

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

پایان نامه

جهت دریافت درجه کارشناسی ارشد رشته سنجش و اندازه گیری

موضوع:

مقایسه دقت برآورد توانایی در سؤالات چند گزینه ای با بکارگیری

مدلهای سؤال - پاسخ دو و چند ارزشی

استاد راهنما:

دکتر محمدرضا فلسفی نژاد

استاد مشاور:

دکتر علی دلاور

استاد داور:

دکتر یحیی مهاجر

پژوهشگر:

مهرناز معلّمی اوره

زمستان ۱۳۸۷

لم يشكر المخلوق لم يشكر الخالق

به حکم ادب و وظیفه

از همراهی بی دریغ اساتید ارجمند

جناب آقای دکتر فلسفی نژاد و جناب آقای دکتر دلاور

که در طی این مسیر همواره از مشاوره و راهنمایی‌های ایشان بهره‌مند بودم و

جناب آقای دکتر مهاجر که داوری این پایان‌نامه را بر عهده داشتند

سپاس‌گزاری می‌کنم.

همچنین از

پدر و مادر عزیزم

که همه هستی‌ام بواسطه وجود و حضور پرفروغ آنهاست

و همسر مهربانم

که در گذر از دشواری‌های دوران تحصیل من در این مقطع، با بردباری مرا همراهی کرد

و نیز همه دوستانی

که در به سرانجام رساندن این پژوهش، هر یک به نوعی مرا یاری رساندند،

به ویژه سرکار خانم صدیقه جلیلی

که در زمینه بکارگیری نرم‌افزار تخصصی مالتی‌لوگ، تجربیات خود را در اختیار بنده قرار داد،

صمیمانه قدردانی و تشکر می‌نمایم.

تقدیم به

همه تلاش‌گران عرصه علم و دانش

که هر یک به سهم خود چراغی بر می‌فروزند

و راهگشای پویندگان این مسیر می‌باشند.

چکیده:

مدل‌های چند ارزشی IRT، برخلاف مدل کلاسیک اندازه‌گیری و مدل‌های دو ارزشی، از این قابلیت برخوردارند که در قدرسنجی سؤالات و آزمودنی‌ها در آزمون‌های توانایی متشکل از سؤالات چندگزینه‌ای، برای گزینه‌های پاسخ، بیش از دو ارزش در نظر بگیرند و لذا مدعی هستند که با مورد توجه قرار دادن اطلاعات موجود در گزینه‌های پاسخ نادرست می‌توانند ارزش‌های متفاوتی برای آنها در نظر بگیرند و بدین وسیله موجب افزایش دقت برآورد توانایی آزمودنی‌ها گردند. رسالت اصلی این پژوهش در واقع معرفی و بکارگیری یکی از جامع‌ترین مدل‌های چند ارزشی IRT (مدل پاسخ اسمی) به‌عنوان یک روش دقیق‌تر برآورد پارامترهای سؤال و توانایی در سؤالات چندگزینه‌ای در مقایسه با مدل‌های دو ارزشی IRT است.

برای این منظور، داده‌های سؤال-پاسخ یک نمونه تصادفی به‌حجم ۲۰۰۰ نفر از میان جامعه پاسخ‌دهندگان (۶۸۴۰۵ نفر) به سؤالات خرده‌آزمون ۵۰ سؤالی زیست‌شناسی کنکور سراسری رشته علوم تجربی سال تحصیلی ۸۵-۸۴، انتخاب شد. همچنین مدل پاسخ اسمی (NRM) که تنها مدل زیرساختی مناسب جهت تحلیل پاسخ‌های اسمی (نظیر پاسخ سؤالات چندگزینه‌ای در اغلب آزمون‌های پیشرفت تحصیلی) می‌باشد، با توجه به ویژگی‌های ساختاری آن و با عنایت به نتایج حاصل از برخی پژوهش‌های پیشین، به‌عنوان مدل چند ارزشی مناسب برای داده‌های سؤال-پاسخ این پژوهش، در نظر گرفته شد.

پس از بررسی شاخص‌های توصیفی سؤالات خرده‌آزمون مذکور (با بکارگیری مدل کلاسیک اندازه‌گیری)، از میان آنها ۱۲ سؤال که از شاخص‌های توصیفی ضعیف‌تری (ضریب تشخیص و همبستگی سؤال-آزمون بسیار پایین) برخوردار بودند و به‌تبع آن عدم برقراری یکی از مفروضه‌های اساسی IRT مبنی بر تک بعدی بودن مجموعه سؤالات را موجب می‌شدند، کنار گذاشته شد و ۳۸ سؤال باقیمانده، پس از اطمینان یافتن از برقراری مفروضات IRT، توسط نرم افزار MULTILOG مورد تحلیل قرار گرفت.

نتایج حاصل از برازاندن مدل‌های دو ارزشی ۲،۱ و ۳ پارامتری بر داده‌های سؤال-پاسخ مذکور نشان داد، مدل دو پارامتری که از نظر ساختار پارامتری در میان مدل‌های دو ارزشی، نزدیکترین مدل به NRM است، در این میان از برازندگی بهتری با داده‌های فوق‌الذکر برخوردار است و لذا این مدل به‌عنوان مدل دو ارزشی مناسب جهت مقایسه دقت برآورد توانایی با مدل چند ارزشی برگزیده (NRM) انتخاب شد.

به‌منظور برازاندن این دو مدل بر داده‌های سؤال-پاسخ، مجموع تفاضل نسبت احتمال مشاهده شده و مورد انتظار گزینه‌های پاسخ هر یک از سؤالات در هر دو مدل، محاسبه شد و ۶ سؤال که با در نظر گرفتن معیار فوق، برازش کمتری برای آنها محتمل بود، حذف شدند و در نهایت، همگرایی مورد نظر برای مجموعه ۳۲ سؤال برگزیده حاصل شد. پس از برآورد پارامترهای سؤال و میزان آگاهی‌دهندگی هر سؤال (و مجموع سؤالات) در سطوح مختلف توانایی و برآورد پارامتر توانایی هر یک از آزمودنی‌ها در هر دو مدل، به‌منظور مقایسه دقت برآورد توانایی افراد، کارایی نسبی این دو مدل در سطوح مختلف توانایی با توجه به شرایط ویژه هر یک از سؤالات پژوهش، محاسبه شد و مورد بررسی و مقایسه قرار گرفت.

نتایج این بررسی‌ها حاکی از برتری دقت برآورد توانایی NRM در غالب سطوح توانایی، در مقایسه با 2PL می‌باشد و همان‌گونه که انتظار می‌رود، برای آزمودنی‌هایی که از سطح توانایی پایین‌تری برخوردارند و همچنین در سؤالات دشوارتر، برتری دقت برآورد توانایی پاسخ‌دهندگان در مدل پاسخ اسمی نسبت به مدل دو پارامتری، بیشتر مشهود است. لذا می‌توان به‌منظور افزایش دقت برآورد توانایی آزمودنی‌ها در آزمون‌هایی با شرایط مشابه، بکارگیری مدل پاسخ اسمی را با در نظر گرفتن برخی ملاحظات، توصیه نمود. به‌ویژه در شرایطی که اکثر آزمودنی‌ها از سطح توانایی نسبتاً پایینی برخوردارند و از طرف دیگر، مقتضیات خاص آزمون استفاده از سؤالات نسبتاً دشوار را ایجاب نماید.

واژه‌های کلیدی: نظریه سؤال-پاسخ (IRT)، مدل‌های دو و چند ارزشی، مدل پاسخ اسمی (NRM)، مدل دو پارامتری (2PL)، سؤالات چندگزینه‌ای، کارایی نسبی، MULTILOG

فهرست مطالب

فصل اول: کلیات پژوهش

۱	مقدمه
۴	بیان مسأله
۷	اهمیت و ضرورت انجام پژوهش
۸	اهداف و سؤالات پژوهش
۸	اهداف پژوهش
۸	سؤالات پژوهش
۹	تعریف مفاهیم و واژگان اختصاصی طرح

فصل دوم: مبانی نظری و پیشینه پژوهش

۱۲	مقدمه
۱۳	بخش اول: مبانی نظری
۱۳	نظریه‌های اندازه‌گیری
۱۴	نظریه کلاسیک اندازه‌گیری
۱۷	نارسایی‌های نظریه کلاسیک
۲۰	نظریه سؤال-پاسخ
۲۰	شکل‌گیری IRT
۲۱	مبانی، مفروضات و ویژگی‌های IRT
۲۱	مبانی ریاضی IRT
۲۴	مفروضات IRT
۲۴	الف: تک بعدی بودن
۲۶	ب: استقلال موضعی
۲۸	ج: برآزش و مناسبت مدل- داده‌ها
۲۸	ویژگی‌های IRT
۲۸	مفاهیم زیربنایی IRT
۲۸	الف) توابع سؤال-پاسخ
۳۰	ب) توابع آگاهی
۳۳	محاسبه توابع آگاهی سؤال و آزمون
۳۵	تفسیر تابع آگاهی آزمون
۳۶	ج) ویژگی نامتغیر بودن
۳۷	مزایای IRT

۳۹ انواع مدل‌های IRT
۴۲ مدل‌های دو ارزشی IRT
۴۲ الف: مدل لجستیک تک‌پارامتری
۴۴ ب: مدل لجستیک دو پارامتری
۴۶ ج: مدل لجستیک سه پارامتری
۴۸ د: مدل لجستیک چهار پارامتری
۴۹ مدل‌های چند ارزشی IRT
۴۹ شکل‌گیری مدل‌های چند ارزشی و دلایل بکارگیری آنها
۵۳ ساختار کلی مدل‌های چند ارزشی
۵۶ طبقه‌بندی ساختاری مدل‌های چند ارزشی
۵۸ الف) رویکرد ترستون/سامیجیما
۵۹ ب) رویکرد راش
۶۰ معرفی متداول‌ترین مدل‌های چند ارزشی
۶۰ الف) مدل‌های چند ارزشی غیر رتبه‌ای
۶۰ الف- ۱: مدل پاسخ اسمی
۶۷ الف- ۲: مدل باک- سامیجیما
۶۹ الف- ۳: مدل چند گزینه‌ای تیسن و استینبرگ
۷۰ ب) مدل‌های چند ارزشی راش
۷۲ ب - ۱: مدل اعتبار سهمی
۷۳ ب - ۲: مدل تعمیم یافته اعتبار سهمی
۷۴ ب - ۳: مدل تقسیم بندی رتبه‌ای
۷۴ ب - ۴: مدل مقیاس ترتیبی
۷۷ ج) مدل‌های درجه‌بندی شده سامیجیما
۷۹ ج - ۱: مدل پاسخ درجه‌بندی شده
۸۳ ج - ۲: مدل پاسخ درجه‌بندی شده مقیاس ترتیبی
۸۴ د) مدل‌های پاسخ پیوسته
۸۵ د- ۱: مدل پاسخ پیوسته
۸۶ طبقه‌بندی ساختاری تیسن و استینبرگ برای مدل‌های IRT
۸۸ مدل‌های دیگر IRT
۸۹ الف) مدل‌های چند بعدی
۹۰ ب) مدل‌های غیر پارامتریک
۹۱ انتخاب مدل و ارزیابی برازش مدل- داده‌ها
۹۱ انتخاب یک مدل سؤال- پاسخ مناسب
۹۲ ارزیابی برازش مدل با داده‌ها
۹۹ روش‌های برآورد پارامترهای سؤال و توانایی در IRT
۱۰۰ کاربردهای IRT
۱۰۰ ساخت آزمون
۱۰۲ تشخیص کنش افتراقی سؤال
۱۰۳ همتراز سازی نمرات یا مرتبط ساختن مقیاس‌ها
۱۰۴ سنجش انطباقی با استفاده از کامپیوتر
۱۰۷ گزارش نمرات
۱۰۸ بخش دوم: پیشینه پژوهش
۱۰۸ روش‌های وزندهی افتراقی گزینه‌ها بر مبنای CTT
۱۱۰ روش‌های وزندهی افتراقی گزینه‌ها بر مبنای IRT

فصل سوم: روش‌شناسی تحقیق

۱۱۶	روش تحقیق
۱۱۶	جامعه، نمونه و روش نمونه‌گیری
۱۱۸	روش گردآوری اطلاعات و داده‌ها
۱۱۸	ابزار اندازه‌گیری
۱۱۹	اعتبار و پایایی اندازه‌گیری
۱۱۹	تجزیه و تحلیل داده‌ها
۱۱۹	تحلیل‌های کلاسیک
۱۲۰	تحلیل‌های عمومی IRT
۱۲۰	بررسی مفروضه‌های زیربنایی IRT
۱۲۰	الف) تک بعدی بودن
۱۲۱	ب) استقلال موضعی
۱۲۱	انتخاب مدل مناسب و ارزشیابی برآزش مدل با داده‌ها
۱۲۲	برآورد پارامترهای کلی مدل در مدل‌های سوال-پاسخ برگزیده
۱۲۳	تحلیل‌های IRT ویژه سوالات پژوهش

فصل چهارم: تجزیه و تحلیل داده‌ها

۱۲۴	مقدمه
۱۲۵	تحلیل‌های کلاسیک
۱۲۵	الف: برآورد شاخص‌های توصیفی سوالات
۱۳۱	ب: برآورد توانایی آزمودنی‌ها
۱۳۳	تحلیل‌های عمومی IRT
۱۳۳	بررسی برقراری مفروضه‌های زیربنایی IRT
۱۳۴	برآورد پارامترهای سوال و توانایی و بررسی دقت پیش‌بینی‌های مدل در IRT
۱۳۸	برآورد پارامترهای سوال و توانایی آزمودنی‌ها با استفاده از مدل دو پارامتری
۱۴۶	برآورد پارامترهای سوال و توانایی آزمودنی‌ها با استفاده از مدل پاسخ اسمی
۱۵۴	بررسی سوالات پژوهش

فصل پنجم: بحث و نتیجه‌گیری

۱۶۲	مقدمه
۱۶۳	بررسی فرایند انجام کار و پاسخ‌گویی به سوالات پژوهش
۱۶۷	محدودیت‌های پژوهش
۱۶۷	پیشنهادات

۱۶۹	منابع فارسی
۱۷۱	منابع لاتین
۱۷۴	ضمائم

فهرست شکل‌ها

- شکل ۲-۱: نمودار تابع احتمال پاسخ سؤال ۳۹ در مدل دو پارامتری ۲۳
- شکل ۲-۲: IRF های سه سؤال A و B و C که به صورت دو ارزشی نمره‌گذاری شده‌اند ۲۹
- شکل ۲-۳: توابع آگاهی سه سؤال A و B و C که به صورت دو ارزشی نمره‌گذاری شده‌اند ۳۲
- شکل ۲-۴: تابع آگاهی آزمون سه سؤالی متشکل از سؤالات A و B و C ۳۴
- شکل ۲-۵: منحنی‌های ویژگی سؤال با استفاده از مدل لجستیک تک پارامتری ۴۴
- شکل ۲-۶: منحنی‌های ویژگی سؤال با استفاده از مدل لجستیک دو پارامتری ۴۵
- شکل ۲-۷: منحنی‌های ویژگی سؤال با استفاده از مدل لجستیک سه پارامتری ۴۶
- شکل ۲-۸: توابع سؤال-پاسخ یک سؤال چند ارزشی مشتمل بر ۵ طبقه پاسخ ۵۲
- شکل ۲-۹: توابع پاسخ مرز طبقه برای یک سؤال چند ارزشی مشتمل بر ۵ طبقه پاسخ ۵۵
- شکل ۲-۱۰: توابع پاسخ طبقات یک سؤال ۵ گزینه‌ای با استفاده از NRM ۶۳
- شکل ۲-۱۱: توابع پاسخ طبقات یک سؤال ریاضی ۴ گزینه‌ای با استفاده از NRM ۶۴
- شکل ۲-۱۲: توابع پاسخ طبقات یک سؤال ۴ گزینه‌ای با استفاده از مدل چندگزینه‌ای ۶۹
- شکل ۲-۱۳: توابع پاسخ طبقات یک سؤال ۵ طبقه‌ای (از نوع لیکرت) با استفاده از دو مدل RSM و PCM ۷۷
- شکل ۲-۱۴: منحنی‌های ویژه عملکرد یک سؤال ۵ طبقه‌ای ۸۰
- شکل ۲-۱۵: منحنی‌های پاسخ طبقات سؤال قبل با استفاده از GRM ۸۱
- شکل ۴-۱: نمودار توزیع فراوانی پاسخ‌های درست افراد گروه نمونه به خرده‌آزمون ۳۲ سؤالی ۱۳۲
- شکل ۴-۲: نمودار توزیع فراوانی نمره کل (با احتساب امتیاز منفی) افراد گروه نمونه در خرده‌آزمون ۳۲ سؤالی ۱۳۲
- شکل ۴-۳: ماتریس منحنی ویژگی سؤالات با استفاده از مدل دو پارامتری برای نمونه ۳۲ سؤالی منتخب ۱۳۸
- شکل ۴-۴: منحنی ویژگی سؤال ۴ ۱۴۰
- شکل ۴-۵: تابع آگاهی سؤال ۴ ۱۴۰
- شکل ۴-۶: منحنی ویژگی سؤال ۴۸ ۱۴۱
- شکل ۴-۷: تابع آگاهی سؤال ۴۸ ۱۴۱
- شکل ۴-۸: منحنی ویژگی سؤال ۳۴ ۱۴۲
- شکل ۴-۹: تابع آگاهی سؤال ۳۴ ۱۴۲
- شکل ۴-۱۰: منحنی ویژگی سؤال ۱ ۱۴۳
- شکل ۴-۱۱: تابع آگاهی سؤال ۱ ۱۴۳
- شکل ۴-۱۲: منحنی تابع آگاهی (و خطای استاندارد) کل آزمون در مدل دو پارامتری ۱۴۴
- شکل ۴-۱۳: نمودار توزیع فراوانی توانایی آزمودنی‌ها در مدل دو پارامتری ۱۴۵
- شکل ۴-۱۴: ماتریس منحنی توابع احتمال گزینه‌های پاسخ سؤالات در مدل پاسخ اسمی برای نمونه ۳۲ سؤالی ۱۴۶
- شکل ۴-۱۵: نمودار منحنی‌های تابع احتمال گزینه‌های پاسخ (ORF) سؤال ۴ ۱۴۹
- شکل ۴-۱۶: تابع آگاهی سؤال ۴ ۱۴۹
- شکل ۴-۱۷: نمودار منحنی‌های تابع احتمال گزینه‌های پاسخ (ORF) سؤال ۴۸ ۱۵۰

- شکل ۴-۱۸ : تابع آگاهی سؤال ۴۸ ۱۵۰
- شکل ۴-۱۹ : نمودار منحنی های تابع احتمال گزینه های پاسخ (ORF) سؤال ۳۴ ۱۵۱
- شکل ۴-۲۰ : تابع آگاهی سؤال ۳۴ ۱۵۱
- شکل ۴-۲۱ : نمودار منحنی های تابع احتمال گزینه های پاسخ (ORF) سؤال ۱ ۱۵۲
- شکل ۴-۲۲ : تابع آگاهی سؤال ۱ ۱۵۲
- شکل ۴-۲۳ : منحنی تابع آگاهی (و خطای استاندارد) کل آزمون در مدل پاسخ اسمی ۱۵۳
- شکل ۴-۲۴ : نمودار توزیع فراوانی توانایی آزمودنی ها در مدل پاسخ اسمی ۱۵۴
- شکل ۴-۲۵ : نمودار منحنی توابع آگاهی در مدل پاسخ اسمی و دوپارامتری ۱۵۷
- شکل ۴-۲۶ : کارایی نسبی مدل پاسخ اسمی به مدل دوپارامتری ۱۵۷
- شکل ۴-۲۷ : نمودار منحنی توابع آگاهی در دو مدل برای بخش دشوارتر آزمون ۱۶۱
- شکل ۴-۲۸ : نمودار منحنی توابع آگاهی در دو مدل برای بخش آسان تر آزمون ۱۶۱

فهرست جداول

- جدول ۲-۱: توابع ریاضی $I_i(\theta)$ ، $I_{\max}(\theta)$ و θ_{\max} برای مدل‌های دو ارزشی منطقی ۱، ۲ و ۳ پارامتری ۴۹
- جدول ۲-۲: مقادیر عددی پارامترهای یک سؤال ریاضی ۴ گزینه‌ای با استفاده از NRM ۶۵
- جدول ۲-۳: چارچوب کلی سامی‌جیما برای داده‌های چند ارزشی ۸۵
- جدول ۲-۴: نمودار طبقه‌بندی تیسن و استینبرگ برای مدل‌های پارامتریک و تک‌بعدی IRT ۸۸
- جدول ۴-۱: توزیع فراوانی گزینه‌های پاسخ سؤالات خرده‌آزمون زیست‌شناسی ۱۲۶
- جدول ۴-۲: آماره‌های مربوط به همبستگی سؤال-آزمون کلیه سؤالات خرده‌آزمون زیست‌شناسی ۱۲۷
- جدول ۴-۳: آماره‌های مربوط به همبستگی سؤال-آزمون ۴۳ سوال باقیمانده پس از حذف سؤالات در مرحله اول ۱۲۸
- جدول ۴-۴: آماره‌های مربوط به همبستگی سؤال-آزمون ۳۸ سوال باقیمانده پس از حذف سؤالات در مرحله دوم ۱۲۹
- جدول ۴-۵: آماره‌های مربوط به آزمون ۳۲ سوالی منتخب پس از حذف سؤالات در مرحله سوم ۱۳۰
- جدول ۴-۶: شاخص‌های برازش و پایایی مدل‌های سؤال-پاسخ مورد بررسی در نمونه ۳۸ سوالی ۱۳۴
- جدول ۴-۷: تفاضل نسبت احتمال مشاهده شده و مورد انتظار گزینه‌های پاسخ نمونه ۳۸ سوالی در 2pl ۱۳۶
- جدول ۴-۸: تفاضل نسبت احتمال مشاهده شده و مورد انتظار گزینه‌های پاسخ نمونه ۳۸ سوالی در NRM ۱۳۷
- جدول ۴-۹: شاخص‌های برازش و پایایی مدل‌های سؤال-پاسخ مورد بررسی در نمونه ۳۲ سوالی برگزیده ۱۳۷
- جدول ۴-۱۰: پارامترهای برآورد شده سؤالات در مدل دو پارامتری ۱۳۹
- جدول ۴-۱۱: پارامترهای برآورد شده گزینه‌های پاسخ سؤالات در مدل پاسخ اسمی ۱۴۷
- جدول ۴-۱۲: مقایسه میزان آگاهی‌دهندگی مدل پاسخ اسمی و مدل دو پارامتری ۱۵۶
- جدول ۴-۱۳: مقایسه میانگین خطای معیار برآورد توانایی در دو مدل پاسخ اسمی و دو پارامتری ۱۵۸
- جدول ۴-۱۴: مقایسه میانگین کارایی نسبی NRM به مدل 2pl در دو گروه آزمودنی‌های ضعیف و قوی ۱۵۹
- جدول ۴-۱۵: مقایسه میانگین کارایی نسبی NRM به مدل 2pl در دو بخش دشوارتر و آسان‌تر آزمون ۱۵۹
- جدول ۴-۱۶: کارایی نسبی NRM به مدل 2pl در دو بخش دشوارتر و آسان‌تر آزمون ۱۶۰

کلیات پژوهش

مقدمه

یکی از نیازهای اساسی جوامع، شناخت استعدادها، توانایی‌ها، رفتارها و شخصیت افراد است و بکارگیری آزمون‌های مناسب در حوزه‌های مختلف، از سودمندترین و اقتصادی‌ترین راهکارها در جهت مرتفع ساختن این نیاز می‌باشد. این آزمون‌ها به منظور ارزیابی رفتار، توانایی‌های ذهنی و دیگر خصوصیات شخصی طراحی می‌شوند و جهت کمک به قضاوت، پیش‌بینی و بالاخره تصمیم‌گیری دربارهٔ افراد، مورد استفاده قرار می‌گیرند.

اگرچه آزمون‌ها به دلایل متعددی غالباً نمی‌توانند اساس و معیار کاملاً دقیقی جهت اتخاذ تصمیمات کاملاً درست دربارهٔ افراد فراهم آورند، با این حال می‌توان ادعا کرد که آنها بهترین، عادلانه‌ترین و اقتصادی‌ترین روش بدست آوردن اطلاعات لازم جهت تصمیم‌گیری‌های مهم دربارهٔ افراد هستند (مورفی و شوfer^۱، ۱۹۹۱، به نقل از مینایی، ۱۳۷۶).

با توجه به اهمیت کاربردی آزمون‌ها در تصمیم‌گیری‌های اساسی، استفاده از آزمون‌هایی که بتوانند توانایی‌ها و شایستگی‌های افراد را با دقت هرچه بیشتر شناسایی و برآورد نمایند، امری ضروری و غیر-قابل اجتناب است.

یکی از متداول‌ترین کاربردهای آزمون‌ها، گزینش و یا سرند کردن افراد است که با توجه به ملاک‌های معین صورت می‌گیرد. آزمون‌های ورودی دانشگاه‌ها و برخی مدارس و مؤسسات آموزشی، از این گروه آزمون‌ها می‌باشند که به واسطهٔ آنها می‌توان با در نظر گرفتن ملاک‌های معین، به اندازه‌گیری سطح

^۱ Morphy & shofer

توانایی داوطلبین ورود به این مراکز، جهت گزینش پذیرفته‌شدگان اقدام نمود. بدیهی است که ضرورت و اهمیت دقت و صحت چنین اندازه‌گیری‌هایی، با توجه به تأثیرات مستقیمی که تصمیمات اتخاذ شده بر پایه آنها، در آینده افراد و به تبع آن جامعه خواهد داشت، غیر قابل انکار است. لذا پیوسته سعی بر این بوده و خواهد بود که اولاً با بکارگیری ملاک‌های صحیح و معتبر و رعایت اصول لازم در طراحی و ساخت این آزمون‌ها، روایی و پایایی هرچه بیشتر سؤالات آزمون تأمین گردد و ثانیاً با استفاده از تئوری‌های اندازه‌گیری و مدل‌های مناسب جهت تحلیل پاسخ‌ها، بتوان دقت برآورد سطح توانایی آزمودنی‌ها را افزایش داد.

مبنا و اصول کلی ساخت و نمره‌گذاری سؤالات آزمون‌ها را نظریه‌ها و مدل‌های اندازه‌گیری زیرساختی آنها تعیین می‌کنند و این مدل‌ها نیز مبتنی بر مفروضات زیربنایی خاص خود هستند. بدیهی است که در صورت فراهم نبودن این پیش‌فرض‌ها و شرایط لازم، بکارگیری این مدل‌ها مشکل‌ساز بوده و دستاوردهای مبتنی بر آنها صحیح نخواهد بود. اگر چه مسلماً رعایت دقیق و کامل مفروضات نیز در غالب موارد، بسیار مشکل و گاه غیر ممکن است. لذا انتخاب بهینه یک مدل اندازه‌گیری، با توجه به شرایط اندازه‌گیری و ویژگی‌های داده‌ها و میزان برآزش داده‌ها با مدل، مسأله‌ای اساسی است که مستلزم شناخت دقیق مدل‌های مختلف اندازه‌گیری و شرایط و محدودیت‌های بکارگیری آنها می‌باشد.

مبانی و مفروضات " **نظریه کلاسیک اندازه‌گیری** ^۱ " که از اوایل قرن بیستم، مبنا و زیربنای ساخت و تحلیل غالب آزمون‌های روانی- تربیتی بوده و از نقش تعیین‌کننده و بسیار مهمی در رشد و توسعه آزمون‌ها برخوردار است، سالهاست که مورد انتقاد متخصصان روان‌سنجی جدید قرار گرفته است. زیرا بنابر نظر این متخصصان، این نظریه قادر نیست پارامترهایی را که آزمودنی به‌وسیله آنها به سؤالات پاسخ می‌دهد، به‌درستی توصیف نماید و با محدودیت‌های اساسی در زمینه ساخت و ارزشیابی آزمون‌ها مواجه است. این محدودیت‌ها منجر به شکل‌گیری نظریه دیگری در رابطه با ساخت و تفسیر نمرات آزمون گردید که از آن به‌عنوان " **نظریه خصیصه مکنون** ^۲ " نام برده می‌شود و از قابلیت تعیین جایگاه سؤالات و آزمودنی‌ها در یک مقیاس مشترک برخوردار است. این نظریه معمولاً هنگامی که در مورد آزمون‌های توانایی و یا پیشرفت بکار می‌رود، تحت عنوان " **نظریه سؤال- پاسخ** ^۳ " نامیده می‌شود. در طول سه دهه اخیر، مدل‌های IRT، همگام با گسترش کاربرد انواع مختلف آزمون، رشد و توسعه قابل ملاحظه‌ای یافته است.

^۱ Classical test theory (CCT)

^۲ Latent Trait Theory

^۳ Item Response Theory (IRT)

یکی از ابعاد این توسعه، بسط و توسعه "مدل‌های سؤال-پاسخ دوارزشی"^۱ به "مدل‌های سؤال-پاسخ چند ارزشی"^۲ می‌باشد. قدیمی‌ترین و شاید شناخته‌شده‌ترین این مدل‌ها، "مدل پاسخ درجه‌بندی شده" (GRM)^۳ است که در اواخر دهه شصت قرن گذشته توسط سامی‌جیما^۴ (۱۹۶۹) ارائه شد.

نوع دیگری از مدل‌های سؤال-پاسخ چند ارزشی، "مدل پاسخ اسمی" (NRM)^۵ است که توسط باک^۶ (۱۹۷۲) مطرح و سپس توسط پژوهشگران متعددی بسط و توسعه یافته است. بررسی مدل‌های چند ارزشی IRT، نشان می‌دهد که صرفاً این دو مدل مذکور، رویکردهای منحصر به فردی را دنبال می‌کنند و سایر مدل‌های چند ارزشی عمدتاً از تعمیم و یا تخصیص یافتن این دو مدل بدست می‌آیند. شاید بتوان گفت، مهم‌ترین عامل احساس نیاز به توسعه و بکارگیری مدل‌های چند ارزشی، وجود آزمون‌هایی با سؤالات دارای پاسخ چندگزینه‌ای و ضرورت استفاده از این نوع سؤالات به جای سؤالات دوگزینه‌ای در زمینه‌های متفاوت به‌ویژه در زمینه متغیرهای شخصیتی و اجتماعی بوده است. زیرا در چنین حوزه‌هایی، سنجش تفاوت‌های جزئی ظریف‌تر از آنچه سؤالات دو ارزشی قادر به سنجیدن آن هستند، مورد نیاز می‌باشد و به عبارتی سؤالاتی که صرفاً دو گزینه برای انتخاب دارند، نمی‌توانند اطلاعات کافی را در زمینه صفت مورد بررسی، منتقل نمایند و بعضاً موجب دل‌سردی پاسخ‌دهندگان می‌شوند. لذا در چنین مواردی بکارگیری سؤالات چند ارزشی که به دلیل امکان ایجاد طبقات بیشتر می‌تواند طیف وسیع‌تری از پیوستار صفت را پوشش دهد، ضرورت پیدا می‌کند. البته قبل از شکل‌گیری مدل‌های سؤال-پاسخ نیز انواع مختلف سؤالات چندگزینه‌ای در آزمون‌ها مورد استفاده قرار می‌گرفت و با استفاده از رویکردهای مقیاس‌پردازی ترستون^۷ و یا لیکرت^۸ ارزیابی می‌شد. اگرچه برخی از پیش-فرض‌های این دو رویکرد (پیش‌فرض توزیع نرمال صفت در جامعه برای مقیاس ترستون و پیش‌فرض وجود رابطه خطی بین احتمال پاسخ و صفت زیرساختی در مقیاس لیکرت) معمولاً محقق نمی‌شوند.

در این پژوهش، پس از انتخاب یک مدل چند ارزشی مناسب برای آزمون‌های چندگزینه‌ای، مقایسه‌ای میان دقت برآورد توانایی پاسخ‌دهندگان به سؤالات خرده‌آزمون زیست‌شناسی کنکور سراسری سال تحصیلی ۸۴-۸۵، با استفاده از مدل چند ارزشی مذکور و یک مدل دو ارزشی که از برازش قابل قبول با

^۱ Dichotomous IRT Models

^۲ Polytomous IRT Models

^۳ Graded Response Model

^۴ Samejima

^۵ Nominal Response Model

^۶ Bock

^۷ Thurstone

^۸ Likert

داده‌ها برخوردار باشد، صورت می‌گیرد و از این طریق امکان بازیابی اطلاعات مربوط به سطح توانایی افراد از روی پاسخ‌های نادرست آنها، نشان داده می‌شود. علاوه بر این، دقت برآورد توانایی آزمودنی‌ها با استفاده از مدل‌های مذکور، برای سؤالات دشوارتر در مقایسه با سؤالات آسان‌تر و همچنین برای آزمودنی‌های ضعیف‌تر (در مقایسه با آزمودنی‌های قوی‌تر) به تفکیک مورد بررسی و مقایسه قرار خواهد گرفت.

بیان مسأله

طراحی و به‌اجرا درآوردن آزمون‌ها، اصولاً به‌منظور شناسایی صفات و توانایی‌های افراد و اتخاذ تصمیماتی بر مبنای نتایج آن صورت می‌گیرد. از آنجایی که در صورت عدم برخوردارگی یک آزمون از دقت برآورد کافی، اعتبار نتایج حاصل از آن مورد تردید قرار می‌گیرد، هر اندازه تصمیمات اتخاذ شده بر مبنای نتایج یک آزمون از اهمیت و حساسیت بالاتری برخوردار باشد، دقت برآورد توانایی یا صفت مورد سنجش در آن، بیشتر مورد توجه قرار می‌گیرد. یکی از مهمترین و فراگیرترین آزمون‌هایی که از حساسیت و جایگاه تعیین‌کننده‌ای در جامعه ما برخوردار است، آزمون ورودی دانشگاه‌هاست که در مقاطع مختلف تحصیلی، هر ساله صدها هزار نفر از افراد جامعه را تحت پوشش قرار می‌دهد.

با توجه به اینکه آزمون‌های ورودی دانشگاه‌ها و برخی دیگر مراکز آموزشی بزرگ، عمدتاً به‌صورت فراگیر و برای تعداد بسیار زیادی آزمودنی مورد استفاده قرار می‌گیرند (مانند کنکور سراسری)، استفاده از سؤالات چندگزینه‌ای نسبت به دیگر انواع سؤالات عینی و همینطور تشریحی، با وجود نقاط ضعف موجود در این گونه سؤالات، برتری داشته و رایج می‌باشد. از میان مزایای عمده سؤالات چندگزینه‌ای، می‌توان به قابلیت گنجایش تعداد بیشتری سؤال که معرف نمونه جامع‌تری از رفتار و یا محتوای مورد بررسی باشد و همچنین عینی بودن و سهولت نمره‌گذاری آن اشاره کرد.

مدلی که تئوری کلاسیک جهت تحلیل نتایج آزمون‌های توانایی متشکل از سؤالات چندگزینه‌ای ارائه می‌دهد، مدلیست که از آن به‌عنوان "مدل دانش- یا- حدس تصادفی"^۱ نام برده می‌شود. ضعف اساسی این مدل در این است که فرض می‌کند در صورتی که آزمودنی قادر به بازشناسی پاسخ صحیح نباشد، در زمینه آن سؤال فاقد هرگونه آگاهی بوده و بنابراین برای انتخاب از میان گزینه‌های پاسخ موجود هیچ مبنایی در اختیار نداشته و ناگزیر به‌صورت تصادفی پاسخی را بر می‌گزیند (لرد و نوک، ۱۹۶۸). مسلماً چنین فرضی چندان معقول به‌نظر نمی‌رسد. زیرا پاسخ‌های غلط در سؤالات چندگزینه‌ای (گزینه‌های انحرافی)^۲ طوری طراحی می‌شوند که برای آزمودنی‌های با سطوح مختلف

^۱ Knowledge- or- Random guessing model

^۲ Lord & Novick

^۳ Distractors

توانایی قدرت جذب متفاوتی داشته باشند. بنابراین آزمودنی‌هایی که اطلاعات نادرست و یا ناقصی در زمینه آن سؤال دارند، معمولاً به‌طور یکسان جذب همه گزینه‌ها نمی‌شوند و تنها در شرایطی که آزمودنی بسیار ضعیف (فاقد آگاهی) و یا سؤال بسیار دشوار باشد، ممکن است انتخاب گزینه پاسخ، صرفاً از روی حدس تصادفی صورت پذیرد. به‌همین ترتیب، هرچه آزمودنی قوی‌تر باشد (در صورت طراحی مناسب گزینه‌ها) قادر به حذف تعداد بیشتری از گزینه‌های نادرست است.

بنابراین در چنین شرایطی مدل واقع‌گرایانه‌تر این است که فرض کنیم آزمودنی پس از حذف یک یا چند گزینه انحرافی که به نادرست بودن آن آگاهی دارد، اقدام به حدس از میان سایر گزینه‌ها خواهد کرد. لذا مطابق نظر لرد و نوپک (۱۹۶۸)، تعداد گزینه‌های نادرست که با علم و آگاهی حذف می‌شوند، تابعی از سطح توانایی آزمودنی است و می‌توان برای گزینه‌های نادرست در یک سؤال نمرات متفاوتی در نظر گرفت.

بر اساس نظریه کلاسیک، در نمره‌گذاری سوالات چندگزینه‌ای، نمره کل هر فرد برابر با جمع نمرات پاسخ‌های صحیح اوست و یا بر اساس مفروضه حدس تصادفی، قدرت جذب همه گزینه‌ها برای شخصی که پاسخ را نمی‌داند، یکسان فرض می‌شود و نمره کل با استفاده از فرمول تصحیح برای جبران حدس (احتساب امتیاز منفی برای پاسخ‌های غلط) محاسبه می‌شود.

نظریه سؤال - پاسخ (IRT) شامل خانواده‌ای از مدل‌های ریاضی است که روابط متغیرهای مشاهده - پذیر و سازه‌های صفات فرضی زیربنایی این متغیرها را نمایش می‌دهد. این نظریه از لحاظ مفهومی، نیرومندتر از CTT است و در طول نیم قرن گذشته گسترش و کاربرد شگرفی یافته است و در حال حاضر از مهمترین ابزارهایی است که جهت تهیه و تحلیل آزمون‌ها بکار گرفته می‌شود و تا حدی فراگیر شده است که به‌جرات می‌توان گفت دوره نظریه کلاسیک اندازه‌گیری سر آمده است.

تحلیل سوالات در آزمون‌های توانایی متشکل از سوالات چندگزینه‌ای با استفاده از مدل‌های دو ارزشی IRT نیز همانند CTT با این محدودیت روبرو است که در آن، برای گزینه‌های نادرست هویت مستقلی لحاظ نمی‌شود. در این مدل‌ها معمولاً برای پاسخ درست ارزش "یک" و برای پاسخ به گزینه‌های نادرست ارزش "صفر" در نظر گرفته شده و پارامترهای سؤال صرفاً برای گزینه درست برآورد می‌شوند.

در صورتی که گزینه‌های انحرافی سوالات، حاوی اطلاعاتی در رابطه با سطح توانایی پاسخ‌دهندگان باشد، احتمال انتخاب هر یک از گزینه‌های نادرست، توسط آزمودنی‌ها در سطوح توانایی متفاوت، یکسان نخواهد بود و نمی‌توان همه پاسخ‌های نادرست را هم‌ارزش دانست. مورد توجه قرار ندادن اطلاعات بالقوه موجود در گزینه‌های نادرست انتخابی افراد، موجب از دست رفتن اطلاعات مفیدی خواهد شد که می‌توانست نقش مؤثری در افزایش دقت برآورد سطح توانایی آزمودنی‌ها ایفا نماید.

لذا استفاده از یک مدل چند ارزشی مناسب، می‌تواند آگاهی‌دهنده‌تر از مدل دو ارزشی باشد و افزایش

دقت برآورد سطح توانایی آزمودنی‌ها را موجب خواهد شد. زیرا در صورت استفاده از مدل دو ارزشی، همه پاسخ‌های نادرست در یک طبقه در نظر گرفته شده و اطلاعات موجود در آنها مورد ارزیابی قرار نمی‌گیرد. اما یک مدل چند ارزشی (مانند مدل پاسخ اسمی) از این مزیت برخوردار است که احتمال انتخاب هر گزینه را به‌عنوان تابعی از سطح توانایی برآورد می‌کند و پاسخ‌های غلط را هم‌ارزش نمی‌داند. و از آنجا که انتخاب هر گزینه تابعی از سطح توانایی است، بطور بالقوه می‌تواند اطلاعاتی هم درباره سطح توانایی پاسخ‌دهنده در بر داشته باشد. از این رو مدل‌های چند ارزشی مدعی هستند که دقت برآورد توانایی را نسبت به مدل‌های دو ارزشی افزایش می‌دهند و این افزایش دقت مدل‌های چند ارزشی پیامد معمول در کاربرد این مدل‌هاست (همبلتون^۱، ۱۹۸۹).

از سوی دیگر، با توجه به اینکه پاسخ‌های نادرست، اغلب در مورد سؤالات دشوارتر مشاهده می‌شود و همچنین افرادی که از سطح توانایی پایین‌تری برخوردارند، پاسخ‌های نادرست بیشتری خواهند داشت، انتظار می‌رود که مدل‌های چند ارزشی، دقت برآورد توانایی را در سؤالات دشوارتر و در سطوح پایین‌تر توانایی، بیشتر ارتقاء بخشند. البته طبیعی است که دقت و آگاهی‌دهندگی بیشتر مدل چند ارزشی نسبت به مدل دو ارزشی، در صورتی محقق خواهد شد که در گزینه‌های انحرافی سؤالات مذکور اطلاعاتی در رابطه با سطح توانایی پاسخ‌دهندگان موجود باشد و در غیر این صورت نمی‌توان در مورد داده‌های بکار گرفته شده، دقت بیشتر مدل‌های چند ارزشی را انتظار داشت.

انتخاب یکی از خرده‌آزمون‌های کنکور سراسری به‌عنوان آزمون مورد استفاده در این پژوهش، به‌دلیل نقش پررنگ و تعیین‌کننده این آزمون در جامعه ما و همچنین ضرورت دسترسی به بانک اطلاعاتی نمونه‌ای با حجم نسبتاً بزرگ (فراخور مدل‌های چند ارزشی) بوده است. انتخاب خرده‌آزمون درس زیست‌شناسی نیز، به‌دلیل اهمیت بسیار زیاد آن در تعیین جایگاه آزمودنی‌های گروه علوم تجربی و در نتیجه، وجود انگیزه قوی آزمودنی‌ها در رابطه با مطالعه و پاسخ‌گویی دقیق (با صرف زمان کافی) به سؤالات این خرده‌آزمون می‌باشد. بدیهی است، در صورتی که می‌توانستیم آزمونی را انتخاب کنیم که سؤالات آن از ابتدا طوری طراحی شده باشند که همه گزینه‌ها شامل اطلاعات مفیدی در زمینه توانایی آزمودنی‌ها باشد، مزیت بکارگیری مدل چند ارزشی بیشتر نمایان می‌شد.

در رابطه با انتخاب مدل چند ارزشی مناسب باید گفت، اگرچه برای سؤالات چندگزینه‌ای می‌توان از طیف گسترده‌ای از مدل‌های چند ارزشی استفاده کرد، اما از آنجاییکه آزمون‌های پیشرفت تحصیلی نظیر سؤالات کنکور معمولاً طوری طراحی نمی‌شوند که بتوان گزینه‌های پاسخ را برحسب درجه حصول به پاسخ صحیح و یا گذراندن مراحل از حل مسأله مرتب نمود (همانند آنچه در مدل‌های GRM و یا

PCM^۱ مورد نظر است)، بنابراین می‌بایست از مدلی استفاده شود که برای هر نوع پاسخ اسمی مناسب باشد. لذا بکارگیری مدل پاسخ اسمی باک (BNM^۲ یا NRM) و یا برخی مدل‌های چندگزینه‌ای که از بسط و توسعه این مدل حاصل شده‌اند، مانند مدل C چندگزینه‌ای سامیجیما (SMCM^۳، ۱۹۷۹) و یا مدل چندگزینه‌ای تیسن و استینبرگ (TSMCM^۴، ۱۹۸۴)، به‌عنوان مدل چند ارزشی مناسب برای این پژوهش امکان‌پذیر می‌باشد. بنابر نتایج حاصل از برخی پژوهش‌ها که در فصل آتی به آنها خواهیم پرداخت، چنین به نظر می‌رسد که افزودن پارامترهای اضافه به مدل اسمی در دو مدل فوق‌الذکر، تأثیر قابل توجهی در میزان برازش مدل و دقت برآورد نخواهد داشت. از طرف دیگر علاوه بر پیچیدگی مدل و روش‌های برازاندن آن در مدل چندگزینه‌ای تیسن و استینبرگ، بنابر اظهارات تیسن، پارامترهای این مدل نیز به‌سهولت قابل تفسیر نیستند (تیسن و استینبرگ، ۱۹۸۴). لذا در این پژوهش مدل پاسخ اسمی باک به‌عنوان مدل چند ارزشی مناسب انتخاب می‌شود. در این مدل، پارامتری برای عامل حدس در نظر گرفته نشده‌است (مانند مدل لجستیک دو پارامتری)، و در واقع می‌توان مدل لجستیک دو-پارامتری (2PL) را حالت خاصی از مدل پاسخ اسمی دانست که در آن تعداد طبقات پاسخ، به دو طبقه محدود می‌شود. بنابراین در صورتی که این مدل از برازش قابل قبولی با داده‌های این پژوهش برخوردار باشد، به‌عنوان مدل دو ارزشی مورد نظر در این پژوهش انتخاب خواهد شد. برآورد پارامترهای دو مدل با استفاده از نرم افزار MULTILOG (۷) (تیسن، ۱۹۹۵)، صورت می‌پذیرد.

اهمیت و ضرورت انجام پژوهش

با وجود اینکه قریب به ۴۰ سال از شکل‌گیری مقدماتی‌ترین مدل‌های چند ارزشی تئوری سؤال-پاسخ می‌گذرد و کارایی این مدل‌ها در مقایسه با مدل‌های دو ارزشی و به‌تبع آن، مدل کلاسیک آزمون، در زمینه‌های متفاوت و انواع مختلف سؤال، در طول این سالها، بارها در پژوهش‌های متعدد و متفاوتی، مورد بررسی قرار گرفته و عمدتاً به‌اثبات رسیده‌است و علی‌رغم رواج رو به افزایش کاربرد این مدل‌ها در ساخت و تحلیل نتایج انواع مختلف آزمون‌ها در حیطه‌های مختلف سنجش، متأسفانه تا به‌امروز استفاده از مدل‌های چند ارزشی IRT به حوزه‌های کاربردی سنجش و اندازه‌گیری در جامعه ما راه نیافته است. یکی از مهمترین آزمون‌های جامعه علمی ما، کنکور ورودی دانشگاه‌ها است که سعی دارد، تواناترین افراد را از روی پاسخ‌های آنها به سؤالات مختلف آزمون، انتخاب و راهی دانشگاه‌ها سازد. با فرض اینکه سؤالات این آزمون‌ها از دقت و صلاحیت کافی برای تعیین سطح توانایی آزمودنی‌ها برخوردار باشند،

^۱ Partial Credit Model

^۲ Bock Nominal Model

^۳ Samejima's multiple-choice Model C

^۴ Thissen & Steinberg multiple-choice model

استفاده از یک مدل چند ارزشی مناسب به منظور تحلیل سؤالات و پاسخ‌های افراد، تلاش در جهت افزایش دقت انتخاب سؤالات مناسب و به دنبال آن تصمیم‌گیری دقیق‌تر در مورد آزمودنی‌هاست. اگرچه در این پژوهش مدل چند ارزشی مورد نظر در مقابل یک مدل دو ارزشی معرفی و مقایسه خواهد شد باید در نظر داشت که تا به امروز، مدل‌های دو ارزشی سؤال-پاسخ نیز در جامعه علمی ما جز در پژوهش‌های دانشجویی معدود، جایگاهی نداشته‌است.

بطور کلی می‌توان گفت، رسالت عمده این پژوهش، نشان دادن یک روش دقیق‌تر قدرسنجی سؤالات و افراد در آزمون‌های چندگزینه‌ای است. در واقع سؤال اصلی که ذهن پژوهشگر را به سمت انجام این پژوهش ترغیب کرده است این است که: آیا (و چگونه) می‌توان از اطلاعات موجود در پاسخ‌های غلط به سؤالات چندگزینه‌ای به منظور افزایش دقت برآورد توانایی آزمون‌شوندگان استفاده کرد؟ امید است این پژوهش و پژوهش‌هایی از این دست، گام مثبتی در راستای معرفی و بکارگیری مدل‌های چند ارزشی و به‌طور کلی تئوری سؤال-پاسخ در ارزیابی نتایج آزمون‌ها محسوب گشته و بتواند زمینه‌ساز بکارگیری این مدل‌ها در ساخت و تحلیل نتایج آزمون‌های تعیین‌کننده و سرنوشت‌ساز جامعه ما بوده و در نتیجه رشد کیفی و افزایش دقت اندازه‌گیری این آزمون‌ها را به همراه داشته باشد.

اهداف و سؤالات پژوهش

اهداف پژوهش:

هدف اصلی این پژوهش، معرفی مدلی مناسب برای استفاده از اطلاعات موجود در پاسخ‌های غلط به سؤالات یک آزمون چندگزینه‌ای به منظور افزایش دقت برآورد توانایی آزمودنی‌هاست که اهداف جنبی ذیل را نیز به دنبال دارد:

- بررسی و مقایسه برآزش داده‌ها در مدل‌های دو و چند ارزشی IRT
- انتخاب مدل دو و چند ارزشی مناسب که از برآزش کافی با داده‌های سؤال-پاسخ این پژوهش برخوردار باشند.
- بررسی و مقایسه میزان دقت برآورد توانایی آزمودنی‌ها و میزان آگاهی‌دهندگی کل سؤالات در سطوح مختلف توانایی با استفاده از مدل‌های دو و چند ارزشی منتخب.
- بررسی میزان تأثیر سطح دشواری سؤالات و سطح توانایی آزمودنی‌ها در افزایش دقت برآورد توانایی آنها در مدل چند ارزشی مورد نظر نسبت به مدل دو ارزشی برگزیده.

سؤالات پژوهش:

در رابطه با داده‌های سؤال-پاسخ خرده‌آزمون زیست‌شناسی کنکور سراسری سال تحصیلی ۸۴-۸۵ رشته علوم تجربی: