انطباق و استانداردسازی آزمون غیرکلامی هوش اسنایدرز- اومان برای کودکان ۷- ۲/۵ سال (SON-R ۲ ½ ۷-): گزارش پایانی

اصغر مینائی[۱]

تاریخ پذیرش: تاریخ وصول:

چکیده

پژوهش حاضر با هدف انطباق، استانداردسازی و هنجاریابی آزمون غیرکلامی هوش اسنایدرز-اومان برای کودکان ۷-۲/۵ سال (SON-R ۲ ½ ۷-) روی یک گروه از کودکان تهران با حجم ۲۸۵۲ نفر (۱۴۵۶ پسر و۱۳۹۶ دختر) که با روش نمونه‌برداری تصادفی چند مرحله‌ای از بین کودکان مهدهای کودک تحت نظارت بهزیستی و آموزش و پرورش انتخاب شدند، اجرا گردید. برای تحلیل سؤالات، مدل کلاسیک آزمون و مدل ۲ پارامتری مبتنی بر نظریه سؤال- پاسخ (IRT) مورداستفاده قرار گرفت. اعتبار خرده آزمونها (موزاییکها، طبقه‌بندیها، موقعیتها، پازلها، قیاسها و الگوها) با استفاده از فرمول λ_4 و اعتبار مقیاسهای عملی و استدلال و کل آزمون از طریق فرمول آلفای طبقه‌ای، به تفکیک جنسیت و کل گروه برآورد گردید. دامنه ضرایب اعتبار خرده آزمونها از ۰/۸۱ تا ۰/۸۸ و مقیاسها و کل آزمون از۰/۸۸ تا ۰/۹۵ بدست آمد. در خصوص روایی نیز تحلیلهای مختلفی صورت گرفت. از جمله این تحلیل‌ها می‌توان به تمایزگذاری سنی، تمایزگذاری گروهی و تحلیل عاملی ساختار آزمون اشاره کرد. همچنین جداول نرم برای نمره‌های خرده آزمونها، مقیاسها و IQ در ۱۰ گروه سنی به فاصله ۶ ماه از یکدیگر از (۲،۳ تا ۷ سال تمام) تهیه گردید. بطور کلی یافته‌های حاصل از این پژوهش نشان داد که آزمون موردبحث از اعتبار و روایی بالایی برخوردار است و با اطمینان می‌توان از آن، جهت سنجش هوش کلی کودکان ۷-۲/۵ سال استفاده کرد.

۱- عضو هیات علمی پژوهشکده کودکان استثنایی و دانشجوی دکتری سنجش و اندازه گیری، دانشگاه علامه طباطبایی.

واژگان کلیدی: آزمون غیرکلامی، هوش، اعتبار، روایی.

مقدمه

هوش کلمه‌ای است که نه تنها معنای روان‌شناختی دارد بلکه وارد واژگان عامیانه و غیرحرفه‌ای جامعه نیز شده است. به نظر می‌آید هر یک از ما سعی می‌کند بداند که هوش چیست؟ علی‌رغم تقریباً یک قرن مشاهده، تعریف و نظریه پردازی در خصوص هوش هنوز اختلافات عمیقی در چگونگی تعریف و مفهوم سازی آن وجود دارد. بخشی از مشکل ناشی از این است که هوش در معنای فیزیکی وجود ندارد. برخلاف وزن یا قد، هوش را نمی‌توانیم بطور مستقیم مشاهده بکنیم. وجود یا فقدان و میزان آن را بایستی بطور غیرمستقیم استنباط کرد که حتی انجام این کار نیز چندان سهل و آسان نیست (براون، شربنو و جانسن، ۱۹۹۷).

شیوهٔ مدرن اندازه‌گیری هوش و به تبع آن نظریه پردازی در خصوص ماهیت آن از کار بینه و سیمون که مقیاس هوشی خود را در سال ۱۹۰۵ تهیه کردند شروع شد. عقاید آنها یک دهه بعد با انتشار کتابی تحت عنوان رشد هوش در کودکان[1] (۱۹۱۶) در ایالات متحده آمریکا، به دنیای انگلیسی زبان معرفی گردید. این آزمون یک آزمون کوتاه ۳۰ سؤالی بود که سؤالها از آسان به دشوار مرتب شده بود و مهارتهای گوناگونی مانند هماهنگی چشم و دست، توانایی تعریف اصطلاحات و تمایزگذاری دیداری- انتزاعی را اندازه می‌گرفت. این آزمون ۳ سال بعد، زمانی که سازندگان آزمون آن را برای گروههای سنی استاندارد کردند و مهمتر از همه مفهوم سن ذهنی را وارد زبان هوش‌آزمایی کردند، مورد تجدیدنظر قرار گرفت.

سرانجام، مقیاسهای بینه ابتدا توسط گودارد (۱۹۱۰، ۱۹۱۱) و سپس توسط ترمن و مریل (۱۹۱۶) جهت استفاده در آمریکا به زبان انگلیسی ترجمه گردید و به آزمون استنفرد- بینه معروف گشت. این آزمون ۹۰ سؤال داشت و دامنه سنی آن افراد ۳ تا ۱۴ ساله را در برمی‌گرفت. متأسفانه نُرمهای فرم ۱۹۱۶ آزمون به کودکان کالیفرنیایی محدود

۱- The Development of Intelligence in Children

می‌شد. این نقص در تجدیدنظری که در سال ۱۹۳۷ بـر روی آزمـون صـورت گرفـت رفع گردید. فرم ۱۹۳۷ آزمون بر روی یک نمونه معرف از کل جامعه آمریکا اجرا گردید.

این آزمون در سـال ۱۹۶۰ موردتجدیـد نظـر قـرار گرفـت. در ایـن فـرم، مفهـوم نمـره هوشبهر نسبتی در معنای سنتی آن بـرای همیشـه کنـار گذاشـته شـد و بـه جـای آن هوشـبهر انحرافی[1] که قبلاً توسط دیوید وکسلر در مقیاسهایش بیان شده بود به کار رفت.

از سال ۱۹۶۰ به این سو، این آزمون دوبار مورد تجدید نظـر قـرار گرفتـه کـه آخـرین تجدیدنظر آن در سال ۱۹۸۶ توسط ثرندایک و همکـارانش بـوده اسـت. در یـک انحـراف بحث‌انگیز از عقاید بینه، فرم جدید استنفورد- بینه تنها یک نمره بدست نمی‌دهد بلکـه در ۴ حوزه شناختی استدلال کلامی، استدلال کمی، استدلال انتزاعی- دیـداری و حافظـه کوتـاه مدت، ۱۵ نمره سنی استاندارد بدست می‌دهد.

مقیاسهای وکسلر که با **مقیاس هوشی وکسلر برای کودکان (WISC)** در سـال ۱۹۴۹، **مقیاس هوشی وکسلر برای بزرگسالان (WAIS)** در سال ۱۹۵۵، و مقیاس **هوشـی وکسـلر برای پیش دبستانی و ابتـدایی (WPPSI)** در سـال ۱۹۶۷ شـروع گردیـد تـأثیر عمیقـی بـر چشم‌انداز هوش‌آزمایی در ایالات متحده داشته است. تفاوت مقیاسهای وکسلر با مقیاسهای بینه در این بود که مقیاسهای وکسلر شامل خرده آزمونهایی بود کـه در دو مقولـه کلامـی و عملی قرار گرفته بود و امکان رسم نیمرخ آنها وجود داشت. تفاوت دیگر در این است کـه مقیاسهای وکسلر، امتیازی هستند اما مقیاسهای بینه سنی است. آزمونهای اصـلی وکسـلر، بـه غیر از آزمون وکسلر- بلویو که از نظر روان‌سنجی دارای تعـدادی نقـص اسـت، دارای نُـرم ملی هسـتند و اضـافه بـر هوشـبهر کـل مقیـاس، بـرای مقولـه‌هـای کلامـی و عملکـردی نیـز نمره‌های هوشبهر انحرافی بدست می‌دهنـد. WAIS در ۱۹۸۱، WPPSI در ۱۹۸۹ و WISC در ۱۹۹۱ مورد تجدید نظر قرار گرفتند.

وکسلر و بینه، به رغم اختلافاتی که با یکدیگر داشتند، هر دو در ساخت و تفسیر آزمونهایشان اساساً از مفهوم g اسپیرمن دنباله روی کردند. با این حال، دهۀ ۱۹۶۰ شاهد انتشار تعدادی آزمون بود که عمدتاً بر عوامل اختصاصی S تأکید داشتند تا عاملg.

! - Deviation IQ

مشهورترین این آزمونها، شاید آزمون تواناییهای روانی- زبانی ایلینویز[1] است که توسط کرک، مک کارتی و کرک (۱۹۸۶) تهیه شده بود و تقریباً بطور انحصاری بر شاخصهای (اندازه‌های) مبتنی بر زبان تمرکز داشت. قصد مؤلفین این بود که نتایج حاصل از آزمونشان بطور مستقیم در طرح ریزی برنامه‌های آموزشی خاص و جبرانی مورد استفاده قرار گیرد.

آزمونهای سنتی هوش کلی[2](GI)، مانند آزمونهای هوش وکسلر و استنفورد – بینه توسط طرفداران آزمونهای ظرفیت یادگیری[3] (LP) مورد انتقاد قرار گرفته است، زیرا این آزمونها به جای ظرفیت یادگیری، نتیجه نهایی یادگیریهای پیشین را اندازه می‌گیرند. آزمونهای GI صرفاً نتیجه نهایی یادگیریهای پیشین را منعکس می‌سازند و از این رو، توانایی یادگیری افراد متعلق به طبقه اجتماعی – اقتصادی پایین و افراد مبتلا به مشکلات یادگیری و اختلالات رفتاری و همچنین افرادی که فرصت‌های اندکی جهت کسب دانش و مهارتهای موردنیاز برای موقعیت آزمون داشته‌اند را کم برآورد می‌کنند. نکته‌ای که تلویحاً با این انتقاد مرتبط است، این است که آزمونهای GI هیچ اطلاعاتی در خصوص اینکه در شرایط بهینه یادگیری تا چه میزان می‌توان انتظار رشد در عملکرد را داشت، فراهم نمی‌کنند. در نتیجه، این آزمونها نخواهند توانست به اندازه کافی بین کودکان عقب مانده ذهنی و کودکان مبتلا به ناتوانیهای یادگیری تمایز قائل شوند.

ویژگی مشترک آزمونهای LP، ارائه آموزش در زمان اجرای آزمون است. هدف از آموزش، حذف تفاوتهای ناشی از فرصتهای فرهنگی و آموزشی پیشین و بهینه کردن شرایط یادگیری است. اگرچه تحقیق درخصوص ظرفیت یادگیری برای چندین دهه است که ادامه دارد، اما ابزارهای عملی برای سنجش، فقط در چند سال اخیر است که ارائه شده‌اند.

آزمونهای GI همچنین از سوی طرفداران آزمونهای هوش نابسته به فرهنگ به خاطر محتوایشان مورد انتقاد قرار گرفته‌اند. به خاطر اینکه این آزمونها از لحاظ محتوا و

! - Illinois Test of Psycholinguistic Abilities
! - General Intelligence
! - Learning Potential

دستورالعمل‌ها اغلب به مهارتهای زبانی خاصی وابسته هستند، لذا باعث می‌گردند تا افراد اقلیّت‌های فرهنگی و قومی و همچنین افراد مبتلا به مشکلات شنوایی، گفتاری و زبانی در وضعیت نامطلوب قرار گیرند. عملکرد پایین این گروهها در یک آزمون GI ممکن است اساساً انعکاس دانش کلامی ضعیف باشد تا ضعف در استدلال یا ناتوانی یادگیری (تلخن و لاروس، ۱۹۹۳). می‌توان گفت کـه ایـن آزمونها بیشتر بر هوش متبلور تأکید دارند تا هوش سیال(کتل،۱۹۷۱). این انتقادها نسبت به آزمونهای GI باعث گردید تا آزمونهای غیرکلامی هوش مانند ماتریسهای پیشرونده ریون (ریون، ۱۹۳۸) و آزمون هوش نابسته به فرهنگ کتل (۱۹۵۰) با هدف به حداقل رساندن اتکا به دانش اکتسابی و توانایی کلامی پرورش یابند.

در اوایل دهه چهل میلادی، خانم نان اسنایدرز – اومان (۱۹۴۳) که در یک مؤسسه مربوط به کودکان ناشنوا در هلند کار می‌کرد اولین آزمون غیرکلامی هوش با نامSON[1] را پرورش داد. او هوش را توانایی یادگیری؛ یعنی میزان توانایی کودکان در بهره‌گیری از آموزشها در مدرسه تعریف کرد. این آزمون، اولین آزمونی بود که حوزه وسیعی از هوش را بدون اینکه وابسته به استفاده از زبان باشد، پوشش می‌داد. آزمونهای غیرکلامی موجود در آن زمان، برای بررسی طیف وسیعی از تواناییهای یادگیری مناسب نبودند. زیرا آنها اساساً از آزمونهای عملی که به توانایی‌های فضایی مربوط می‌شدند، تشکیل می‌یافتند. اولین آزمون SON از خرده آزمونهای غیرکلامی‌تشکیل می‌یافـت که به استدلال انتزاعی و عینی مربوط می‌شدند و دارای نُرمهایی برای کودکان ناشنوای ۱۴-۴ سال بود. در حال حاضر، نسل چهارم آزمونهای SON در دو نـسخـه وجـود دارند. یـک نسـخه بـرای کودکـان سنیـن پـایین بـا عنـوان ۲ SON-R ½ ۷- و نـسخهٔ دیـگر برای کودکان سنین بالا با عنوان ۵ SON-R ½ ۱۷-. پژوهش حاضر مطالعه‌ای را توصیف می‌کند که با آزمون - ۲ SON-R ½ ۷ صورت گرفته است.

!. Snijders-Oomen Nonverbal Intelligence Test

آزمون۲ SON-R ½ ۷- سه تفاوت اساسی با آزمونهای سنتی هوش دارد. اولاً، این آزمون مستلزم تواناییهای زبانی خاصی نیست؛ ثانیاً، اجرای آن به شیوه انطباقی، صورت می‌گیرد. هدف شیوه‌های انطباقی محدود نمودن تعداد سؤالهای ارائه شده همراه با کاهش نسبتاً کم اعتبار[1] است (ویس، ۱۹۸۲). در ۲ SON-R ½ ۷- نقطه شروع یا سطح ورودی هر خرده آزمون براساس سن کودک و سالهای حضور در مهدکودک، تعیین می‌شود و اجرای هر خرده آزمون نیز پس از ۳ پاسخ نادرست متناوب و در بخش دوم برخی از خرده آزمونها نیز پس از ۲ پاسخ نادرست متوالی متوقف می‌گردد. ثالثاً، در مورد درست یا نادرست بودن پاسخ آزمودنی به او بازخورد داده می‌شود. مزیّت اصلی ارائهٔ بازخورد این است که به آزمودنی فرصت داده می‌شود تا در طول اجرای آزمون، دست به یادگیری بزند. با توجه به این ویژگیها، می‌توان گفت که آزمون ۲ SON-R ½ ۷- بیشتر به آزمونهای هوش نابسته به فرهنگ و آزمونهای ظرفیت یادگیری شباهت دارد تا آزمونهای سنتی هوش کلی.

در ایران نیاز زیادی به آزمونهای روانی روا، مناسب و استاندارد، بویژه آزمونهای هوش برای کودکان سنین پایین وجود دارد. به منظور برآورده ساختن این نیاز و با توجه به مطالبی که در خصوص مزایای آزمون ۲ SON-R ½ ۷- و انتقادات وارده بر آزمونهای سنتی هوش کلی ذکر گردید، آزمون ۲ SON-R ½ ۷- جهت استاندارد سازی، رواسازی و هنجاریابی برای کودکان ایرانی انتخاب گردید.مقاله حاضر به مراحل و نتایج حاصل از استاندارد کردن آزمون ۲ SON- R ½ ۷- می‌پردازد.

روش

نمونه و روش نمونه گیری

نمونه مورد مطالعه در این تحقیق، ۲۸۵۲ کودک تهرانی (۱۴۵۶ پسر و ۱۳۹۶ دختر) با دامنه سنی ۳، ۲ تا ۳ ، ۷ (۲ سال و ۳ ماه تا ۷ سال و ۳ ماه) با میانگین ۴/۴۱ و انحراف استاندارد ۱/۲۰ است که از اردیبهشت تا پاییز سال ۱۳۸۵ در مهدهای کودک تحت نظارت بهزیستی

! . Reliability

و آموزش و پرورش تهران مشغول به آموزش بودند. برای انتخاب کودکان از روش نمونه‌گیری طبقه ای چندمرحله‌ای تصادفی استفاده گردید. ۱۱ گروه سنی از سن ۲،۳ تا ۳، ۷ (۲ سال و ۳ ماه تا ۷ سال و ۳ ماه) با فاصله سنی ۶ ماه تشکیل گردید. سعی ما بر این بود تا در هر گروه سنی، ۱۵۰ دختر و ۱۵۰ پسر مورد آزمون قرار گیرد، اما در عمل، بروز برخی مشکلات از جمله قطع همکاری تعدادی از آزمونگران بخاطر مشکلات شخصی و عدم همکاری تعدادی از مهدهای کودک در ارائه تاریخ تولّد کودکان پیش از اجرای آزمون، باعث گردید تا حجم نمونه در گروههای سنی مختلف، یکسان نباشد.

فرایند انتخاب کودکان به این صورت بود که ابتدا فهرست مهدهای کودک فعال در سطح شهر تهران از سازمان بهزیستی شهرستان تهران، سازمان آموزش و پرورش تهران و سازمان آموزش و پرورش شهرستانهای تهران اخذ گردید. آنگاه از بین مهدهای کودک موجود در مناطق ۲۲ گانه تهران تعداد ۲۰۰ مهد کودک بطور تصادفی انتخاب شدند. بعد از انجام هماهنگیهای لازم با سازمانهای مربوطه و کسب مجوز، آزمونگران به مهدکودک مراجعه و از اولیای مهدکودک درخواست می‌کردند تا فهرست اسامی و همچنین تاریخ تولّد کودکان را در اختیار آنها قرار دهند. آزمونگران پس از محاسبه سن دقیق کودک، دختر و پسری را که سن آنها به سن موردنیاز برای گروه سنی، نزدیک بود انتخاب و آزمون SON-R ۲ ½-۷ در یک اطاق ساکت که در اختیار آزمونگر قرار داده شده بود، اجرا می‌گردید. در نهایت، آزمون بر روی ۲۸۵۲ کودک (۲۷۷۵ کودک نرمال، ۳۳ کودک ناشنوا و ۴۴ کودک عقب مانده ذهنی) اجرا گردید. اجرای میدانی آزمون از اواسط اردیبهشت ۱۳۸۵ شروع و تا اواخر آبان همان سال ادامه یافت. کودکان ناشنوا و عقب مانده ذهنی از مدرسه باغچه‌بان شماره ۵ و ۷ و آموزشگاههای شهید اصلانی، نیر، ایمان، پیام نور، مهرطلبان و ساریخانی که تحت نظارت مدیریت آموزش و پرورش استثنایی شهر تهران قرار دارند موردآزمون قرار گرفتند. ویژگیهای نمونه نهایی برای تشکیل جداول هنجار به تفکیک متغیرهای جمعیت شناختی در جدول ۱ ارائه شده است.

در جدول۱، منطقه جغرافیایی شمال، شامل مناطق ۱، ۳ و بخشهایی از مناطق ۲ و ۵ است؛ منطقه جغرافیایی جنوب شامل مناطق ۱۵، ۱۶، ۱۷، ۱۸، ۱۹ و بخشهایی از منطقه ۱۴

است؛ منطقه جغرافیایی مرکز، از مناطق ۶، ۷، ۸، ۹، ۱۰، ۱۱ و ۱۲ ترکیب یافته است؛ منطقه جغرافیایی شرق شامل مناطق ۴، ۸، ۱۳ و بخشهایی از منطقه ۱۴ است؛ و منطقه جغرافیایی غرب نیز، مناطق ۵، ۹، ۲۱ و بخشهایی از منطقه ۲ را دربر می‌گیرد.

سطح تحصیلات والدین براساس آخرین مدرک تحصیلی آنها تعیین گردید. برای طبقه‌بندی مشاغل والدین از یک طرح ۶ مقوله ای استفاده گردید. طبقه متخصَص، مشاغلی مانند وکیل، استاد دانشگاه، خلبان، مهندس، و... را شامل می‌شود؛ طبقه کارمند میانی، دربرگیرنده مشاغلی چون معلم، مدیر مدرسه، کتابدار، مشاور مدرسه و... است؛ در طبقه کارمند جزء مشاغلی مانند آبدارچی اداره، نامه رسان اداره، پستچی، تایپیست، مسئول دبیر خانه، و... قرار می‌گیرد؛ طبقه کارگر ماهر، مشاغلی مانند مکانیک، خیَاط، نجّار، برق‌کار و... را در برمی‌گیرد؛ طبقه آزاد از مشاغلی مانند بازرگان، مغازه‌دار، بنکدار، کتابفروش، طلافروش و... تشکیل یافته است. والدینی که بازنشسته بودند و یا در زمان گردآوری اطلاعات، فاقد شغل بودند، آخرین شغل آنها ملاک طبقه بندی قرار گرفت.

طبقه اجتماعی خانواده‌ها براساس یک شاخص ۴ عاملی که توسط هالینگ شید (۱۹۷۵) پیشنهاد شده است برآورد گردید. دامنه نمره ها ۸ تا ۵۰ و میانگین و انحراف استاندارد آنها به ترتیب ۳۵/۶۷ و ۵/۸۲ است که با استفاده از مدل توزیع نرمال به ۵ طبقه پایین، متوسط پایین، متوسط، متوسط بالا و بالا تقسیم گردید.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

جدول۱: ویژگیهای جمعیت شناختی گروه هنجار (N =۲۸۵۲)

ویژگی	فراوانی	درصد
جنسیت کودک		
پسر	۱۴۵۶	٪۵۱
دختر	۱۳۹۶	٪۴۹
منطقه جغرافیایی محل سکونت		
شمال	۵۴۴	٪۱۹
غرب	۵۵۰	٪۱۹/۵
شرق	۵۷۰	٪۲۰
جنوب	۵۵۴	٪۱۹/۵
مرکز	۶۳۴	٪۲۲
وضعیت جسمانی / روانی کودک		
عادی	۲۷۷۵	٪۹۷/۳
ناشنوا	۳۳	٪۱/۲
عقب مانده ذهنی	۴۴	٪۱/۵
گروه سنی		
۲/۶ -۲/۳	۱۴۲	٪۵
۲/۱۲-۲/۷	۲۳۶	٪۸/۳
۳/۶-۳/۱	۲۹۵	٪۱۰/۳
۳/۱۲-۳/۷	۳۰۰	٪۱۰/۵
۴/۶-۴/۱	۳۴۱	٪۱۲
۴/۱۲-۴/۷	۳۴۴	٪۱۲/۱
۵/۶-۵/۱	۳۸۴	٪۱۳/۵
۵/۱۲-۵/۷	۳۴۸	٪۱۲/۲
۶/۶-۶/۱	۳۰۲	٪۱۰/۶
۷/۳-۶/۷	۱۶۰	٪۵/۵
سطح تحصیلات پدر		
بیسواد	۱۴	٪۰۰۵
ابتدایی و راهنمایی	۱۸۵	٪۶/۵
دیپلم	۱۰۰۵	٪۳۵
فوق دیپلم و لیسانس	۱۲۵۴	٪۴۴

فوق لیسانس و بالاتر	۳۹۴	٪۱۴
سطح تحصیلات مادر		
بیسواد	۸	٪۰۰۲
ابتدایی و راهنمایی	۱۴۵	٪۵
دیپلم	۱۰۴۷	٪۳۶/۸
فوق دیپلم و لیسانس	۱۴۴۲	٪۵۱
فوق لیسانس و بالاتر	۲۱۰	٪۷
شغل پدر		
متخصص	۳۴۶	٪۱۲
کارمند میانی	۱۲۲۹	٪۴۳
کارمند جزء	۳۲۲	٪۱۱
کارگر ماهر	۲۰۹	٪۷
کارگر غیرماهر	۱۰۳	٪۴
آزاد	۶۴۳	٪۲۳
شغل مادر		
خانه دار	۸۴۰	٪۲۹/۵
متخصص	۱۶۹	٪۵/۹
کارمند میانی	۱۳۲۵	٪۴۶/۵
کارمند جزء	۲۵۱	٪۸/۸
کارگر ماهر	۶۰	٪۲/۱
کارگر غیر ماهر	۳۷	٪۱/۳
آزاد	۳۲	٪۱/۱
طبقه اجتماعی-اقتصادی		
پایین	۶۰	٪۲/۱
متوسط پایین	۳۹۳	٪۱۳/۸
متوسط	۱۹۲۲	٪۶۷/۴
متوسط بالا	۴۲۸	٪۱۵
بالا	۴۹	٪۱/۷

ابزار

ابزار مورداستفاده در این پژوهش، آزمون غیرکلامی هوش اسنایدرز- اومان برای کودکان ۷-۲/۵ سال است (تلخن و همکاران، ۱۹۹۸). این آزمون دارای دو فرم کوتاه و بلند است. فرم بلند آزمون از ۶ خرده آزمون تشکیل یافته است که به ترتیب اجرا عبارتند از:

۱- موزاییکها[۱] با ۱۵ سؤال؛

۲- طبقه‌بندیها[۲] با ۱۵ سؤال؛

۳- پازلها[۳] با ۱۴ سوال؛

۴- قیاسها[۴] با ۱۷ سوال؛

۵- موقعیتها[۵] با ۱۴ سوال؛

۶- الگوها[۶] با ۱۶ سوال.

استانداردسازی این آزمون در هلند براساس یک نمونه ملی از کودکان هلند با حجم ۱۱۲۴ نفر صورت گرفته است که دامنه سنی آنها از ۳، ۲ تا ۳، ۷ است.

این آزمون در مجموع دارای ۹۱ سؤال است و در مدت ۶۰- ۵۰ دقیقه در یک یا دو جلسه اجرا می‌گردد. خرده آزمونهای ۲ SON-R ½۷- به دو گروه کلی طبقه‌بندی می‌گردند. گروه اول شامل خرده‌آزمونهای الگوها، موزاییکها و پازلها است که مقیاس عملی[۷] (PS) نامیده می‌شود و گروه دیگر شامل خرده آزمونهای موقعیتها، طبقه‌بندیها و قیاسها است که مقیاس استدلال[۸] (RS) نام دارد. بطورکلی، در این آزمون برای هر کودک، علاوه بر نمره خرده‌آزمونها، ۳ نمره دیگر با نامهای SON-IQ (هوشبهر)، SON-PS (بهرعملی) و SON-RS (بهراستدلال) بدست می‌آید.

! . Mosaics
! . Categories
! . Puzzles
! . Analogies
! . Situations
! . Patterns
! . Performance Scale
! . Reasoning Scale

خرده آزمونهای SON-R ۲ ½ ۷- را می‌توان براساس نوع موادی که مورداستفاده قرار می‌گیرد به دو دسته تقسیم کرد:

الف) خرده آزمونهایی که از مواد تصویری معنادار[۱] استفاده می‌کنند (طبقه‌بندیها، موقعیتها و پازلها)؛

ب- خرده آزمونهایی که از مواد تصویری بی‌معنا[۲] مانند اشکال هندسی استفاده می‌نمایند (موزاییکها، الگوها و قیاسها).

هر دو فرم این آزمون توسط کمیته آزمون مؤسسه روان‌شناسان هلند[۳] (COTAN) از ۸ بعد و در یک مقیاس ۳ طبقه‌ای ناکافی، کافی و خوب ارزیابی شده است. نتایج این ارزیابی در زیر ارائه شده است (اقتباس از تلخن و همکاران، ۱۹۹۸).

اصول و پایه‌های ساخت آزمون	خوب
قابلیت اجرایی مواد	خوب
قابلیت اجرایی کتابچه راهنما	خوب
نرمها	خوب
اعتبار	خوب
روایی سازه	خوب
روایی ملاکی	کافی

جدول۲، (اقتباس از تلخن و همکاران، ۱۹۹۸) ضرایب اعتبار خرده‌آزمونها را که مبتنی بر همسانی درونی سؤالات هر خرده‌آزمون است و با استفاده از فرمول λ_4 گاتمن (۱۹۴۵) و براساس نمونه استانداردسازی هلند برآورد شده است به تفکیک گروههای سنی نشان می‌دهد.

پرتال جامع علوم انسانی

! . Meaningful

! . Non-Meaningful

۳. the test commission of the NetherLands Institute for Psychologists

جدول ۲: ضرایب اعتبار خرده آزمونها، مقیاسها و کل آزمون به تفکیک گروههای سنی

سن	خرده‌آزمونها								مقیاسها
	موزاییکها	طبقه بندیها	پازلها	قیاسها	موقعیتها	الگوها	عملی	استدلال	کل آزمون
۶، ۲	۰/۴۱	۰/۸۱	۰/۴۵	۰/۷۵	۰/۷۹	۰/۷۹	۰/۶۸	۰/۸۹	۰/۸۶
۶، ۳	۰/۷۶	۰/۷۳	۰/۷۵	۰/۷۳	۰/۶۶	۰/۷۳	۰/۸۶	۰/۸۴	۰/۹۰
۶، ۴	۰/۷۷	۰/۷۰	۰/۷۵	۰/۷۴	۰/۶۲	۰/۷۲	۰/۸۸	۰/۸۱	۰/۹۰
۶، ۵	۰/۷۴	۰/۶۸	۰/۷۰	۰/۷۸	۰/۶۲	۰/۷۴	۰/۸۷	۰/۸۱	۰/۹۰
۶، ۶	۰/۷۸	۰/۶۸	۰/۶۹	۰/۸۳	۰/۶۶	۰/۷۶	۰/۸۷	۰/۸۴	۰/۹۱
۶، ۷	۰/۸۴	۰/۶۹	۰/۶۹	۰/۸۵	۰/۶۹	۰/۷۹	۰/۸۸	۰/۸۶	۰/۹۲
میانگین	۰/۷۳	۰/۷۱	۰/۶۹	۰/۷۸	۰/۶۷	۰/۷۵	۰/۸۵	۰/۸۴	۰/۹۰

ضرایب اعتبار نمرات مقیاس عملی، مقیاس استدلال و کل آزمون با استفاده از فرمول آلفای طبقه‌ای[1] برآورد شده است. این فرمول جهت برآورد ضرایب اعتبار ترکیبات خطی مورد استفاده قرار می‌گیرد (نانالی، ۱۹۷۸؛ نانالی و برنشتاین، ۱۹۹۴؛ اُسبرن، ۲۰۰۰).

تلخن و همکارانش (۱۹۹۸) مطالعات متعددی در زمینه روایی آزمون۲ SON-R ½ ۷- انجام داده‌اند. نتایج این مطالعات حاکی از روایی خوب نمرات آزمون است. ازجمله این مطالعات می‌توان به همبستگی بین این آزمون با سایر آزمونهای شناختی اشاره کرد. جدول ۳، خلاصه‌ای از ضرایب همبستگی بین ۲ SON-R ½ ۷- با سایر آزمونهای شناختی را نشان می‌دهد.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

! . Stratified Alpha

جدول۳: ضرایب همبستگی آزمون ۲ SON-R ½ ۷- با سایر آزمونهای شناختی

آزمون ملاک	N	r	آزمون ملاک	N	r
BOS	۵۰	۰/۵۹	TONI-۲[۳] فرم الف	۱۵۳	۰/۶۰
GOS					
DI شناختی	۱۱۵	۰/۶۵	WPPSI		
DI همزمان		۰/۶۴	IQ کلی	۵۳	۰/۶۰
DI زنجیره‌ای		۰/۴۹	مقیاس کلامی		۰/۴۹
RAKIT	۱۶۵	۰/۶۰	مقیاس عملی		۰/۵۹

شیوه اجرا

مراحل استانداردسازی آزمون ۲ SON-R ½ ۷- به شرح زیر است.

الف- ترجمه کتابچه راهنمای آزمون: گام نخست در این پژوهش، ترجمه کتابچه راهنمای آزمون (تلخن و همکاران، ۱۹۹۸) به زبان فارسی بود. به منظور تسلّط بر نحوه اجرای آزمون، دستورالعمل اجرایی آن چندبار مورد مطالعه دقیق قرار گرفت. علاوه بر این یک نوار ویدئویی که نحوه اجرای آزمون را نشان می‌دهد و توسط آقای دکتر تلخن ارسال شده بود به دقت مورد مشاهده قرار گرفت.

ب- اجرای مقدماتی: پس از اطمینان از تسلّط بر اجرای آزمون، تعداد ۱۲۵ کودک تهرانی(۶۶ پسر و ۵۹ دختر) با دامنه سنّی ۲،۹ تا ۷،۹ (۲ سال و نه ماه تا ۷ سال و ۹ ماه) با میانگین ۴/۴۱ و انحراف استاندارد ۱/۲۰ که در تابستان سال ۱۳۸۴ در مهدهای کودک تحت نظارت بهزیستی تهران مشغول به آموزش بودند با استفاده از روش نمونه‌گیری تصادفی چندمرحله‌ای انتخاب گردیدند. برای این منظور از میان مهدهای کودک تحت نظارت بهزیستی مناطق شمال، جنوب، شرق و غرب تهران، تعداد ۱۲ مهد کودک (از هر منطقه ۳ مهد کودک) بصورت کاملاً تصادفی انتخاب گردید. در گام بعد، با استفاده از دفتر ثبت نام از میان کودکان هر مهد تعداد ۱۲-۱۰ کودک دختر و پسر با استفاده از روش تصادفی منظم انتخاب شدند.

اجرای آزمون بر روی این ۱۲۵ کودک و گردآوری داده ها توسط خود پژوهشگر صورت گرفت و به مدت تقریباً دو ماه بطول انجامید. برای مطالعه بیشتر در زمینه مرحله اجرای مقدماتی و نتایج حاصل از آن به مینائی (۱۳۸۴) مراجعه نمایید.

ج- اجرای نهایی: براساس نتایج حاصل از اجرای مقدماتی، برخی از تصاویر سؤالهای A ، ۸، ۹، ۱۰، ۱۲، ۱۳، ۱۴ و ۱۵ خرده آزمون طبقه‌بندیها با تصاویر متناسب با فرهنگ ایران که آشنا برای کودکان بودند جایگزین گردیدند. علاوه بر این، ترتیب سؤالها نیز براساس سطح دشواری آنها اصلاح گردید. پس از انجام اصلاحات فوق، صورت نهایی آزمون شکل گرفت و با مساعدت مؤسسه تحقیقات علوم و فناوری سینا (روان تجهیز) تعداد ۲۰ نسخه از آزمون ساخته شد.

پس از آماده سازی مواد و لوازم آزمون، تعداد ۲۰ آزمونگر که همه آنها خانم و دارای حداقل مدرک لیسانس روان‌شناسی یا رشته های مرتبط بودند، جهت اجرای میدانی آزمون و گردآوری داده‌ها دعوت به همکاری شدند. آزمونگران در یک کارگاه آموزشی دو روزه شرکت کردند و روش اجرا و نمره‌گذاری آزمون، توسط پژوهشگر به دقت به آنها آموزش داده شد. جهت اطمینان از تسلط آزمونگران بر نحوه اجرای آزمون، از آنها خواسته شد تا هر کدام، آزمون را بطور آزمایشی بر روی ۳ کودک اجرا و نمره‌گذاری نمایند. این عمل باعث گردید تا اشکالات آزمونگران شناسایی و رفع گردد.

تحلیل داده‌ها

برای تحلیل و برآورد ویژگیهای روان‌سنجی سؤالات (درجه دشواری و قدرت تشخیص) از مدل کلاسیک آزمون و مدل ۲ پارامتری مبتنی بر نظریه سؤال- پاسخ (IRT)[1] استفاده گردید. برای برآورد مشخصه‌های سؤالها در مدل کلاسیک و انجام تحلیلهای آماری از ویرایش پانزدهم برنامه SPSS و برای برآورد پارامترهای سؤال در مدلهای مبتنی

! . Item–Response Theory

بر IRT از برنامه BILOG –MG ۳ (زیموسکی، موراکی، میسلوی و باک، ۲۰۰۳) استفاده گردید.

در برخی از تحلیل‌ها از نمرات استاندارد خرده‌آزمونها که دارای میانگین ۱۰ و انحراف استاندارد ۳ هستند و نمرات استاندارد مقیاسها و کل آزمون با میانگین ۱۰۰ و انحراف استاندارد ۱۵، استفاده گردید. جهت برآورد اعتبار نمرات خرده آزمونها از فرمول λ_2 (لاندای۲) گاتمن (فلدت و برنان، ۱۹۸۸) استفاده گردید. زیرا این فرمول، اعتبار را بویژه در آزمونهای کوتاه به اندازه فرمول آلفای کرونباخ، کم برآورد نمی‌کند (تنبرگ و زگرس، ۱۹۷۸). جهت برآورد اعتبار نمرات مقیاسها و کل آزمون از فرمول آلفای طبقه‌ای (نانالی و برنشتاین، ۱۹۹۴؛ اُسبرن، ۲۰۰۰) استفاده گردید.

تفاوت بین میانگین نمره‌های دختران و پسران در خرده‌آزمونها با استفاده از تحلیل واریانس چند متغیری (MANOVA) و از طریق T^2 هاتلینگ (استیونس، ۲۰۰۴) مورد آزمون قرار گرفت. تفاوت بین میانگین نمره‌های دختران و پسران در دو مقیاس استدلال و مقیاس عملی نیز به همان شیوه موردآزمون قرار گرفت. همچنین برای آزمون تفاوت بین میانگین نمره‌های دختران و پسران در کل آزمون، از آزمون t برای دو گروه مستقل (هاول، ۱۹۹۷) استفاده گردید. در انجام این تحلیل‌ها از نمره‌های خام استفاده شد، زیرا در نمره‌های خام تمام واریانس یا پراکندگی موجود در نمره‌ها به حساب می‌آید (آخنباخ و رسکورلا، ۲۰۰۱). در توصیف و تحلیل داده‌ها از روشها و شاخص‌های آماری دیگری، مانند ضریب همبستگی گشتاوری پیرسون، میانگین، انحراف استاندارد و درصد، نیز استفاده گردید.

نتایج

از آنجا که اجرای این آزمون به شیوه انطباقی صورت می‌گیرد و برای شروع و توقف آن از قواعدی پیروی می‌گردد، از این رو ترتیب سؤالهای خرده‌آزمونها براساس درجهٔ دشواری از اهمیّت خاصی برخوردار است. جدول ۴، درجه دشواری سؤالها را در مدل کلاسیک نشان می‌دهد. سؤالهایی که پایین‌تر از نقطه شروع قرار داشتند و از این رو، بدون

پاسخ بودند، صحیح در نظر گرفته شده و نمره ۱ به آنها تعلق گرفت و سؤالهایی که پس از توقف آزمون اجرا نشده بودند، بعنوان نادرست تلقی شده و نمره صفر به آنها اختصاص داده شد.

قدرت تشخیص سؤال در مدل کلاسیک، که گاهی اوقات روایی سؤال نیز نامیده می‌شود، همواره به صورت همبستگی دو رشته‌ای یا دو رشته‌ای نقطه‌ای سؤال با نمره کل آزمون یا خرده آزمون بیان می‌شود (ثرندایک، ترجمه هومن، ۱۳۷۵). از آنجا که در برنامه SPSS، همبستگی دو رشته‌ای نقطه‌ای وجود ندارد، جهت برآورد همبستگی سؤال با نمره کل خرده‌آزمون از همبستگی گشتاوری پیرسون استفاده گردید. همبستگی دو رشته‌ای نقطه‌ای به لحاظ عددی، دقیقاً معادل با همبستگی گشتاوری پیرسون است(کلاین، ۲۰۰۰). از آنجا که نمره سؤال در نمره کل خرده آزمون لحاظ می‌شود و این امر باعث افزایش غیرواقعی همبستگی سؤال با نمره کل می‌گردد، از این رو جهت حذف تأثیر نمره سؤال، بعنوان بخشی از نمره کل خرده آزمون، یک اصلاح صورت گرفت (کلاین، ۲۰۰۰؛ نانالی و برنشتاین، ۱۹۹۴).

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

جدول ۴: درجه دشواری و قدرت تشخیص سؤالها در مدل کلاسیک

سؤال	خرده آزمونها											
	موزاییکها		طبقه بندیها		پازلها		قیاسها		موقعیتها		الگوها	
	P	r_{pbis}	P	r_{pbis}	P	r_{pbis}	P	r_{pbis}	P	r_{pbis}	P	r_{pbis}
۱	۰/۹۷	۰/۲۴	۰/۹۵	۰/۳۵	۰/۹۴	۰/۲۸	۰/۹۶	۰/۲۸	۰/۹۷	۰/۲۵	۰/۹۴	۰/۴۰
۲	۰/۹۰	۰/۴۷	۰/۹۳	۰/۴۳	۰/۸۹	۰/۴۱	۰/۹۳	۰/۴۳	۰/۹۴	۰/۳۷	۰/۹۳	۰/۴۲
۳	۰/۸۱	۰/۵۵	۰/۹۰	۰/۴۹	۰/۷۸	۰/۵۴	۰/۸۳	۰/۵۲	۰/۸۶	۰/۴۷	۰/۹۲	۰/۴۹
۴	۰/۸۱	۰/۶۶	۰/۸۱	۰/۵۹	۰/۷۳	۰/۶۳	۰/۸۳	۰/۵۵	۰/۸۵	۰/۵۰	۰/۹۲	۰/۵۰
۵	۰/۷۵	۰/۷۰	۰/۷۱	۰/۶۰	۰/۶۸	۰/۶۷	۰/۵۵	۰/۴۸	۰/۴۰*	۰/۵۱	۰/۸۸	۰/۵۳
۶	۰/۶۸	۰/۷۵	۰/۶۷	۰/۶۵	۰/۵۹	۰/۶۹	۰/۵۳	۰/۵۶	۰/۵۲	۰/۵۸	۰/۷۵	۰/۶۷
۷	۰/۵۸	۰/۷۶	۰/۶۲	۰/۶۱	۰/۴۷	۰/۷۲	۰/۵۲	۰/۵۹	۰/۴۸	۰/۶۲	۰/۷۰	۰/۶۹
۸	۰/۵۱	۰/۷۳	۰/۴۴	۰/۶۵	۰/۳۹	۰/۷۱	۰/۴۷	۰/۶۳	۰/۴۲	۰/۶۳	۰/۶۵	۰/۷۱
۹	۰/۲۹	۰/۵۹	۰/۳۹	۰/۶۵	۰/۲۴	۰/۶۲	۰/۳۴	۰/۶۰	۰/۲۶	۰/۵۶	۰/۵۵	۰/۷۲
۱۰	۰/۱۴	۰/۴۷	۰/۳۱	۰/۵۸	۰/۲۰	۰/۶۲	۰/۲۱	۰/۵۲	۰/۲۱	۰/۵۷	۰/۳۶	۰/۶۶
۱۱	۰/۰۸	۰/۳۹	۰/۱۷	۰/۵۲	۰/۱۳	۰/۵۵	۰/۰۹	۰/۴۰	۰/۱۱	۰/۴۴	۰/۲۲	۰/۵۸
۱۲	۰/۰۴	۰/۳۲	۰/۰۷*	۰/۳۳	۰/۱۰	۰/۵۲	۰/۰۸	۰/۴۱	۰/۰۶	۰/۳۶	۰/۱۸	۰/۵۶
۱۳	۰/۰۱	۰/۲۳	۰/۱۱	۰/۴۹	۰/۰۷*	۰/۴۶	۰/۰۳	۰/۳۰	۰/۰۲	۰/۲۲	۰/۱۰	۰/۴۷
۱۴	۰/۰۱	۰/۲۳	۰/۰۷	۰/۴۲	۰/۰۸	۰/۵۰	۰/۰۲	۰/۲۵	۰/۰۱	۰/۱۵*	۰/۰۷	۰/۴۱
۱۵	۰/۰۱	۰/۱۷*	۰/۰۵	۰/۳۵	–	–	۰/۰۱	۱۸*	–	–	۰/۰۴	۰/۳۳
۱۶	–	–	–	–	–	–	۰/۰۱	۱۴*	–	–	۰/۰۱	۰/۱۸*
۱۷	–	–	–	–	–	–	۰/۰۱	۱۲*	–	–	–	–
میانگین	۰/۴۴	۰/۴۸	۰/۴۸	۰/۵۱	۰/۴۵	۰/۵۷	۰/۳۸	۰/۴۱	۰/۴۳	۰/۴۵	۰/۵۲	۰/۵۲
میانه	۰/۵۱	۰/۴۷	۰/۴۴	۰/۵۲	۰/۴۳	۰/۵۹	۰/۳۴	۰/۴۲	۰/۴۰	۰/۴۷	۰/۶۰	۰/۵۱

در مدل ۲ پارامتری، هر سؤال براساس ۲ پارامتر a و b توصیف می‌گردد. نماد b بیانگر پارامتر دشواری سؤال است و جایگاه سؤال را در مقیاس توانایی توصیف می‌کند (بیکر، ۲۰۰۱). به زبان فنی و ریاضی می‌توان گفت که این پارامتر نقطه عطف تابع ویژگی سؤال را توصیف می‌کند و معمولاً در دامنه بین ۲/۵- تا ۲/۵+ مقیاس پردازی می‌شود.

پارامتر قدرت تشخیص در مدلهای IRT با نماد a نمایش داده می‌شود و بیانگر این است که سؤال تا چه اندازه می‌تواند بین افرادی که توانایی آنها پایین‌تر از جایگاه سؤال است با افرادی که توانایی آنها بالاتر از این جایگاه قرار دارد، تمایز ایجاد کند. به بیان دیگر، پارامتر a نشان می‌دهد که وقتی سطح توانایی مکنون بالا می‌رود، احتمال موفقیت در

سؤال با چه سرعتی افزایش می‌یابد (ثرندایک، ترجمه هومن، ۱۳۷۵). جدول ۵ پارامترهای سؤالها را براساس مدل دو پارامتری، نمایش می‌دهد.

جدول ۵: پارمتر b و a سؤالها براساس مدل منطقی دو پارامتری

سؤالها	خرده آزمونها												
	موزاییکها		طبقه بندیها		پازلها		قیاسها		موقعیتها		الگوها	
	b	a	b	a	b	a	b	a	b	a	b	a
۱	-۳/۱۴	۰/۹۲	-۲/۶۵	۱/۰۱	-۲/۳۰	۱/۰۶	-۲/۹۴	۰/۹۷	-۲/۹۶	۱/۱۷	-۲/۳۴	۱/۱۵
۲	-۱/۷۵	۱/۲۲	-۲/۱۸	۱/۳۴	-۱/۶۲	۱/۵۶	-۲/۱۰	۱/۵۲	-۲/۱۰	۱/۴۶	-۲/۲۱	۱/۱۴
۳	-۱/۱۹	۱/۱۰	-۱/۷۳	۱/۵۲	-۱/۰۵	۱/۳۱	-۱/۲۵	۱/۲۰	-۱/۴۰	۱/۳۷	-۱/۸۰	۱/۸۹
۴	-۰/۹۶	۲/۵۱	-۰/۹۶	۱/۵۱	-۰/۷۴	۱/۸۳	-۱/۱۶	۱/۵۳	-۱/۲۵	۱/۵۷	-۱/۷۲	۲/۲۰
۵	-۶۷۰	۲/۱۹	-۰/۵۰	۱/۴۳	-۰/۵۴	۱/۹۵	-۰/۰۶	۰/۹۰	۰/۴۶	۱/۰۵	-۱/۴۳	۱/۴۸
۶	-۰/۳۶	۲/۵۶	-۰/۲۶	۱/۹۷	-۰/۱۸	۱/۸۶	۰/۰۹	۱/۳۵	۰/۰۷	۱/۲۶	-۰/۶۳	۲/۷۴
۷	۰/۰۳	۳/۵۱	-۰/۱۳	۱/۵۹	۰/۲۴	۲/۵۳	۰/۱۵	۱/۷۲	۰/۲۱	۱/۷۰	-۰/۴۵	۲/۷۵
۸	۰/۲۱	۳/۷۹	۰/۳۶	۲/۷۵	۰/۴۷	۲/۸۰	۰/۳۰	۱/۳۵	۰/۴۲	۲/۱۱	-۰/۲۸	۳/۶۲
۹	۰/۶۶	۴/۲۸	۰/۴۸	۳/۲۲	۰/۸۷	۲/۳۵	۰/۵۸	۲/۸۴	۰/۸۰	۲/۳۷	-۰/۰۳	۴/۰۳
۱۰	۰/۹۶	۴/۵۸	۰/۶۳	۲/۶۱	۰/۹۵	۳/۴۴	۰/۸۳	۳/۱۳	۰/۸۹	۴/۳۲	۰/۴۸	۳/۰۴
۱۱	۱/۱۵	۴/۷۶	۰/۸۷	۳/۹۶	۱/۱۶	۳/۴۵	۱/۱۱	۳/۷۲	۱/۱۰	۴/۱۳	۰/۸۶	۲/۹۰
۱۲	۱/۳۳	۴/۹۳	۱/۲۲	۲/۸۸	۱/۲۶	۴/۵۴	۱/۰۹	۵/۲۸	۱/۲۳	۴/۱۵	۰/۹۶	۴/۲۰
۱۳	۱/۶۰	۴/۵۴	۰/۹۸	۵/۸۰	۱/۳۷	۶/۰۸	۱/۳۱	۴/۳۵	۱/۴۵	۴/۹۷	۱/۲۰	۵/۰۹
۱۴	۱/۵۷	۵/۸۰	۱/۱۰	۹/۴۷	۱/۳۱	۶/۰۸	۱/۳۵	۷/۵۶	۱/۶۳	۴/۲۰	۱/۳۰	۵/۲۰
۱۵	۲/۴۴	۱/۶۲	۱/۱۷	۵/۳۸	–	–	۱/۵۱	۵/۴۴	–	–	۱/۴۳	۵/۹۳
۱۶	–	–	–	–	–	–	۱/۷۱	۳/۶۷	–	–	۱/۷۶	۴/۶۵
۱۷	–	–	–	–	–	–	۲/۱۴	۲/۱۲	–	–	–	–
میانگین	۰/۱۳	۳/۲۲	-۰/۱۱	۳/۱۰	۰/۰۹	۲/۹۱	۰/۲۷	۲/۹۲	۰/۰۴	۲/۵۶	-۰/۱۸	۳/۲۵

اعتبار نمره ها

اعتبار (همسانی درونی) نمره های خرده آزمونها با استفاده از فرمول λ_2 (لاندای دو؛ فلدت و برنان، ۱۹۸۸؛ دی گروجتر و ون در کمپ، ۲۰۰۲) برآورد گردید. این فرمول نیز همانند فرمول آلفای کرونباخ (کرونباخ، ۱۹۵۱) شاخصی برای همسانی درونی سؤالهاست. با این حال، زمانی که تعداد سؤالها اندک است و یا زمانی که کوواریانس بین سؤالها ثابت (یکسان) نیست، فرمول λ_2 بر فرمول آلفای کرونباخ برتری دارد.

ضرایب همسانی درونی نمره‌های خرده‌آزمونها به تفکیک جنس و کل گروه در جدول ۸ ارائه شده است. در این جداول، همچنین ضرایب اعتبار نمره های مقیاس عملی[۱] (PS) و مقیاس استدلال[۲] (RS) و کل آزمون (IQ) که با استفاده از فرمول آلفای طبقه‌ای (اسبرن، ۲۰۰۰؛ نانالی و برنشتاین، ۱۹۹۴) برآورد شده اند نشان داده شده است. از آنجا که ضرایب λ_3 براساس نمونه‌هایی که از لحاظ سن ناهمگن هستند برآورد شدند، از این رو تأثیر سن با استفاده از فرمول زیر (تلخن، ارتباط شخصی) حذف گردید.

$$rel^*=(rel-cor^2)/(1-cor^2)$$

در فرمول مذکور *rel بیانگر ضریب اعتبار اصلاح شده برای سن است؛ rel بیانگر ضریب اعتبار اصلاح نشده؛ و cor^2 نشانگر مجذور همبستگی نمره کل خرده‌آزمون با سن است. ضرایب ارائه شده در جدول ۶، ضرایب اصلاح شده هستند.

جدول ۶: ضرایب اعتبار نمره‌های خرده آزمونها، مقیاسها و کل آزمون به تفکیک جنس و کل گروه

خرده آزمونها و مقیاسها	دختران	پسران	کل گروه
موزاییکها	۰/۸۳	۰/۸۶	۰/۸۳
طبقه بندیها	۰/۸۳	۰/۸۴	۰/۸۳
پازلها	۰/۸۴	۰/۸۸	۰/۸۵
قیاسها	۰/۸۴	۰/۸۶	۰/۸۶
موقعیتها	۰/۸۳	۰/۸۲	۰/۸۱
الگوها	۰/۸۴	۰/۸۵	۰/۸۵
مقیاس عملی	۰/۸۹	۰/۹۱	۰/۹۰
مقیاس استدلال	۰/۸۸	۰/۸۹	۰/۸۸
کل آزمون	۰/۹۴	۰/۹۵	۰/۹۴

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

! . Exploratory Factor Analysis
! . Confirmatory Factor Analysis

اعتبار آزمون- بازآزمون

اعتبار آزمون- بازآزمون به میزان ثبات عملکرد فرد در طول زمان اشاره دارد. جهت برآورد ثبات زمانی نمره های آزمون مورد بحث، از میان کل کودکانی که در مرحله اول (مرحله استاندارد سازی و هنجاریابی) مورد آزمون قرار گرفته بودند، تعداد ۱۲۵ کودک (۶۹ پسر و ۵۶ دختر) بصورت کاملاً تصادفی، انتخاب و مجدداً توسط همان آزمونگران مرحله اول، مورد آزمون قرار گرفتند. میانگین سن کودکان در اجرای دوم، ۴،۴ (چهار سال و چهار ماه) و انحراف استاندارد آن ۱،۲ (یکسال و دوماه) بود. دامنه سن کودکان در اجرای دوم، از ۲،۳ تا ۶،۱۱ بود. میانگین و انحراف استاندارد سن کودکان در مرحله اول اجرا، به ترتیب ۴،۳ و ۱،۳ بود. میانگین فاصله زمانی بین دوبار اجرای آزمون، ۵۸ روز و انحراف استاندارد آن ۲۵ روز بود. نتایج حاصل در جدول ۷ ارائه گردیده است. در این جدول، میانگین و انحراف استاندارد نمره‌های خام خرده آزمونها، مقیاسها و کل آزمون در هر بار اجرای آزمون و همچنین ضرایب همبستگی که نشان دهندهٔ میزان رابطه بین نمره های حاصل از دوبار اجرای آزمون است ارائه شده است. اندازهٔ ضرایب همبستگی نشان دهنده این است که آزمون حاضر از ثبات زمانی خوبی برخوردار است و خطای نمونه برداری زمانی آن اندک است.

جدول ۷: نتایج ثبات زمانی (N =۱۲۵)

خرده آزمونها و مقیاسها	اجرای اول		اجرای دوم		
	M	SD	M	D	r
موزاییکها	۶/۷۰	۲/۹۲	۷/۲۳	۲/۷۷	۰/۸۵
طبقه بندیها	۷/۳۵	۳/۲۹	۸/۱۳	۳/۴۴	۰/۷۹
پازلها	۶/۲۶	۳/۴۴	۷/۲۴	۳/۶۵	۰/۸۴
قیاسها	۶/۳۱	۲/۸۴	۶/۴۶	۲/۸۰	۰/۶۳
موقعیتها	۵/۹۸	۲/۹۴	۷/۱۴	۸/۱۶	۰/۸۱
الگوها	۸/۲۳	۳/۴۰	۸/۸۲	۳/۱۲	۰/۸۵
مقیاس عملی	۲۱/۲۰	۸/۹۸	۲۳/۳۰	۸/۹۴	۰/۹۳
مقیاس استدلال	۱۹/۶۴	۸/۳۰	۲۱/۷۴	۱۱/۳۳	۰/۸۳
کل آزمون	۴۰/۸۴	۱۶/۷۵	۴۵/۰۳	۱۸/۴۸	۰/۹۳

روایی نمره های آزمون

مقصود از روایی آزمون، پاسخ دادن به این سؤال است که آزمون چه چیزی را اندازه‌گیری می‌کند و تا چه اندازه از این لحاظ کارآیی دارد. در خصوص آزمون موردبحث، تحلیلهای مختلفی در زمینه روایی نمره ها صورت گرفته است، اما به دلیل این که مجلّه‌ها فضای محدودی به هرمقاله اختصاص می‌دهند، از اینرو در اینجا صرفاً خلاصه‌ای از تحلیل‌های صورت گرفته در مورد روایی نمره‌ها ارایه می‌شود. جهت مطالعه بیشتر در مورد روایی نمره ها به مینائی (زیرچاپ) مراجعه نمایید.

روایی سازه با این مسأله سروکار دارد که یک آزمون خاص، تا چه اندازه سازه یا صفت بخصوصی را اندازه می‌گیرد. گرونلند و لین (۱۹۹۰)برای نشان دادن این نوع روایی، یک روش ۳ مرحله‌ای را پیشنهاد می‌کنند. ابتدا چندسازه که فرض می‌شود عملکرد در آزمون را تبیین می‌کنند مشخص می‌شود. در مرحله دوم، براساس این سازه‌ها فرضیه‌هایی شکل می‌گیرد و در مرحله سوم، صحت و سقم این فرضیه‌ها با استفاده از شیوه‌های منطقی با تجربی مورد بررسی قرار می‌گیرد. سازه‌های اساسی که فرض می‌شود زیربنای عملکرد در آزمون SON-R ۲ ½ ۷- را تشکیل می‌دهند به صورت فرضیه‌هایی به شکل زیر مورد بررسی قرار گرفتند.

۱- از آنجا که هوش ماهیتاً وابسته به رشد است و نشان داده شده است که تا سن ۱۷ سالگی افزایش می‌یابد، لذا عملکرد در خرده آزمونها، مقیاسها و کل آزمون بایستی با سن تقویمی قویاً همبستگی داشته باشد.

۲- از آنجا که این آزمون، هوش را اندازه می‌گیرد بنابراین، نتایج آن باید بین کودکان عادی و کودکان عقب ماندهٔ ذهنی تمایز قائل شود.

تمایز گذاری سنی

در جدول ۸ میانگین نمره‌های خام خرده آزمونها، مقیاسها و کل آزمون در ۱۰ گروه سنی و همچنین ضرایب همبستگی (r) نمره‌های خرده آزمونها، مقیاسها و کل آزمون با سن ارائه گردیده است. محتوای جدول نشان می‌دهد که عملکرد در آزمون با سن رابطه دارد زیرا با

افزایش سن، میانگین خرده آزمونها، مقیاسها و کل آزمون افزایش می‌یابد. این نتیجه‌گیری با این واقعیت که تمامی‌ضرایب همبستگی به لحاظ آماری در سطح کمتر از ۰/۰۰۱ معنادار هستند مورد تأیید قرار می گیرد.

علاوه بر تحلیلهای فوق، میانگین نمره‌های گروههای سنی در خرده آزمونها، مقیاسها و کل آزمون با استفاده از تحلیل واریانس یکراهه (ANOVA) و دو تحلیل واریانس چند متغیری (MANOVA) بصورت جداگانه برای خرده آزمونها و مقیاسها مورد آزمون قرار گرفت.

نتیجه MANOVA با استفاده از آمارهٔ اثر هاتلینگ- لاولی (T) نشان داد که بین میانگین نمره‌های گروههای سنی مختلف در خرده آزمونها تفاوت آماری معنادار وجود دارد [$\eta^{\square}$=۰/۳۰، P<۰/۰۰۱، $T_{(۱۶۵۵۶،۵۴)}$ =۲/۵۱]. در ارتباط با مقیاسها نیز نتیجه MANOVA نشان داد که بین میانگین نمره‌های گروههای سنی، تفاوت آماری معنادار وجود دارد [$\eta^{۲}$ =۰/۵۴ ، P< ۰/۰۰۱، $T_{(۵۵۲۸،۱۸)}$ =۲/۳۵]. نتیجه آزمون ANOVA نیز نشان داد که تفاوت بین میانگین گروههای سنی مختلف در کل آزمون، از لحاظ آماری معنادار است [$\eta^{۲}$ =۰/۷۰ ، P<۰/۰۰۱ ، $F_{(۲۷۶۶،۹)}$ =۶۹۹/۴۰].

جدول ۸: میانگین نمره های خام خرده آزمونها، مقیاسها و کل آزمون در ۱۰ گروه سنی و همبستگی آنها با سن

خرده آزمونها و مقیاسها	گروه سنی										r
	۲/۵	۳/۰	۳/۵	۴/۰	۴/۵	۵/۰	۵/۵	۶/۰	۶/۵	۷/۰	
موزاییکها	۱/۸۳	۲/۴۴	۳/۶۶	۵/۰۰	۶/۴۸	۷/۱۷	۸/۱۷	۸/۷۷	۹/۳۴	۹/۸۱	۰/۷۸
طبقه بندیها	۲/۰۷	۳/۱۶	۴/۳۷	۵/۶۵	۶/۷۰	۷/۶۱	۸/۶۲	۹/۴۸	۱۰/۳۲	۱۰/۹۶	۰/۷۳
پازلها	۲/۰۱	۲/۱۸	۳/۲۵	۴/۱۰	۵/۵۹	۶/۶۸	۷/۸۴	۸/۶۸	۹/۶۱	۱۰/۹۷	۰/۷۵
قیاسها	۲/۰۲	۳/۰۴	۴/۴۳	۵/۶۰	۶/۳۰	۶/۸۰	۷/۳۷	۷/۹۷	۸/۶۷	۹/۱۲	۰/۶۲
موقعیتها	۲/۲۳	۲/۹۰	۳/۹۹	۴/۷۲	۵/۷۲	۶/۱۶	۷/۳۱	۷/۹۱	۸/۶۰	۹/۲۷	۰/۷۰
الگوها	۲/۳۲	۳/۹۷	۵/۵۱	۶/۹۵	۷/۸۶	۸/۶۷	۹/۵۱	۱۰/۴۶	۱۱/۶۶	۱۲/۱۳	۰/۷۶
مقیاس عملی	۶/۱۶	۸/۵۹	۱۲/۴۲	۱۶/۰۳	۱۹/۹۲	۲۲/۵۳	۲۵/۵۲	۲۷/۹۰	۳۰/۶۱	۳۲/۹۱	۰/۸۲
مقیاس استدلال	۶/۳۳	۹/۱۰	۱۲/۷۹	۱۵/۹۸	۱۸/۷۱	۲۰/۵۶	۲۳/۳۰	۲۵/۳۶	۲۷/۶۰	۲۹/۳۵	۰/۷۷
کل آزمون	۱۲/۴۹	۱۷/۶۹	۲۵/۲۰	۳۲/۰۱	۳۸/۶۴	۴۳/۰۹	۴۸/۸۲	۵۳/۲۶	۵۸/۲۰	۶۲/۲۶	۰/۸۲

تمایزگذاری گروهی

یکی از راههای ارزیابی روایی سازه یک آزمون، بررسی عملکرد گروههای مختلف در آزمون است. بایستی بین عملکرد گروههایی که فرض می‌شود در خصیصه مورد اندازه گیری آزمون، متفاوت هستند اختلاف وجود داشته باشد. بنابراین در مورد آزمون ۲ SON-R ۷½- انتظار می‌رود که عملکرد کودکان عقب ماندهٔ ذهنی در مقایسه با کودکان عادی، ضعیف‌تر باشد. جدول ۹ میانگین نمره‌های خام کودکان عقب ماندهٔ ذهنی و عادی را در خرده آزمونها، مقیاسها و کل آزمون نشان می‌دهد. در این جدول، همچنین ضریب همبستگی بین عضویت گروهی و نمره در خرده آزمونها، مقیاسها و کل آزمون را که با استفاده از همبستگی دورشته‌ای نقطه‌ای (گیلفورد و فروچتر، ۱۹۷۸) محاسبه گردید نشان داده شده است.

لازم به ذکر است با توج به تفاوت فاحش بین حجم نمونه کودکان عادی (N=۲۷۷۵) و حجم نمونه کودکان کم توان ذهنی (N=۴۴) و با عنایت به اینکه مقایسه گروهها زمانی دقیق‌تر خواهد بود که حجم گروهها یکسان باشد، لذا از بین کل نمونه کودکان عادی، تعداد ۴۴ کودک (۲۴ پسر و ۲۰ دختر) بطور کاملاً تصادفی انتخاب گردید. دامنه سنی این کودکان از ۲;۱۰ تا ۶;۸ و میانگین و انحراف استاندارد سن آنها به ترتیب برابر با ۴/۳۸ و ۱/۱۶ است. کلیه تحلیلهای مربوط به تمایزگذاری گروهی که در اینجا ارائه شده است براساس عملکرد این ۴۴ کودک انجام گرفته است.

جدول۹، میانگین، حداقل و حداکثر نمره خام کودکان عادی و کودکان عقب مانده ذهنی را در خرده آزمونها، مقیاسها و کل آزمون نشان می‌دهد. با نگاه به ارقام ارائه شده در جدول، می‌توان دریافت که میانگین نمره‌های کودکان عقب‌مانده ذهنی بطور قابل ملاحظه‌ای پایین‌تر از میانگین نمره‌های کودکان عادی است. همچنین کلیه ضرایب همبستگی در سطح کمتر از ۰/۰۰۱ به لحاظ آماری معنادار هستند. علاوه بر این، نتیجه تحلیل واریانس چند متغیری دو گروهی با استفاده از آمارهٔ T^2 هاتلینگ نشان داد که میانگین نمره‌های کودکان عقب‌مانده‌ذهنی در خرده‌آزمونها [η^2=۰/۳۳، P<۰/ ۰۰۱، ۰/۴۹= $T^2_{(۸۱،۶)}$] و مقیاسها [η^2=۰/۲۶، P<۰/۰۰۱، ۰/۴۰= $T^2_{(۲،۳۸)}$] بطور معناداری پایین‌تر از میانگین

نمره‌های کودکان عادی است. به همین نحو، نتیجه آزمون t برای دو گروه مستقل نیز نشان داد که میانگین نمره‌های کودکان عقب مانده ذهنی در کل آزمون بطور معناداری پایین‌تر از میانگین نمره‌های کودکان عادی است [r^2 =۰/۳۱ ، P< ۰/۰۰۱ ، $t_{(۸۶)}$=۵/۵۱].

جدول۹: میانگین، حداقل و حداکثر نمره خام کودکان عادی و عقب مانده ذهنی در خرده آزمونها، مقیاسها و کل آزمون

خرده آزمونها و مقیاسها	کودکان عادی (N =۴۴)			کودکان عقب مانده ذهنی (N =۴۴)			
	میانگین	حداقل	حداکثر	میانگین	حداقل	حداکثر	r_{pbis}
موزاییکها	۶/۶۶	۰	۱۱	۳/۹۸	۰	۱۱	۰/۴۰
طبقه بندیها	۷/۶۸	۲	۱۴	۴/۱۸	۰	۱۲	۰/۴۹
پازلها	۶/۳۲	۱	۱۴	۳/۵۰	۰	۱۲	۰/۴۰
قیاسها	۶/۷۸	۲	۱۳	۳/۷۰	۰	۱۱	۰/۴۷
موقعیتها	۶/۲۵	۲	۱۲	۳/۲۵	۰	۱۰	۰/۴۸
الگوها	۸/۷۷	۲	۱۴	۴/۸۰	۰	۱۳	۰/۵۱
مقیاس عملی	۲۱/۷۵	۳	۳۸	۱۲/۲۷	۰	۳۳	۰/۴۷
مقیاس استدلال	۲۰/۷۰	۹	۳۷	۱۱/۱۴	۰	۲۶	۰/۵۳
کل آزمون	۴۲/۴۵	۱۲	۷۵	۲۳/۴۱	۰	۵۷	۰/۵۴

ساختار عاملی

روایی سازه، همچنین به میزان قابلیت تشخیص صفات زیربنایی آزمون و اینکه این صفات تا چه اندازه، مدل نظری را که آزمون بر آن مبتنی است منعکس می‌کنند، مرتبط است. به منظور شناسایی و تعیین عامل یا عاملهایی که زیربنای عملکرد در خرده آزمونها را تشکیل می‌دهند، ابتدا کل گروه کودکان عادی (N = ۲۷۷۵) بطور تصادفی به دو زیر نمونه تقسیم گردید. زیر نمونه اول، شامل ۱۳۷۸ کودک (۶۷۷ پسر و ۷۰۱ دختر) با میانگین سنی ۴،۳ و انحراف استاندارد ۱،۳ (یک سال و ۳ ماه) است. زیر نمونه دوم نیز ۱۳۹۷ کودک (۷۳۷ پسر و ۶۶۰ دختر) با میانگین سنی ۴;۳ و انحراف استاندارد ۱;۳، تشکیل یافته است. هدف از تقسیم نمونه استاندارد سازی به دو زیر نمونه این بود که با استفاده از یکی از زیرنمونه‌ها،

ساختار عاملی آزمون با استفاده از تحلیل عاملی اکتشافی (EFA)[1] مورد تحلیل قرار گیرد و عامل یا عاملهای زیربنایی، شناسایی و تعیین گردد و آنگاه مدل عاملی حاصل از زیر نمونه اول با استفاده از روش تحلیل عاملی تأییدی، در زیر نمونه دوم مورد آزمون قرار گیرد.

برای شناسایی و تعیین تعداد عاملهای مورد نیاز برای تبیین واریانس مشترک خرده آزمونها، ماتریس همبستگی حاصل از زیر نمونه اول با استفاده از روش عامل‌یابی محورهای اصلی (PAF)[2] که از رایج‌ترین روشهای استخراج عاملها در تحلیل عاملی اکتشافی است (راسل، ۲۰۰۲؛ هنسون و رابرتس، ۲۰۰۶) مورد تحلیل قرار گرفت. البته، پیش از اجرای تحلیل عاملی، آزمونهای KMO و کرویت بارتلت محاسبه گردید. نتیجه حاصل ازاین دوآزمون، نشان دادکه اجرای تحلیل عاملی، قابل توجیه است [KMO=۰/۹۳، P<۰/۰۰۱، ۷۰۹۴/۷۴= آزمون کرویت بارتلت].

جهت تعیین تعداد عاملها از روش تحلیل موازی (PA)[3] و روش حداقل میانگین همبستگی تفکیکی (MAP)[4] که جزو دقیقترین روشهای تعیین تعداد عاملها به حساب می‌آیند (تامپسون، ۲۰۰۴؛ راسل، ۲۰۰۲؛ هنسون و رابرتس، ۲۰۰۶) استفاده شد. روش تحلیل موازی توسط هورن (۱۹۶۵) و روش حداقل میانگین همبستگی تفکیکی، توسط ولیسر (۱۹۷۶) ارائه شده است. روش MAP در سال ۲۰۰۰ توسط ولیسر، اتیدن و فاوا (۲۰۰۰) اصلاح گردید. روش PA و روش MAP به ترتیب از طریق برنامه parallel.sps و map.sps (اوکانر، ۲۰۰۰) قابل اجرا است.

اگرچه نتیجه حاصل از هر دو روش PA و MAP حاکی از این بود که تنها یک عامل قابل استخراج است، اما بنا به پیروی از پیشنهاد گورساچ (۱۹۹۷) یک عامل دیگر نیز استخراج گردید.

! . Exploratory factor analysis
! . Principle Axis Factoring
! . Parallel Analysis
! . Minimum Average Partial

برای چرخش عاملها، بنا به توصیه فابریگر و همکاران (۱۹۹۹) و راسل (۲۰۰۲) از چرخش متمایل و از روش پروماکس[1] که روش بسیار خوبی است (تامپسون، ۲۰۰۴) استفاده گردید. مقدار kappa برابر با ۴ قرار داده شد. میانگین، انحراف استاندارد و همبستگی درونی خرده‌آزمونها در زیر نمونه اول، در جدول ۱۰ ارائه شده است.

جدول۱۰: میانگین، انحراف استاندارد و همبستگی درونی خرده آزمونها در زیر نمونه اول (N=۱۳۷۸)

خرده آزمونها	خرده آزمونها					
	Mos	Cat	Puz	Ana	Sit	Pat
موزاییکها(Mos)	۱	۰/۷۹	۰/۸۲	۰/۶۸	۰/۷۵	۰/۸۰
طبقه بندیها(Cat)		۱	۰/۷۴	۰/۶۷	۰/۷۳	۰/۷۴
پازلها(Puz)			۱	۰/۶۵	۰/۷۵	۰/۷۵
قیاسها(Ana)				۱	۰/۶۵	۰/۶۲
موقعیتها(Sit)					۱	۰/۷۱
الگوها(Pat)						۱
میانگین	۶/۵۵	۷/۱۱	۶/۲۸	۶/۳۵	۶/۱۱	۸/۲۳
انحراف استاندارد	۳/۰۳	۳/۳۸	۳/۵۳	۲/۹۶	۲/۸۲	۳/۳۷

ماتریس الگوی عاملی و ماتریس ساختار عاملی چرخش یافته، در جدول ۱۱ ارائه شده است. اگر چه درک و تفسیر ماتریس ساختار عاملی آسانتر است، اما تاباکنایک و فیدل (۲۰۰۱) اعتقاد دارند زمانی که از چرخش متمایل استفاده می‌شود، همبستگی بین عاملها و متغیرها (ضرایب ماتریس ساختار عاملی) به خاطر وجود همبستگی بین عاملها متورم می‌شود و هر چقدر همبستگی بین عاملها بالا باشد، مقدار این توَرم بیشتر خواهد بود و تعیین این که کدام متغیر به کدام عامل، مربوط می‌شود بسیار دشوار می‌گردد. از سوی دیگر، ماتریس الگوی عاملی، سهم یگانه هر عامل را در واریانس متغیرها نشان می‌دهد. در پژوهش حاضر، با توجه به اینکه همبستگی بین دو عامل، ۰/۸۶ است که همبستگی بسیار

!. Promax

قوی به حساب می‌آید (نصفت، ۱۳۷۱)، از این رو در تفسیر عاملها از ماتریس الگوی عاملی استفاده گردید.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

جدول ۱۱: ماتریس الگوی عاملی و ساختار عاملی چرخش یافته با روش پروماکس

خرده آزمونها	ماتریس الگوی عاملی		ماتریس ساختار عاملی	
	عامل ۱	عامل ۲	عامل ۱	عامل ۲
الگوها	۰/۷۹	۰/۰۹	۰/۸۷	۰/۷۷
موزاییکها	۰/۷۵	۰/۲۰	۰/۹۲	۰/۸۴
پازلها	۰/۶۲	۰/۲۹	۰/۸۷	۰/۸۲
قیاسها	۰/۱۱	۰/۶۸	۰/۷۰	۰/۷۸
موقعیتها	۰/۳۳	۰/۵۵	۰/۸۰	۰/۸۳
طبقه بندیها	۰/۳۹	۰/۵۱	۰/۸۳	۰/۸۵

ضرایب ماتریس الگوی عاملی، تمایز روشنی را بین خرده‌آزمونهای عملی (الگوها، موزاییکها و پازلها) و خرده‌آزمونهای استدلال (قیاسها، موقعیتها و طبقه‌بندیها) نشان می‌دهد. عامل اول، ۷۲٪ و عامل دوم ۵٪ از واریانس کل را تبیین می‌کنند.

لازم به ذکر است، تلخن و همکاران (۱۹۹۸) نیز با استفاده از تحلیل مؤلفه‌های اصلی (PCA)[1] به دو عامل دست یافتند. عامل اول، ۴۷٪ و عامل دوم، ۱۳٪ از واریانس کل را تبیین می‌کرد.

جهت تعیین میزان تفاوت بارهای عاملی در ۳ گروه سنی (۳-۲ سال؛ ۵-۴ سال؛ ۷-۶ سال) با استفاده از داده‌های زیر نمونه اول، یک تحلیل مؤلفه‌های همزمان[2] (SCA)؛ میل سپ و مردیت، ۱۹۸۸؛ کایرس و تنبرگ، ۱۹۸۹) صورت گرفت. هدف از این تحلیل، این بود که ببینیم آیا راه حل یکنواختی از وزنهای عاملی، در مقایسه با راه حلهای بهینه برای گروههای سنی مجزاً، بطور اساسی واریانس کمتری را تبیین می‌کند. این تحلیل که با استفاده از برنامه SCA (کایرس، ۱۹۹۰) صورت گرفت نشان داد که این موضوع، صادق نیست. در واقع راه حل یکنواخت بر روی ۳ گروه سنی، ۶۸/۴٪ از واریانس کل و راه حلهای بهینه جداگانه، ۶۸/۴۴٪ از واریانس کل را تبیین می‌کردند.

---

1. Principle Component Analysis
2. Simultaneous Component Analysis

تحلیل عامل تأییدی

برازش مدل ۲ عاملی حاصل از زیر نمونه اول با داده‌های زیر نمونه دوم، با استفاده از تحلیل عاملی تأییدی موردآزمون قرار گرفت. زیر نمونه دوم شامل ۱۳۹۷ کودک (۷۳۷ پسر و ۶۶۰ دختر) بود. میانگین و انحراف استاندارد سن آنها به ترتیب برابر با ۴،۳ و ۱،۴ است. جدول ۱۲، همبستگی درونی، انحراف استاندارد، چولگی و کشیدگی خرده‌آزمونها را در زیر نمونه دوم نشان می‌دهد. اگر چه تحلیلهای توصیفی نشان داد که توزیع نمره‌های تک تک خرده آزمونها تقریباً نرمال است، اما آزمون نرمال بودن چندمتغیری، چولگی و کشیدگی معناداری را نشان داد (χ^2 = ۲۰۵/۹۵ ،P< ۰/۰۰۰۱). بنابراین برای برآورد پارامترها و برازش مدل از ماتریس کوواریانس و روش کمترین مجذورات وزن‌دار (WLS؛ جارزکاگ و سوربوم، ۲۰۰۱) استفاده شد. از ماتریس واریانس – کوواریانس مجانبی که توسط برنامه PRELIS ۲ (ویرایش ۲/۷؛ جارزکاگ و سوربوم، ۲۰۰۲) ایجاد شد، به عنوان ماتریس وزنی استفاده گردید.

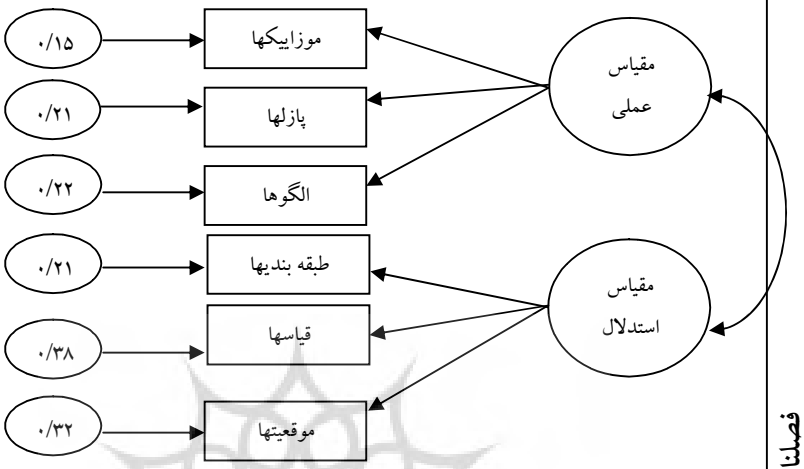
جدول۱۲: میانگین، انحراف استاندارد و همبستگی درونی خرده آزمونها در زیر نمونه دوم (N=۱۳۹۷)

خرده آزمونها	خرده آزمونها					
	Mos	Cat	Puz	Ana	Sit	Pat
موزاییکها(Mos)	۱	۰/۷۹	۰/۸۲	۰/۶۹	۰/۷۳	۰/۸۱
طبقه بندیها(Cat)		۱	۰/۷۶	۰/۷۰	۰/۷۳	۰/۷۷
پازلها(Puz)			۱	۰/۶۷	۰/۷۶	۰/۷۷
قیاسها(Ana)				۱	۰/۶۶	۰/۶۸
موقعیتها(Sit)					۱	۰/۷۱
الگوها(Pat)						۱
انحراف استاندارد	۳/۰۷	۳/۳۷	۳/۵۲	۳/۰۰	۲/۸۰	۳/۴۰
چولگی	-۰/۲۶	-۰/۰۶	۰/۱۹	۰/۰۶	۰/۱۱	-۰/۳۰
کشیدگی	-۰/۵۰	-۰/۴۹	-۰/۷۴	-۰/۳۴	-۰/۶۵	-۰/۱۹

برای سنجش کفایت برازش مدل، آماره‌ها و شاخصهای مختلفی ارائه شده است. از آنجا که هر یک از این شاخصها تنها جنبهٔ خاصی از برازش مدل را منعکس می‌سازند (کلاین، ۲۰۰۵)، از این رو برای سنجش برازش مدل، معمولاً از چندین شاخص استفاده می‌شود؛ برای مثال، کلاین (۲۰۰۵) آماره‌ها و شاخصهای زیر را پیشنهاد می‌کند: آماره خی دو (χ^2)، ریشه دوم برآورد واریانس خطای تقریب (RMSEA)، شاخص برازش تطبیقی (CFI) و ریشه دوم میانگین مجذورات پس‌مانده‌های استاندارد شده (SRMR). سان (۲۰۰۵) نیز برای سنجش برازش مدل در مطالعاتی از نوع مطالعه حاضر که هدف آن بررسی روایی سازه ابزارهای روان سنجی است، شاخصهای زیر را پیشنهاد می‌کند: ریشه دوم برآورد واریانس خطای تقریب (RMSEA)، شاخص تاکر- لویز (TLI) یا شاخص نرم نشده برازش (NNFI)، شاخص برازش تطبیقی (CFI)، شاخص مرکزی مک دونالد (MC) و ریشه دوم میانگین مجذورات پس‌مانده‌های استاندارد شده (SRMS). در مطالعه حاضر، برای سنجش برازش مدل، به غیر از آماره χ^2 که به حجم نمونه و انحراف از نرمال بودن چندمتغیری بسیار حسّاس است (جارزکاگ و سوربوم، ۲۰۰۱) و شاخص مرکزی مک دونالد (MC) که LISREL ۸ آن را ارائه نمی‌دهد از همه شاخصهای پیشنهادی کلاین (۲۰۰۵) و سان (۲۰۰۵) استفاده گردید. علاوه بر شاخصهای فوق الذکر از شاخص خوبی برازندگی (GFI) و شاخص تعدیل یافته خوبی برازندگی (AGFI) (جارزکاگ و سوربوم، ۱۹۸۹) نیز استفاده شد. این دو شاخص به حجم نمونه بستگی ندارند و نشان می‌دهند که مدل تا چه حد نسبت به عدم وجود آن، برازندگی بهتری دارد (هومن، ۱۳۸۴).

در تفسیر شاخصهای برازندگی، قواعد پیشنهادی زیر به کار برده شد. براونی و کادک (۱۹۹۳) معتقدند که مقادیر بالاتر از ۰/۹۰ برای شاخصهای TLI و CFI حاکی از برازش خوب مدل است. در مورد شاخص RMSEA نیز از ملاک پیشنهادی هیو و بنتلر (۱۹۹۹) استفاده شد. آنها معتقدند اگر مقدار این شاخص کمتر از ۰/۰۶ باشد می‌توان نتیجه گرفت که برازش مدل با داده‌ها خوب است. در ارتباط با شاخص SRMR نیز مقادیر کمتر از ۰/۱۰ به طورکلی به صورت مقادیر مطلوب در نظر گرفته می‌شود (کلاین، ۲۰۰۵).

شکل ۱، مدل مورد آزمون و برآوردهای استاندارد شدهٔ پارامترها را نشان می‌دهد. بار عاملی همه خرده آزمونها در سطح P< ۰/۰۱ به لحاظ آماری، معنادار است.



شکل ۱: مدل مورد آزمون و برآوردهای استاندارد شدهٔ پارامترها

مقدار شاخصهای AGFI، GFI، SRMR، TLI، CFI، و RMSEA به ترتیب برابر با ۰/۹۹، ۰/۹۹، ۰/۰۰۵، ۰/۹۹، ۰/۹۹ و ۰/۰۲ است. فاصله اطمینان ۹۰٪ برای شاخص RMSEA برابر با ۰ تا ۰/۰۴، است. مقدار تمام این شاخصها با ملاکها و قواعد سرانگشتی که در تفسیر آنها بکار می‌رود (کلاین، ۲۰۰۵) مطابقت دارد. بنابراین می‌توان نتیجه گرفت که مدل ۲ عاملی از برازش خوبی با داده ها برخوردار است.

جداول نرم

جداول نرم برای نمره‌های خرده آزمونها، مقیاسها و IQ در ۱۰ گروه سنی به فاصله ۶ ماه از یکدیگر از (۲;۳ تا ۷ سال تمام) تهیه گردیده است. نتایج MANOVA با استفاده از آزمون T^2 هاتلینگ نشان داد که بین میانگین نمره‌های دختران و پسران در خرده آزمونها و مقیاسها، تفاوت آماری معنادار وجود ندارد. مقادیر مشاهده شدهٔ T^2 هاتلینگ برای خرده آزمونها و مقیاسها به ترتیب: [P= ۰/۰۹، P= ۰/۰۲، $T^2_{(6, 2769)}$] و [P= ۰/۱۷، P= ۰/۰۰۲، $T^2_{(2773,2)}$]. نتیجه آزمون t نیز نشان داد که بین میانگین نمره‌های دختران و پسران در کل

آزمون، تفاوت آماری معنادار وجود ندارد. مقدار مشاهده شده t برابر با [۰/۱۵= P , ۱/۴۶ = $t_{(۲۷۷۴)}$] است.

با توجه به این واقعیت که تفاوت بین نمره‌های گروه دختران و پسران در خرده آزمونها برای هیچ یک از گروههای سنی مورد مطالعه از لحاظ آماری معنادار نبود از این رو، گروههای دختر و پسر با هم ترکیب و نمره‌های مربوط به کل گروه نمونه، در هر یک از سطوح سنی مبنای محاسبه نمره های هنجار قرار گرفت.

جهت به دست آوردن نمره‌های مقیاسی خرده‌آزمونها ابتدا توزیع فراوانی و توزیع فراوانی تراکمی نمره‌های خام شکل گرفت. آنگاه رتبه‌های درصدی متناظر با نمره‌های خام محاسبه گردید. در گام بعد مقادیر Z متناظر با رتبه‌های درصدی با استفاده از جدول توزیع نرمال استاندارد بدست آمد. در گام آخر، نمره‌های Z بدست آمده برای خرده آزمون با استفاده از یک تبدیل خطی به توزیعی با میانگین ۱۰ و انحراف استاندارد ۳ تبدیل گردید.

دلیل اینکه توزیع نمره‌های خام به توزیع نرمال تبدیل گردید این است که «در توزیع نمره‌های خام یا در توزیع دیگری که با تبدیل خطی توزیع نمره‌های خام بدست آمده، نمره‌ها معنی دقیق آماری ندارند... وضع نسبی نمره معیار بخصوصی در توزیعی که بهنجار نشده، کاملاً تابع شکل توزیع است... برای رفع معایب نمره های خام یا نمره‌های دیگری که از تبدیل خطی آنها بدست آمده‌اند می‌توان شکل توزیع را به نحوی تغییر داد که نمره‌های حاصل، توزیع بهنجار داشته باشد. به عبارت دیگر باید توزیع مورد نظر را بهنجار کرد» (مگنوسون، ۱۹۶۷؛ ترجمه براهنی، ۱۳۷۰). علاوه بر این، نرمال سازی توزیعهای چولهٔ نمره‌های خام برای هدفهای مقایسه، مطلوب است (کوهن و اسوردلیک، ۲۰۰۲).

برای محاسبه نمره‌های استاندارد و رتبه‌های درصدی مقیاس عملی و استدلال، مجموع نمره‌های مقیاسی خرده آزمونهای مربوط به هر کدام از مقیاسها بدست آمد. مجموع نمره‌های مقیاسی ۳ خرده آزمون طبقه بندیها، موقعیتها و قیاسها مبنای تشکیل جدول نرم مقیاس استدلال را تشکیل می‌دهند ومجموع نمره‌های مقیاسی ۳ خرده آزمون موزاییکهای، پازلها و الگوها پایه و اساس تشکیل جدوم نرم مقیاس عملی را تشکیل می دهند.

پس از محاسبه مجموع نمره‌های مقیاسی مقیاسها، توزیع فراوانی و توزیع تراکمی این نمره ها به تفکیک گروه سنی شکل گرفت. آنگاه رتبه‌های درصدی نمره‌های مقیاسی بدست آمد. در نهایت نمره‌های Z متناظر با رتبه‌های درصدی با استفاده از جدول نرمال استاندارد بدست آمد و این نمره های Z با استفاده از یک تبدیل خطی به توزیعی با میانگین ۱۰۰ و انحراف استاندارد ۱۵ تبدیل گردید.

برای محاسبه هوشبهر و رتبه‌های درصدی متناظر، مجموع نمره‌های مقیاسی تمام ۶ خرده‌آزمون بدست آمد. پس از محاسبه مجموع نمره‌های مقیاسی ۶ خرده آزمون، توزیع فراوانی و توزیع تراکمی این نمره‌ها به تفکیک گروه سنی شکل گرفت. آنگاه رتبه درصدی هر نمره محاسبه گردید. در نهایت نمره‌های Z متناظر با رتبه‌های درصدی با استفاده از جدول توزیع نرمال استاندارد، بدست آمد. این نمره‌های Z با استفاده از یک تبدیل خطی به توزیعی با میانگین ۱۰۰ و انحراف استاندارد ۱۵ تبدیل گردید.

بحث و نتیجه‌گیری

در این پژوهش، ویژگیهای روان سنجی آزمون ۲ SON-R ½ ۷- برای کودکان ایرانی مورد بررسی قرار گرفت. نتایج این مطالعه نشان داد که افزایش دشواری سؤالها به همان صورتی است که انتظار می‌رفت (جدول ۴). از بین ۹۱ سؤال، تنها ۳ سؤال وجود دارد که دشوارتر از سؤال بعدی هستند، که البته تفاوت بین دشواری ۲ تا از این سؤالها با سؤال بعدی، تنها ۰/۰۱ است. این ۳ سؤال با علامت * در جدول ۴ نشان داده شده اند.

بطور کلی، نتایج حاکی از این است که درجه دشواری سؤالها در سطح مطلوب قرار دارند. آناستازی (۱۹۹۰) و یوربینا (۲۰۰۴) اظهار می‌دارند که متوسط دشواری سؤالها باید تقریباً ۰/۵۰ باشد و همچنین باید پراکندگی نسبتاً زیادی در دشواری سؤالها وجود داشته باشد.

ضرایب ارائه شده در جدول ۴ حاکی از این است که اکثر سؤالها از قدرت تشخیص مطلوب و بالایی برخوردارند و صرفاً ۶ سؤال که با علامت * مشخص شده اند قدرت تشخیص نسبتاً ضعیفی دارند. نانالی و برنشتاین (۱۹۹۴) اظهار می‌دارند که اکثر ضرایب

همبستگی سؤال– نمره کل، در دامنه ۰ تا ۰/۴۰ قرار دارند. گارت (۱۹۶۵؛ به نقل از نیوکامر و هامیل، ۱۹۹۷) اظهار کرده است که سؤالهای دارای ضریب همبستگی ۰/۲۰ یا بالاتر، در صورتی که آزمون نسبتاً طولانی باشد می‌توانند دارای روایی تلقی شوند. آناستازی (۱۹۹۰) و آناستازی و یوربینا (۱۹۹۷) توصیه می‌کنند که ضرایب همبستگی ۰/۲۰ یا ۰/۳۰ می‌توانند قابل قبول تلقی گردند.

نتایج این مطالعه نشان می‌دهد که نمرات خرده آزمونها، مقیاسها و کل آزمون از همسانی درونی بالایی برخوردارند. همانطور که در جدول ۶ ملاحظه می‌کنید ضرایب همسانی درونی نمره‌های خرده آزمونها، مقیاسها و کل آزمون برای هر دو جنس، کل گروه و گروههای سنی مختلف، تماماً در سطح مطلوبی قرار دارند. این یافته که با یافته‌های حاصل از مطالعات تلخن و همکاران (۱۹۹۸)، و تلخن و لاروس (۲۰۰۴) همخوانی دارد بیانگر این است که اعتبار نمرات آزمون به اندازه‌ای است که بتوان با اطمینان از آنها استفاده کرد. در واقع، ضریب اعتبار ۰/۸۰ بوسیله اغلب صاحب‌نظران برای تمیز بین آزمونهای معتبر و غیرمعتبر پذیرفته شده است (آناستازی، ۱۹۹۰) و اکثر کاربران نیز به دنبال آزمونهایی با ضریب اعتبار حداقل ۰/۸۰ و بالاتر هستند (یوربینا، ۲۰۰۴).

براساس یافته‌های مربوط به تمایزگذاری سنی (جدول ۸)، تمایز گذاری گروهی (جدول ۹)، و تحلیل عاملی (جدول ۱۱) می‌توان چنین نتیجه گرفت که این آزمون یک ابزار روا جهت سنجش هوش کلی است و می‌توان با اطمینان خاطر از آن استفاده کرد. علاوه براین، یافته‌های حاصل از مطالعات بین‌فرهنگی که درخصوص روایی این آزمون در کشورهای آمریکا، استرالیا، و انگلستان صورت گرفته است (لاروس و تلخن، ۱۹۹۹) حاکی از این است که همبستگی این آزمون با مقیاسهای عملی تعدادی از آزمونها، مانند آزمون WPPSI-R، BAS و MSCA، قویتر از همبستگی آن با مقیاسهای کلامی این آزمونها است. این همبستگی‌ها از روایی واگرا و همگرای آزمون مورد مطالعه حمایت می‌کند.

بطورخلاصه می‌توان گفت که آزمون SON-R ۲ ½-۷ با در نظر گرفتن یافته‌های مربوط به اعتبار، روایی، مدت اجرا، سهولت نمره‌گذاری و تفسیر که از مهمترین جنبه‌های

عملی بودن آزمون به حساب می‌آید (هومن، ۱۳۸۱) یک ابزار کاملاً مناسب برای سنجش هوش کلی کودکان۷ – ۲/۵ سال ایران است و از آن می‌توان بعنوان وسیله‌ای معتبر و روا برای: الف) شناسایی کودکانی که بطور قابل ملاحظه‌ای از لحاظ شناختی، ضعیف‌تر از همسالان خود هستند، ب) مستند کردن پیشرفت کارکرد شناختی کودکان در نتیجه برنامه‌های مداخله‌ای ویژه، ج) مشخص کردن نقاط قوت و ضعف کودکان در مهارتهای شناختی، و د) اندازه‌گیری هوش کلی در مطالعات پژوهشی استفاده کرد.

منابع

آناستازی، آنا.(۱۳۶۴). *روان آزمایی*. ترجمهٔ محمدتقی براهنی. تهران: انتشارات دانشگاه تهران.

ثرندایک، آر. ال. (۱۳۷۵). *روان سنجی کاربردی*. ترجمهٔ حیدرعلی هومن. تهران: انتشارات دانشگاه تهران (تاریخ انتشار به زبان اصلی ۱۹۸۲).

مگنوسون، داوید.(۱۳۷۵). *مبانی نظری آزمونهای روانی*. ترجمهٔ محمدتقی براهنی. تهران: انتشارات دانشگاه تهران. (سال انتشار به زبان اصلی، ۱۹۶۷).

مینائی، اصغر.(۱۳۸۴). انطباق و استانداردسازی آزمون غیرکلامی هوش اسنایدرز– اومان برای کودکان ۷–۲/۵ سال: یک مطالعه مقدماتی. *پژوهش در حیطه کودکان استثنایی، ۱۷*، ۳۲۲–۲۹۵.

مینائِی، اصغر.(زیر چاپ). انطباق و استانداردسازی آزمون غیرکلامی هوش اسنایدرز– اومان برای کودکان ۷–۲/۵ سال: کتابچه راهنما و گزارش پژوهش. تهران: پژوهشکده کودکان استثنایی.

نصفت، مرتضی.(۱۳۷۱). *اصول و روشهای آمار*. جلد اول. تهران: انتشارات دانشگاه تهران.

هومن، حیدرعلی. (۱۳۸۴). *مدل یابی معادلات ساختاری با کاربرد نرم افزار لیزرل*. تهران: سمت.

هومن، حیدرعلی (۱۳۸۱). *اندازه‌گیریهای روانی و تربیتی: فن تهیه تست و پرسشنامه*. تهران: نشر پارسا.

Achenbach,T.M., & Rescorla, L. A. (۲۰۰۱). *Manual for the ASEBA School-Age Forms*
& *Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.

Anastasi, A.(۱۹۹۰).*Psychological testing*. (۶th Ed). New York: Macmillan.

Anastasi. A; & Urbina, S. (۱۹۹۷). *Psychological testing.* (۷th ed). Upper Saddle River, NJ. Prentice - Hall

Baker, F.B. (۲۰۰۱). *The basics of item response theory*. Eric clearinghous on assessment and evaluation. University of Maryland, college park.

Binet, A. & Simon, T. (۱۹۱۶). *The development of intelligence in children* (E. S. Kit, Tras.). Baltimor: Williams & Wilkins.(Original work published ۱۹۰۵).

Brown, L., Sherbenou, R. J., & Johnson, S. K.(۱۹۹۷). *Test of nonverbal intelligence (۳rd ed.)*. Austin, TX: PRO-ED.

Browne, M.W., & Cudeck, R. (۱۹۹۳). Alternative ways of assessing model fit. In K.A. Bollen & J. S. Long (Eds). *Testing Structural Equation Model (pp. ۱۳۶-۱۶۲).* Newbury Park: Sage

Cattell, R.B. (۱۹۷۱). *Abilities: their structure, growth, and action.* Boston: Houghton Mifflin.

Cattell, R.B. (۱۹۵۰). *Handbook for the individual of group culture Fair Intelligence Test. Scale I.* Champaign,IL: Institute for PERSONALITY and Ability testing.

Cohen, R. J. & Swerdilk, M. R.(۲۰۰۲). *Psychological testing and assessment: An introduction to test and measurement(۵th ed.)*. New York: McGraw-Hill.

Cronbach, L. J.(۱۹۵۱). Coefficient alpha and the internal structure of tests. *Psychometrika, ۱۶*, ۲۹۷-۳۳۴.

De Gruijter, D. N. M. & Van Der Kamp, L. J. Th. (۲۰۰۲). *Statistical test theory for education and psychology.* Available in PDF format at: http://www.tu-dresden.de/erzwiae/ewwm/lehre/q۲_ss۰۶/material/measurement.pdf.[Accessed ۲۰ September ۲۰۰۴]

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E. J.(۱۹۹۹). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods,۴*, ۲۷۲-۲۹۹.

Feldt, L. S; & Brennan, R.L.(۱۹۸۸). Reliability. In R.L. Linn (Ed), *Educational measurement (۳rd ed.)* New York: American Council on Education/ Macmillan.

Goddard, H. H. (۱۹۱۰). A measuring scale for intelligence. *The Training School, ۶*, ۱۴۶-۱۵۵

Goddard, H. H. (۱۹۱۱). Two thousand normal children measured by the Binet measuring scale of intelligence. *Pedagogical Seminary, ۱۸*, ۲۳۲-۲۵۹.

Gorsuch, R. L.(۱۹۹۷). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment, ۶۸*, ۵۳۲-۵۶۰.

Gronlund, N. E. & Linn, R. L.(۱۹۹۰). *Measurement and evaluation in teaching* (۶th ed.). New York: Macmillan.

Guilford, J. P., & Fruchter, B. (۱۹۷۸). Fundamental statistics in psychology and education.New York: McGraw- Hill.

Guttman, L.(۱۹۴۵). A basis for analyzing test-retest reliability. *Psychometrika, ۱۰*, ۲۵۵-۲۸۲.

Henson, R. K.& Roberts, J.K.(۲۰۰۶). Use of exploratory factor analysis in published research: Common errors and some comment of improved practice. *Educational and Psychological Measurement, ۶۶*, ۳۹۳-۴۱۶.

Hollinshaead, A. B.(۱۹۷۵). *Four factor index of social status.* Unpublished paper: New Haven, CT: Yale University, Department of Sociology.

Horn, J. L. (۱۹۶۵) A rational and test for the number of factors in factor analysis. *Psychometrika. ۳۰*, ۹۷۹-۹۸۵.

Joreskog, K., & Sorbom, D.(۲۰۰۲). *PRELIS ۲: User's Reference Guide.* Chicago, IL: Scientific Software International .

Howell, D.C.(۱۹۹۷).Statistical methods for psychology.(۴th ed.) Belmont, CA: Duxbury.

Hu, L., & Bentler, P. (۱۹۹۹). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, ۶*, ۱-۵۵ .

Joreskog, K.G,. & Sorbom, B.(۲۰۰۱). *LISREL ۸: User's Reference Guide.* Chicago, IL: Scientific Software International.

Joreskog, K. G., & Sorbom, D.(۱۹۸۹). *LISREL ۷: A guide to the program and applications (۲nd ed.).* Joreskog and Sorbom/SPSS Inc., Chicago IL.

Kiers, H. A. L. & ten Berg, J. M. F. (۱۹۸۹). Alternative least squares algorithms for simultaneous component analysis with equal weight matrices in two or more populations. *Psychometrika, ۵۴*, ۴۶۷-۴۷۳.

Kiers, H. A. L.(۱۹۹۰). *SCA: een programma voor simultane component analysis.* Groningen: IEC, ProGamma.

Kline, P. (۲۰۰۰). *The handbook of psychological testing.(۲nd ed).*London: Routledge.

Kline, R.B. (۲۰۰۵). *Principles and practice of structural equation modeling ۲nd ed*). New York: Guilford Press.

Laros, J.A; & Tellegen, P.J. (n.d.).Cross-cultural research with Snijders-Oomen Nonverbal Intelligence Tests. [Available at http://www.testreseart. Nl /sonroe / crssculte. html]. Accessed ۱۵ Agust ۲۰۰۵.

Millsap, R. E. & Meredith, W. M. (۱۹۸۸). Component analysis in cross-sectional and longitudinal data. *Psychometrika, ۵۳,* ۱۲۳-۱۳۴.

Newcomer, P.L; & Hammill, D.D. (۱۹۹۷). *Test of language development- primary. (۳rd ed.)* Austin, TX: PROED.

Nunnally, J.C. (۱۹۷۸). Psychometric theory. (۲nd ed.) New York: McGraw–Hill.

Nunnally, J.C;& Bernstein, I. H. (۱۹۹۴). *Psychometric theory. (۳rd ed.)* New York: McGraw-Hill.

O'Connor, B. P. (۲۰۰۰). SPSS and SAS programs for determining the number of component using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrument, and Computers, ۳۲,* ۳۹۶-۴۰۲.

Osburn, H. G. (۲۰۰۰). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Bulletin, ۵,* ۳۴۳-۳۵۵.

Raven, J.C. (۱۹۳۸) *Progressive matrices: A perceptual test of intelligence.* London: H. K. Lewis.

Snijders- Oomen,N .(۱۹۴۳). *Intellinentieonderzoekvan doofstomme kinderen.* Nigmegen: Berkhout.

Russell. D. W. (۲۰۰۲) In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and social Psychology Bulletin, ۲۸,* ۱۶۲۹-۱۶۵۶.

Stevens, J. (۲۰۰۴). Applied multivariate statistic for the social sciences (۴th ed). Mahwah, NJ: Erlbaum.

Sun, J. (۲۰۰۵). Assessing goodness of fit in confirmatory factor analysis. *Measurement and Evaluation in Counseling and Development, ۳۷,* ۲۴۰-۲۵۶.

Thompson, B. (۲۰۰۴). *Exploratory and confirmatory factor analysis: Understanding concepts and application.*Washington, DC: American Psychological Association.

Tabachnick, B. G. & Fidell, L. S.(۲۰۰۱). *Using multivariate statistics (۵th ed.).*Needham Heights, MA: Allyn & Bacon.

Tellegen, P. J., Winkel, M; Wijnbeg-Williams, B; & Laros, J. A (۱۹۹۸). *Snijders-Oomen nonverbal intelligence test. SON-R ۲½ -۷ Manual and research report*. Lisse: Swets & Zeitlinger.

Tellegen, P.J; & Laros, J. A. (۱۹۹۳). The Snijders-Oomen nonverbal intelligence tests: general intelligence test or tests for learning potential? In J.H. M. Hamers; K. Sijtsma; & J.J. M. Ruijsenaars (Eds). *Learning potential assessment: theoretical,methodolgical and practical Issues.* Amsterdam: Swets & Zeitliner.

Tellegen, P.J; & laros, J. A. (۲۰۰۴). Cultural Bias in the SON-R Test: Comparative study of brazilian and dutch children. *psicologia: Teoria e pesquisa,* ۲۰, ۱۰۳-۱۱۱.

Ten Berge, J. M. F. & Zegers, F. E.(۱۹۷۸). A series of lower bound to the reliability of a test. *Psychoetrika,* ۴, ۵۷۵-۵۷۹.

Terman, L. M. & Merrill, M. M. (۱۹۱۶). *Measuring intelligence.* Boston: Houghton Mifflin.

Urbina, S.(۲۰۰۴). *Essentials of psychological testing.* Hoboken, New Jersey: John Wiley & Sons.

Velicer, W. F.(۱۹۷۶). Determining the number of components from the matrix of partial correlations. *Psychometrika, ۴, ۳۲۱-۳۲۷.*

Velicer, W. F., Eaton, C. A., & Fava, J. L. (۲۰۰۰). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In Goffin, R. D., & Helmes, E. (Eds.), p*roblems and solutions in human assessment: Honoring Douglas Jackson at Seventy*. Boston: Kluwer. (pp. ۴۱-۷۱).

Wechsler, D. (۱۹۴۹). *Wechsler intelligence scale for children.* San Antonio: Psychological Cooperation.

Wechsler, D. (۱۹۵۵). *Wechsler adult intelligence scale.* San Antonio: Psychological Cooperation.

Wechsler, D. (۱۹۶۷). *Wechsler preschool and primary scale of intelligence.* San Antonio: Psychological Cooperation.

Weiss, D. J. (۱۹۸۲). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement, ۶ (۴),* ۴۷۳-۴۹۲.

Zimowski, M. F., Muraki, E., Mislevey, R. J. & Bock, R. D.(۲۰۰۳). *BILOG-MG ۳ [computer program]*. Chicago, IL: Scientific Software.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی