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 روا، سنجی و عملی بودن فرم کوتاه آزمون غیرکلامی هوش اسنایدرز اومان برای کودکان ۲ ۱ سال( SON-R 2 ½ 7- روی یک گروه از کودکان تهران با حجم ۱۲۵ نفر ۶۶ پسر، ۵۹ دختر) که با روش نمونه برداری تصادفی چند مرحله ای از بین کودکان مهدهای کودک تحت نظارت بهزیستی انتخاب شدند، اجرا گردید. برای تحلیل سؤالات از مدل کلاسیک آزمون و مدل ۲، ۳ پارامتری مبتنی بر نظریه سوال - پاسخ (IRT) استفاده گردید. اعتبار خرده آزمونها (موزاییکها، طبقه بندیها، موقعیتها و الگوها) با استفاده از فرمول $\lambda_2$ و اعتبار مقیاسهای عملی و استدلالی و کل آزمون از طریق فرمول آلفای طبقه ای برآورد گردید. ضریب اعتبار خرده آزمونها برای کل گروه به ترتیب ۰/۷۲، ۰/۷۱، ۰/۷۲، ۰/۷۵ و ضریب اعتبار مقیاس عملی، استدلالی و کل آزمون نیز به ترتیب ۰/۹۱، ۰/۹۶ و ۰/۹۴ است. در خصوص روایی نیز تحلیلهای مختلفی صورت گرفت. از جمله تحلیلها می توان 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 یافته های حاصل از این پژوهش نشان داد که آزمون مورد بحث از اعتبار و روایی بالایی برخوردار است و با اطمینان می توان از آن جهت سنجش هوش کلی کودکان ۲ ۱ سال استفاده کرد. با این حال، قبل از استانداردسازی و هنجاریابی آن در مقیاس وسیع باید برخی از تصاویر سؤالات A ۴، ۳، ۲، ۱۰، ۱۰، و ۱۵ خرده آزمون طبقه بندیها با تصاویری که برای کودکان ایرانی آشنا هستند، جایگزین گردند. علاوه بر این ترتیب سوالات نیز باید بر اساس دشواری آنها اصلاح گردد.

**واژه‌های کلیدی:** *آزمون غیرکلامی، هوش، اعتبار، روایی*

---

* عضو هیأت علمی پژوهشکده کودکان استثنایی ( Email: as_minaei@yahoo.com )

در اینجا لازم می دانم از آقای دکتر تلخن بخاطر ارسال یک نسخه از آزمون و همچنین نوار ویدئو صمیمانه تشکر نمایم.



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

## مقدمه

آزمونهای سنتی هوش کلی[1](GI ، مانند آزمونهای هوش وکسلر و اس نفورد – بینه توسط طرفداران آزمونهای ظرفیت یادگیری[2] (LP) مورد انتقاد قرار گرفته است، زیرا این آزمونها به جای ظرفیت یادگیری، نتیجه نهایی یادگیریهای پیشین را اندازه می گیرند. آزمونهای GI صرفاً نتیجه نهایی یادگیریهای پیشین را منعکس می سازند و از این رو، توا یی یادگیری افراد متعلق به طبقه اجتماعی – اقتصادی پایین و افراد مبتلا به مشکلات یادگیری و اختلالات رفتاری و همچنین افرادی که فرصت های اندکی جهت کسب دانش و مهارتهای مورد نیاز برای موقعیت آزمون داشت اند را کم برآورد می کنند. نکت ای که تلویحاً با این انتقاد مرتبط است، این است که آزمونهای GI هیچ اطلاعاتی در خصوص اینکه در شرایط بهینه یادگیری تا چه میزان می توان انتظار رشد در عملکرد را داشت، فراهم نمی کنند. 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 د رد، لکن ابزارهای عملی برای سنجش، فقط در چند سال اخیر است که ارائه شد اند.

آزمونهای GI همچنین از سوی طرفداران آزمونهای هوش نابسته به فرهنگ به خاطر محتوایشان مورد انتقاد قرار گرفت اند. به خاطر اینکه این آزمونها از لحاظ محتوا و دستورالعمل ها اغلب به مهارتهای زبانی خاصی وابسته هستند. لذ ، باعث می گردند تا افراد اقلیت های فرهنگی و قومی و همچ ین 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، ۹۹۳ ). می توان گفت کـه ایـن آزمونها بیشتر بر هوش متبلور تاکید دارند تا هوش سیال (کتل[4] ۹۷۱ ). این انتقادها نسبت به آزمونهای GI باعث گردید تا آزمونهای غیرکلامی هوش مانند ماتریسهای پیشرونده ریون (ریون[5]،



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

۹۳۸ ) و آزمون هوش نابسته به فرهنگ کتل ۹۵۰ ) با هدف به حداقل رساندن اتکا به دانش اکتسابی و توانایی کلامی پرورش یابند.

در اوایل دهه چهل میلادی ، خانم نان اسنایدر – اومان[۶] ۹۴۳ ) که در یک موسسه مربوط به کودکان ناشنوا در هلند کار می کرد اولین آزمون غیرکلامی ، وش با نام SON[۷] را پرورش داد. او هوش را توانایی یادگیری؛ یعنی میزان توانایی کودکان در بهر 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 موج د در آن زمان، برای بررسی طیف وسیعی از تواناییهای یادگیری مناسب نبودند. زیرا آنها اساساً از آزمونهای عملی که به تواناییِ های 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 نرمهایی برای کودکان ناشنوای ۴ 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 ½ 7- و نسخه دیگر برای کودکان سنین بالا با عنوان SON-R 5 ½ 17- . مقاله حاضر مطالعه ای را توصیف می کند که با آزمون SON-R 2 ½ 7- صورت گرفته است.

آزمونSON-R 2 ½ 7- سه تفاوت اساسی با آزمونهای سنتی هوش دارد. اولا، این آزمون مستلزم تواناییهای زبانی خاصی نیست؛ ثانیا، اجرای آن به شیوه انطباقی ، صورت می گیرد. هدف شیو های انطباقی محدود نمودن تعداد سوالات ارائه شده همراه با کاهش نسبتاً کمِ اعتبار[۱] است(ویس ، ۹۸۲ ). درSON-R 2 ½ 7- نقطه شروع یا سطح ورودی هر خرده آزمون بر اساس سن کودک و سالِ ی 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. ثالثا، در مورد درست یا نادرست بودن پاسخ آزمودنی به او بازخورد داده می شود. مزیت اصلی ار ه بازخورد این است که به آزمودنی فرصت داده می شود تا در طول اجرای آزمون، دست به یادگیری بزند. با توجه به این ویژگیه می توان گفت که آزمونSON-R 2 ½ 7- بیشتر به آزمونهای هوش نابسته به فرهنگ و آزمونهای ظرفیت یادگیری شباهت دارد تا آزمونهای سنتی هوش کلی.



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

در ایران نیاز زیادی به آزمونهای روانی رو، 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 ½ 7- و انتقادات وارده بر آزمونهای سنتی هوش کلی ذکر گردید، آزمون SON-R 2 ½ 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[10] 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 بین المللی آزمون[11] در خصوص استفاده از آزمون، هماهنگ و مطابق است (ون دی ویجور و همبلتون[2] ۹۹۶ ). یکی از این رهنمودها مطرح می کند که پرورش دهندگان 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 نمونه گیری

نمونه مورد مطالعه در این تحقیق ۱۲۵ کودک تهرانی ۶۶ پسر و ۵۹ دختر) با دامنه سنی ۱ ′ تا ۱ ′ ۲ سال و ۹ ماه تا ۷ سال و ۹ ماه) با میانگین ۱ ۴ و انحراف استاندارد ۰ ′ ۱ است که در تابستان سال ۱۳۸۴ در مهدهای کودک تحت نظارت بهزیستی مشغول به آموزش بودند. جهت انتخاب نمونه از روش نمونه گیری تصادفی چند مرحله ای[13] 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 ′ مهد کودک) بصورت کاملاً تصادفی انتخاب گردید. در گام بعد با استفاده از دفتر ثبت نام، از میان کودکان هر مهد تعداد ۲ ۱۰ کودک دختر و پسر با استفاده از روش تصادفی منظم انتخاب شدند.



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

## ابزار

ابزار مورد استفاده در این پژوهش، آزمون غیرکلامی هوش اسنایدر – اومان برای کودکان ۲ تا ۷ 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: موزاییکها[۱۴] با ۱۵ سؤال؛ طبق بندیها[۱۵] با ۱۵ سؤال؛ پازلها[۱۶] با ۱۴ سوال؛ قیاسها[۱۷] با ۱۷ سوال؛ موقعیتها[۱۸] با ۱۴ سوال؛ الگوها[۱۹] با ۱۶ سوال. 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 تا ۷ است.

این آزمون در مجموع دارای ۹۱ سؤال است و در مدت ۴۰ تا ۵۰ دقیقه در یک یا دو جلسه اجرا می گردد. خرده آزمونهای SON-R 2 ½-7 به دو گروه کلی طبق بندی می گردند. گروه اول شامل خرده آزمونهای الگوها، موزاییکها و پازلها است که مقیاس عملی[۲۰] (PS) نامیده می شود و گروه دیگر شامل خرده آزمونهای موقعیتها، طبق بند ها و قیاسها است که مقیاس استدلال[۲۱] (RS) نام دارد. بطو کلی، 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 می آید.

هر دو فرم این آزمون توسط کمیته آزمون موسسه روان شناسان هلند[۲] (COTAN) از ۸ بعد و در یک مقیاس ۴ طبقه ای ناکافی، کافی و خوب ارزیابی شده است. نتایج این ارزیابی در زیر ارائه شده است (اقتباس از تلخن و همکاران ۱۹۹۸).

| | |
|---|---|
| اصول و پایه های ساخت آزمون | خوب |
| قابلیت اجرایی مواد | خوب |
| قابلیت اجرایی کتابچه راهنما | خوب |
| نرمها | خوب |
| اعتبار | خوب |
| روایی سازه | خوب |
| روایی ملاکی | کافی |



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

جدوا ، (اقتباس از تلخن و همکاران ، ۹۹۸ ) ضرایب اعتبار خرده آزمونها را که مبتنی بر همسانی درونی سوالات هر خرده آزمون است و با استفاده از فرمول $\lambda_2$ گاتمن[۳] ۹۴۵ ) و بر اساس نمونه استانداردسازی هلند برآورد شده است به تفکیک گروههای سنی نشان می دهد.

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

| | خرده‌آزمونها | | | | | | مقیاسها | | |
|---|---|---|---|---|---|---|---|---|---|
| سن | موزاییکها | طبقه‌بندیها | پازلها | قیاسها | موقعیتها | الگوها | عملی | استدلال | کل آزمون |
| ۲;۶ | ۰٫۱ | ۰٫۱ | ۰٫۵ | ۰٫۵ | ۰٫۹ | ۰٫۹ | ۰٫۸ | ۰٫۹ | ۰٫۶ |
| ۳;۶ | ۰٫۶ | ۰٫۳ | ۰٫۵ | ۰٫۳ | ۰٫۶ | ۰٫۳ | ۰٫۶ | ۰٫۴ | ۰٫۰ |
| ۴;۶ | ۰٫۷ | ۰٫۰ | ۰٫۵ | ۰٫۴ | ۰٫۲ | ۰٫۲ | ۰٫۸ | ۰٫۱ | ۰٫۰ |
| ۵;۶ | ۰٫۴ | ۰٫۸ | ۰٫۰ | ۰٫۸ | ۰٫۲ | ۰٫۴ | ۰٫۷ | ۰٫۱ | ۰٫۰ |
| ۶;۶ | ۰٫۸ | ۰٫۸ | ۰٫۹ | ۰٫۳ | ۰٫۶ | ۰٫۶ | ۰٫۷ | ۰٫۴ | ۰٫۱ |
| ۷;۶ | ۰٫۴ | ۰٫۹ | ۰٫۹ | ۰٫۵ | ۰٫۹ | ۰٫۹ | ۰٫۸ | ۰٫۶ | ۰٫۲ |
| میانگین | ۰٫۳ | ۰٫۱ | ۰٫۹ | ۰٫۸ | ۰٫۷ | ۰٫۵ | ۰٫۵ | ۰٫۴ | ۰٫۰ |

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 ½ 7-2 انجام داده‌اند. نتایج این مطالعات حاکی از روایی خوب نمرات آزمون است. ازجمله این مطالعات می‌توان به همبستگی بین این آزمون باسایر آزمونهای شناختی اشاره کرد. جدول ۲، خلاصه‌ای از ضرایب همبستگی بین SON-R ½ 7-2 با سایر آزمونهای شناختی را نشان می‌دهد.

خرده آزمونهای SON-R ½ 7-2 را می توان بر اساس نوع موادی که مورد استفاده قرار می گیرد به دو دسته تقسیم کرد: الف – خرده آزمونهایی که از مواد تصویری معنادار[۲۸] استفاده می کنند (طبق بندیها موقعیتها و پازلها)؛ ب - خرده آزمونهایی که از مواد تصویری بی معنی[۲۹] مانند اشکال هندسی استفاده می نمایند (موزاییکها، الگوها و قیاسها).



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

**جدول۲- ضرایب همبستگی آزمون SON-R 2 ½ 7- با سایر آزمونهای شناختی**

| آزمون ملاک | N | r | آزمون ملاک | N | r |
|---|---|---|---|---|---|
| BOS[۳۰] | ۵۰ | ۰/۹۰ | TONI-2[۳۳] فرم الف | ۱۵۳ | ۰/۰۰ |
| GOS[۳۱] | | | | | |
| DI شناختی | ۱۱۵ | ۰/۵۰ | WPPSI | | |
| DI همزمان | | ۰/۴۰ | IQ کلی | ۵۳ | ۰/۰۰ |
| DI زنجیر ای | | ۰/۹۰ | مقیاس کلامی | | ۰/۹۰ |
| RAKIT[۳۴] | ۱۶۵ | ۰/۰۰ | مقیاس عملی | | ۰/۹۰ |

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 روان سنجی و عملی بودن فرم کوتاه SON-R 2 ½ 7- که از چهار خرده آزمون موزاییکها، طبقه بندیها، موقعیتها 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 بر 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 می پردازد.

## شیوه اجرا

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

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. جمع آوری داده ها تقریباً دو ماه بطول انجامید.

همانطور که گفته شد اجرای این آزمون به شیوه انطباقی صورت می گیرد. در این مطالعه نیز از شیوه انطباقی استفاده گردید. برای اجرای آزمون بر روی هر کودک، بطور متوسط ۳۵/۰ دقیقه وقت صرف گردید. بعد از هر سوال به کودک گفته می شد که آیا پاسخ او درست است یا نادرست. ارائه بازخورد به کودک یک بخش مهم از شیوه اجرای


[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

استاندارد آزمون SON-R 2 ½ 7 است، زیرا باعث می گردد تا دستورالعمل ها روشن و واضح گردد و به کودک فرصت می دهد 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 می شد که آیا تصاویر بکار رفته در سؤال را می شناسد و می تواند نام آنها را بگوید. اجرای هرخرده آزمون بعد از ۳ پاسخ نادرست متناوب و یا ۲ پاسخ نادرست متوالی متوقف می گردید. از مربیان مهد کودک خواسته می شد تا کودکان را براساس میزان انگیزش، همکاری، توجه و درک آموزشها در طول ساعات کلاس در یک مقیاس ۴ امتیازی از = خوب ' = متغیر ' = متوسط ' = ضعیف؛ ارزشیابی نمایند.

## تحلیل داده‌ها

برای تحلیل و برآورد ویژگیهای روان سنجی 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 روان سنجی سؤالات در مدل کلاسیک و انجام تحلیلهای آماری از نرم افزار SPSS و برای برآورد پارامترهای سؤال در مدلهای مبتنی بر IRT از برنامه CL[۳۵] (هانسون[۳۶] ۲۰۰۲) استفاده گردید.

در برخی از تحلیل ها از نمرات استاندارد خرد آزمونها 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 $r_{tt}$ (الفای کرونباخ) گاتمن (فلدت و برنان[۳۷] ۹۸۸) استفاده گردید. زیرا این فرمول، اعتبار را بویژه در آزمونهای کوتاه به اندازه فرمول آلفای کرونباخ، کم برآورد نمی کند (تنبرگ و زگرس[۳۸]، ۹۷۸). جهت برآورد اعتبار نمرات مقیاسها و کل آزمون از فرمول آلفای طبقه ای (نانالی و برنشتاین ۹۹۴؛ اُسبرن 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 استدلالی) و کل آزمون از آزمون t برای دو گروه مستقل (هاول[۴۲] ۹۹۷) استفاده گردید. در انجام این تحلیل ها از نمرات خام استفاده شد. زیرا در نمرات خام تمام واریانس یا پراکندگی موجود در نمرات به حساب می آید (آخنباخ و


[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

رسکورلا [۳] ۲۰۰۱). در توصیف و تحلیل داد ها از روشها و شاخص های 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 ۲ و ۳) و کل گروه (جدول ۱) ارائه شده است.

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

| خرده آزمونها و مقیاسها | نمرات خام | | نمرات‌استاندارد | | | |
|---|---|---|---|---|---|---|
| | میانگین | انحراف‌استاندارد | میانگین | انحراف‌استاندارد | چولگی | کشیدگی |
| موزاییکها | ۱۲ ۶ | ۱۰ ۲ | ۳ ۹ | ۱۸ ۲ | ۰۷ - | ۰ ۵ |
| طبق بند ها | ۸ ۷ | ۷ ۲ | ۲ ۹ | ۵ ۳ | ۶ - | ۵ - |
| موقعیتها | ۳ ۶ | ۲ ۲ | ۴ ۹ | ۱۴ ۲ | ۴ - | ۱۱ - |
| الگوها | ۶ ۹ | ۵ ۳ | ۷ ۱۰ | ۳ ۳ | ۴ ۰ | ۶ - |
| مقیاس عملی | ۱۸ ۱۶ | ۸ ۵ | ۰ ۹۹ | ۸ ۱۵ | ۴ - | ۱۰ - |
| مقیاس استدلال | ۱ ۱۴ | ۷ ۴ | ۵ ۹۹ | ۶ ۱۵ | ۱ - | ۷ - |
| کل آزمون | ۹ ۳۰ | ۵ ۹ | ۸ ۹۹ | ۴ ۱۵ | ۴ - | ۱ - |

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

| خرده آزمونها و مقیاسها | نمرات خام | | نمرات‌استاندارد | | | |
|---|---|---|---|---|---|---|
| | میانگین | انحراف‌استاندارد | میانگین | انحراف‌استاندارد | چولگی | کشیدگی |
| موزاییکها | ۷ ۷ | ۱۶ ۲ | ۹ ۱۰ | ۴ ۳ | ۵ - | ۵ ۰ |
| طبق بند ها | ۱۰ ۷ | ۰ ۲ | ۰ ۱۰ | ۰ ۲ | ۰ - | ۱۲ ۰ |
| موقعیتها | ۲ ۷ | ۱ ۲ | ۹ ۱۰ | ۵ ۳ | ۳ - | ۱۳ - |
| الگوها | ۲ ۹ | ۹ ۲ | ۳ ۹ | ۶ ۲ | ۱ ۰ | ۲ - |
| مقیاس عملی | ۹ ۱۶ | ۳ ۴ | ۹ ۱۰۰ | ۵ ۱۴ | ۱ - | ۹ - |
| مقیاس | ۲ ۱۵ | ۳ ۵ | ۴ ۱۰۱ | ۶ ۱۴ | ۵ - | ۲ ۰ |
| استدلال | ۱ ۳۱ | ۰ ۹ | ۱ ۱۰۱ | ۰ ۱۴ | ۹ - | ۳ ۰ |
| کل آزمون | | | | | | |



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

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

| خرده آزمونها و مقیاسها | نمرات خام | | نمرات‌استاندارد | | | |
|---|---|---|---|---|---|---|
| | میانگین | انحراف‌استاندارد | میانگین | انحراف‌استاندارد | چولگی | کشیدگی |
| موزاییکها | ۹ ۷ | ۲،۲ | ۱۰ | ۳ | ۰٫۴ - | ۰ ۷ |
| طبق بند ها | ۷ ٫۳ | ۲ ،۶ | ۱۰ | ۳ | ۰٫۰ - | ۰ ٫۷ |
| موقعیتها | ۱ ۷ | ۲ ٫۷ | ۱۰ | ۳ | ۱،۲ - | ۱،۰ - |
| الگوها | ۹ ۹ | ۲ ،۳ | ۱۰ | ۳ | ۰ ۸ | ٫۸ - |
| مقیاس عملی | ۱۶ ٫۸ | ۵ ،۵ | ۱۰۰ | ۱۵ | ۸ - | ۰٫۱ - |
| مقیاس استدلال | ۱۴ ٫۴ | ۴ ٫۸ | ۱۰۰ | ۱۵ | ۰٫۴ - | ۹ - |
| کل آزمون | ۳۱ ۲ | ۹ ٫۴ | ۱۰۰ | ۱۵ | ۰٫۱ - | ٫۳ - |

از اطلاعات ارائه شده در جداول ۳ و ۴ می توان چنین استنباط کرد که بین مشخصه های آماری خرده آزمونها، مقیاسها و کل آزمون در دو گروه دختران و پسران، تفاوت قابل ملاحظه ای 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 [$T^2$( ۲۰ ،۰۴ = $p$= ۰٫۲ ۰۴ =].

نتیجه آزمون t نشان دادکه 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 :[ ۰۴، =p، ۰٫۱ = ۲۳ )t] ؛ [t( ۲۲ = ۰٫۴ $p$ = ۰٫۶ و [t( ۲۳ = ۰٫۶ $p$ = ۰٫۱ ) $t$ است.

شاخص های 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 می شود، بیانگر نسبت افرادی است که به سؤال پاسخ درست داد اند. قدرت تشخیص سوال در این مدل، که گاهی اوقات روایی سؤال[۴۴] نیز نامیده می شود، همواره به صورت همبستگی دو رشته ای یا دو رشته ای نقط ای سوال با نمره کل آزمون یا خرده آزمون بیان می شود (ثرندایک، ترجمه هـ ز، ۳۷۵ ). این شاخص معلوم می نماید که سوال تا چه اندازه می تواند گروههای مختلف افراد را از هم جدا سازد. در واقع این شاخص نشار دهنده همسویی و هماهنگی سؤال با آزمون یا خرده آزمون است. از آنجا که در برنامه SPSS، همبستگی دو رشته ای نقط ای وجود



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

ندارد، جهت برآو د همبستگی سوال با نمره کل خرده آزمون از همبستگی گشتاوری پیرسون استفاده گردید. همبستگی دو رشت ای نقط ای به لحاظ عددی ، دقیقاً معادل با همبستگی گشتاوری پیرسون است (کلاین[5]، ۲۰۰۰). از آنجا که نمره سؤال در نمره کل خرده آزمون لحاظ می شود و این امر باعث افزایش غیرواقعی همبستگی سوال با نمره کل می گردد، 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**

| سوال | خرده آزمونها | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | موزاییکها | | طبقه‌بندیها | | موقعیتها | | الگوها | |
| | درجه دشواری | قدرت تشخیص | درجه دشواری | قدرت تشخیص | درجه دشواری | قدرت تشخیص | درجه شواری | قدرت تشخیص |
| ۱ | ۰،۸ | ۰،۰ | ۰،۷ | ۰،۷ | ۰،۹ | ۰،۵* | ۰،۹ | ۰،۳* |
| ۲ | ۰،۶ | ۰،۲ | ۰،۸ | ۰،۹ | ۱ ۰ | ۰،۰* | ۰،۷ | ۰،۵ |
| ۳ | ۰،۲ | ۰،۴ | ۰،۸ | ۰،۵ | ۰،۴ | ۰،۸ | ۰،۸ | ۰،۱ |
| ۴ | ۰،۹ | ۰،۰ | ۰،۶ | ۰،۳ | ۰،۶ | ۰،۵ | ۰،۸ | ۰،۷ |
| ۵ | ۰،۵ | ۰،۹ | ۰،۶ | ۰،۵ | ۰،۷ | ۰،۷ | ۰،۴ | ۰،۵ |
| ۶ | ۰،۱ | ۰،۸ | ۰،۲ | ۰،۴ | ۰،۶ | ۰،۵ | ۰،۸ | ۰،۳ |
| ۷ | ۰،۱ | ۰،۸ | ۰،۲ | ۰،۴ | ۰،۶ | ۰،۵ | ۰،۸ | ۰،۸ |
| ۸ | ۰،۸ | ۰،۶ | ۰،۸ | ۰،۱ | ۰،۸ | ۰،۹ | ۰،۶ | ۰،۴ |
| ۹ | ۰،۷ | ۰،۳ | ۰،۰ | ۰،۷ | ۰،۸ | ۰،۱ | ۰،۹ | ۰،۷ |
| ۱۰ | ۰،۵ | ۰،۲ | ۰،۴ | ۰،۳ | ۰،۳ | ۰،۴ | ۰،۱ | ۰،۳ |
| ۱۱ | ۰،۲ | ۰،۵ | ۰،۸ | ۰،۷ | ۰،۴ | ۰،۰ | ۰،۶ | ۰،۴ |
| ۱۲ | ۰،۶ | ۰،۸ | ۰،۸ | ۰،۱ | ۰،۶ | ۰،۶ | ۰،۳ | ۰،۸ |
| ۱۳ | ۰،۱ | ۰،۵* | ۰،۸ | ۰،۹ | ۰،۰۲ | ۰،۹* | ۰،۶ | ۰،۰ |
| ۱۴ | ۰،۳ | ۰،۲ | ۰،۶ | ۰،۰ | ۰،۰۸ | ۰،۵* | ۰،۰ | ۰،۹ |
| ۱۵ | ۰،۲ | ۰،۹ | ۰،۳ | ۰،۱ | — | — | ۰،۸ | ۰،۷ |
| ۱۶ | — | — | — | — | — | — | ۰،۰۸ | ۱۸/* |
| **میانگین** | ۰،۷ | ۰،۷ | ۰،۲ | ۰،۹ | ۰،۰ | ۰،۴ | ۰،۸ | ۰،۵ |



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

نتایج جدول فوق نشان می دهدکه درجه دشواری و قدرت تشخیص سوالات در سطح مطلوب قرار دارند. آناستازی ۹۹۰ ) و آنا تازی و یوربینا ۹۹۷ ) اظهار می دارند که متوسط درجه دشواری سوالات باید تقریباً ۰٫۵۰ 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 بر- وردارند و صرفاً ۸ سوال که با علامت* مشخص شد اند قدرت تشخیص نسبتاً ضعیفی دارند. نانالی و برنشتاین ۹۹۴ ) اظهار می دارند که اکثر ضرایب همبستگی سوال - نمره کل، در دامنه ۰ تا ۰٫۶۰ قرار دارند. گارت [۴۶] ۹۶۵ ؛ به نقل از نیوکامر و هامیل[۴۷] ۹۹۷ ) اظهار کرده است که سوالات دارای ضریب همبستگی ۰٫۲۰ یا بالاتر در صورتی که آزمون نسبتاً طولانی باشد می توانند دارای روایی تلقی شوند. آناستازی ۹۹۰ ) و آناستازی و یوربینا ۹۹۷ ) توصیه می کنند که ضرایب همبستگی ۰٫۱۵ یا ۰٫۱۰ می توانند قابل قبول تلقی گردند .

در مدل ۳ پارامتری، هر سوال براساس ۳ پارامتر a ، b، و c توصیف می گردد. 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 می شود. به بیان دقیق تر، در این مدل پارامتر c برای سوال برآورد نمی شود و سؤال صرفاً بر اساس ۲ پارامتر a و b توصیف می شود.

نماد b در IRT بیانگر پارامتر دشواری سؤال است و جایگاه سؤال را در مقیاس توانایی توصیف می کند (بیکر[۸]، ۲۰۰۱ ). به زبان فنی و ریاضی می توان گفت که این پارامتر نقطه عطف تابع ویژگی سؤال را توصیف می کند و معمولاً در دامنه بین ۲٫۰ - تا ۲٫۰ + مقیاس پردازی می شود.

پارامتر قدرت تشخیص در مدل های IRT با نماد a نمایش داده می شود و بیانگر این است که سؤال تا چه اندازه می تواند بین افرادی که توانایی آنها پایین تر از جایگاه سؤال است با افرادی که توانایی آنها بالاتر از این جایگاه قرار دارد، تمایز ایجاد کند. به بیان دیگر پارامتر a نشان می دهدکه وقتی سطح توانایی مکنون[۴۹] بالا می رود احتمال موفقیت در سؤال با چه سرعتی افزایش می یابد (ثرندایک، ترجمه هومن ۱۳۷۵). به لحاظ نظری این پارامتر در مقیاس $-\infty$ تا $+\infty$ تعریف می شود. ولی سؤالاتی که دارای قدرت تشخیص منفی هستند از آزمونها حذف می گردد. همچنین بدست آوردن



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

مقادیر بزرگتر از ۲ برای این پارامتر غیر معمول است. بنابراین، دامنه این پارامتر معمولاً بین ۰ تا ۲ تغییر می کند.

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 ۰/۵ برای آن قابل قبول نیست (بیکر، ۲۰۰۱). جدول ۷ پارامترهای سؤالات را بر اساس مدلهایIRT که با استفاده از برنامه BILOG (هانسون و همکاران، ۲۰۰۲) برآورد شده اند نمایش می دهد.

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

| سوال | خرده آزمونها | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | موزاییکها | | طبقه‌بندیها | | | موقعیتها | | | الگوها | |
| | a | b | a | b | c | a | b | c | a | b |
| ۱ | ۰/۹۳ | -۳/۵۰ | ۱/۶۰ | -۲/۲۸ | ۰/۲۲ | ۱/۲۱ | -۳/۲۵ | ۰/۲۳ | ۰/۹۳ | -۳/۹۵ |
| ۲ | ۱/۲۷ | -۱/۵۱ | ۱/۴۶ | -۲/۷۷ | ۰/۲۲ | ۰/۹۷ | -۵/۸۱ | ۰/۲۳ | ۱/۰۹ | -۲/۷۵ |
| ۳ | ۱/۸۲ | -۱/۱۴ | ۰/۶۴ | -۳/۷۴ | ۰/۲۲ | ۱/۱۸ | -۲/۰۴ | ۰/۲۲ | ۱/۱۸ | -۳/۰۹ |
| ۴ | ۱/۹۵ | -۱/۵۱ | ۱/۸۷ | -۱/۱۹ | ۰/۱۹ | ۱/۰۴ | -۲/۳۹ | ۰/۲۳ | ۱/۴۵ | -۲/۶۴ |
| ۵ | ۱/۱۸ | -۰/۹۳ | ۰/۹۶ | -۰/۸۱ | ۰/۱۹ | ۰/۸۸ | ۰/۰۵ | ۰/۱۶ | ۱/۱۸ | -۲/۱۹ |
| ۶ | ۲/۰۲ | -۱/۰۳ | ۰/۸۷ | -۱/۲۲ | ۰/۱۸ | ۱/۴۱ | -۰/۳۶ | ۰/۱۴ | ۱/۱۲ | -۱/۷۳ |
| ۷ | ۱/۸۹ | -۰/۶۵ | ۰/۸۳ | -۰/۷۱ | ۰/۱۶ | ۱/۲۹ | ۰/۳۵ | ۰/۱۲ | ۱/۵۳ | -۱/۰۱ |
| ۸ | ۱/۹۵ | -۰/۲۲ | ۱/۹۸ | -۰/۱۵ | ۰/۰۹ | ۱/۳۰ | -۰/۱۰ | ۰/۱۰ | ۲/۳۱ | -۰/۸۷ |
| ۹ | ۱/۶۵ | ۰/۷۷ | ۱/۰۳ | ۰/۵۴ | ۰/۰۹ | ۱/۰۸ | ۰/۵۸ | ۰/۰۸ | ۲/۲۳ | -۰/۶۰ |
| ۱۰ | ۱/۲۰ | ۱/۴۳ | ۰/۴۸ | ۲/۵۹ | ۰/۱۳ | ۱/۰۷ | ۱/۲۰ | ۰/۰۶ | ۱/۴۷ | ۰/۲۹ |
| ۱۱ | ۱/۸۰ | ۱/۴۶ | ۱/۷۷ | ۱/۲۱ | ۰/۰۴ | ۱/۵۲ | ۱/۴۰ | ۰/۰۴ | ۱/۸۵ | ۰/۷۸ |
| ۱۲ | ۱/۷۸ | ۱/۹۰ | ۱/۰۹ | ۲/۱۸ | ۰/۰۴ | ۰/۹۷ | ۲/۶۲ | ۰/۰۴ | ۲/۲۸ | ۰/۸۸ |
| ۱۳ | ۱/۱۶ | ۳/۴۸ | ۱/۷۹ | ۱/۷۳ | ۰/۰۳ | ۰/۸۸ | ۴/۸۲ | ۰/۰۳ | ۲/۰۵ | ۱/۲۵ |
| ۱۴ | ۱/۷۸ | ۲/۳۱ | ۱/۹۲ | ۱/۹۲ | ۰/۰۳ | ۰/۹۱ | ۵/۵۸ | ۰/۰۳ | ۱/۷۰ | ۱/۷۰ |
| ۱۵ | ۱/۷۷ | ۲/۵۰ | ۱/۸۰ | ۲/۲۵ | ۰/۰۲ | — | — | — | ۱/۹۳ | ۱/۷۹ |
| ۱۶ | — | — | — | — | — | — | — | — | ۱/۳۷ | ۳/۳۲ |
| میانگین | ۱/۶۱ | ۰/۲۲ | ۱/۳۴ | -۰/۰۳ | ۰/۱۲ | ۱/۱۲ | ۰/۱۹ | ۰/۱۲ | ۱/۶۰ | -۰/۵۵ |

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



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

**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،۱ | -۱۷ |
| طبق بند ها | ۰،۱ | -۱۵ |
| موقعیتها | ۰،۱ | -۱۳ |
| الگوها | ۰،۱ | -۱۷ |

بین نمره خام افراد و پارامتر توانایی آنها در خرده آزمونها همب تگی گشتاوری پیرسون محاسبه گردید. در مدلهای RT پارامتر توانایی افراد با نماد θ (تتا) نشان داده می شود. در برنامه CL، پارامتر θ به دو روش بیشینه احتمال (MLE) و بیر[۱۱]، که با عنوان برآورد بعدی مورد انتظار[۵۲](EAP) مورد اشاره قرار می گیرد، برآورد می گردد. ضرایب همبستگی بدست آمده که همگی در سطح ۰،۱ معنادار هستند در جدول ۱، 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 | MLE و EAP |
|---|---|---|---|
| موزاییکها | ۰،۱۵ | ۰،۱۹ | ۰،۱۴ |
| طبق بند ها | ۰،۱۷ | ۰،۱۸ | ۰،۱۸ |
| موقعیتها | ۰،۱۷ | ۰،۱۹ | ۰،۱۹ |
| الگوها | ۰،۱۷ | ۰،۱۸ | ۰،۱۹ |

اعتبار نمرات خرده آزمونها که بر اساس فرمول $\lambda_2$ (فلدت و برناز ۹۸۸ ؛ اسبرز، ۰۰۰) برآورد شد اند به تفکیک جنس (جدول ۰ ) و گروه سنی (جدول ۱۰) ارائه شد اند. در این جداول همچنین ضرایب اعتبار نم ت مقیاس عملی (PS، مقیاس استدلال (RS) و کل آزمون (IQ) که با استفاده از فرمول آلفای طبق ای (نانالی و برنشتاین، ۹۹۴ ؛ اُسبرز، ۰۰۰) برآورد شد اند نشان داده شده است. از آنجا که ضرایب اعتبار $\lambda_2$ براساس نمونه هایی که از لحاظ سن ناهمگن هستند برآورد شدند،



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

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 | ۰/۷ | ۰/۹ | ۰/۲ |
| طبق بند ها | ۰/۰ | ۰/۶ | ۰/۱ |
| موقعیتها | ۰/۹ | ۰/۴ | ۰/۷ |
| الگوها | ۰/۶ | ۰/۶ | ۰/۵ |
| مقیاس عملی | ۰/۱ | ۰/۰ | ۰/۱ |
| مقیاس استدلال | ۰/۲ | ۰/۸ | ۰/۶ |
| کل آزمون | ۰/۱ | ۰/۶ | ۰/۴ |

همانطور که در جداول ۱۰ و ۱۱ ملاحظه می کنید، 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 می توان 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 | ۰/۵ | ۰/۹ | ۰/۵ |
| طبق بند ها | ۰/۳ | ۰/۶ | ۰/۱ |
| موقعیتها | ۰/۱ | ۰/۹ | ۰/۷ |
| الگوها | ۰/۸ | ۰/۵ | ۰/۱ |
| مقیاس عملی | ۰/۹ | ۰/۹ | ۰/۷ |
| مقیاس استدلال | ۰/۳ | ۰/۹ | ۰/۷ |
| کل آزمون | ۰/۲ | ۰/۱ | ۰/۰ |



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

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 ½-7 را می توان بعنوان نمونه ای از حیطه خرده آزمونهای غیرکلامی 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 غیرکلامی 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**

| مقیاس‌ها و کل آزمون | جنس | | گروه سنی | | | |
|---|---|---|---|---|---|---|
| | پسران | دختران | ۳-۲ | ۵-۴ | ۷-۶ | کل گروه |
| مقیاس عملی | ۰/۱۰ | ۰/۷۰ | ۰/۳۰ | ۰/۹۰ | ۰/۳۰ | ۰/۹۰ |
| مقیاس استدلال | ۰/۱۰ | ۰/۶۰ | ۰/۶۰ | ۰/۱۰ | ۰/۳۰ | ۰/۹۰ |
| کل آزمون | ۰/۱۰ | ۰/۹۰ | ۰/۱۰ | ۰/۴۰ | ۰/۵۰ | ۰/۹۰ |

رابطه بین نمرات خرده آزمونها با یکدیگر و نمرات هر خرده آزمون بامقیاس 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



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

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 پاپوییچ[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 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mos | Cat | Sit | Pat | Mos | Cat | Sit | Pat | Mos | Cat | Sit | Pat |
| موزاییکها (Mos) | - | ۰/۱۰ | ۰/۲۰ | ۰/۷۱ | – | ۰/۸۰ | ۰/۴۰ | ۰/۱۰ | – | ۰/۰۰ | ۰/۰۰ | ۰/۵۱ |
| طبق بندیها(Cat) | | - | ۰/۶۰ | ۰/۱۰ | | – | ۰/۲۱ | ۰/۵۰ | | – | ۰/۱۰ | ۰/۸۰ |
| موقعیتها (Sit) | | | - | ۰/۵۰ | | | – | ۰/۰۰ | | | – | ۰/۳۰ |
| 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 ۰/۵۰، ۰/۸۰ و ۰/۳۰ 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 پاپوییچ 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 می دهد. 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 پاپوییچ، 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 پاپوییچ، 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.



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

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

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 درجدول ۱۵، ارائه شده است. ستون آخر این جدول، همبستگی بین سن با نمره استاندارد را نشان می دهد. محتوای جدول نشان می دهد 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 می یابد. این نتیجه گیری با این واقعیت که تمامی ضرایب همبستگی ارائه شده در ستون سمت چپ جدول به لحاظ آماری در سطح ۱ ۰ معنادار هستند، مورد تایید قرار می گیرد. این یافته ها 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 |
|---|---|---|---|---|
| | ۳-۲ | ۵-۴ | ۷-۶ | |
| موزاییکها | ۷ ۰ | ۱۰ ۴ | ۱۲ ۱ | ۰ ۳ |
| طبقه بندیها | ۷ ۴ | ۱۰ ۴ | ۱۲ ۵ | ۰ ۸ |
| موقعیتها | ۷ ۹ | ۱۰ ۵ | ۱۲ ۴ | ۰ ۳ |
| الگوها | ۷ ۰ | ۱۰ ۹ | ۱۲ ۳ | ۰ ۲ |
| مقیاس عملی | ۸۶ ۰ | ۱۰۱ ۷ | ۱۱۳ ۰ | ۰ ۳ |
| مقیاس استدلال | ۸۴ ۱۴ | ۱۰۲ ۳ | ۱۱۲ ۲ | ۰ ۱ |
| کل آزمون | ۸۵ ۱ | ۱۰۲ ۷ | ۱۱۳ ۴ | ۰ ۵ |



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

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 =خوب ' = متغیر ' = متوسط؛ و = ضعیف، ارزیابی نمایند. علاوه بر این، پژوهشگر نیز بر اساس مشاهده خود در طول اجرای آزمون هر کودک ا به همان شیوه مورد ارزیابی قرار می داد. 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 ا) برخوردار نیستند، از اینرو جهت کنترل تأثیر بی اعتباری از فرمول اصلاح برای کاهش[8] (گیلفورد و فروچتر، ۹۷۸ ) 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 ۶۰ است که یک ضریب همبستگی متوسط (نصفت ۳۷۱ ) محسوب می گردد. علاوه بر این، بغیر از ضریب همبستگی خرده آزمون موقعیتها با میزان همکاری که ۵ ۰ است، کلیه ضرایب همبستگی در سطح ۱ ۰ ۰ به لحاظ آماری معنادار هستند.

متوسط ضرایب همبستگی مربوط به ارزیابیهای مربی ۴ ۰ ا است که یک ضریب همبستگی بسیار ضعیف (نصفت ۳۷۱ ) به حساب می آید. این یافته با انتظار ما و با یافته های حاصل از تحقیقات تلخن و همکاران ۹۹۸ ) 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 دریافته ها 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 می گردد تا مربی فرصت و انرژی کافی جهت شناخت هر چه بهتر کودکان و ارزیابی دقیق آنها نداشته باشد.



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

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

| خرده آزمونها/ مقیاسها/نمره‌کل | ارزیابی مربی | | | | ارزیابی پژوهشگر | | | |
|---|---|---|---|---|---|---|---|---|
| | انگیزش | توجه | همکاری | درک آموزشها | انگیزش | توجه | همکاری | درک دستورالعملها |
| وزاییکها | ۰/۲ | ۰/۴ | ۰/۴ | ۰/۸ | ۰/۴ | ۰/۱۰ | ۰/۲ | ۰/۱۰ |
| طبق بندیها | ۰/۸ | ۰/۵ | ۰/۴ | ۰/۶ | ۰/۹ | ۰/۳ | ۰/۴ | ۰/۷ |
| موقعیتها | ۰/۵ | ۰/۵ | ۰/۲ | ۰/۲ | ۰/۹ | ۰/۸ | ۰/۵ | ۰/۸ |
| الگوها | ۰/۵ | ۰/۱ | ۰/۵ | ۰/۴ | ۰/۷ | ۰/۱ | ۰/۵ | ۰/۸ |
| مقیاس عملی | ۰/۸ | ۰/۳ | ۰/۴ | ۰/۸ | ۰/۴ | ۰/۲ | ۰/۰ | ۰/۱۰ |
| مقیاس استدلال | ۰/۶ | ۰/۵ | ۰/۸ | ۰/۴ | ۰/۴ | ۰/۱۰ | ۰/۰ | ۰/۱۱ |
| کل آزمون | ۰/۱۰ | ۰/۴ | ۰/۷ | ۰/۳ | ۰/۹ | ۰/۶ | ۰/۶ | ۰/۱۲ |

آخرین شاهدی که در مورد روایی مورد بررسی قرار گرفت، قدرت تمایزگذاری[۵۹] ست. این شاخص که به واریانس یا پراکندگی نمرات مربوط است زمانی به حداکثر می رسد 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 آزمون، دلتای فرگسون است که مقدار آن بین ۱ تا ۰ تغییر می کند (فرگسون ۱۹۴۹؛ به نقل از کلاین، ۲۰۰۰). مقدار این شاخص در توزیع های نرمال ۱/۳ است. بطور کلی آزمون رو، آزمونی است که قدرت تمایزگذاری آن بالاتر از ۱/۰ 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**

| خرده‌آزمونها/مقیاسها/ کل آزمون | دختران | پسران | کل گروه |
|---|---|---|---|
| موزاییکها | ۱/۴ | ۱/۲ | ۱/۴ |
| طبق بند ها | ۱/۴ | ۱/۶ | ۱/۳ |
| موقعیتها | ۱/۳ | ۱/۴ | ۱/۴ |
| الگوها | ۱/۴ | ۱/۱ | ۱/۹ |
| مقیاس عملی | ۱/۷ | ۱/۵ | ۱/۷ |
| مقیاس استدلال | ۱/۶ | ۱/۵ | ۱/۶ |
| کل آزمون | ۱/۸ | ۱/۷ | ۱/۸ |


[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

**شناخت تصاویر**

همانطور که قبلاً نیز گفته شد، خرده آزمونهای طبقه بندها و موقعیتها از تصاویر معنادار تشکیل یافته اند. از این رو احتمال تورش فرهنگی[۶۰] در این نوع خرده آزمونها بیشتر از خرده آزمونهایی است که از تصاویر بی معنا، مانند اشکال هندسی، تشکیل یافته اند (جنسن[۱] ۹۸۰ ). 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 طبقه بندیها و موقعیتها 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 نمی شناسد در دادن پاسخ درست به سؤال دچار شکست گردد. جدول ۱۸ ۸ تا از سوالات خرده آزمون طبقه بندها را نشان می دهد که شامل تصاویری هستند که برای حداقل ۵ ۰ از کودکان مورد مطالعه، که به سؤال پاسخ نادرست دادند، ناآشنا هستند. ستون اول، شماره سوال و شماره تصویر را نشان می دهد. در ستون دوم فراوانی پاسخهای نادرست به سؤال ارائه شده است و در ستون سوم نیز درصدی از این کودکان که تصویر را نمی شناختند نشان داده شده است.

بر طبق نتایج جدول ۱۸، تصاویر مورد استفاده در سوالات ۹، ۱۰، ۱۱، ۱۲، ۱۳، ۱۴ و ۱۵ خرده آزمون طبقه بندها، بطور متوسط برای ۰٬۰۱ ۰ 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 می آید تعداد ۸ تا از سؤالات خرده آزمون طبقه بندیها از لحاظ فرهنگی دارای تورش هستند و لازم است که تصاویر آشنا، جایگزین آنها گردد.



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

**جدول ۱۸- سؤالات خرده آزمون طبقه‌بندیها، شامل تصاویری است که برای حداقل۰/۰۵ از کودکان ناآشنا هستند**

| سوال | فراوانی پاسخهای نادرست | فراوانی پاسخهای نمی شناسم | درصد پاسخهای نمی شناسم |
|---|---|---|---|
| A (تصویر شمار ۱ اصلی) | ۲۷ | ۸ | ٪ ۰ |
| A (تصویر شمار ۵ جایگزین) | ۲۷ | ۱۲ | ٪ ۴ |
| ۸ (تصویر شمار ۱ جایگزین) | ۵۲ | ۳۴ | ۰ ٫۰۵ |
| ۸ (تصویر شمار ۵ جایگزین) | ۵۲ | ۲۲ | ٪ ۲ |
| ۹ (تصویر شمار ۱ جایگزین) | ۷۵ | ۴ | ۰ ۵ |
| ۹ (تصویر شمار ۲ جایگزین) | ۷۵ | ۴ | ۰ ۵ |
| ۱۰ (تصویر شمار ۱ اصلی) | ۹۵ | ۲۰ | ۰ ٫۰۱ |
| ۱۰ (تصو ر شمار ۳ اصلی) | ۹۵ | ۲۲ | ٪ ۳ |
| ۱۲ (تصویر شمار ۲ اصلی) | ۱۱۵ | ۷ | ٪ |
| ۱۲ (تصویر شمار ۳ اصلی) | ۱۱۵ | ۷ | ٪ |
| ۳ (تصویر شمار ۳ اصلی) | ۱۱۵ | ۱۲ | ٪ ۰ |
| ۱۳ (تصویر شمار ۵ جایگزین) | ۱۱۵ | ۷ | ٪ |
| ۱۴ (تصویر شمار ۲ اصلی) | ۱۱۸ | ۲۲ | ٪ ۹ |
| ۱۵ (تصویر شمار ۱ جایگزین) | ۱۲۱ | ۲۵ | ۰ ٫۰۱ |
| ۵ (تصویر شه ر ۳ جایگزین) | ۱۲۱ | ۲۰ | ٪ ۷ |

## بحث و نتیجه‌گیری

در این پژوهش، ویژگیهای روان سنجی و عملی بودن آزمون SON-R 2 ½ -7 برای کودکان ایرانی مورد بررسی قرار گرفت. نتایج این مطالعه (جدول ۰ ) نشان می دهد که نمرات خرده آزمونها، مقیاسها و کل آزمون ز همسانی درونی بالایی برخوردارند. این یافته که با یافته های حـاصل از مطالعات تلخن و همکاران ۱۹۹۸، و تلخن و لاروس ۲۰۰۴ ) همخوانی دارد بیانگر این است که اعتبار نمرات آزمون به انداز ای است که بتوان با اطمینان از آنها استفاده کرد.

بر اساس یافته های مربوط به تمایزگذاری سنی (جدول ۵ ) همبستگی سوال - نمره کل(جدول ۶ )، همبستگی درونی خرده آزمونها (جدول ۳ ) همبستگی خرده آزمونها با



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

مقیاسها و نمره کل (جدول ۴ ) و همچنین همبستگی بین نمرات کودکان در خرده آزمونها، مقیاسها و کل آزمون با ارزشیابیهای پژوهشگر و مربی از میزان انگیزش، درک دستورالعملها و آموزش ها، همکاری و توجه آنها (جدول ۶) می توان چنین نتیجه گرفت که این آزمون یک ابزار روا جهت سنجش هوش کلی است و می توان با اطمینان خاطر از آن استفاده کرد. شاهد دیگری که در خصوص روایی نمرات آزمون مورد بررسی قرار گرفت، قدرت تمایزگذاری (جدول ۷ ) 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 براین، یافته های حاصل از مطالعات بین فرهنگی که درخصوص روایی این آزمون در کشورهای آمریکا، استرالیا، و انگلستان صورت گرفته است (لاروس و تلخن ۹۹۹ ) حاکی از این است که همبستگی این آزمون با مقیاسهای عملی تعدادی از آزمونها، مانند آزمون WPPSI-R ، [۶۲] BAS و [۶۳]MSCA، قویتر از همبستگی آن با مقیاسهای کلامی این آزمونها است. این همبستگی ها از روایی واگرا و همگرای آزمون مورد مطالعه حمایت می کند.

تحلیلهایی که در خصوص شاخص های آماری سؤالات (درجه دشواری و قدرت تشخیص) با استفاده از مدل کلاسیک اعتبار (جدول ۱ ) و مدل ۲ و ۳ پارامتری مبتنی بر IRT صورت گرفت (جدول ۱ ) نشان داد که کلیه سؤالات از شاخصهای آماری مطلوبی برخوردارند. بر وردهای مربوط به پارامتر b در مدلهای IRT نشان می دهد که سؤالات از لحاظ دشواری در دامنه ۲ – تا ۲ + بطور تقریباً یکنواخت توزیع شد اند. علاوه براین بغیر از تعداد اندکی از سوالات، با بالا رفتن ترتیب سؤالات، دشواری آنها نیز افزایش می یابد. برآور های مربوط به پارامتر a نشان می دهد که سؤالات از قدرت تشخیص مطلوبی برخوردارند.

بطورخلاصه می توان گفت که آزمون SON-R 2 ½ -7 با در نظر گرفتن یافته های مربوط به اعتبار، روایی، مدت اجرا، سهولت نمره گذاری و تفسیر که از مهمترین جنبه های عملی بودن آزمون به حساب می آید (هومن ۳۸۱ ) یک ابزار کاملاً مناسب برای سنجش هوش کلی کودکان ۷ – ۲ سال ایران است و از آن می توان بعنوان وسیله ای معتبر و روا برای : الف) شناسایی کودکانی که بطور قابل ملاحظه ای از لحاظ شناختی، ضعیف تر از همسالان خود هستند، ب) مستند کردن پیشرفت کارکرد



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

شناختی کودکان در نتیجه برنامه های مداخله ای ویژه، ج) مشخص کردن نقاط قوت و ضعف کودکان در مهارتهای شناختی، و د) انداز گیری 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 سوالات A، ۸، ۹، ۱۰، ۱۲، ۱۳، ۱۴ و ۱۵ خرده آزمون طبق بند ها با تصاویری که متناسب با فرهنگ ایران و آشنا برای کودکان ایرانی هستند، جایگزین گردند. علاوه برایز، ترتیب سوالات نیز باید بر اساس دشواری آنها اصلاح گردد. به نظر می رسد که با انجام این اصلاحات، می توان 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) Cattel, 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



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

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

ثرندایک، آر. ال. ۳۷۵ ). روان سنجی کاربردی. ترجمه حیدرعلی هومن. تهران: انتشارات دانشگاه تهران (تاریخ انتشار به زبان اصلی ۹۸۲ )

نصفت، مرتضی ۳۷۱ ). اصول ورشهای آه ر. جلد اول. تهران: انتشارات دانشگاه تهران

هومز، حیدرعلی ۱۳۸۱). انداز گیریهای روانی و تربیتی: فن تهیه تست و پرسشنامه. تهران نشر پارسا.



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

Achenbach, T.M; & Rescorla, L.A. (2001). Mannual 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 clearinghous 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.



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

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.testreseart. Nl/sonroe/crssculte. html]. Accessed 15 Agust 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).Intellinentieonderzoekvan doofstomme kinderen. Nigmegen: Berkhout.

Stevens, J.P. (2002). Applied multivariate statistics for the social sciences. (4th ed.) Mahwah, NJ: lawrence Erlbaum.



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]

Tellegen, P. J., Winkel, M; Wijnbeg-Williams, B; & Laros, J. A (1998). Snijders-Oomen nonverbal intelligence test. SON-R 2 ½ -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,methodolgical and practical Issues. Amsterdam: Swets & Zeitliner.

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 practiecal 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.



[ DOR: 20.1001.1.16826612.1384.5.3.4.4 ]