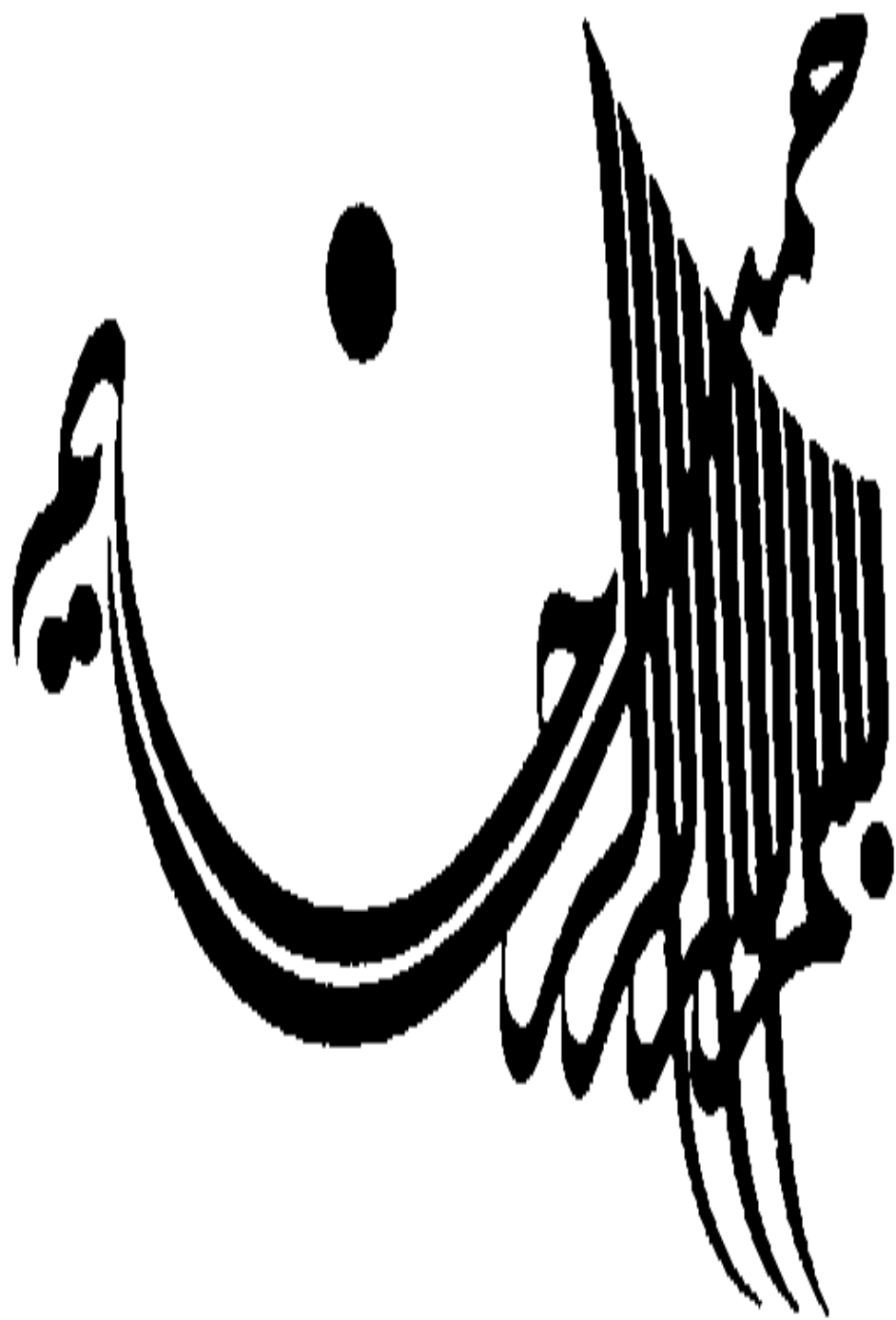
اساتید مشاور:

دکتر علی دلاور

دکتر نورعلی فرخی

پژوهشگر:

مسعود گرامی پور



تفکر و سیاسی خالصانه خود را تقدیم به استاد ارجمند و فرهیخته
جناب آقای دکتر محمد رضا فلسفی نزلو می نمایم که در هدایت این جانب و حل مشکلات ،
بسیار صبور و با درایت بودند. از دانشمند فرزانه و استاد ارجمند جناب آقای دکتر ولاد که
با مشاورت ارزنده و نظرات سازنده منطبق بر انجام این مهم بوده اند ، سپاسگزارم.
امید آنکه کشور ماندگاری این جهره ی ارزشمند علمی تعالی بخش جامعه ی دانشگاهی و آموزشی
عالی کشور ما باشد و نیز تفکر و سیاسی خود را تقدیم به استاد فرزانه
جناب آقای دکتر نور علمی فرخمی می نمایم که از راهنماییها و درایت استاد در این پژوهش
برخوردار بودم.

– چکیده

اگر چه روشهای متعددی جهت شناسایی سؤالات سودار معرفی شده است، اما تحقیقات اندکی بصورت تجربی قدرت و کارایی هر یک از این روشها را مورد بررسی قرار داده اند. هدف از تحقیق حاضر مقایسه آزمون نسبت درستنمایی مبتنی بر مدل پرسش-پاسخ (IRT)، روش تحلیل عاملی تأییدی (CFA) و روش رگرسیون لوجستیک در شناسایی کنش افتراقی سؤالات آزمون های سرنوشت ساز است. برای پاسخگویی به سؤالات تحقیق از روش مطالعات شبیه سازی مونت کارلو استفاده شد. داده های مورد نیاز با استفاده از نرم افزار WINGEN² در قالب ۱۰۰ آزمون ۳۰ سؤالی که با مدل دو پارامتری برازش داشتند شبیه سازی شد. توزیع دشواری و قدرت تمیز سؤالات همه آزمون ها، نرمال بود. اما آزمون ها از حیث نوع DIF، میزان DIF و همچنین پارامترهای سؤال با هم متفاوت بودند. برای هر آزمون، پاسخهای ۱۰۰۰ آزمودنی با توزیع توانایی (θ) نرمال شبیه سازی، و از روش برآورد بیشینه درستنمایی حاشیه ای و روش برآورد حداقل مربعات وزنی در تعیین نوع و میزان DIF استفاده شد. تجزیه و تحلیل داده ها در تکرارهای متوالی نشان داد که روش آزمون نسبت درستنمایی مبتنی بر IRT در شناسایی DIF نسبت به روشهای CFA و رگرسیون لوجستیک برتری دارد. این برتری در تمامی مقادیر DIF (کم، متوسط و زیاد) مشاهده شد. با این همه، تفاوت سه روش در حجمهای نمونه ۱۰۰۰ آزمودنی، جزئی و قابل اغماض است. همچنین میان سه روش از نظر شناسایی نوع DIF تفاوتی دیده نشد.

کلید واژه ها: کنش افتراقی سؤال-آزمون نسبت درستنمایی – تحلیل عاملی تأییدی – رگرسیون لوجستیک – آزمون های سرنوشت ساز

فهرست مطالب

فصل اول: کلیات تحقیق..... ۱

- ۱-۱- مقدمه ۲
- ۱-۲- بیان مسأله..... ۳
- ۱-۳- اهداف تحقیق ۸
- ۱-۴- سؤال های تحقیق ۱۰
- ۱-۵- فرضیه های تحقیق ۱۰
- ۱-۶- تعریف مفاهیم تحقیق..... ۱۲
- ۱-۶-۱- تعریف نظری ۱۲
- ۱-۶-۲- تعریف عملیاتی ۱۳

فصل دوم: پیشینه تحقیق..... ۱۵

۱-۲- مقدمه ۱۶

- ۱-۱-۲- سوگیری سؤال و رویکردهای کلی بررسی کنش افتراقی ۱۷
- ۱-۲-۱-۲- آزمون های سرنوشت ساز ۲۰
- ۱-۲-۲- سوگیری در اندازه گیری..... ۲۲
- ۱-۲-۲-۱- جامعه های چندگانه ۲۴
- ۱-۲-۲-۲- مفهوم آماری سوگیری سؤال..... ۲۵
- ۱-۲-۲-۳- تک بعدی بودن و تغییر ناپذیری..... ۲۷
- ۱-۲-۳- تحلیل عاملی تأییدی ۲۸
- ۱-۳-۲- اعمال محدودیت های تغییر ناپذیری ۳۱
- ۱-۳-۲- آزمون های تغییر ناپذیری..... ۳۲
- ۱-۴-۲- نظریه پرسش-پاسخ ۳۴
- ۱-۴-۲-۱- مدل دو پارامتری..... ۳۶
- ۱-۴-۲-۲- مدل پاسخ درجه بندی شده ۳۷
- ۱-۴-۲-۳- آزمون های نسبت درستی برای بررسی کنش افتراقی سؤال..... ۳۷
- ۱-۵-۲- رگرسیون لوجستیک و کنش افتراقی سؤال ۳۹

۲-۶- مطالعات شبیه سازی داده

ها.....۴۳

۲-۷- رابطه میان روشهای بررسی کنش افتراقی سؤال مبتنی بر IRT ، CFA و رگرسیون لجستیک۴۶

۲-۸- پیشینه مطالعاتی روشهای بررسی سوگیری سؤالات آزمون۴۷

فصل سوم: روش تحقیق۵۱

۳-۱- روش تحقیق۵۲

۳-۲- ابزار تحقیق ، شیوه شبیه سازی داده های پرسش- پاسخ و پایایی۵۴

۳-۳- جامعه آماری، گروه نمونه و شیوه گزینش آن۵۶

۳-۴- روش اجرایی۵۶

۳-۵- روش تجزیه و تحلیل داده ها۵۷

فصل چهارم: تجزیه و تحلیل داده های آماری۶۱

۴-۱- مقدمه۶۲

۴-۲- یافته های تحقیق۶۲

۴-۳- آزمون فرضیه های تحقیق۹۶

فصل پنجم: بحث و نتیجه گیری۱۱۰

۵-۱- مقدمه۱۱۱

۵-۲- بحث۱۱۲

۵-۳- نتیجه گیری۱۳۱

۵-۴- محدودیتهای تحقیق۱۳۴

۵-۵- پیشنهاد های تحقیق۱۳۵

- منابع فارسی۱۳۸

- منابع انگلیسی۱۳۸

- پیوست ها

فصل اول: کلیات تحقیق

آزمون های سرنوشت ساز^۱ آزمون هایی هستند که در مقیاس وسیع اجرا شده و بر اساس نتایج آنها تصمیماتی خطیر در مورد افراد اتخاذ می شود. چنین تصمیماتی دارای تبعات مهم شخصی، اجتماعی و سیاسی هستند. لذا، اعتبار تصمیم گیری و کارایی مدل های سنجش در این مورد از اهمیت فوق العاده ای برخوردار بوده و در سالیان گذشته جزء مهمترین چالش های روان سنجی باقی مانده اند. یکی از مهمترین تهدیدها برای اعتبار^۲ آزمون، سوگیری سؤال^۳ یا تغییر ناپذیری اندازه گیری^۴ است که از آن تحت عنوان کنش افتراقی سؤال (DIF)^۵ نیز یاد می شود. کنش افتراقی سؤال شرط لازم ولی ناکافی برای سوگیری سؤال است (رامسی^۶، ۱۹۹۳). هم ارزی^۷ اندازه گیری ویژگی ای است که در آن سؤالهای آزمون در شرایط مختلف ویژگی های یکسانی را اندازه گیری می کنند (هورن^۸ و مک آردل^۹، ۱۹۹۲). این شرایط گروهها، مراکز اجرایی و روشهای مختلف آزمون (مداد-کاغذی یا رایانه ای) را شامل می شوند. یک مقیاس موقعی دارای تغییر پذیری اندازه گیری است که در آن بیش از یک گروه نمرات یکسانی در یک سازه پنهان داشته باشند و نمرات خام مشاهده شده و مورد انتظار آنان در گروههای مختلف در سطح سؤال و آزمون یکسان باشد (دراسگو و کانفر^{۱۰}، ۱۹۸۵).

روانشناسان اهمیت تغییر پذیری اندازه گیری را به عنوان پیش نیازی برای مقایسه های گروهی گوشزد کرده اند (بطور مثال دراسگو، ۱۹۸۴؛ راجو^{۱۱} و همکاران، ۲۰۰۲؛ ریس^{۱۲} و همکاران، ۱۹۹۳؛ وندربرجک^{۱۳}، ۲۰۰۲). هر چقدر که آزمون کارکرد متفاوتی برای گروههای مختلف داشته باشد، مقایسه های گروهی نیز غیر قابل اعتماد خواهد بود. بنابراین فقدان تغییر پذیری اندازه گیری و کنش افتراقی سؤال بعنوان تهدیدی جدی برای اعتبار آزمون تلقی می شود، بطوریکه استانداردهای انجمن روانشناسی آمریکا (APA)^{۱۴} به بررسی DIF سؤالهای آزمون برای ارزیابی منصفانه بودن آن تأکید دارد (برن^{۱۵} و استوارت^۱، ۲۰۰۶).

^۱ - high stakes tests

^۲ - validity

^۳ - Item Bias

^۴ - measurement invariance

^۵ - Differential Item Functioning

^۶ - Ramsey

^۷ - equivalence

^۸ - Horn

^۹ - McArdle

^{۱۰} - Kanfer

^{۱۱} - Raju

^{۱۲} - Reise

^{۱۳} - Vanderberg

^{۱۴} - American Psychological Association

^{۱۵} - Byrne

روشهای مختلفی برای بررسی DIF سؤالهای آزمون معرفی و بکار گرفته شده اند، اما علی رغم ریشه دار بودن موضوع و فعالیت های زیادی که در این مورد صورت پذیرفته است هنوز در مورد روش های مطلوب و بهینه تعیین DIF میان صاحب نظران اتفاق نظر وجود ندارد (میلسپ، ۲۰۱۱).

۱-۲- بیان مسأله

شاید مهمترین اتهام وارده به سنجش و مطمئناً اولی ترین مسئله برای مردم، سنجش عادلانه^۱ است. دعوی این که آزمون ها در برابر اقلیت های عمومی و نژادی سوگیری دارند، منجر به لوایح قانونی متعدد شده است. آزمون ها و حرفه آزمونگری زیر تیزبینی مردم قرار گرفته، و اکنون منتشر کنندگان و بکار برندگان آزمون ها باید نشان دهند که آزمون هایشان در برابر اقلیتها، عاری از سوگیری هستند (همبلتون^۲ و همکاران، ۱۹۹۱ به ترجمه فلسفی نژاد، ۱۳۸۹).

از طرفی عادلانه بودن برای آزمون های سرنوشت ساز بعنوان آزمون هایی که جهت تصمیم گیری برای آینده آزمودنی ها بکار می روند از اهمیت بیشتری برخوردار است و حساسیت بیشتری نسبت به عادلانه بودن سؤالهای این نوع آزمون ها نشان داده می شود (پاپهام^۳، ۲۰۰۵).

اصطلاح کنش افتراقی سؤال (DIF) بعنوان گواهی تجربی برای وجود یا عدم وجود سوگیری در سؤال مورد استفاده قرار می گیرد. یک سؤال موقعی دارای DIF است که افرادی با توانایی یکسان اما با گروهی متفاوت، احتمال یا بختی متفاوت در دادن پاسخ درست به سؤال داشته باشند. این گروهها در ادبیات DIF اصطلاحاً گروههای مرجع و کانونی^۴ نامیده می شوند. معمولاً گروه مرجع، گروه اکثریت و گروه کانونی گروه اقلیت یا محروم در نظر گرفته می شوند (سوامیناتان^۵ و راجرز^۶، ۱۹۹۰). با کنترل توانایی در دو گروه اگر آزمودنی ها احتمال یا بخت متفاوت در پاسخ درست به یک سؤال داشته باشند آن سؤال دارای کنش افتراقی است. باید توجه داشت که تفاوت موقعی معنی دار است که توانایی مورد نظر در دو گروه کنترل شده باشد. این بدین معنی است که تفاوت در عملکرد بنهایی نشانه سوگیری سؤال نیست. ممکن است در بعضی مواقع آزمودنی ها در گروه های متفاوت واقعاً در توانایی مورد نظر تفاوت داشته باشند. تعیین سؤالهای سوگیرانه گامی مهم در

^۱ - Stewart

^۲ - test fairness

^۳ - Hambleton

^۴ - Popham

^۵ -reference and focal groups

^۶ - Swaminathan

^۷ - Rogers

سنجش معتبر است. یک سؤال سوگیرانه دارای خطای نظامدار است که باعث نامعتبر شدن نتایج آزمون می شود (کمیلی^۱ و شپارد^۲، ۱۹۹۴). روشهای آماری مختلفی جهت تعیین DIF وجود دارد که می توان آنها را به دو دسته کلی روشهای مبتنی بر نظریه پرسش-پاسخ (IRT)^۳ و روشهای غیر مبتنی بر نظریه پرسش-پاسخ تقسیم کرد. روشهای موجود در گروه اول شامل روشهایی برای مقایسه پارامترهای سؤال و مساحت میان منحنی های ویژگی سؤال (ICC)^۴ است (همبلتون^۵ و همکاران، ۱۹۸۵). روشهایی که شامل گروه دوم می شوند مبتنی بر نظریه نظریه اندازه گیری کلاسیک (CTT)^۶، روشهای تحلیل عاملی و روشهای مبتنی بر تحلیل داده های طبقه ای هستند (گومز^۷ و ناواس^۸، ۲۰۰۰).

تیشن^۹، استینبرگ^{۱۰} و جرارد^{۱۱} (۱۹۸۶) و تیشن، استینبرگ و وینر^{۱۲} (۱۹۹۳) آزمون نسبت درستنمایی را مطرح کرده اند که مبتنی بر نظریه IRT است. در این روش، برازش مدل افزایشی^{۱۳} A که در آن پارامترهای IRT سؤال مورد مطالعه اجازه دارند در گروههای مورد مقایسه تغییر کنند با برازش مدل فشرده^{۱۴} C که در آن پارامترهای سؤال در گروهها محدود می شوند، مقایسه می شود. مقدار درستنمایی L در مدل فشرده در یک نوبت برآورد محاسبه می شود. در محاسبات بعدی، درستنمایی مدل افزایشی A محاسبه می شود که در آن از ترکیب پارامترهای ضریب تمیز (a)، ضریب دشواری (b) و حدس (c) برای یک سؤال در گروه کانونی صرفنظر می شود. در این روش سؤال مورد مطالعه در هر یک از گروهها متفاوت در نظر گرفته می شود. فرض صفر (مدل C: H₀ عدم DIF) به نفع فرض خلاف (مدل A: H_a) با استفاده از آزمون نسبت درستنمایی رد می شود. کیم^{۱۵} و همکاران (۱۹۹۶) در یک مطالعه شبیه سازی شده مونت کارلو برای آزمون های سرنوشت ساز کارکرد روش مبتنی بر IRT نسبت درست نمایی را در خطای نوع اول در DIF مورد بررسی قرار دادند. نتایج

^۱ - Camilli
^۲ - Shepard
^۳ - Item Response Theory
^۴ - Item Characteristic Curve
^۵ - Hambleton
^۶ - Classical Test Theory
^۷ - Gomez
^۸ - Navas
^۹ - Thissen
^{۱۰} - Steinberg
^{۱۱} - Gerrard
^{۱۲} - Wainer
^{۱۳} - augmented
^{۱۴} - compact
^{۱۵} - Kim

نشان داد که نرخ خطای نوع اول برای مدل دو پارامتری مطابق با نظریه های موجود است اما برای مدل ۳ پارامتری نتایج کاملاً با نظریات موجود تفاوت معنی داری دارد.

روش بعدی مورد مطالعه در تحقیق حاضر که بر اساس نظریه پرسش پاسخ نیست روش تحلیل عاملی تأییدی است. در این روش همسانی میان بارهای عاملی و مقادیر ثابت مدل بصورت مستقیم آزمون می شود. ابتدا باید خط پایه مدل تعیین شود، جایی که همه پارامترها آزادانه می توانند تغییر کنند، به جز آنهایی که برای تعیین مدل ثابت نگه داشته می شوند. مرحله بعد تعیین مجموعه ای از مدل‌های محدود شده است که در آن بارهای عاملی یا مقادیر ثابت توسط اعمال محدودیتهایی در گروه‌های مرجع و کانونی مساوی در نظر گرفته می شوند و سپس مدل‌های محدود شده با مدل‌های خط پایه با استفاده از آماره خی دو مقایسه می شوند. به دلیل اینکه هر کدام از مدل‌های محدود شده در درون مدل‌های خط پایه قرار گرفته اند می توان تغییر در مقدار خی دو را با توجه به خط پایه محاسبه کرد. برای هر مقایسه یک مقدار معنی دار از نظر آماری گواهی بر کنش افتراقی سؤال است (سوربوم^۱، ۱۹۷۴). در یک مطالعه شبیه سازی برای آزمون های سرنوشت ساز توسط مید^۲ و لوچنلاجر^۳ (۲۰۰۴) نتایج مقایسه روش نسبت درستنمایی مبتنی بر IRT و روش تحلیل عاملی تأییدی نشان می دهد که روش CFA نسبت به روش نسبت درستنمایی IRT در شناسایی کنش افتراقی سؤالهایی که در پارامتر دشواری (b) و تشخیص (a) در گروه‌های مرجع و کانونی متفاوتند ضعیف تر عمل می کند ولی روش IRT در نمونه های کمتر از ۱۵۰ آزمودنی در مقایسه با روش CFA در تشخیص درست کنش افتراقی سؤال ضعیفتر عمل می کند. سوامیناتان^۴ و راجرز^۵ (۱۹۹۰) روش کنش افتراقی سؤال بر اساس رگرسیون لجستیک را معرفی کرده اند. رگرسیون لجستیک به خانواده بزرگتری از تحلیل ها تعلق دارد که مدل‌های خطی کلی نامیده می شوند (اگرستی^۶، ۲۰۰۷). در این رویکرد مجموع نمرات یا نمرات صحیح بعنوان متغیری پیوسته و مستقل بعنوان توانایی در نظر گرفته می شود. در رگرسیون لجستیک پاسخ سؤال بعنوان متغیر وابسته با توزیع برنولی دو وجهی در نظر گرفته می شود و یک متغیر گروهی دیگر بعنوان متغیری که قصد بررسی پاسخ ها در آن را داریم بعنوان گروه‌های مرجع و کانونی در نظر گرفته می شوند که معنی داری اثر آن در این مدل بعنوان گواهی برای کنش افتراقی سؤال در نظر گرفته می شود. این در حالی است که اثر متغیر توانایی و تعامل آن با متغیر

^۱-Sorbom

^۲- Meade

^۳- Lautenschlager

^۴- Swaminathan

^۵- Rogers

^۶-Agresti

گروهی کنترل شده است. معنی داری تعامل متغیر گروهی و توانایی نیز خود گواهی بر نوعی سوگیری سؤال است که کنش افتراقی ناهماهنگ نامیده می شود. مطالعات قبلی نشان می دهند که روش رگرسیون لجستیک مانند روشهای قبلی در تشخیص DIF اثربخش است. مطالعات قبلی نشان می دهند که این روش به حجم نمونه حساس است و در نمونه های کوچک برآورد دقیق پارامترها میسر نیست (اگرستی، ۲۰۰۷). همچنین سایر مطالعات با داده های شبیه سازی شده برای آزمون های سرنوشت ساز نشان می دهند که عواملی نظیر مقدار DIF و دشواری سؤال بر کارکرد روش رگرسیون لجستیک در تشخیص DIF سؤال مؤثر است (میزور^۱ و همکاران، ۱۹۹۲).

همانطور که ملاحظه می شود تعاریف مفهومی از سوگیری بسیار کلی است و برای مدل‌های مختلف مورد استفاده قرار می گیرد. از دهه های گذشته روان سنج هایی که درباره تغییر ناپذیری اندازه گیری تحقیق می کنند، روشهای آماری متفاوتی را برای اندازه گیری کنش افتراقی سؤال (DIF) پیشنهاد کرده اند. این روشها معمولاً مفروضات آماری متفاوتی را در مورد شکل رابطه میان متغیرهای پنهان و مشاهده شده و شکل توزیع متغیر پنهان در جامعه در بر دارند و راهبردهای مدل یابی متفاوتی را برای تعیین معنی داری کنش افتراقی سؤال بکار می برند. بنابراین تعیین اینکه کدامیک از روشها نتایج بهتری را بدست می دهد بسیار مشکل است، بعلاوه معلوم نیست در چه مواقعی استفاده از روشهای متفاوت، نتایج متفاوتی را در بر دارد. بطوریکه متغیرهایی از قبیل حجم نمونه، DIF هماهنگ، DIF ناهماهنگ، درصد سؤالات دارای DIF و تعداد گزینه های سؤال بعنوان عوامل مداخله گر عنوان شده اند (کرین^۲ و همکاران، ۲۰۰۶).

روشهای تشخیص DIF که در این مطالعه مد نظر است بصورت گسترده ای مورد استفاده قرار گرفته و به نتایج آنها در مطالعات روان سنجی استناد می شود، اما به نظر می رسد که مطالعه ای در زمینه مقایسه همزمان روش رگرسیون لجستیک با روش های CFA و IRT برای تعیین قدرت نسبی آن در شناسایی DIF انجام نشده است و علی رغم اهمیت مطالعه DIF این ویژگی مهم روان سنجی کمتر مورد مطالعه قرار گرفته است و کمتر بصورت عملی در آزمون های سرنوشت ساز انجام و نتایج آن گزارش شده است. تحقیقات انجام شده غالباً نتایج ناهمسانی را گزارش کرده اند و هنوز بطور قطع برتری این روشها در مقابل یکدیگر به اثبات نرسیده است.

^۱ - Mazor

^۲ - Crane

تحقیق حاضر از روشهای مبتنی بر IRT، روش نسبت درست‌نمایی^۱ و از روشهای غیر مبتنی بر IRT، روشهای رگرسیون لجستیک دو وجهی^۲ و روش تحلیل عاملی تأییدی (CFA)^۳ را بعنوان مجموعه ای جدید برای پژوهش انتخاب کرده است و قصد دارد با استفاده از مطالعات شبیه سازی داده های آزمون های سرنوشت ساز که به مطالعات مونت کارلو^۴ معروف هستند، قدرت هر کدام از این روشها را در تشخیص کنش افتراقی سؤال بررسی کنند.

همچنین بیشتر مطالعات انجام شده در حوزه DIF صرفاً بصورت نظری و در قالب پژوهشهای بنیادی مسأله را بررسی نموده اند. به نظر می رسد یکی از عوامل اصلی فقدان مطالعات لازم در شرایط واقعی، اجرای آزمون ها، جدید بودن، نامأنوس بودن و همچنین پیچیدگی نسبی روشهای شبیه سازی و نرم افزارهای انجام مطالعات DIF است. در مورد آزمون های سرنوشت ساز مطالعات شبیه سازی شده بستری را فراهم می کنند تا محقق بتواند مجموعه بزرگی از داده های پرسش-پاسخ را در تکرار های مختلف ایجاد کند و با شبیه سازی سؤالات سوگیرانه و دستکاری پارامترهای مختلف کارایی روشهای آماری مختلف را در آشکارسازی سؤالات سوگیرانه در آزمون های سرنوشت ساز مورد بررسی قرار دهد (هان^۵، ۲۰۰۷).

قربان و نزدیکی زیادی میان مدل های CFA، IRT و رگرسیون لجستیک وجود دارد و تاکنون این سه روش بدین شکل و با استفاده از داده های شبیه سازی شده مورد بررسی و مقایسه قرار نگرفته اند. همانطور که در مطالعات قبلی نیز اشاره شد، عوامل مداخله گر متفاوتی در ارتباط با این روشها مورد بررسی قرار گرفته است، اما هنوز بطور قطع نقش این عوامل در شرایط مختلف آزمایشی و شبیه سازی شده مشخص نشده است. بنابراین سؤال اساسی در تحقیق حاضر این است که در شرایط مختلف و با کنترل عوامل مداخله گر کدامیک از روشهای مذکور بررسی کنش افتراقی سؤال با قدرت بیشتری می توانند سوگیری سؤال را در آزمون های سرنوشت ساز بصورت معنی داری آشکار کنند؟

۱-۳- اهداف تحقیق

پژوهشگران مدتهاست که به اندازه گیری تفاوت میان خصیصه ها و نگرش ها در گروههای مختلف و در طول زمان علاقمند هستند. در طول سالهای مختلف پژوهشگران بر این عقیده اند که به منظور

^۱- Likelihood Ratio

^۲- Binary Logistic Regression

^۳- Confirmatory Factor Analysis

^۴- Monte Carlo Studies

^۵- Han

اندازه گیری درست این خصیصه ها باید مقیاس های اندازه گیری را در وضعیتهای مختلف اجرا و تفاوت میان نمره های مشاهده شده را محاسبه کرد (کرونباخ^۱ و فربسی^۲، ۱۹۷۰). ولی اساساً مشکلات بالقوه ای در استفاده از نمرات اختلافی وجود دارد. به منظور سنجش تفاوت میان نمره مشاهده شده دو مقیاس ضروری است که آن مقیاس در گروههای متفاوتی از افراد بطور یکسان درک شود. به زبان دیگر باید نشان داد که دو اندازه گیری از لحاظ روانسنجی موازی هستند تا بتوان در طول زمان یا میان گروههای مختلفی از پاسخگویان مقایسه های معتبری انجام داد.

اندازه گیری موازی بعنوان عملیاتی تلقی می شود که در شرایط مختلف اندازه های یکسانی را تولید می کند (هورن^۳ و مک آردل، ۱۹۹۲). این شرایط شامل ثبات اندازه گیری در زمان، مراکز مختلف اندازه گیری یا گروههای قضاوت کننده مختلف و یا عدم سوگیری سؤاها به نفع گروه خاصی می شوند (الن^۴، ۱۹۷۹ به ترجمه دلاور، ۱۳۷۴). این در حالی است که اصولاً استفاده از ابزارهای اندازه گیری سوگیرانه منجر به خدشه دار شدن اعتبار درونی تحقیقات می شود (کرک^۵، ۱۹۹۵).

از سوی دیگر تعیین روش آماری مناسب برای شناسایی سؤالهای سوگیرانه در آزمون های سرنوشت ساز روانی و تحصیلی و مقیاسهای اندازه گیری دارای اهمیت زیادی است بطوریکه استفاده از روشهای مختلف آماری برای شناسایی کنش افتراقی سؤال ممکن است نتایج متفاوتی بدست دهد (کمیلی و شپارد، ۱۹۹۴).

روش مبتنی بر IRT آزمون نسبت درستی، تحلیل عاملی تأییدی و رگرسیون لوجستیک از جمله روشهای آماری تشخیص کنش افتراقی سؤال هستند که بطور گسترده مورد استفاده قرار می گیرند. پیرامون این روشها و قدرت آنها در شناسایی کنش افتراقی سؤال بصورت مجزا با داده های آزمون های سرنوشت ساز واقعی و اخیراً با داده های شبیه سازی شده تحقیقات زیادی انجام گرفته است (کرین و همکاران، ۲۰۰۶). مطالعات مونت کارلو برای محققان بستری را فراهم کرده تا بتوانند روشهای مختلف آماری را با یکدیگر مقایسه کنند. تحقیق حاضر نیز بعنوان یک مطالعه جدید با داده های شبیه سازی شده برای آزمون های سرنوشت ساز قصد دارد کارکرد سه روش متداول بررسی DIF شامل نسبت درست نمایی مبتنی بر مدل های پرسش-پاسخ، روشهای رگرسیون لوجستیک و تحلیل عاملی تأییدی را مورد بررسی قرار دهد.

بنابراین اهداف، سؤالات و فرضیه های تحقیق عبارتند از:

^۱ - Cronbach
^۲ - Furby
^۳ - Horn
^۴ - Allen
^۵ - Kirk

۱- بکارگیری روش مبتنی بر IRT نسبت درست نمایی در تشخیص کنش افتراقی سؤال در آزمون های سرنوشت ساز

۲- بکارگیری روش تحلیل عاملی تأییدی در تشخیص کنش افتراقی سؤال در آزمون های سرنوشت ساز

۳- بکارگیری روش رگرسیون لوجستیک در تشخیص کنش افتراقی سؤال در آزمون های سرنوشت ساز

۴- مقایسه روش مبتنی بر IRT نسبت درست نمایی و روشهای تحلیل عاملی تأییدی و رگرسیون لوجستیک در تشخیص کنش افتراقی سؤال در آزمون های سرنوشت ساز

۵- تعیین عوامل مداخله گر در تشخیص کنش افتراقی سؤال در هر یک از روش های آزمون درستنمایی مبتنی بر IRT، تحلیل عاملی تأییدی و رگرسیون لوجستیک در آزمون های سرنوشت ساز

۱-۴- سؤال های تحقیق

همچنین سؤالهایی که از اهداف تحقیق ناشی می شوند عبارتند از:

- ۱- آیا شناسایی سوگیری سؤال در روشهای مختلف در آزمون های سرنوشت ساز تحت تأثیر حجم نمونه آزمودنی ها قرار می گیرد؟
- ۲- آیا شناسایی سوگیری سؤال در روشهای مختلف در آزمون های سرنوشت ساز تحت تأثیر نوع DIF (همانگ یا ناهمانگ) قرار می گیرد؟
- ۳- آیا شناسایی سوگیری سؤال در روشهای مختلف در آزمون های سرنوشت ساز تحت تأثیر درصد سؤالات دارای DIF در مجموعه سؤالات قرار می گیرد؟
- ۴- آیا شناسایی سوگیری سؤال در روشهای مختلف در آزمون های سرنوشت ساز تحت تأثیر تعداد گزینه های سؤال قرار می گیرد؟

۱-۵- فرضیه های تحقیق

بر اساس تحقیقات گذشته و نظریات موجود می توان برای قدرت آزمون های نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ و آزمون های تحلیل عاملی تأییدی و رگرسیون لجستیک در شناسایی کنش افتراقی سؤال (DIF) در آزمون های سرنوشت ساز فرضیه های اصلی و فرعی زیر را مطرح کرد:

- ۱- حجم نمونه بر قدرت آزمون های آشکارسازی DIF نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ ، تحلیل عاملی تأییدی و رگرسیون لجستیک مؤثر است.
- ۱-۱- در حجم نمونه بالا (۱۰۰۰ آزمودنی) ، آزمون نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ در مقایسه با آزمون های تحلیل عاملی تأییدی و رگرسیون لجستیک در شناسایی کنش افتراقی سؤال (DIF) در آزمون های سرنوشت ساز قدرت بیشتری دارد.
- ۱-۲- در حجم نمونه متوسط (۵۰۰ آزمودنی) ، قدرت آزمون نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ در مقایسه با آزمون های تحلیل عاملی تأییدی و رگرسیون لجستیک در شناسایی کنش افتراقی سؤال (DIF) در آزمون های سرنوشت ساز تفاوتی ندارد.

۱-۳- در حجم نمونه پایین (۱۵۰)، آزمون های تحلیل عاملی تأییدی و رگرسیون لجستیک در مقایسه با آزمون نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ در شناسایی کنش افتراقی سؤال (DIF) در آزمون های سرنوشت ساز قدرت بیشتری دارند.

۲- نوع DIF بر قدرت آزمون های آشکارسازی DIF نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ، تحلیل عاملی تأییدی و رگرسیون لجستیک مؤثر است.

۲-۱- در کنش افتراقی هماهنگ، آزمون های تحلیل عاملی تأییدی و رگرسیون لجستیک در مقایسه با آزمون نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ در شناسایی کنش افتراقی سؤال (DIF) در آزمون های سرنوشت ساز قدرت بیشتری دارند.

۲-۲- در کنش افتراقی ناهماهنگ، آزمون رگرسیون لجستیک در مقایسه با آزمون های تحلیل عاملی تأییدی و آزمون نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ در شناسایی کنش افتراقی سؤال (DIF) در آزمون های سرنوشت ساز قدرت بیشتری دارد.

۲-۳- در کنش افتراقی ناهماهنگ، قدرت آزمون های تحلیل عاملی تأییدی و آزمون نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ در شناسایی کنش افتراقی سؤال (DIF) در آزمون های سرنوشت ساز تفاوتی ندارد.

۳- تعداد سؤالات DIF در آزمون بر قدرت آزمون های آشکارسازی DIF نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ، تحلیل عاملی تأییدی و رگرسیون لجستیک مؤثر است.

۳-۱- در هر سه سطح DIF سؤال ۱۰ درصد (۳ سؤال)، ۲۰ درصد (۶ سؤال) و ۳۰ درصد (۹ سؤال)، آزمون نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ در مقایسه با آزمون های تحلیل عاملی تأییدی و رگرسیون لجستیک در شناسایی کنش افتراقی سؤال (DIF) در آزمون های سرنوشت ساز قدرت بیشتری دارد.

۴- تعداد گزینه های سؤال بر قدرت آزمون های آشکارسازی DIF نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ، تحلیل عاملی تأییدی و رگرسیون لجستیک مؤثر است.

۴-۱- در داده های دو وجهی، آزمون نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ در مقایسه با آزمون های تحلیل عاملی تأییدی و رگرسیون لجستیک در شناسایی کنش افتراقی سؤال (DIF) در آزمون های سرنوشت ساز قدرت بیشتری دارد.

۴-۲- در داده های چند وجهی، آزمون تحلیل عاملی تأییدی در مقایسه با آزمون نسبت درستنمایی مبتنی بر نظریه پرسش-پاسخ در شناسایی کنش افتراقی سؤال (DIF) در آزمون های سرنوشت ساز قدرت بیشتری دارد.

۵- شدت DIF بر قدرت آزمون های آشکارسازی DIF نسبت درستی مبتنی بر نظریه پرسش-پاسخ، تحلیل عاملی تأییدی و رگرسیون لجستیک مؤثر است.

۵-۱- در تمامی مقادیر شدت DIF (۰.۲۵، ۰.۵۰، ۰.۷۵ و ۱) آزمون نسبت درستی مبتنی بر نظریه پرسش-پاسخ در مقایسه با آزمون های تحلیل عاملی تأییدی و رگرسیون لجستیک در شناسایی کنش افتراقی سؤال (DIF) در آزمون های سرنوشت ساز قدرت بیشتری دارد.

۱-۶- تعریف مفاهیم تحقیق

۱-۶-۱- تعریف نظری

- **کنش افتراقی سؤال (DIF):** زمانیکه یک سؤال آزمون به نفع یک گروه از آزمودنی ها عمل کند می توان گفت که آن سؤال سوگیری دارد. چنین سؤالهایی دارای کنش افتراقی (DIF) هستند (رامسی، ۱۹۹۳).

- **روشهای تشخیص کنش افتراقی سؤال:** روشهایی که جهت تشخیص DIF مورد استفاده قرار می گیرد که بطور کلی به دو دسته روشهای مبتنی بر IRT و روشهای نمره مشاهده شده (غیر مبتنی بر IRT) تقسیم می شوند (کمیلی و شپارد، ۱۹۹۴).

- **DIF مثبت درست^۱ و مثبت غلط^۲:** در صورتی که سؤال مورد بررسی در آزمون دارای DIF باشد و روشهای بررسی DIF بدرستی کنش افتراقی سؤال را آشکار کند، در این حالت DIF مثبت درست تشخیص داده شده است. اما هنگامی که سؤال مورد بررسی در آزمون دارای کنش افتراقی نباشد ولی روشهای بررسی DIF به غلط حکم به سوگیری سؤال کند، در این حالت DIF مثبت غلط تشخیص داده شده است (تیشن^۳ و همکاران، ۱۹۹۳).

- **حجم نمونه:** تعدادی معین که زیر مجموعه ای از یک مجموعه بزرگتر بعنوان جامعه است و می تواند جامعه را نمایندگی کند (کوهن^۴ و همکاران، ۲۰۰۷).

- **DIF هماهنگ:** تفاوت گروه مرجع و کانونی در پارامتر دشواری سؤال (b) در تابع پرسش-پاسخ (فلاورز و همکاران، ۱۹۹۹).

^۱ - True Positive

^۲ - False Positive

^۳ - Thissen

^۴ - Cohen

- **DIF ناهماهنگ:** تفاوت گروه مرجع و کانونی در پارامتر قدرت تمیز سؤال (a) در تابع پرسش-پاسخ (فلاورز و همکاران، ۱۹۹۹).

- **شدت DIF:** مقدار تفاوت میان پارامترهای دشواری و قدرت تمیز سؤال میان دو گروه مرجع و کانونی هر چقدر بیشتر باشد، اصطلاحاً شدت DIF بیشتر است (تیشن^۱ و همکاران، ۱۹۹۳).

- **درصد سؤالات دارای DIF:** نسبت سؤالاتی که در یک مجموعه سؤال^۲ دارای DIF است (وینر^۳، ۱۹۹۳).

- **گزینه های سؤال:** سؤالهای بسته پاسخ انتخابهایی را در اختیار پاسخگو می گذارند که گزینه سؤال نامیده می شود. گزینه ها می توانند دو یا بیشتر از دو گزینه داشته باشند (نیتکو^۴، ۲۰۰۱).

- **آزمون های سرنوشت ساز:** آزمون های سرنوشت ساز آزمون هایی هستند که در مقیاس های با حجم آزمودنی بزرگ اجرا می شوند و معمولاً برای تصمیم گیری در مورد جایگزینی، ارتقاء و اعطای مدرک تحصیلی مورد استفاده قرار می گیرند. چنین تصمیم هایی دارای تبعات مهم شخصی، اجتماعی و سیاسی هستند. بنابراین خیلی مهم است که تصمیم های معتبری بر اساس این آزمون ها اتخاذ شود (پایهام، ۲۰۰۵).

۱-۶-۲- تعریف عملیاتی

از نظر عملیاتی کنش افتراقی سؤال DIF، حجم نمونه برای آزمون های سرنوشت ساز، DIF هماهنگ، DIF ناهماهنگ، درصد سؤالات دارای DIF و گزینه های سؤال آنچیزی است که در تحقیق حاضر در نرم افزار WINGEN شبیه سازی می شوند و روشهای تشخیص DIF از نظر عملیاتی در تحقیق حاضر عبارتند از روش مبتنی بر IRT آزمون نسبت درستمایی و روشهای نمره مشاهده شده (غیر مبتنی بر IRT) تحلیل عاملی تأییدی و رگرسیون لجستیک

^۱ - Thissen

^۲ - Test Battery

^۳ - Wainer

^۴ - Nitko

فصل دوم: پیشینه تحقیق