

پژوهش در حیطه کودکان استثنایی ۱۷/ سال پنجم، شماره ۳/ پاییز ۱۳۸۴/ ۳۲۲-۲۹۵
Research on Exceptional Children, Fall 2005, Vol.17, No.3, 295-322

انطباق و استانداردسازی آزمون غیرکلامی هوش اسنایدرز- اومان برای کودکان ۷-۲/۵ سال: یک مطالعه مقدماتی

اصغر مینایی*

(دریافت: ۸۴/۷/۱۳ تجدید نظر: ۸۴/۸/۲۸ پذیرش نهایی: ۸۴/۹/۹)

چکیده

پژوهش حاضر با هدف بررسی ویژگیهای روان‌سنجی و عملی بودن فرم کوتاه آزمون غیرکلامی هوش اسنایدرز-اومان برای کودکان ۷-۲/۵ سال (SON-R 2 1/2-7) روی یک گروه از کودکان تهران با حجم ۱۲۵ نفر (۶۶ پسر و ۵۹ دختر) که با روش نمونه‌برداری تصادفی چند مرحله‌ای از بین کودکان مهدهای کودک تحت نظارت بهزیستی انتخاب شدند، اجرا گردید. برای تحلیل سؤالات از مدل کلاسیک آزمون و مدل ۲ و ۳ پارامتری مبتنی بر نظریه سوال- پاسخ (IRT) استفاده گردید. اعتبار خرده آزمونها (موزاییکها، طبقه‌بندیها، موقعیتها و الگوها) با استفاده از فرمول λ_2 و اعتبار مقیاسهای عملی و استدلال و کل آزمون از طریق فرمول آلفای طبقه‌ای برآورد گردید. ضریب اعتبار خرده آزمونها برای کل گروه به ترتیب ۰/۷۲، ۰/۷۱، ۰/۶۷ و ۰/۷۵ و ضریب اعتبار مقیاس عملی، استدلال و کل آزمون نیز به ترتیب ۰/۸۱، ۰/۷۶، ۰ و ۰/۸۴ است. در خصوص روایی نیز تحلیلهای مختلفی صورت گرفت. از جمله این تحلیلهای می‌توان به همبستگی درونی خرده آزمونها، همبستگی سوال - نمره کل، همبستگی خرده آزمونها با مقیاسها و کل آزمون، تمایزگذاری سنی و قدرت تمایزگذاری اشاره کرد. بطور کلی یافته‌های حاصل از این پژوهش نشان داد که آزمون مورد بحث از اعتبار و روایی بالایی برخوردار است و با اطمینان می‌توان از آن، جهت سنجش هوش کلی کودکان ۷-۲/۵ سال استفاده کرد. با این حال، قبل از استانداردسازی و هنجاریابی آن در مقیاس وسیع باید برخی از تصاویر سؤالات A، ۱۰، ۹، ۱۲، ۱۳، ۱۴، ۸، و ۱۵ خرده آزمون طبقه‌بندیها با تصاویری که برای کودکان ایرانی آشنا هستند، جایگزین گردند. علاوه بر این ترتیب سؤالات نیز باید بر اساس دشواری آنها اصلاح گردد.

واژه‌های کلیدی: آزمون غیرکلامی، هوش، اعتبار، روایی

(Email: as_minai@yahoo.com)

* عضو هیات علمی پژوهشکده کودکان استثنایی
در اینجا لازم می‌دانم از آقای دکتر تلخ بخاطر ارسال یک نسخه از آزمون و همچنین نوار ویدئو صمیمانه تشکر نمایم.

مقدمه

آزمونهای سنتی هوش کلی^۱ (GI)، مانند آزمونهای هوش وکسلر و استنفورد - بینه توسط طرفداران آزمونهای ظرفیت یادگیری^۲ (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 مورد نیاز 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 در خصوص ظرفیت یادگیری برای چندین دهه است که ادامه دارد، لکن ابزارهای عملی برای سنجش، فقط در چند سال اخیر است که ارائه شده‌اند.

آزمونهای GI همچنین از سوی طرفداران آزمونهای هوش وابسته به فرهنگ به خاطر محتوایشان مورد انتقاد قرار گرفته‌اند. به خاطر اینکه این آزمونها از لحاظ محتوا و دستورالعمل‌ها اغلب به مهارت‌های زبانی خاصی وابسته هستند. لذا، باعث می‌گردند تا افراد اقلیت‌های 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^۳، ۱۹۹۳). می‌توان گفت که این آزمونها بیشتر بر هوش متبلور تاکید دارند تا هوش سیال (کتل^۴، ۱۹۷۱). این انتقادها نسبت به آزمونهای GI باعث گردید تا آزمونهای غیرکلامی هوش مانند ماتریسهای پیشرونده ریون (ریون^۵،

انطباق و استانداردسازی آزمون غیرکلامی.../۲۹۷

(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^۶ (۱۹۴۳) که در یک موسسه مربوط به کودکان ناشنوا در هلند کار می کرد اولین آزمون غیرکلامی هوش با نام SON^۷ را پرورش داد. او هوش را توانایی یادگیری؛ یعنی میزان توانایی کودکان در بهره گیری 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 می داد. 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 می شدند، تشکیل می یافتند. اولین آزمون SON از خرده آزمونهای غیرکلامی تشکیل می یافت که به استدلال انتزاعی و عینی مربوط می شدند و دارای نمرههایی برای کودکان ناشنوی ۱۴-۴ سال بود. در حال حاضر، نسل چهارم آزمونهای SON در دو نسخه وجود دارند. یک نسخه برای کودکان سنین پایین با عنوان SON-R 2 1/2-7 و نسخه دیگر برای کودکان سنین بالا با عنوان SON-R 5 1/2-17. مقاله حاضر مطالعه ای را توصیف می کند که با آزمون SON-R 2 1/2-7 صورت گرفته است.

آزمون SON-R 2 1/2-7 سه تفاوت اساسی با آزمونهای سنتی هوش دارد. اولاً، این آزمون مستلزم تواناییهای زبانی خاصی نیست؛ ثانیاً، اجرای آن به شیوه انطباقی، صورت می گیرد. هدف شیوههای انطباقی محدود نمودن تعداد سوالات ارائه شده همراه با کاهش نسبتاً کم اعتبار^۸ است (ویس^۹، ۱۹۸۲). در SON-R 2 1/2-7 نقطه شروع یا سطح ورودی هر خرده آزمون بر اساس سن کودک و سالهای حضور در مهد کودک، تعیین می شود 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 می گردد. ثالثاً، در مورد درست یا نادرست بودن پاسخ آزمودنی به او بازخورد داده می شود. مزیت اصلی ارائه بازخورد این است که به آزمودنی فرصت داده می شود تا در طول اجرای آزمون، دست به یادگیری بزند. با توجه به این ویژگیها، می توان گفت که آزمون SON-R 2 1/2-7 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 SON-R 2 1/2-7 و انتقادات وارده بر آزمونهای سنتی هوش کلی ذکر گردید، آزمون SON-R 2 1/2-7 جهت رواسازی و استانداردسازی برای کودکان ایرانی انتخاب گردید. با این حال، پیش از استانداردسازی و هنجاریابی، لازم است بررسی شود که آیا مواد مورد استفاده در این آزمون برای کودکان ایرانی آشنا است؟ آیا آزمون از اعتبار و روایی^{۱۰} مطلوب برخوردار است؟ و آیا کاربرد این آزمون در ایران، عملی است؟ به منظور بدست آوردن چنین شواهدی است که مطالعه حاضر صورت گرفت. بنابراین، هدف مطالعه حاضر بررسی این موضوع است که به منظور سنجش عادلانه و منصفانه سازه هوش در کودکان ایرانی با این آزمون، آیا و تا چه اندازه لازم است که انطباقهایی در مواد آزمون صورت گیرد. این هدف با رهنمودهای کمیسیون بین‌المللی آزمون^{۱۱} در خصوص استفاده از آزمون، هماهنگ و مطابق است (ون دی ویجور و همبلتون^{۱۲}، ۱۹۹۶). یکی از این رهنمودها مطرح می‌کند که پرورش‌دهندگان و ناشران آزمون باید شواهدی در خصوص اینکه سوالات و مواد آزمون برای جامعه مورد نظر، مناسب و آشنا است فراهم آورند.

روش

نمونه و روش نمونه‌گیری

نمونه مورد مطالعه در این تحقیق، ۱۲۵ کودک تهرانی (۶۶ پسر و ۵۹ دختر) با دامنه سنی ۲;۹ تا ۷;۹ (۲ سال و ۹ ماه تا ۷ سال و ۹ ماه) با میانگین ۴/۴۱ و انحراف استاندارد ۱/۲۰ است که در تابستان سال ۱۳۸۴ در مهدهای کودک تحت نظارت بهزیستی مشغول به آموزش بودند. جهت انتخاب نمونه از روش نمونه‌گیری تصادفی چند مرحله‌ای^{۱۳} استفاده گردید. 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.

ابزار

ابزار مورد استفاده 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^{۱۴} با ۱۵ سؤال؛ ۲- طبقه‌بندیها^{۱۵} با ۱۵ سؤال؛ ۳- پازلها^{۱۶} با ۱۴ سؤال؛ ۴- قیاسها^{۱۷} با ۱۷ سؤال؛ ۵- موقعیتها^{۱۸} با ۱۴ سؤال؛ ۶- الگوها^{۱۹} با ۱۶ سؤال. استانداردسازی این آزمون در هلند بر اساس یک نمونه ملی از کودکان هلند با حجم ۱۱۲۴ نفر صورت گرفته است که دامنه سنی آنها از ۲;۳ تا ۷;۳ است.

این آزمون در مجموع دارای ۹۱ سؤال است و در مدت ۶۰-۵۰ دقیقه در یک یا دو جلسه اجرا می‌گردد. خرده آزمونهای 7-2 1/2 SON-R به دو گروه کلی طبقه‌بندی می‌گردند. گروه اول شامل خرده آزمونهای الگوها، موزاییکها و پازلها است که مقیاس عملی^{۲۰} (PS) نامیده می‌شود و گروه دیگر شامل خرده آزمونهای موقعیتها، طبقه‌بندیها و قیاسها است که مقیاس استدلال^{۲۱} (RS) نام دارد. بطورکلی، در این آزمون برای هر کودک، علاوه بر نمره خرده آزمونها، ۳ نمره دیگر با نامهای SON-IQ (هوشبهر)، SON-PS (بهرعملی) و SON-RS (بهر استدلال) بدست می‌آید.

هر دو فرم این آزمون توسط کمیته آزمون موسسه روان شناسان هلند^{۲۲} (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).

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

جدول ۱، (اقتباس از تلخن و همکاران ، ۱۹۹۸) ضرایب اعتبار خرده آزمونها را که مبتنی بر همسانی درونی سوالات هر خرده آزمون است و با استفاده از فرمول λ_2 گاتمن^{۲۳} (۱۹۴۵) و براساس نمونه استانداردسازی هلند برآورد شده است به تفکیک گروههای سنی نشان می‌دهد.

جدول ۱- ضرایب اعتبار خرده آزمونها، مقیاسها و کل آزمون به تفکیک گروههای سنی

سن	خرده آزمونها							مقیاسها	
	موزاییکها	طبقه‌بندیها	پازلها	قیاسها	موقعیتها	الگوها	عملی	استدلال	کل آزمون
۲;۶	۰/۴۱	۰/۸۱	۰/۴۵	۰/۷۵	۰/۷۹	۰/۷۹	۰/۶۸	۰/۸۹	۰/۸۶
۳;۶	۰/۷۶	۰/۷۳	۰/۷۵	۰/۷۳	۰/۶۶	۰/۷۳	۰/۸۶	۰/۸۴	۰/۹۰
۴;۶	۰/۷۷	۰/۷۰	۰/۷۵	۰/۷۴	۰/۶۲	۰/۷۲	۰/۸۸	۰/۸۱	۰/۹۰
۵;۶	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
۷;۶	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^{۲۴} برآورد شده است. این فرمول جهت برآورد ضرایب اعتبار ترکیبات خطی مورد استفاده قرار می‌گیرد. (نانالی^{۲۵}، ۱۹۷۸؛ نانالی و برنشتاین^{۲۶}، ۱۹۹۴؛ اُسبرن^{۲۷}، ۲۰۰۰).

تلخن و همکارانش (۱۹۹۸) مطالعات متعددی در زمینه روایی آزمون SON-R 2 1/2-7 انجام داده‌اند. نتایج این مطالعات حاکی از روایی خوب نمرات آزمون است. از جمله این مطالعات می‌توان به همبستگی بین این آزمون با سایر آزمونهای شناختی اشاره کرد. جدول ۲، خلاصه‌ای از ضرایب همبستگی بین SON-R 2 1/2-7 با سایر آزمونهای شناختی را نشان می‌دهد.

خرده آزمونها SON-R 2 1/2-7 را می‌توان بر اساس نوع موادی که مورد استفاده قرار می‌گیرد به دو دسته تقسیم کرد: الف - خرده آزمونهایی که از مواد تصویری معنادار^{۲۸} استفاده می‌کنند (طبقه‌بندیها، موقعیتها و پازلها)؛ ب- خرده آزمونهایی که از مواد تصویری بی معنا^{۲۹} مانند اشکال هندسی استفاده می‌نمایند (موزاییکها، الگوها و قیاسها).

جدول ۲- ضرایب همبستگی آزمون 7- 2 1/2 SON-R با سایر آزمونهای شناختی

r	N	آزمون ملاک	r	N	آزمون ملاک
۰/۶۰	۱۵۳	TONI-2 ^{۳۳} فرم الف	۰/۵۹	۵۰	BOS ^{۳۰}
					GOS ^{۳۱}
۰/۶۰	۵۳	WPSI IQ کلی	۰/۶۵	۱۱۵	DI شناختی
۰/۴۹		مقیاس کلامی	۰/۶۴		DI همزمان
۰/۵۹		مقیاس عملی	۰/۴۹		DI زنجیره‌ای
			۰/۶۰	۱۶۵	RAKIT ^{۳۲}

مطلب آخر در مورد ابزار پژوهش اینکه بنا به توصیه دکتر تلخن و جهت جلوگیری از اتلاف وقت و هزینه، تصمیم بر این شد که ابتدا در یک مطالعه مقدماتی، ویژگیهای روان‌سنجی و عملی بودن فرم کوتاه 7- 2 1/2 SON-R که از چهار خرده آزمون موزاییکها، طبقه‌بندیها، موقعیتهای و الگوها ترکیب یافته و جمعاً دارای ۶۰ سوال است مورد بررسی قرار گیرد و چنانچه نتایج بدست آمده، رضایتبخش بود آنگاه فرم بلند آزمون در مقیاس وسیع استانداردسازی و هنجاریابی گردد. از این رو در این مطالعه، فرم کوتاه آزمون مورد استفاده قرار گرفت و گزارش حاضر به ارائه نتایج حاصل از فرم کوتاه آزمون می‌پردازد.

شیوه اجرا

گام نخست در این مطالعه، ترجمه دستورالعمل اجرای آزمون به زبان فارسی بود. به منظور تسلط بر نحوه اجراء، دستورالعمل مورد نظر چند بار مورد مطالعه قرار گرفت. علاوه براین یک نوار ویدئو که اجرای آزمون را نشان می‌داد به دقت مورد مشاهده قرار گرفت.

پس از انتخاب مهدهای کودک، مجوز لازم برای حضور در مهدها از بهزیستی اخذ گردید. کلیه کودکان گروه نمونه توسط پژوهشگر مورد آزمون قرار گرفتند. جمع‌آوری داده‌ها تقریباً دو ماه بطول انجامید.

همانطور که گفته شد اجرای این آزمون به شیوه انطباقی صورت می‌گیرد. در این مطالعه نیز از شیوه انطباقی استفاده گردید. برای اجرای آزمون بر روی هر کودک، بطور متوسط ۳۵-۴۰ دقیقه وقت صرف گردید. بعد از هر سوال به کودک گفته می‌شد که آیا پاسخ او درست است یا نادرست. ارائه بازخورد به کودک یک بخش مهم از شیوه اجرای

استاندارد آزمون SON-R 2 1/2-7 است، زیرا باعث می‌گردد تا دستورالعمل‌ها روشن و واضح گردند و به کودک فرصت می‌دهد تا از خطاها و موفقیت‌های خود درس بگیرد و استراتژی حل مسأله خود را تعدیل نماید. بعد از هر پاسخ نادرست، بلافاصله از کودک پرسیده می‌شد که آیا تصاویر بکار رفته در سؤال را می‌شناسد و می‌تواند نام آنها را بگوید. اجرای هر خرده آزمون بعد از ۳ پاسخ نادرست متناوب و یا ۲ پاسخ نادرست متوالی متوقف می‌گردید. از مریبان مهد کودک خواسته می‌شد تا کودکان را براساس میزان انگیزش، همکاری، توجه و درک آموزشها در طول ساعات کلاس در یک مقیاس ۴ امتیازی از ۱=خوب؛ ۲=متغیر؛ ۳=متوسط؛ ۴=ضعیف؛ ارزشیابی نمایند.

تحلیل داده‌ها

برای تحلیل و برآورد ویژگی‌های روان‌سنجی سؤالات (درجه دشواری و قدرت تشخیص) از مدل کلاسیک آزمون و مدل ۲ و ۳ پارامتری مبتنی بر نظریه سؤال-پاسخ^{۳۴} (IRT) استفاده گردید. جهت برآورد ویژگی‌های روان‌سنجی سؤالات در مدل کلاسیک و انجام تحلیل‌های آماری از نرم افزار SPSS و برای برآورد پارامترهای سؤال در مدل‌های مبتنی بر IRT از برنامه^{۳۵} ICL (هانسون^{۳۶}، ۲۰۰۲) 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^{۳۷} λ_2 (لاندای ۲) گاتمن (فلدت و برنان^{۳۷}، ۱۹۸۸) استفاده گردید. زیرا این فرمول، اعتبار را بویژه در آزمون‌های کوتاه به اندازه فرمول آلفای کرونباخ، کم برآورد نمی‌کند (تنبرگ و زگرس^{۳۸}، ۱۹۷۸). جهت برآورد اعتبار نمرات مقیاسها و کل آزمون از فرمول آلفای طبقه‌ای (نانالی و برنشتاین، ۱۹۹۴؛ آسبرن، ۲۰۰۰) استفاده گردید.

برای بررسی تفاوت بین میانگین نمرات دختران و پسران در خرده آزمون‌ها از تحلیل واریانس چند متغیری (MANOVA) دو گروهی^{۳۹} و از آزمون T^2 هتلینگ^{۴۰} (استیونس^{۴۱}، ۲۰۰۲) استفاده شد و برای بررسی اختلاف بین میانگین نمرات این دو گروه در مقیاسها (عملی و استدلال) و کل آزمون از آزمون t برای دو گروه مستقل (هاول^{۴۲}، ۱۹۹۷) استفاده گردید. در انجام این تحلیل‌ها از نمرات خام استفاده شد. زیرا در نمرات خام تمام واریانس یا پراکندگی موجود در نمرات به حساب می‌آید (آخنباخ و

رسکورلا^{۴۳}، (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

میانگین و انحراف استاندارد نمرات خام و نمرات استاندارد و همچنین چولگی و کشیدگی خرده آزمونها، مقیاسها و کل آزمون به تفکیک جنس (جداول ۳ و ۴) و کل گروه (جدول ۵) ارائه شده است.

جدول ۳- مشخصه‌های آماری خرده آزمونها، مقیاسها و کل آزمون در گروه پسران (N= ۶۶)

مقیاسها	خرده آزمونها و مقیاسها	نمرات خام		نمرات استاندارد	
		میانگین	انحراف استاندارد	میانگین	انحراف استاندارد
موزاییکها	۶/۹۲	۲/۹۰	۹/۸۳	۲/۹۸	-۰/۳۷
طبقه‌بندیها	۷/۵۸	۲/۷۷	۹/۸۲	۳/۳۵	-۰/۳۶
موقعیتها	۶/۷۳	۲/۳۲	۹/۶۴	۲/۹۴	-۰/۱۴
الگوها	۹/۲۶	۳/۱۵	۱۰/۰۷	۳/۲۳	۰/۰۴
مقیاس عملی	۱۶/۱۸	۵/۷۸	۹۹/۲۰	۱۵/۴۸	-۰/۰۴
مقیاس استدلال	۱۴/۳۱	۴/۶۷	۹۹/۰۵	۱۵/۶۶	-۰/۴۱
کل آزمون	۳۰/۴۹	۹/۸۵	۹۹/۰۸	۱۵/۶۴	-۰/۲۴

جدول ۴- مشخصه‌های آماری خرده آزمونها، مقیاسها و کل آزمون در گروه دختران (N= ۵۹)

مقیاسها	خرده آزمونها و مقیاسها	نمرات خام		نمرات استاندارد	
		میانگین	انحراف استاندارد	میانگین	انحراف استاندارد
موزاییکها	۷/۲۷	۲/۹۶	۱۰/۱۹	۳/۰۴	-۰/۵۵
طبقه‌بندیها	۷/۹۰	۲/۳۰	۱۰/۲۰	۲/۷۰	-۰/۱۰
موقعیتها	۷/۳۲	۲/۴۱	۱۰/۳۹	۳/۰۵	-۰/۱۳
الگوها	۹/۱۲	۲/۶۹	۹/۹۳	۲/۷۶	۰/۱۱
مقیاس عملی	۱۶/۳۹	۴/۲۳	۱۰۰/۸۹	۱۴/۵۵	-۰/۱۱
مقیاس استدلال	۱۵/۲۲	۵/۳۳	۱۰۱/۰۴	۱۴/۲۶	-۰/۴۵
کل آزمون	۳۱/۶۱	۹/۰۰	۱۰۱/۰۱	۱۴/۳۰	-۰/۳۹

جدول ۵- مشخصه‌های آماری خرده آزمونها، مقیاسها و کل آزمون در کل گروه (N= ۱۲۵)

خرده آزمونها و مقیاسها	نمرات خام		نمرات استاندارد		کشیدگی
	میانگین	انحراف استاندارد	میانگین	انحراف استاندارد	
موزاییکها	۷/۰۹	۲/۹۲	۱۰	۳	۰/۰۷
طبقه بندیها	۷/۷۳	۲/۵۶	۱۰	۳	۰/۲۷
موقعیتها	۷/۰۱	۲/۳۷	۱۰	۳	۰/۹۰
الگوها	۹/۱۹	۲/۹۳	۱۰	۳	۰/۲۸
مقیاس عملی	۱۶/۲۸	۵/۵۵	۱۰۰	۱۵	۰/۶۱
مقیاس استدلال	۱۴/۷۴	۴/۴۸	۱۰۰	۱۵	۰/۱۹
کل آزمون	۳۱/۰۲	۹/۴۴	۱۰۰	۱۵	۰/۳۳

از اطلاعات ارائه شده در جداول ۳ و ۴ می‌توان چنین استنباط کرد که بین مشخصه‌های آماری خرده آزمونها، مقیاسها و کل آزمون در دو گروه دختران و پسران، تفاوت قابل ملاحظه‌ای وجود ندارد. علاوه بر این، نتایج MANOVA با استفاده از آزمون T^2 هتلینگ نشان داد که بین میانگین نمرات دختران و پسران در خرده آزمونها، تفاوت آماری معنادار وجود ندارد [$T^2(۴,۱۲۰) = ۰/۰۴, p = ۰/۳۲$].

نتیجه آزمون t نشان داد که بین میانگین نمرات دختران و پسران در مقیاس عملی، مقیاس استدلال و کل آزمون تفاوت آماری معنادار وجود ندارد. مقادیر مشاهده شده t برای مقیاس عملی، مقیاس استدلال و کل آزمون به ترتیب: [$t(۱۲۳) = ۰/۲۱, p = ۰/۸۴$]؛ [$t(۱۲۳) = ۰/۶۶, p = ۰/۵۱$]؛ و [$t(۱۲۳) = ۱/۱۴, p = ۰/۲۶$] است.

شاخص‌های آماری سوالات بر اساس مدل کلاسیک اعتبار و مدل ۲ و ۳ پارامتری مبتنی بر IRT به ترتیب در جدول ۶ و ۷ ارائه شده است. درجه دشواری سؤال در مدل کلاسیک، که با نماد P نشان داده می‌شود، بیانگر نسبت افرادی است که به سؤال پاسخ درست داده‌اند. قدرت تشخیص سوال در این مدل، که گاهی اوقات روایی سؤال^{۴۴} نیز نامیده می‌شود، همواره به صورت همبستگی دو رشته‌ای یا دو رشته‌ای نقطه‌ای سوال با نمره کل آزمون یا خرده آزمون بیان می‌شود (ثرن‌دایک، ترجمه هومن، ۱۳۷۵). این شاخص معلوم می‌نماید که سوال تا چه اندازه می‌تواند گروه‌های مختلف افراد را از هم جدا سازد. در واقع این شاخص نشان‌دهنده همسویی و هماهنگی سؤال با آزمون یا خرده آزمون است. از آنجا که در برنامه SPSS، همبستگی دو رشته‌ای نقطه‌ای وجود

ندارد، جهت برآورد همبستگی سوال با نمره کل خرده آزمون از همبستگی گشتاوری پیرسون استفاده گردید. همبستگی دو رشته‌ای نقطه‌ای به لحاظ عددی، دقیقاً معادل با همبستگی گشتاوری پیرسون است (کلاین^{۴۵}، ۲۰۰۰). از آنجا که نمره سؤال در نمره کل خرده آزمون لحاظ می‌شود و این امر باعث افزایش غیرواقعی همبستگی سوال با نمره کل می‌گردد، از این رو جهت حذف تاثیر نمره سوال، بعنوان بخشی از نمره کل خرده آزمون، یک اصلاح صورت گرفت (کلاین، ۲۰۰۰؛ نانالی و برنشتاین، ۱۹۹۴).

جدول ۶- درجه دشواری و قدرت تشخیص سؤالات بر اساس مدل کلاسیک اعتبار

خرده آزمونها								سوال
الگوها		موقعیتها		طبقه‌بندیها		موزاییکها		
قدرت	درجه دشواری تشخیص	قدرت تشخیص	درجه دشواری	قدرت تشخیص	درجه دشواری	قدرت	درجه دشواری تشخیص	
*.۰/۱۳	.۰/۹۹	*.۰/۱۵	.۰/۹۹	.۰/۳۷	.۰/۹۷	.۰/۳۰	.۰/۹۸	۱
.۰/۲۵	.۰/۹۷	*.۰/۰۰	۱/۰۰	.۰/۲۹	.۰/۹۸	.۰/۵۲	.۰/۸۶	۲
.۰/۲۱	.۰/۹۸	.۰/۲۸	.۰/۹۴	.۰/۱۵	.۰/۹۸	.۰/۶۴	.۰/۸۲	۳
.۰/۲۷	.۰/۹۸	.۰/۲۵	.۰/۹۶	.۰/۵۳	.۰/۸۶	.۰/۶۰	.۰/۸۹	۴
.۰/۳۵	.۰/۹۴	.۰/۴۷	.۰/۵۷	.۰/۴۵	.۰/۷۶	.۰/۵۹	.۰/۷۵	۵
.۰/۴۳	.۰/۸۸	.۰/۵۵	.۰/۶۶	.۰/۴۴	.۰/۸۲	.۰/۶۸	.۰/۸۱	۶
.۰/۵۸	.۰/۷۸	.۰/۵۵	.۰/۴۶	.۰/۴۴	.۰/۷۲	.۰/۶۸	.۰/۷۱	۷
.۰/۶۴	.۰/۷۶	.۰/۵۹	.۰/۵۸	.۰/۶۱	.۰/۵۸	.۰/۶۶	.۰/۵۸	۸
.۰/۶۷	.۰/۶۹	.۰/۵۱	.۰/۳۸	.۰/۴۷	.۰/۴۰	.۰/۵۳	.۰/۲۷	۹
.۰/۶۳	.۰/۴۱	.۰/۴۴	.۰/۲۳	.۰/۲۳	.۰/۲۴	.۰/۴۲	.۰/۱۵	۱۰
.۰/۶۴	.۰/۲۶	.۰/۴۰	.۰/۱۴	.۰/۴۷	.۰/۱۸	.۰/۴۵	.۰/۱۲	۱۱
.۰/۶۸	.۰/۲۳	.۰/۲۶	.۰/۰۶	.۰/۳۱	.۰/۰۸	.۰/۳۸	.۰/۰۶	۱۲
.۰/۶۰	.۰/۱۶	*.۰/۱۹	.۰/۰۰۲	.۰/۳۹	.۰/۰۸	*.۰/۱۵	.۰/۰۱	۱۳
.۰/۴۹	.۰/۱۰	*.۰/۱۵	.۰/۰۰۸	.۰/۴۰	.۰/۰۶	.۰/۳۲	.۰/۰۳	۱۴
.۰/۴۷	.۰/۰۸	—	—	.۰/۳۱	.۰/۰۳	.۰/۲۹	.۰/۰۲	۱۵
*.۰/۱۸	.۰/۰۰۸	—	—	—	—	—	—	۱۶
.۰/۴۵	.۰/۵۸	.۰/۳۴	.۰/۵۰	.۰/۳۹	.۰/۵۲	.۰/۴۷/۵	.۰/۴۷	میانگین

نتایج جدول فوق نشان می‌دهد که درجه دشواری و قدرت تشخیص سوالات در سطح مطلوب قرار دارند. آناستازی (۱۹۹۰) و آناستازی و یوربینا (۱۹۹۷)، اظهار می‌دارند که متوسط درجه دشواری سوالات باید تقریباً $0/50$ باشد و همچنین باید پراکندگی نسبتاً زیادی در دشواری سوالات وجود داشته باشد.

ارقام ارائه شده در جدول ۶ حاکی از این است که اکثر سوالات از قدرت تشخیص مطلوب و بالایی برخوردارند و صرفاً ۸ سوال که با علامت* مشخص شده‌اند قدرت تشخیص نسبتاً ضعیفی دارند. نانالی و برنشتاین (۱۹۹۴) اظهار می‌دارند که اکثر ضرایب همبستگی سوال - نمره کل، در دامنه ۰ تا $0/40$ قرار دارند. گارت^{۴۶} (۱۹۶۵)؛ به نقل از نیوکامر و هامیل^{۴۷}، (۱۹۹۷) اظهار کرده است که سوالات دارای ضریب همبستگی $0/20$ 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 $0/20$ یا $0/30$ می‌توانند قابل قبول تلقی گردند.

در مدل ۳ پارامتری، هر سوال براساس ۳ پارامتر a، b، و c توصیف می‌گردد. در مدل ۲ پارامتری فرض بر این است که امکان پاسخ درست از طریق حدس وجود ندارد و مقدار آن برابر با صفر در نظر گرفته می‌شود. به بیان دقیق‌تر، در این مدل پارامتر c برای سوال برآورد نمی‌شود و سؤال صرفاً بر اساس ۲ پارامتر a و b توصیف می‌شود. نماد b در IRT بیانگر پارامتر دشواری سؤال است و جایگاه سؤال را در مقیاس توانایی توصیف می‌کند (بیکر^{۴۸}، ۲۰۰۱). به زبان فنی و ریاضی می‌توان گفت که این پارامتر نقطه عطف تابع ویژگی سؤال را توصیف می‌کند و معمولاً در دامنه بین $2/5 -$ تا $2/5 +$ مقیاس پردازش 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^{۴۹} بالا می‌رود احتمال موفقیت در سؤال با چه سرعتی افزایش می‌یابد (ثرندایک، ترجمه هومن، ۱۳۷۵). به لحاظ نظری این پارامتر در مقیاس $-\infty$ تا $+\infty$ تعریف می‌شود. ولی سؤالاتی که دارای قدرت تشخیص منفی هستند از آزمونها حذف می‌گردد. همچنین بدست آوردن

مقادیر بزرگتر از ۲/۵ برای این پارامتر غیر معمول است. بنابراین، دامنه این پارامتر معمولاً بین ۰/۵ تا ۲/۵ تغییر می‌کند.

پارامتر c بیانگر احتمال بدست آوردن پاسخ درست سوال از طریق حدس محض است. به بیان دیگر، این پارامتر بیانگر احتمال پاسخ درست به سؤال توسط افرادی است که در سطح خیلی پایین از توانایی قرار دارند. هر چند دامنه نظری پارامتر حدس برابر با $0 \leq C \leq 1$ است اما در عمل مقدار بالاتر از ۰/۳۵ برای آن قابل قبول نیست (بیکر، ۲۰۰۱). جدول ۷ پارامترهای سؤالات را بر اساس مدل‌های IRT که با استفاده از برنامه ICL (هانسون، ۲۰۰۲) برآورد شده‌اند، نمایش می‌دهد.

جدول ۷- پارامترهای a، b و c سؤالات بر اساس مدل منطقی ۲ و ۳ پارامتری (N = ۱۲۵)

خرده آزمونها										سوال	
الگوها		موقعیتها				طبقه‌بندیها			موزاییکها		
b	a	c	b	a	c	b	a	b	a		
-۲/۹۵	۰/۹۳	۰/۲۳	-۲/۲۵	۱/۲۱	-۰/۲۲	-۲/۲۸	۱/۶۰	-۳/۵۰	۰/۹۳	۱	
-۲/۷۵	۱/۰۹	۰/۲۳	-۵/۸۱	۰/۹۷	-۰/۲۲	-۲/۷۷	۱/۴۶	-۱/۵۱	۱/۲۷	۲	
-۳/۰۹	۱/۱۸	۰/۲۲	-۲/۰۴	۱/۱۸	-۰/۲۲	-۳/۷۴	۰/۶۴	-۱/۱۴	۱/۸۲	۳	
-۲/۶۴	۱/۴۵	۰/۲۳	-۲/۳۹	۱/۰۴	۰/۱۹	-۱/۱۹	۱/۸۷	-۱/۵۱	۱/۹۵	۴	
-۲/۱۹	۱/۱۸	۰/۱۶	۰/۰۵	۰/۸۸	۰/۱۹	-۰/۸۱	۰/۹۶	-۰/۹۳	۱/۱۸	۵	
-۱/۷۳	۱/۱۲	۰/۱۴	-۰/۳۶	۱/۴۱	-۰/۱۸	-۱/۲۲	۰/۸۷	-۱/۰۳	۲/۰۲	۶	
-۱/۰۱	۱/۵۳	۰/۱۲	۰/۳۵	۱/۲۹	۰/۱۶	-۰/۷۱	۰/۸۳	-۰/۶۵	۱/۸۹	۷	
-۰/۸۷	۲/۳۱	۰/۱۰	-۰/۱۰	۱/۳۰	۰/۰۹	-۰/۱۵	۱/۹۸	-۰/۳۲	۱/۹۵	۸	
-۰/۶۰	۲/۲۳	۰/۰۸	۰/۵۸	۱/۰۸	۰/۰۹	۰/۵۴	۱/۰۳	۰/۷۷	۱/۶۵	۹	
۰/۲۹	۱/۴۷	۰/۰۶	۱/۲۰	۱/۰۷	-۰/۱۳	۲/۵۹	۰/۴۸	۱/۴۲	۱/۲۰	۱۰	
۰/۷۸	۱/۸۵	۰/۰۴	۱/۴۰	۱/۵۲	۰/۰۴	۱/۲۱	۱/۷۷	۱/۴۶	۱/۸۰	۱۱	
۰/۸۸	۲/۲۸	۰/۰۴	۲/۶۲	۰/۹۷	۰/۰۴	۲/۱۸	۱/۰۹	۱/۹۰	۱/۷۸	۱۲	
۱/۲۵	۲/۰۵	۰/۰۳	۴/۸۲	۰/۸۸	-۰/۰۳	۱/۷۳	۱/۷۹	۳/۴۸	۱/۱۶	۱۳	
۱/۷۰	۱/۷۰	۰/۰۳	۵/۵۸	۰/۹۱	-۰/۰۳	۱/۹۲	۱/۹۲	۲/۳۱	۱/۷۸	۱۴	
۱/۷۹	۱/۹۳	—	—	—	-۰/۰۲	۲/۲۵	۱/۸۰	۲/۵۰	۱/۷۷	۱۵	
۳/۳۲	۱/۳۷	—	—	—	—	—	—	—	—	۱۶	
-۰/۵۵	۱/۶۰	۰/۱۲	۰/۱۹	۱/۱۲	-۰/۱۲	-۰/۰۳	۱/۳۴	-۰/۲۲	۱/۶۱	میانگین	

جدول ۸ ضرایب همبستگی گشتاوری پیرسون بین شاخص‌های آماری سؤالات در مدل کلاسیک و مدل‌های مبتنی بر IRT را نشان می‌دهد. کلیه ضرایب همبستگی در سطح ۰/۰۱ به لحاظ آماری معنادار هستند.

جدول ۸- همبستگی بین شاخص‌های آماری سوالات در مدل کلاسیک و مدل‌های IRT

خرده آزمونها	همبستگی بین قدرت تشخیص در مدل کلاسیک و IRT	همبستگی بین درجه دشواری در مدل کلاسیک و IRT
موزاییکها	۰/۸۱	-۰/۹۷
طبقه‌بندیها	۰/۵۱	-۰/۹۵
موقعیتها	۰/۵۱	-۰/۹۳
الگوها	۰/۸۱	-۰/۹۷

بین نمره خام افراد و پارامتر توانایی آنها در خرده آزمونها همبستگی گشتاوری پیرسون محاسبه گردید. در مدل‌های IRT پارامتر توانایی افراد با نماد θ (تتا) نشان داده می‌شود. در برنامه ICL، پارامتر θ به دو روش بیشینه احتمال^{۵۰} (MLE) و بیز^{۵۱}، که با عنوان برآورد بعدی مورد انتظار^{۵۲} (EAP) مورد اشاره قرار می‌گیرد، برآورد می‌گردد. ضرایب همبستگی بدست آمده که همگی در سطح ۰/۰۱ معنادار هستند در جدول ۹، ارائه شده است.

جدول ۹- همبستگی بین نمره خام و برآوردهای θ در خرده آزمونها

خرده آزمونها	نمره خام و MLE	نمره خام و EAP	EAP و MLE
موزاییکها	۰/۹۵	۰/۹۹	۰/۹۴
طبقه‌بندیها	۰/۹۷	۰/۹۸	۰/۹۸
موقعیتها	۰/۹۷	۰/۹۹	۰/۹۹
الگوها	۰/۹۷	۰/۹۸	۰/۹۹

اعتبار نمرات خرده آزمونها که بر اساس فرمول λ_2 (فلدت و برنان، ۱۹۸۸؛ اسپرن، ۲۰۰۰) برآورد شده‌اند به تفکیک جنس (جدول ۱۰) و گروه سنی (جدول ۱۱) ارائه شده‌اند. در این جداول همچنین ضرایب اعتبار نمرات مقیاس عملی (PS)، مقیاس استدلال (RS) و کل آزمون (IQ) که با استفاده از فرمول آلفای طبقه‌ای (نانالی و برنشتاین، ۱۹۹۴؛ اسپرن، ۲۰۰۰) برآورد شده‌اند نشان داده شده است. از آنجا که ضرایب اعتبار λ_2 براساس نمونه‌هایی که از لحاظ سن ناهمگن هستند برآورد شدند،

بنابراین جهت حذف تاثیر سن، یک اصلاح صورت گرفت. ضرایب ارایه شده در جدول ۱۰ ضرایب اصلاح شده هستند.

جدول ۱۰- ضرایب اعتبار نمرات خرده آزمونها، مقیاسها و کل آزمون به تفکیک جنس و کل گروه

خرده آزمونها	دختران	پسران	کل گروه
موزاییکها	۰/۷۷	۰/۶۹	۰/۷۲
طبقه‌بندیها	۰/۶۰	۰/۷۶	۰/۷۱
موقعیتها	۰/۶۹	۰/۶۴	۰/۶۷
الگوها	۰/۷۶	۰/۷۶	۰/۷۵
مقیاس عملی	۰/۸۱	۰/۸۰	۰/۸۱
مقیاس استدلال	۰/۷۲	۰/۷۸	۰/۷۶
کل آزمون	۰/۸۱	۰/۸۶	۰/۸۴

همانطور که در جداول ۱۰ و ۱۱ ملاحظه می‌کنید، ضرایب اعتبار نمرات خرده آزمونها، مقیاسها و کل آزمون برای هر دو جنس، کل گروه و گروههای سنی مختلف تماماً در سطح مطلوبی قرار دارند. بالا بودن مقدار این ضرایب حاکی از این است که آزمون مورد بحث آزمون معتبری است و می‌توان نتایج آن را با اطمینان مورد استفاده قرار داد.

جدول ۱۱- ضرایب اعتبار نمرات خرده آزمونها، مقیاسها و کل آزمون به تفکیک گروه سنی

خرده آزمونها و مقیاسها	گروه سنی		
	۲-۳ (N=۳۵)	۴-۵ (N=۵۵)	۶-۷ (N=۳۵)
موزاییکها	۰/۸۵	۰/۷۹	۰/۷۵
طبقه‌بندیها	۰/۷۳	۰/۶۶	۰/۸۱
موقعیتها	۰/۸۱	۰/۶۹	۰/۷۷
الگوها	۰/۷۸	۰/۸۵	۰/۸۱
مقیاس عملی	۰/۸۹	۰/۸۹	۰/۸۷
مقیاس استدلال	۰/۸۳	۰/۷۹	۰/۸۷
کل آزمون	۰/۹۲	۰/۹۱	۰/۹۰

ضرایب تعمیم پذیری^{۵۳} نمرات کل آزمون و مقیاسهای عملی و استدلال نیز مورد بررسی قرار گرفت. این ضریب نشان می دهد که تا چه اندازه می توان نتایج خرده آزمونها را به کل حیطه مورد سنجش خرده آزمونها تعمیم داد (تلخن و همکاران، ۱۹۹۸). تعمیم پذیری با استفاده از فرمول ضریب آلفا محاسبه گردید که در آن به جای نمرات سؤال از نمرات خرده آزمونها بعنوان واحد تحلیل استفاده شد. از آلفا که شاخصی از همسانی درونی برای آزمونها یا خرده آزمونهای همگن است می توان برای برآورد اعتبار استفاده نمود. با این حال ضریب آلفا برای نمره کل خرده آزمونها که هر کدام از آنها واریانس معتبر خاص خود را دارند دارای معنای متفاوتی است. در این مورد ضریب آلفا را می توان بعنوان شاخص تعمیم پذیری تفسیر نمود. ۴ خرده آزمون SON-R 2 1/2-7 را می توان بعنوان نمونه ای از حیطه خرده آزمونهای غیر کلامی مشابه، تصور کرد. ضریب آلفا همبستگی مورد انتظار بین نمره کل آزمون و نمره کل در یک ترکیب متفاوت اما با حجم یکسان (مساوی) از خرده آزمونهای آن حیطه را نشان می دهد. ریشه دوم ضریب آلفا حاکی از همبستگی نمره کل آزمون با نمره یک آزمون فرضی با تعداد زیادی از خرده آزمونهای غیر کلامی مشابه است که اگر اجرا می گردید بدست می آمد. این مطلب در مورد مقیاس عملی و استدلال نیز کاربرد دارد. ضرایب تعمیم پذیری به تفکیک جنس و گروه سنی و کل گروه در جدول ۱۲، ارائه شده است.

جدول ۱۲- ضرایب تعمیم پذیری نمرات کل آزمون، مقیاس عملی و استدلال به تفکیک جنس، گروه سنی و کل گروه

مقیاس ها و کل آزمون	جنس		گروه سنی		
	پسران	دختران	۲-۳	۴-۵	۶-۷
مقیاس عملی	۰/۹۰	۰/۸۷	۰/۷۳	۰/۷۹	۰/۸۳
مقیاس استدلال	۰/۸۰	۰/۷۶	۰/۵۶	۰/۷۱	۰/۷۳
کل آزمون	۰/۹۰	۰/۸۹	۰/۸۱	۰/۸۴	۰/۷۵

رابطه بین نمرات خرده آزمونها با یکدیگر و نمرات هر خرده آزمون با مقیاس عملی، استدلال و کل آزمون و همچنین رابطه بین نمرات هر خرده آزمون با مجموع ۳ خرده آزمون دیگر مورد بررسی قرار گرفت. همبستگی درونی نمرات خرده آزمونها با یکدیگر

برای کل گروه، و برای دختران و پسران در جدول ۱۳، ارائه شده است. به منظور کنترل تأثیر سن از روش همبستگی تفکیکی^{۵۴} استفاده گردید (گیلفورد و فروچتر،^{۵۵} ۱۹۷۸؛ چن و پاپویچ،^{۵۶} ۲۰۰۲).

جدول ۱۳- ضرایب همبستگی درونی نمرات خرده آزمونها در کل گروه و به تفکیک پسران و دختران

خرده آزمونها	کل گروه				پسران				دختران			
	Pat	Sit	Cat	Mos	Pat	Sit	Cat	Mos	Pat	Sit	Cat	Mos
موزاییکها (Mos)	۰/۵۷	۰/۴۲	۰/۵۰	-	۰/۶۱	۰/۴۴	۰/۵۸	-	۰/۵۵	۰/۴۰	۰/۴۰	-
طبقه‌بندیها (Cat)	۰/۴۱	۰/۴۶	-	-	۰/۴۵	۰/۵۲	-	-	۰/۳۸	۰/۴۱	-	-
موقعیتها (Sit)	۰/۳۵	-	-	-	۰/۳۰	-	-	-	۰/۴۳	-	-	-
الگوها (Pat)	-	-	-	-	-	-	-	-	-	-	-	-

میانگین ضرایب همبستگی بین خرده آزمونها که با استفاده از روش تبدیل Z فیشر^{۵۷} بدست آمد. برای کل گروه و گروه پسران و دختران به ترتیب، ۰/۴۵، ۰/۴۳ و ۰/۴۳ است. کلیه ضرایب همبستگی در سطح ۰/۰۱ به لحاظ آماری معنادار هستند. تفاوت بین ضرایب همبستگی درونی خرده آزمونها در دو گروه دختران و پسران با استفاده از آزمون Z (چن و پاپویچ، ۲۰۰۲) مورد بررسی معناداری آماری قرار گرفت. نتایج نشان داد که تفاوت بین هیچ یک از ضرایب همبستگی در سطح ۰/۰۵ به لحاظ آماری معنادار نیست.

جدول ۱۴ همبستگی بین نمرات خام خرده آزمونها با نمرات خام مقیاس عملی (PS)، مقیاس استدلال (RS) و نمره کل آزمون را برای کل گروه و به تفکیک برای پسران و دختران نشان می‌دهد. همبستگی نمرات خرده آزمونها با نمره کل آزمون از طریق محاسبه همبستگی بین نمرات هر خرده آزمون با مجموع ۳ خرده آزمون دیگر صورت گرفت. به منظور کنترل تأثیر سن از روش همبستگی تفکیکی استفاده گردید (گیلفورد و فروچتر، ۱۹۷۸؛ چن و پاپویچ، ۲۰۰۲). کلیه ضرایب در سطح ۰/۰۱ به لحاظ آماری معنادار هستند. تفاوت بین ضرایب همبستگی در دو گروه دختران و پسران با استفاده از آزمون Z (چن و پاپویچ، ۲۰۰۲) مورد بررسی معناداری آماری قرار گرفت. نتایج نشان داد که در سطح ۰/۰۵ بین هیچ یک از ضرایب همبستگی تفاوت آماری معنادار وجود ندارد.

جدول ۱۴- ضرایب همبستگی بین خرده آزمونها با مقیاسها و نمره کل آزمون برای کل گروه و به تفکیک جنس

خرده آزمونها	کل گروه			پسران			دختران		
	نمره کل	RS	PS	نمره کل	RS	PS	نمره کل	RS	PS
موزاییکها	۰/۶۴	۰/۸۴	۰/۵۹	۰/۶۹	۰/۸۵	۰/۶۵	۰/۵۸	۰/۸۵	۰/۵۴
طبقه‌بندیها	۰/۵۸	۰/۸۹	۰/۵۳	۰/۶۴	۰/۹۲	۰/۶۰	۰/۵۰	۰/۸۳	۰/۴۷
موقعیتها	۰/۵۰	۰/۸۴	۰/۵۱	۰/۵۱	۰/۵۵	۰/۸۱	۰/۵۱	۰/۴۸	۰/۸۹
الگوها	۰/۵۶	۰/۷۹	۰/۵۶	۰/۵۵	۰/۵۸	۰/۵۸	۰/۵۹	۰/۵۶	۰/۷۹

از آنجا که هوش ماهیتاً وابسته به سن است لذا بایستی عملکرد در این آزمون با سن تقویمی قویاً همبستگی داشته باشد. میانگین نمرات استاندارد گروههای سنی مختلف در خرده آزمونها، مقیاسها و کل آزمون در جدول ۱۵، ارائه شده است. ستون آخر این جدول، همبستگی بین سن با نمره استاندارد را نشان می‌دهد. محتوای جدول نشان می‌دهد که عملکرد در خرده آزمونها، مقیاسها و کل آزمون با سن رابطه دارد، زیرا که با افزایش سن، میانگین نمرات خرده آزمونها، مقیاسها و کل آزمون افزایش می‌یابد. این نتیجه‌گیری با این واقعیت که تمامی ضرایب همبستگی ارائه شده در ستون سمت چپ جدول به لحاظ آماری در سطح ۰/۰۱ معنادار هستند، مورد تایید قرار می‌گیرد. این یافته‌ها بیانگر این است که آزمون مورد بحث از قدرت تمایزگذاری سنی بسیار خوبی برخوردار است.

جدول ۱۵- میانگین نمرات استاندارد خرده آزمونها، مقیاسها و نمره کل آزمون به تفکیک گروههای سنی

خرده آزمونها و مقیاسها	گروه سنی			همبستگی با سن
	۲-۳	۴-۵	۶-۷	
موزاییکها	۷/۰۰	۱۰/۵۴	۱۲/۶۱	۰/۷۳
طبقه‌بندیها	۷/۴۴	۱۰/۵۴	۱۲/۰۵	۰/۵۸
موقعیتها	۷/۸۹	۱۰/۲۵	۱۲/۱۴	۰/۷۳
الگوها	۷/۲۰	۱۰/۳۹	۱۲/۷۳	۰/۷۲
مقیاس عملی	۸۶/۴۰	۱۰۱/۷۷	۱۱۳/۵۰	۰/۷۳
مقیاس استدلال	۸۴/۹۴	۱۰۲/۹۳	۱۱۲/۶۲	۰/۷۱
کل آزمون	۸۵/۰۱	۱۰۲/۴۷	۱۱۳/۶۴	۰/۷۵

همانطور که پیش از این نیز گفته شد به هنگام اجرای آزمون از مربیان مهدهای کودک درخواست کردیم تا کودک را از لحاظ میزان انگیزش، توجه، همکاری و درک دستورات و آموزشها، در یک مقیاس ۴ امتیازی از ۱=خوب؛ ۲=متغیر؛ ۳=متوسط؛ و ۴=ضعیف، ارزیابی نمایند. علاوه بر این، پژوهشگر نیز بر اساس مشاهده خود در طول اجرای آزمون هر کودک را به همان شیوه مورد ارزیابی قرار می‌داد. بین عملکرد کودکان در خرده آزمونها، مقیاسها و کل آزمون با ارزیابی مربیان و پژوهشگر، ضریب همبستگی محاسبه گردید. از آنجا که خرده آزمونها، مقیاسها و کل آزمون از اعتبار کامل (یعنی ۱) برخوردار نیستند، از اینرو جهت کنترل تأثیر بی‌اعتباری از فرمول اصلاح برای کاهش^{۵۸} (گیلفورد و فروچتر، ۱۹۷۸) استفاده گردید. ضرایب ارائه شده در جدول ۱۶ ضرایب اصلاح شده هستند.

میانگین ضرایب همبستگی مربوط به ارزیابیهای پژوهشگر، که با استفاده از روش تبدیل Z فیشر بدست آمد، ۰/۴۰ است که یک ضریب همبستگی متوسط (نصفت، ۱۳۷۱) محسوب می‌گردد. علاوه بر این، بغیر از ضریب همبستگی خرده آزمون موقعیتها با میزان همکاری که ۰/۱۵ است، کلیه ضرایب همبستگی در سطح ۰/۰۱ به لحاظ آماری معنادار هستند.

متوسط ضرایب همبستگی مربوط به ارزیابیهای مربی، ۰/۱۴ است که یک ضریب همبستگی بسیار ضعیف (نصفت، ۱۳۷۱) به حساب می‌آید. این یافته با انتظار ما و با یافته‌های حاصل از تحقیقات تلخن و همکاران (۱۹۹۸) و تلخن و لاروس (۲۰۰۴) همخوانی ندارد. آنها در پژوهشهای خود دریافتند که بین عملکرد کودکان در خرده آزمونها با ارزیابیهای معلم، همبستگی متوسط وجود دارد. یکی از دلایل احتمالی عدم همخوانی دریافته‌ها ممکن است این باشد که در ایران تعداد کودکانی که تحت آموزش و نظارت یک مربی قرار دارند نسبتاً زیاد است و این مسأله باعث می‌گردد تا مربی فرصت و انرژی کافی جهت شناخت هر چه بهتر کودکان و ارزیابی دقیق آنها نداشته باشد.

آخرین شاهدهی که در مورد روایی مورد بررسی قرار گرفت، قدرت تمایزگذاری^{۵۹} است. این شاخص که به واریانس یا پراکندگی نمرات مربوط است زمانی به حداکثر

می‌رسد که توزیع نمرات به شکل مستطیلی باشد. حداقل تمایزگذاری (صفر) زمانی به دست خواهد آمد که تمام افراد نمره یکسانی دریافت کنند. شاخص قدرت تمایزگذاری

جدول ۱۶- ضرایب همبستگی بین عملکرد در خرده آزمونها، مقیاسها و کل آزمون با ارزیابیهای مربیان و پژوهشگر

ارزیابی پژوهشگر				ارزیابی مربی				خرده آزمونها / مقیاسها / نمره کل
درک دستورالعملها	همکاری	توجه	انگیزش	درک آموزشها	همکاری	توجه	انگیزش	
۰/۵۰	۰/۳۲	۰/۵۰	۰/۴۴	۰/۰۸	۰/۱۴	۰/۲۴	۰/۱۲	موزاییکها
۰/۴۷	۰/۲۴	۰/۴۳	۰/۳۹	۰/۰۶	۰/۲۴	۰/۲۵	۰/۰۸	طبقه‌بندیها
۰/۴۸	۰/۱۵	۰/۳۸	۰/۲۹	۰/۰۲	۰/۱۲	۰/۲۵	۰/۰۵	موقعیتها
۰/۴۸	۰/۳۵	۰/۴۱	۰/۳۷	۰/۱۴	۰/۱۵	۰/۲۱	۰/۰۵	الگوها
۰/۵۰	۰/۳۰	۰/۴۲	۰/۳۴	۰/۰۸	۰/۱۴	۰/۲۳	۰/۰۸	مقیاس عملی
۰/۵۱	۰/۳۰	۰/۵۰	۰/۴۴	۰/۰۴	۰/۱۸	۰/۲۵	۰/۰۶	مقیاس استدلال
۰/۵۲	۰/۲۶	۰/۴۶	۰/۳۹	۰/۰۳	۰/۱۷	۰/۲۴	۰/۱۰	کل آزمون

آزمون، دلتای فرگسون است که مقدار آن بین ۱ تا ۰ تغییر می‌کند (فرگسون ۱۹۴۹؛ به نقل از کلاین، ۲۰۰۰). مقدار این شاخص در توزیع‌های نرمال ۰/۹۳ است. بطور کلی آزمون روا، آزموننی است که قدرت تمایزگذاری آن بالاتر از ۰/۹۰ باشد. درجدول ۱۷، ضرایب قدرت تمایزگذاری خرده آزمونها، مقیاسها و کل آزمون به تفکیک جنس و کل گروه ارائه شده است. کلیه ضرایب در سطح بالایی قرار دارند و با ملاک ارائه شده توسط کلاین همخوانی دارند.

جدول ۱۷- ضرایب قدرت تمایزگذاری خرده آزمونها، مقیاسها و کل آزمون

کل گروه	پسران	دختران	خرده آزمونها / مقیاسها / کل آزمون
۰/۹۴	۰/۹۲	۰/۹۴	موزاییکها
۰/۹۳	۰/۹۶	۰/۹۴	طبقه‌بندیها
۰/۹۴	۰/۹۴	۰/۹۳	موقعیتها
۰/۹۹	۰/۹۱	۰/۹۴	الگوها
۰/۹۷	۰/۹۵	۰/۹۷	مقیاس عملی
۰/۹۶	۰/۹۵	۰/۹۶	مقیاس استدلال
۰/۹۸	۰/۹۷	۰/۹۸	کل آزمون

شناخت تصاویر

همانطور که قبلاً نیز گفته شد، خرده آزمونهای طبقه‌بندیها و موقعیتها از تصاویر معنادار تشکیل یافته‌اند. از این رو احتمال تورش فرهنگی^۶ در این نوع خرده آزمونها

بیشتر از خرده آزمونهایی است که از تصاویر بی‌معنا، مانند اشکال هندسی، تشکیل یافته‌اند (جنسن^{۶۱}، ۱۹۸۰). بنابراین شیوه دیگری که جهت بررسی مناسب و روا بودن آزمون SON-R 2 ½-7 برای کودکان ایرانی مورد استفاده قرار گرفت بر شناخت تصاویر خرده آزمونهای طبقه‌بندیها و موقعیتهای توسط کودکانی که به سوال پاسخ نادرست دادند، مبتنی بود. اصلی که در پس این شیوه قرار دارد، این است که کودک نباید به خاطر اینکه دو یا چند تا از تصاویر یک سؤال را نمی‌شناسد در دادن پاسخ درست به سؤال دچار شکست گردد. جدول ۱۸، ۸ تا از سوالات خرده آزمون طبقه‌بندیها را نشان می‌دهد که شامل تصاویری هستند که برای حداقل ۰/۰۵ از کودکان مورد مطالعه، که به سؤال پاسخ نادرست دادند، ناآشنا هستند. ستون اول، شماره سوال و شماره تصویر را نشان می‌دهد. در ستون دوم فراوانی پاسخهای نادرست به سؤال ارائه شده است و در ستون سوم نیز درصدی از این کودکان که تصویر را نمی‌شناختند نشان داده شده است.

جدول ۱۸- سوالات خرده آزمون طبقه‌بندیها، شامل تصاویری است که برای حداقل ۰/۰۵ از کودکان ناآشنا هستند

درصد پاسخهای نامی شناسم	فراوانی پاسخهای نامی شناسم	فراوانی پاسخهای نادرست	سوال
٪۳۰	۸	۲۷	A (تصویر شماره ۱ اصلی)
٪۴۴	۱۲	۲۷	A (تصویر شماره ۵ جایگزین)
۰/۶۰۵	۳۴	۵۲	۸ (تصویر شماره ۱ جایگزین)
٪۴۲	۲۲	۵۲	۸ (تصویر شماره ۵ جایگزین)
۰/۰۵	۴	۷۵	۹ (تصویر شماره ۱ جایگزین)
۰/۰۵	۴	۷۵	۹ (تصویر شماره ۲ جایگزین)
۰/۲۰۱	۲۰	۹۵	۱۰ (تصویر شماره ۱ اصلی)
٪۲۳	۲۲	۹۵	۱۰ (تصویر شماره ۳ اصلی)
٪۶	۷	۱۱۵	۱۲ (تصویر شماره ۲ اصلی)
٪۶	۷	۱۱۵	۱۲ (تصویر شماره ۳ اصلی)
٪۱۰	۱۲	۱۱۵	۱۳ (تصویر شماره ۳ اصلی)
٪۶	۷	۱۱۵	۱۳ (تصویر شماره ۵ جایگزین)
٪۱۹	۲۲	۱۱۸	۱۴ (تصویر شماره ۲ اصلی)
۰/۲۰۱	۲۵	۱۲۱	۱۵ (تصویر شماره ۱ جایگزین)
٪۱۷	۲۰	۱۲۱	۱۵ (تصویر شماره ۳ جایگزین)

بر طبق نتایج جدول ۱۸، تصاویر مورد استفاده در سوالات A، ۸، ۹، ۱۰، ۱۲، ۱۳، ۱۴ و ۱۵ خرده آزمون طبقه‌بندیها، بطور متوسط برای ۰/۲۰۱ از کودکان مورد مطالعه ناآشنا هستند. در خرده آزمون موقعیتها، کودکان مورد مطالعه هیچ مشکلی در شناسایی تصاویر مورد استفاده در این خرده آزمون نشان ندادند. بطور کلی، نتایج این روش نشان داد که از بین ۶۰ سوال این آزمون، بنظر می‌آید تعداد ۸ تا از سوالات خرده آزمون طبقه‌بندیها از لحاظ فرهنگی دارای تورش هستند و لازم است که تصاویر آشنا، جایگزین آنها گردد.

بحث و نتیجه‌گیری

در این پژوهش، ویژگیهای روان سنجی و عملی بودن آزمون SON-R 2 1/2-7 برای کودکان ایرانی مورد بررسی قرار گرفت. نتایج این مطالعه (جدول ۱۰) نشان می‌دهد که نمرات خرده آزمونها، مقیاسها و کل آزمون از همسانی درونی بالایی برخوردارند. 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.

بر اساس یافته‌های مربوط به تمایزگذاری سنی (جدول ۱۵)، همبستگی سوال - نمره کل (جدول ۶)، همبستگی درونی خرده آزمونها (جدول ۱۳) همبستگی خرده آزمونها با مقیاسها و نمره کل (جدول ۱۴) و همچنین همبستگی بین نمرات کودکان در خرده آزمونها، مقیاسها و کل آزمون با ارزشیابیهای پژوهشگر و مربی از میزان انگیزش، درک دستورالعملها و آموزشها، همکاری و توجه آنها (جدول ۱۶) می‌توان چنین نتیجه گرفت که این آزمون یک ابزار روا جهت سنجش هوش کلی است و می‌توان با اطمینان خاطر از آن استفاده کرد. شاهد دیگری که در خصوص روایی نمرات آزمون مورد بررسی قرار گرفت، قدرت تمایزگذاری (جدول ۱۷) است. مقدار این شاخص برای کلیه خرده آزمونها، مقیاسها و نمره کل آزمون در سطح بالایی قرار دارد و حاکی از روایی نمرات آزمون است. علاوه بر این، یافته‌های حاصل از مطالعات بین فرهنگی که در خصوص 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^{۶۲}، BAS^{۶۳} و MSCA^{۶۴}، قویتر از همبستگی آن با مقیاسهای کلامی این آزمونها است. این همبستگیها از روایی واگرا و همگرایی آزمون مورد مطالعه حمایت می‌کند.

تحلیلهایی که در خصوص شاخص‌های آماری سؤالات (درجه دشواری و قدرت تشخیص) با استفاده از مدل کلاسیک اعتبار (جدول ۶) و مدل ۲ و ۳ پارامتری مبتنی بر IRT صورت گرفت (جدول ۷) نشان داد که کلیه سؤالات از شاخصهای آماری مطلوبی برخوردارند. برآوردهای مربوط به پارامتر b در مدل‌های IRT نشان می‌دهد که سؤالات از لحاظ دشواری در دامنه ۳- تا ۳+ بطور تقریباً یکنواخت توزیع شده‌اند. علاوه‌براین بغیر از تعداد اندکی از سؤالات، با بالا رفتن ترتیب سؤالات، دشواری آنها نیز افزایش می‌یابد. برآوردهای مربوط به پارامتر a نشان می‌دهد که سؤالات از قدرت تشخیص مطلوبی برخوردارند.

بطور خلاصه می‌توان گفت که آزمون SON-R 2 1/2-7 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 کودکان ۷- تا 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.

با این حال، باید خاطر نشان کرد که قبل از استانداردسازی و هنجاریابی آزمون در مقیاس وسیع باید برخی از تصاویر سؤالات A، ۸، ۹، ۱۰، ۱۲، ۱۳، ۱۴ و ۱۵ خرده آزمون طبقه‌بندیها با تصاویری که متناسب با فرهنگ ایران و آشنا برای کودکان ایرانی هستند، جایگزین گردند. علاوه براین، ترتیب سؤالات نیز باید بر اساس دشواری آنها اصلاح گردد. به نظر می‌رسد که با انجام این اصلاحات، می‌توان مطمئن بود که استانداردسازی و هنجاریابی آزمون مورد بحث در مقیاس وسیع مفید و سودمند خواهد بود.

یادداشت‌ها:

- 1) General Intelligence
- 2) Learning Potential
- 3) Tellegen & Laros
- 4) Cattell, R.B.
- 5) Raven
- 6) Nan Snijders- Oomen
- 7) Snijders-Oomen Nonverbal Intelligence Test
- 8) Reliability
- 9) Weiss
- 10) Validity
- 11) International Test Commission
- 12) Van de Vijver & Hambleton
- 13) Multi – Stage Random Sampling
- 14) Mosaics
- 15) Categories
- 16) Puzzles
- 17) Analogies
- 18) Situations
- 19) Patterns
- 20) Performance Scale
- 21) Reasoning scale
- 22) the test commission of the NetherLands Institute for Psychologists
- 23) Guttman
- 24) Stratified Alpha
- 25) Nunnally
- 26) Nunnally & Bernstein
- 27) Osburn
- 28) Meaningful
- 29) Non-Meaningful
- 30) Bayley Developmental Scales
- 31) Groningen Developmental Scales
- 32) Revision of Amsterdam Intelligence Test for Children
- 33) Test of Nonverbal Intelligence
- 34) Item–Response Theory
- 35) IRT Command Language
- 36) Hanson
- 37) Feldt & Brennan
- 38) Ten Berge & Zegres
- 39) Two-Group Multivariate Analysis of Variance
- 40) Hotelling's T^2
- 41) Stevense
- 42) Howell
- 43) Achenbach & Rescorla
- 44) Item Validity
- 45) Kline.P
- 46) Garrett
- 47) Newcomer & Hammill
- 48) Baker
- 49) Latent Ability
- 50) Maximum Likelihood (MLE)
- 51) Bayes

- | | |
|---|----------------------------------|
| 52) Expected a posterior (EAP) | 53) Generalizability Coefficient |
| 54) Partial Correlation | 55) Guilford & Fruchter |
| 56) Chen & Popovich | 57) Fisher's Z Transformation |
| 58) Correction for Attenuation | 59) Discrimination Power |
| 60) Cultural Bias | 61) Jensen |
| 62) British Ability Scales | |
| 63) McCarthy Scales of Children's Abilities | |

منابع

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

نصفت، مرتضی (۱۳۷۱). اصول ورشهای آمار. جلد اول. تهران: انتشارات دانشگاه تهران

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

Achenbach, T.M; & Rescorla, L.A. (2001). Manual for the ASEBA School –Age Forms & Profiles. Burlington,VT: University of Vermont, Research Center of Children, Youth & Families.

Anastasi, A.(1990).Psychological testing. (6th Ed). New York: Macmillan.

Anastasi. A; & Urbina, S. (1997). Psychological testing. (7th ed). Upper Saddle River, NJ. Prentice - Hall

Baker, F.B. (2001). The basics of item response theory. Eric clearinghouse on assessment and evaluation. University of Maryland, college park.

Cattell, R.B. (1971). Abilities: their structure, growth, and action. Boston: Houghton Mifflin.

- Cattell, R.B. (1950). Handbook for the individual of group culture Fair Intelligence Test. Scale I. Champaign,IL: Institute for PERSONALITY and Ability testing.
- Chen,P.Y;&Popovich, P.M. (2002). Correlation: parametric and nonparametric measures. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-139. Newbury Park, CA: Sage.
- Feldt, L. S; & Brennan, R.L.(1988). Reliability. In R.L. Linn (Ed), Educational measurement (3rd ed.) New York: American Council on Education/ Macmillan.
- Ferguson, G.A. (1949). On the theory of test development . Psychometrika, 14, 61-68.
- Garrett, H. (1965).Testing for teachers. New York: American Book.
- Guilford, J. P., & Fruchter, B. (1978). Fundamental statistics in psychology and education.New York: McGraw- Hill.
- Guttman,L.(1945).A basis for analyzing test–retest reliability. Psychometrika, 10, 255-282.
- Hanson, B. A. (2002). IRT Command Language. [Available at <http://www.b-a-h.com/software/irt/icl/>].Accessed 24 September 2005.
- Howell, D.C. (1997). Statistical methods for psychology.(4th ed.) Belmont, CA: Duxbury.
- Jensen, A. R. (1980). Bias in mental testing. New York: Free Press.
- Kline, P. (2000). The handbook of psychological testing. (2nd ed). London: Routledge.

- Laros, J.A; & Tellegen, P.J. (n.d.). Cross-cultural research with Snijders-Oomen Nonverbal Intelligence Tests. [Available at <http://www.testreart.nl/sonroe/crssculte.html>]. Accessed 15 August 2005.
- Newcomer, P.L; & Hammill, D.D. (1997). Test of language development- primary. (3rd ed.) Austin, TX: PROED.
- Nunnally, J.C. (1978). Psychometric theory. (2nd ed.) New York: McGraw-Hill.
- Nunnally, J.C; & Bernstein, I. H. (1994). Psychometric theory. (3rd ed.) New York: McGraw-Hill
- Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. Psychological Bulletin, 5, 343-355.
- Raven, J.C. (1938) Progressive matrices: A perceptual test of intelligence. London: H. K. Lewis.
- Snijders-Oomen, N. (1943). Intellinentieonderzoek van doofstomme kinderen. Nijmegen: Berkhout.
- Stevens, J.P. (2002). Applied multivariate statistics for the social sciences. (4th ed.) Mahwah, NJ: Lawrence Erlbaum.
- Tellegen, P. J., Winkel, M; Wijnbeg-Williams, B; & Laros, J. A (1998). Snijders-Oomen nonverbal intelligence test. SON-R 2 1/2 -7 Manual and research report. Lisse: Swets & Zeitlinger.
- Tellegen, P.J; & Laros, J. A. (1993). 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, methodological and practical Issues. Amsterdam: Swets & Zeitlinger.

- Tellegen, P.J; & laros, J. A. (2004). Cultural Bias in the SON-R Test: Comparative study of brazilian and dutch children. *psicologia: Teoria e pesquisa*, 20, 103-111.
- Ten Berge, J.M. F; & Zegers, F. E. (1978). A series of lower bounds to the reliability of a test. *Psychometrika*, 43, 257-579.
- Van de Vijver, F.J. R.; & Hambleton, R.K. (1996). Translating tests: some practical guidelines. *European Psychologists*, 1 (2). 89-99 .
- Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, 6 (4), 473-492.