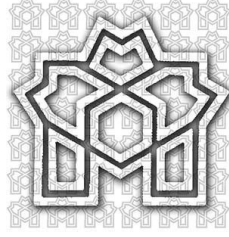


بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه علامه طباطبائی

دانشکده روانشناسی و علوم تربیتی

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

موضوع پایان نامه

تاثیر تعداد گزینه های سوال در ویژگی های روانسنجی آزمون و توانایی برآورد شده

در مدل های پرسش پاسخ و کلاسیک اندازه گیری

استاد راهنما: جناب دکتر محمد رضا فلسفی نژاد

استاد مشاور: جناب دکتر فریبرز درتاج

استاد داور: جناب دکتر علی دلاور

محقق: بهنام کریمی خرامه

گفت استاد مبر درس از یاد
یاد باد آنچه به من گفت استاد
یاد باد آن که مرا یاد آموخت
آدی نان خورد از دوات یاد
بچه یادم نرود بان معنی
که مرا مادر من نادان زاد
پدرم زینر چه استادم دید
گشت از تیرت من آزاد
اس مرا منت از استاد بود
که تله م من استاد استاد
هر چو من دانت آموخت مرا
غیر یک اصل که نگفت نهاد
قدر استاد نکو دلمه تن
حرف استاد به من یاد نداد
خدا یارش باد
(م. آزاد)

چکیده

هدف از پژوهش حاضر بررسی تاثیر تعداد گزینه ها بر ویژگی های روانسنجی آزمون ها و سوالات و همچنین توانایی برآورد شده آزمودنی ها در نظریه کلاسیک و پرسش پاسخ بود. جامعه آماری عبارت است از کلیه دانش آموزان سال سوم دبیرستان شهر شیراز و گروه نمونه عبارت از ۶۰۸ نفر از دانش آموزان سال سوم دبیرستان بود که به شیوه تصادفی از جامعه آماری مورد نظر انتخاب شد. جهت پاسخ گویی به سوالات پژوهش از روش تجربی استفاده گردید. برای جمع آوری اطلاعات از دو آزمون زبان و حسابان که به همین منظور تهیه و تنظیم شده بودند استفاده شد قبل از به کار گیری این آزمون نسبت به مقبولیت روایی و اعتبار آزمون به شیوه مختلف از جمله آلفای کرونباخ و تحلیل عاملی اطمینان حاصل شد. برای کاربرد نظریه پرسش پاسخ مفروضات، تک بعدی بودن و استقلال موضعی نیز چک شد.

تجزیه و تحلیل آماری داده ها نشان داد که تعداد گزینه ها بر پارامتر های سوال اثر ندارد و تاثیر تعداد گزینه ها بر ویژگی های روان سنجی برآورد شده آزمودنی ها، در آزمون های مختلف یکسان است. همچنین بین پارامتر های برآورد شده در نظریه کلاسیک و پرسش پاسخ تفاوت مشاهده شد.

بعد از بررسی مفروضه های نظریه پرسش پاسخ مشخص شد که داده ها با مدل دو پارامتری برازش بهتری دارند، و تفاوتی بین تعداد گزینه ها و برازش با مدل مشاهده نشد. همچنین بین توانایی برآورد شده و تعداد گزینه ها تفاوت مشاهده شد.

فهرست

شماره صفحه

فصل اول.....	۱۰
بیان مسئله.....	۱۴
اهمیت و ضرورت پژوهش.....	۲۰
اهداف تحقیق.....	۲۲
سوالات تحقیق.....	۲۲
فصل دوم.....	۲۴
مبانی نظری پژوهش.....	
نظریه کلاسیک اندازه گیری.....	۲۷
نارسایی نظریه کلاسیک.....	۳۲
نظریه پرسش پاسخ.....	۳۶
پیش فرض های نظریه سوال پاسخ.....	۳۸
مدل های IRT و ویژگی های هر یک.....	۳۹
انتخاب یک مدل سوال پاسخ مناسب.....	۴۳
سه روش برآورد بیشینه درست نمایی برای سطوح صفت نامعلوم.....	۴۷
بیشینه درست نمایی هم زمان (JML).....	۴۰
بیشینه درست نمایی کناری (MML).....	۵۱
بیشینه درست نمایی شرطی (CML).....	۵۲
مقدمه ای بر آزمون های چند گزینه ای.....	۵۳
پیشینه پژوهش در ایران.....	۵۸
پیشینه پژوهش در خارج از ایران.....	۶۲
فصل سوم.....	۷۳
روش اجرای پژوهش.....	۷۴
نمونه و روش نمونه گیری.....	۷۶
روش تجزیه و تحلیل داده ها.....	۸۰
فصل چهارم.....	۸۲

۸۵	بررسی مفروضات
۱۰۶	استفاده از Z برای مقایسه پارامترها
۱۱۳	استفاده از منطق DIF برای مقایسه پارامترها
۱۱۸	استفاده از χ^2 برای بررسی تفاوت توانایی ها
۱۲۴	فصل پنجم
	بحث و نتیجه گیری
۱۲۵	مقدمه
۱۳۱	نتیجه گیری
۱۴۳	محدودیت و پیشنهادات
۱۴۵	منابع فارسی
۱۴۷	منابع انگلیسی
	ضمایم

فهرست جداول

- جدول ۱-۳ تعداد دبیرستان های جامعه ۷۵
- جدول ۲-۳ تعداد افراد جامعه ۷۵
- جدول ۴-۳ نمونه سوال و آزمودنی ها در آزمون حسابان ۷۸
- جدول ۵-۳ نمونه سوال و آزمودنی ها در آزمون زبان ۷۸
- جدول ۱-۴ و ۲-۴ ضریب پایایی آزمون حسابان ۳ گزینه ای ۸۵
- جدول ۳-۴ و ۴-۴ ضریب پایایی آزمون حسابان ۴ گزینه ای ۸۶
- جدول ۵-۴ و ۶-۴ ضریب پایایی آزمون حسابان ۵ گزینه ای ۸۷
- جدول ۷-۴ و ۸-۴ ضریب پایایی آزمون زبان ۳ گزینه ای ۸۸
- جدول ۹-۴ و ۱۰-۴ ضریب پایایی آزمون زبان ۴ گزینه ای ۸۹
- جدول ۱۱-۴ و ۱۲-۴ ضریب پایایی آزمون زبان ۵ گزینه ای ۹۰
- جدول ۱۳-۴ اندازه KMO آزمون حسابان ۵ گزینه ای ۹۲
- جدول ۱۴-۴ اندازه KMO آزمون زبان ۳ گزینه ای ۹۲
- جدول ۱۵-۴ مقادیر ویژه و تحلیل عاملی آزمون حسابان ۳ گزینه ای ۹۳
- جدول ۱۶-۴ مقادیر ویژه و تحلیل عاملی آزمون حسابان ۴ گزینه ای ۹۴
- جدول ۱۷-۴ مقادیر ویژه و تحلیل عاملی آزمون حسابان ۵ گزینه ای ۹۵
- جدول ۱۸-۴ مقادیر ویژه و تحلیل عاملی آزمون زبان ۳ گزینه ای ۹۶
- جدول ۱۹-۴ مقادیر ویژه و تحلیل عاملی آزمون زبان ۴ گزینه ای ۹۷
- جدول ۲۰-۴ مقادیر ویژه و تحلیل عاملی آزمون زبان ۵ گزینه ای ۹۸
- جدول ۲۱-۴ پارامتر های آزمون حسابان ۳ گزینه ای ۱۰۰
- جدول ۲۲-۴ پارامتر های آزمون حسابان ۴ گزینه ای ۱۰۱
- جدول ۲۳-۴ پارامتر های آزمون حسابان ۵ گزینه ای ۱۰۲
- جدول ۲۴-۴ پارامتر های آزمون زبان ۳ گزینه ای ۱۰۳
- جدول ۲۵-۴ پارامتر های آزمون زبان ۴ گزینه ای ۱۰۴
- جدول ۲۶-۴ پارامتر های آزمون زبان ۵ گزینه ای ۱۰۵
- جدول ۲۷-۴ تبدیل ضریب تمیز آزمون حسابان به Z ۱۰۷

- جدول ۲۸-۴ تبدیل ضریب دشواری آزمون حسابان به Z ۱۰۸
- جدول ۲۹-۴ تبدیل ضریب حدس آزمون حسابان به Z ۱۰۹
- جدول ۳۰-۴ تبدیل ضریب تمیز آزمون زبان به Z ۱۱۰
- جدول ۳۱-۴ تبدیل ضریب دشواری آزمون زبان به Z ۱۱۱
- جدول ۳۲-۴ تبدیل ضریب حدس آزمون زبان به Z ۱۱۲
- جدول ۳۳-۴ بررسی تفاوت ضریب دشواری حسابان CCT,IRT ۱۱۶
- جدول ۳۴-۴ بررسی تفاوت ضریب دشواری زبان CCT,IRT ۱۱۶
- جدول ۳۵-۴ بررسی تفاوت ضریب تمیز حسابان CCT,IRT ۱۱۷
- جدول ۳۶-۴ بررسی تفاوت ضریب تمیز زبان CCT,IRT ۱۱۸
- جدول ۳۷-۴ آزمون خبی ۲ برای بررسی تفاوت توانایی آزمون های حسابان ۱۱۹
- جدول ۳۸-۴ آزمون خبی ۲ برای بررسی تفاوت توانایی آزمون های زبان ۱۲۰
- جدول ۳۹-۴ بررسی برازش با مدل حسابان ۳ گزینه ای ۱۲۰
- جدول ۴۰-۴ بررسی برازش با مدل حسابان ۴ گزینه ای ۱۲۰
- جدول ۴۱-۴ بررسی برازش با مدل حسابان ۵ گزینه ای ۱۲۰
- جدول ۴۲-۴ بررسی برازش با مدل زبان ۳ گزینه ای ۱۲۱
- جدول ۴۳-۴ بررسی برازش با مدل زبان ۴ گزینه ای ۱۲۱
- جدول ۴۴-۴ بررسی برازش با مدل زبان ۳ گزینه ای ۱۲۱
- جدول ۴۵-۴ مقدار آگاهی در آزمون های مختلف ۱۲۲
- جدول ۱-۵ پایایی کل آزمون ها ۱۲۷
- جدول ۲-۵ درصد پاسخ صحیح به کل آزمون های حسابان ۱۲۸
- جدول ۳-۵ درصد پاسخ صحیح به کل آزمون های زبان ۱۲۹

فهرست نمودارها

- نمودار ۱-۴ اسکری آزمون حسابان ۳ گزینه ای ۹۳
- نمودار ۲-۴ اسکری آزمون حسابان ۴ گزینه ای ۹۴
- نمودار ۳-۴ اسکری آزمون حسابان ۵ گزینه ای ۹۵
- نمودار ۴-۴ اسکری آزمون زبان ۳ گزینه ای ۹۶
- نمودار ۵-۴ اسکری آزمون زبان ۴ گزینه ای ۹۷
- نمودار ۶-۴ اسکری آزمون زبان ۵ گزینه ای ۹۸
- نمودار ۷-۴ ICC کل آزمون حسابان ۳ گزینه ای ۱۰۰
- نمودار ۸-۴ منحنی توزیع توانایی در آزمون حسابان ۳ گزینه ای ۱۰۰
- نمودار ۹-۴ ICC کل آزمون حسابان ۴ گزینه ای ۱۰۱
- نمودار ۱۰-۴ منحنی توزیع توانایی در آزمون حسابان ۴ گزینه ای ۱۰۱
- نمودار ۱۱-۴ ICC کل آزمون حسابان ۵ گزینه ای ۱۰۲
- نمودار ۱۲-۴ منحنی توزیع توانایی در آزمون حسابان ۵ گزینه ای ۱۰۲
- نمودار ۱۳-۴ ICC کل آزمون زبان ۳ گزینه ای ۱۰۳
- نمودار ۱۴-۴ منحنی توزیع توانایی در آزمون زبان ۳ گزینه ای ۱۰۳
- نمودار ۱۵-۴ ICC کل آزمون زبان ۴ گزینه ای ۱۰۴
- نمودار ۱۶-۴ منحنی توزیع توانایی در آزمون زبان ۴ گزینه ای ۱۰۵
- نمودار ۱۷-۴ ICC کل آزمون زبان ۵ گزینه ای ۱۰۵
- نمودار ۱۸-۴ منحنی توزیع توانایی در آزمون زبان ۵ گزینه ای ۱۱۳
- نمودار ۱۹-۴ بیشترین تفاوت در ICC ها در آزمون حسابان ۳ و ۴ گزینه ای ۱۱۴
- نمودار ۲۰-۴ کمترین تفاوت در ICC ها در آزمون حسابان ۳ و ۵ گزینه ای ۱۱۴
- نمودار ۲۱-۴ بیشترین تفاوت در ICC ها در آزمون زبان ۵ و ۴ گزینه ای ۱۱۵
- نمودار ۲۲-۴ کمترین تفاوت در ICC ها در آزمون زبان ۳ و ۵ گزینه ای ۱۱۵
- نمودار ۲۳-۴ تابع آگاهی کل آزمون های حسابان ۱۲۳
- نمودار ۲۴-۴ تابع آگاهی کل آزمون های زبان ۱۲۳

فصل اول

مقدمه:

دانش آدمی در رشته های مختلف علوم از طریق طرح نظریه های نوین، آزمایش درستی و نادرستی این نظریه ها و تلفیق یافته های حاصل با پیکره دانش قبل پیش میروند. هر نظریه جدید فوراً فرضیه هایی را درباره دنیای واقعی به دنبال می آورد، که آزمایش آنها درخت علم را تناور می سازد. اما هر فرضیه ای در اصل به رابطه بین دو یا چند متغیر اشاره دارد و بنابراین آزمایشگر برای مطالعه این رابطه ها مجبور است، متغیر های مندرج در فرضیه را به دقت تعریف و اندازه گیری کند (آناستازی ترجمه براهنی، ۱۳۷۱).

علم روانسنجی در نتیجه جنبش اندازه گیری روان در سال ۱۸۸۳ به وسیله زیست شناس انگلیسی گالتون، در ضمن پژوهش هایی در باره وراثت پایه ریزی شد (ثراندایک ترجمه هومن، ۱۳۶۹).

داوری درباره پیشرفت هر علم اغلب بر پایه درجه توفیق آن علم در استفاده از ریاضیات صورت می گیرد. روانسنجی نیز شیوه هایی است برای اندازه گیری های روانی و مقصود از اندازه گیری توصیف داده ها به صورت اعداد و ارقام است که این نیز به نوبه خود یعنی برخورداری از امکانات بسیاری که، عملیات ریاضی فراهم آورد.

دو نظریه مهم روانسنجی، نظریه کلاسیک اندازه گیری و نظریه سوال پاسخ می باشد که در قرن اخیر خدمات فراوانی به دیگر شاخه های روانشناسی و علوم رفتاری نموده اند. اساس نظریه

کلاسیک آزمون نخستین بار توسط اسپیرمن در خلال سالهای ۱۹۰۴ و ۱۹۱۳ بیان شد. بر پایه این مدل صفت یا خصیصه ای را که یک آزمون می سنجد ، می توان یک پیوستار مکنون فرض کرد . که هر یک از افراد روی آن محل مخصوصی را اشغال می کند جایگاه فرد روی این پیوستار و با نمره واقعی او در آزمون یا جایگاه او در مقیاس نمره های واقعی همبستگی کامل ولی غیر خطی دارد .

این مدل نمره تست 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 . خطای تصادفی به عنوان عاملی ناپسته به نمره حقیقی و نیز ناپسته به خطای ناشی از اندازه گیری از همان خصیصه مورد سنجش تعریف می گردد(ثراندایک ، ترجمه هومن، ۱۳۶۹).

به طور کلی هر نوع شرایطی که از لحاظ هدف آزمون نامربوط شناخته می شود، نمایشگر پراش خطاست . به این ترتیب هر گاه آزماینده سعی کند که شرایط آزمایش را همسان نگه دارد و برای این منظور محیط آزمایش ، دستورالعمل ، حدود زمانی ، همدلی و عوامل مشابه دیگر را کنترل کند در این صورت او در واقع می کوشد تا پراش خطا را کاهش دهد و نمره های آزمون را پایاتر سازد (آناستازی ، ترجمه براهنی، ۱۳۷۱.ص ۱۱۱).

همانگونه که گفته شد ، در این نظریه اعتقاد بر این است که هر فرد دارای کمیت مشاهده ناپذیری است که آن را نمره حقیقی می نامند این نمره هرگز مستقیماً قابل اندازه گیری نیست . در عوض در این تئوری ، هر فرد نمره مشاهده شده ای دارد که فرض میشود دارای مقداری خطاست و به عنوان برآوردی از نمره حقیقی مورد استفاده قرار می گیرد (ویس، ۱۹۹۴ به نقل از هومن، ۱۳۶۶).

نظریه خصیصه مکنون در بر گیرنده خانواده ای از مدل های ریاضی است . به گفته ثراندایک (۱۹۸۲، ترجمه هومن، ۱۳۶۹) در این مدل می خواهیم بدانیم ارتباط بین خصوصیات سوالهای

متشکل تست و تست واحدی که از ترکیب آنها بدست می آید و روابط بین نمره های تست و جایگاه امتحان شونده در خصیصه بنیادی آن نمره کدام است. علاوه بر این مایلیم بدانیم که در یک سلسله مراتب یا ترتیب کلی که معرف توانایی یا شخصیت آدمی است خصیصه های گوناگون چگونه مرتب میشوند.

نظریه پاسخ پرسش یا سوال پاسخ درصدد یافتن روابط تابعی بین عملکرد در سوال آزمون و صفات زیر بنایی آن عملکرد است. این صفات زیر بنایی در روی یک پیوستاری از $-\infty$ تا $+\infty$ واقع شده است که با حرف یونانی تتا (θ) نمایش داده میشود. نظریه پاسخ پرسش با مدل های سه پارامتری، دو پارامتری، یک پارامتری در واقع جنبه عملی و کاربردی نظریه صفت مکنون است. این نظریه در حقیقت کاربرد نظریه صفت مکنون به منظور مشخص ساختن طبقه ای از مدل های صفت مکنون بویژه برای سوال های دو ارزشی است که توسط لرد تدوین شده است. لرد در سال ۱۹۵۲ مقاله ای را منتشر کرد و در آن نظریه منحنی ویژه سوال را به عنوان یک مدل یا نظریه مطرح کرد وی در همین سال با به کار بردن یک نمونه صد هزار نفری از آزمودنی ها تحقیقی را در این زمینه ترتیب داد. این مطالعه و تحقیق ثابت کرد که استفاده از IRT برای سوالات چند گزینه ای مناسب است. نتایج نشان دادند که مدل سه پارامتری بهترین توصیف را از آزمون های چند گزینه ای با تتای واقعی آزمون به دست می دهند (ستاری، ۱۳۸۲).

در مورد نمرات در آزمون ها می توان گفت دامنه نمرات واقعی در نظریه کلاسیک آزمون، که برآوردی از توانایی فرد در یک حیطة خاص است محدود است. حداقل آن صفر و حداکثر آن، بیشترین مقدار نمره آزمون است. در حالی که در نظریه پاسخ پرسش صفت مکنون دامنه ای از $-\infty$ تا $+\infty$ را می توان پذیرد. در نظریه کلاسیک آزمون نمره هر آزمودنی معمولاً بر اساس تعداد پاسخ صحیح محاسبه میشود و لذا جایگاه فرد بر روی پیوستار مورد سنجش حالت گسسته و منفعل پیدا کرده و افراد الزاماً در فواصل خاص قرار می گیرند در حالی که در نظریه پاسخ پرسش

هر فرد آزمودنی بر حسب صفت مورد سنجش می تواند هر نقطه ای از پیوستار را اشغال کند و فواصل از پیش تعیین شده وجود ندارد (ستاری، ۱۳۸۲).

بیان مسئله:

روشهای سنجش توانایی یا آزمونهای توانایی، آموخته ها، مهارتها، و استعدادهای افراد را می سنجد. دسته ای از این روشها توانایی های شناختی را اندازه می گیرند و تعدادی دیگر توانایی های روانی - حرکتی را اندازه می گیرند. هریک از این دو دسته روش کاربرد خاص خودش را دارد. روش های سنجش توانایی های شناختی بیشتر در زمینه های تحصیلی و فعالیت های فکری - حرکتی غالباً در موقعیت های صنعتی و نظامی و تربیت بدنی به کار میروند. این دو دسته روش، صرف نظر از کاربرد های اختصاصی به طور عمده برای تعیین جنبه های مختلف توانایی های افراد تهیه و اجرا میشوند. ویژگی مهم روش های سنجش توانایی این است که هنگام اجرای آنها از آزمون شونده خواسته می شود تا حداکثر سعی خود را به کاربندد تا بهترین نمره ممکن را به دست آورد.

بی شک آزمون های چند گزینه ای را می شناسیم این آزمون ها تمام محاسن آزمون های صحیح - غلط و جور کردنی را دارند علاوه بر آن بنا به گفته ایبل (۱۹۷۹، به نقل از سیف، ۱۳۸۴): این آزمون ها قادر به اندازه گیری غالب بازده های یادگیری از دانش گرفته تافهمیدن، قضاوت کردن، حل مسئله، ارائه پیشنهاد های علمی، و پیش بینی امور هستند.

تقریباً هر گونه درک و فهم یا توانایی را که بتوان با آزمون های دیگر (کوتاه پاسخ ، کامل کردنی ، صحیح غلط ، جور کردنی ، یا تشریحی) اندازه گرفت می توان با آزمون های چند گزینه ای سنجش کرد .

آزمون های چند گزینه ای که دامنه آنها از ۲ تا ۵ گزینه معمول است از بهترین ابزار های سنجش گزینش پاسخ یا بهترین نوع آزمون های بسته پاسخ هستند و سوال های این آزمون ها می توانند هدف های آموزشی مختلفی را اندازه بگیرند . و با محتوای غالب موضوع های درسی سازگاری دارند (لین و گرانلاند ، ۲۰۰۰ به نقل از سیف ، ۱۳۸۴) .

از آنجا که سوال های چند گزینه ای بسیار انعطاف پذیرند و قابلیت کاربرد متنوعی دارند تا کنون اکثریت آزمون های استاندارد ، از آنها استفاده کرده اند . سوال های چند گزینه ای عموماً برای سنجش هدف های سطح دانش ، فهمیدن و کار بستن مفیدند . با این حال ، اگر در تهیه آنها دقت و ابتکار به خرج داده شود می توان با آن یادگیری های پیچیده تر را نیز سنجید .

در آزمون هایی که به صورت پرسش های چند گزینه ای تهیه می شوند، علاوه بر آنکه ضریب تشخیص ، سطح دشواری پرسش ها باید با هدفهای آزمون هماهنگ باشند گزینه های انحرافی هر پرسش نیز باید از کارایی لازم برخوردار باشند . گزینه ای انحرافی هر پرسش چند گزینه ای در صورتی دارای کارایی نسبی است که دو ویژگی زیر را دارا باشند .

الف . توجه افراد بی اطلاع از موضوع مورد اندازه گیری را به اندازه گزینه درست ، به خود جلب کنند .

ب. برای آزمودنی هایی که در مورد موضوع مورد اندازه گیری ، توانایی و مهارت کافی دارند ، گول زنده نباشد . علاوه بر آن ، افراد گروه بالا باید بیش از افراد گروه پایین گزینه درست را انتخاب کنند، و این تفاوت باید از نظر آماری معنا دار باشد .

به طور کلی آزمون های چند گزینه ای معایبی دارد و انتقاداتی به آن وارد است . نمونه بارز آن کنکور سراسری است. تمام دانشجویان و پشت کنکوری ها به آزمون کنکور سراسری آشنایی کامل دارند. تمام انتقاداتی که ذکر می شود به این آزمون بزرگ نیز وارد است:

ساختن این آزمون ها بسیار دشوار است و غالباً معلمان و اساتید نمی توانند تعدادی گزینه انحرافی خوب برای این سوال های چند گزینه ای انتخاب کنند .

در مقایسه با بعضی آزمون های دیگر ، خواندن این آزمونها و پیدا کردن گزینه درست ، مستلزم صرف وقت نسبتاً زیادتری است ، به ویژه اگر گزینه ها خیلی شبیه به هم ساخته شوند .

زمانی که به پاسخ های غلط نمره منفی داده میشود، دانش آموزانی که خطر می کنند از سایر دانش آموزان نمره بهتری میگیرند .

دانش آموزان قوی بیش از دانش آموزان معمولی قادر به پیدا کردن اشکالات ، پیچیدگی ها و ناظرگاه های واگرا (غیر معمول) در سوال هستند . اما از آنجا که تنها یک پاسخ صحیح در هر سوال وجود دارد بابت این تیز بینی های خودنه تنها تشویق نمی شوند، بلکه با دادن جوابی که مورد نظر طراح سوال نیست ممکن است تنبیه هم شوند (سیف، ۱۳۸۴).

وقتی تعداد گزینه ها کم باشد عامل حدس افزایش پیدا می کند و ممکن است توانایی فرد درست برآورد نشود و از طرفی افزایش تعداد گزینه ها باعث افت کیفیت سوالات خواهد شد.

در آزمونهایی که اصطلاحاً چند گزینه ای نامیده میشوند، در هر سؤال چند پاسخ پیشنهادی ارائه می شود که آزمایش شونده باید از بین آنها یکی را انتخاب کند . در این نوع آزمون ، نمره گذاری عینی پاسخ ها آسانتر است اما از سوی دیگر، در آنها فرد امکان بیشتری دارد که پاسخ درست بعضی از سوالات را از راه حدس زدن پیدا کند ، در حالی که اگر حدس نمی زد موفق به پیدا کردن پاسخ های درست این سوالها نمی بود . حدس زدن سبب میشود که تعداد نمره

های ۱ در ماتریس نمره ها تغییر کند. در نتیجه ، فراوانی پاسخهای درست سوالها ،نمره‌های افراد در آزمون ،ونیز واریانس کل آزمون تغییر می کند. در این بحث ، روشهای رایج برای اصلاح اثر حدس زدن ارائه خواهد شد ، چون هم فراوانی پاسخ های درست سوالها وهم مقدار واریانس کل آزمون در ضریب پایائی آزمون نوع سوالهایی که در جریان آزمون سازی انتخاب می شوند، موثرند.

اثر حدس زدن در مقدار p سوالهای انفرادی کاملاً نمایان است . در ماتریس نمره ها ، در هر سوال تعدادی نمره ۱ خواهیم داشت، گرچه از بین افرادی که در سوال بخصوصی نمره ۱ گرفته اند، فقط بعضی ها حقیقتاً پاسخ درست سوال را می دانند. حدس زدن سبب می شود که فراوانی پاسخهای درست بیشتر از موقعی باشد که عامل حدس در کار نیست. در ارزیابی اثر حدس زدن در ضریب پایائی باید دو عامل مختلف را از هم متمایز ساخت: حدس زدن هم نوعی واریانس منظم وهم نوعی واریانس خطا به واریانس کل آزمون می افزاید . شدت گرایش افراد به حدس زدن در مواردی که پاسخ درست سوال را نمی دانند، یکسان نیست. اگر گروهی از افراد را بایک آزمون چند گزینه ای امتحان کنیم که هیچ یک از آنها پاسخ درست تمام سوالها را نمی دانند، ملاحظه خواهیم کرد که بعضی از افراد بیش از دیگران به حدس زدن متوسل می شوند. در نتیجه حدس زدن نوعی واریانس منظم به واریانس کل آزمون افزوده می شود که اگر همان افراد را با آزمونی همتای آزمون اول امتحان کنیم باز هم این واریانس منظم دیده خواهد شد. در این صورت، این واریانس نوعی واریانس واقعی است که به پایائی آزمون کمک می کند. در عین حال این واریانس ، شاخص تفاوت واقعی بین افراد از لحاظ حدس گزائی خواهد بود. اگر در موقعیت بخصوصی این نوع واریانس را نامربوط بدانیم ، می توان با اصلاح لازم اثر آن را خنثی کرد.

هرگاه بتوانیم تفاوت های فردی را از لحاظ حدس گزائی کنترل کنیم -مثلاً از افراد بخواهیم که وقتی پاسخ درست سوال بخصوصی را نمی دانند از راه حدس جواب دهند -در این صورت نوعی واریانس که کلاً ناشی از خطاست به واریانس کل افزوده خواهد شد . در این

شرایط، درماتریس نمره، بطور تصادفی تعدادی نمره های ۱ وارد خواهد شد که در نتیجه حدس زدن پیدا شده اند. چون در این مورد واریانس خطا افزایش یافته، بی آنکه واریانس واقعی نیز افزایش یابد، این واریانس خطا موجب کاهش پایایی خواهد شد. روش بخصوصی برای اصلاح اثر این واریانس خطا در دست نیست.

در بررسی اثر حدس زدن در اعتبار آزمون نیز باید واریانس منظم را از واریانس خطا تفکیک کرد. آن قسمت از واریانس که منظم تلقی می شود، وفی المثل در آزمونهای همتا دیده می شود، شاخص تفاوت بین افراد از لحاظ یکی از صفات شخصیت است، و به همین جهت می تواند مبنای پیش بینی و تشخیص باشد. این نوع واریانس ممکن است در توزیع متغیی ملاک نیز مشاهده شود و از این راه به افزایش اعتبار آزمون کمک کند. اما اگر چنین نباشد، واریانس مذکور ضریب اعتبار را کاهش خواهد داد. واریانس غیر منظم خطا که در نتیجه حدس زدن پیدا می شود، مسلمانی می تواند به اعتبار آزمون کمک کند، و از این لحاظ عامل مزاحمی محسوب می شود (مگنوسون، ترجمه برهانی، ۱۳۵۱).

با توجه به معایب آزمون های چند گزینه ای چند مدتی است که بحث حذف کنکور سراسری به میان آمده است، حذفی که به گفته صاحب نظران علاوه بر عوامل بالا علل دیگری نیز دارد.

در یک وضعیت افراط گونه همه سرنوشت دانش آموز در یک جلسه سه، چهار ساعته تعیین می گردد. یک دانش آموز ۱۲ سال زحمت می کشد و سوابق متعدد تحصیلی دارد اما ممکن است در آن آزمون خاص به دلایل مختلف نتواند بدرخشد. سیستم فعلی قطعاً یک وضعیت افراطی و بیمار گونه است. اما این هم که فکر کنیم، از همان اول که دانش آموز دارد وارد مدرسه و دبیرستان میشود، تصور کند دارد در نوعی کنکور شرکت میکند و هر امتحانی که از او می گیرند

در سر نوشت تحصیلی و زندگی اش موثر است ، درست نیست .این استرس بسیار شدیدی را به خانواده ها وارد می کند (شیرزاد۱۳۸۸) .

اصولا هر چه ضوابط سنجش دانشجو از منابع متنوع تری انجام شود بهتر است . شاید آزمون های استاندارد آموزش و پرورش بتواند بخشی از این منابع باشد و نه همه آن .

با حذف کنکور آزمون های تستی جای خود را به آزمون های تشریحی خواهد داد . آزمون تستی هر عیبی داشته باشد این مزیت را دارد که با ماشین تصحیح میشود و افراد نظر شخصی خود را درنمره آن دخالت نمی دهند . بسیار دشوار است که ، کاری کرد همه یک جور برگه ها را تصحیح کنند ، درک کنند و نمره دهند . این مشکل مصححان که تصحیح میکنند نیست . این خطای نوع کار است . آزمون تشریحی برای فضای کوچک جواب می دهد . معلم می تواند ورقه ها را به دانش آموزان بدهد و همه را برایش توضیح دهد . اما این آزمون ها برای ورود به دانشگاه دشواری زیادی را ایجاد میکند و با توجه به رشته های زیاد و دارای متقاضی از لحاظ اصولی نمی شود کنکور را حذف کرد و باز نیاز به راهی برای سنجش آشکار می شود . و راهی نیست بجز ، سنجش با آزمون های چند گزینه ای .

یعنی اجرای آزمون های چند گزینه ای در سطح مدارس به صورت سراسری و باز مشکل در توانایی برآورد شده مطرح خواهد شد .

اصولا در سوالات کنکور سراسری در نقطه برش جایی که افراد با توانایی بالاتر از افراد با توانایی کمتر متمایز می شوند تراکم شدید نمرات بوجود می آید . چطور می توان در این نقطه بین افراد تمایز ایجاد کرد ؟

با توجه به سطور بالا مسائلی در ذهن به وجود می آید وقتی سخن از سوالات چند گزینه ای به میان می آید کدام آزمون (سه گزینه ای ، ۴گزینه ای ، ۵ گزینه ای) توانای افراد رابه خوبی برآورد می کند ؟

آیا تعداد گزینه های آزمون های توانایی در برآورد خصوصیات روانسنجی آزمون های توانایی تاثیر دارد؟

کدام تعداد گزینه در مدل کلاسیک و IRT دارای برآزش بهتری خواهد بود؟

در تستهای توانایی نظیر کنکور سراسری کدامین آزمون (۳گزینه ای، ۴گزینه ای، ۵گزینه ای) می تواند در نقطه برش افراد با توانایی بیشتر راز افراد با توانایی کمتر متمایز کند؟

اهمیت و ضرورت پژوهش :

دنیای حاضر دنیای تصمیم است (هومن، ۱۳۷۲) امروزه بیش از هر زمان دیگری پیشرفت و رفاه زندگی به توانایی های افراد در به کار گیری صحیح منابع انسانی، شناخت و پرورش استعدادها و سنجش قابلیت ها و ضعف ها و به منظور کمک به یکایک افراد در تحقق ظرفیت هایشان بستگی دارد. در جوامع پیشرفته بیشتر و در جامعه ما تا اندازه ای برای سنجش قابلیت ها و ضعف های افراد عموماً از آزمون های استاندارد استفاده می کنند. پیش فرض این آزمون ها آن است که افراد انسانی به گونه معنی داری و قابل اندازه گیری از یکدیگر متفاوتند و این تفاوت ها در زمینه توانایی های شناختی، استعداد، انجام تکالیف مختلف، خلق و خو و علائق بروز می کند. اعتقاد بر این است که این تفاوت ها به وضعیت عمومی نوع انسان و به دلیل اختصاص مربوط به نوع انسان و به دلیل اختصاصی تر مربوط به وضعیت هر فرد بستگی دارد (فالوز به نقل از لطف آبادی، ۱۳۷۵).

الزامات جامعه ما، به ویژه در سالهای اخیر نیز ایجاب می کند که هر چه دقیق تر و سریع تر استعداد های مختلف افراد شناسایی شود و افراد بر اساس توانش های ذهنی به کار گمارده شوند. زمینه شناسایی افراد خلاق و نوآور از بین افراد سرآمد فراهم شود (هومن و افروز، ۱۳۷۲).