

پژوهش در حیطه کودکان استثنایی / ۱۶ سال پنجم، شماره ۲/تابستان ۱۳۸۴/۱۳۴-۱۱۵
Research on Exceptional Children, Summer 2005, Vol. 16, No.2, 115-134

استانداردسازی آزمون هوشی ویلیامز برای افراد دچار آسیب بینایی*

اصغر مینایی**، محمد احمد پناه**، دکتر حیدرعلی هومن***

(دریافت: ۸۴/۱/۲۸ تجدید نظر: ۸۴/۳/۳۰ پذیرش نهایی: ۸۴/۶/۷)

چکیده

این پژوهش با هدف انطباق و استاندارد سازی آزمون هوشی ویلیامز برای افراد دچار آسیب بینایی بر روی نمونه‌ای با حجم ۴۳۹ دانش آموز (۲۲۱ پسر و ۲۱۸ دختر) شهرهای تهران و مشهد از ۵؛۷ (پنج سال و هفت ماه) تا ۱۵؛۶ (پانزده سال و شش ماه)، اجرا گردید. این آزمون مشتمل بر آزمون واژگان (رهنمون)، و آزمون اصلی است. همسانی درونی آزمون اصلی و رهنمون با استفاده از فرمول ۲۰ کودر- ریچاردسون، به تفکیک گروههای سنی، برآورد گردید. دامنه ضرایب همسانی درونی آزمون اصلی ۰/۹۴ تا ۰/۹۷ (میانگین ۰/۹۶) و آزمون رهنمون ۰/۸۷ تا ۰/۹۴ (میانگین ۰/۹۱) بود. ثبات زمانی آزمونها با استفاده از روش آزمون- باز آزمون مورد مطالعه قرار گرفت. برای این منظور از بین کل نمونه هنجاری، تعداد ۷۰ نفر به طور تصادفی انتخاب و مجدداً توسط همان آزمونگران، آزمون شدند. فاصله بین دو اجرای آزمون ۳ تا ۴ ماه بود. ضریب همبستگی بین دو بار اجرای آزمون اصلی ۰/۸۸ و آزمون رهنمون ۰/۸۷ بود. اندازه ضرایب نشان دهنده این است که آزمونها از ثبات زمانی قابل قبولی برخوردارند و خطای نمونه‌برداری زمانی آنها اندک است. روایی محتوایی، ملاکی و سازه آزمون اصلی با استفاده از روشهای مناسب مورد بررسی قرار گرفت. نتایج این تحلیلها حاکی است که این آزمون یک ابزار روا در سنجش هوش کلی افراد نابینا و نیمه بینا است. جداول هنجاری برای نمرات خام آزمون اصلی و رهنمون در ۲۰ گروه سنی به فاصله ۶ ماه از یکدیگر (از ۵؛۷ تا ۱۵؛۶) تهیه گردید. علاوه بر این، معادلهای سنی متناظر با نمرات خام آزمون اصلی محاسبه گردید. به طور خلاصه نتایج حاصل از بررسی ویژگیهای روان سنجی آزمون حاکی از معتبر و روا بودن استنباطهای حاصل از نمرات آن است و کاربران می‌توانند با اطمینان خاطر از آن استفاده کنند.

واژگان کلیدی: استانداردسازی؛ هوش کلی؛ آسیب بینایی؛ اعتبار؛ روایی

* این پژوهش در پژوهشکده کودکان استثنایی انجام شده است
** عضو هیات علمی پژوهشکده کودکان استثنایی Email: a_minai@riec.ac.ir
*** استاد دانشگاه

مقدمه

هوش کلمه‌ای است که نه تنها معنای روان‌شناختی دارد بلکه وارد واژگان عامیانه و غیر حرفه‌ای جامعه نیز شده است. به نظر می‌رسد که هر کس سعی می‌کند بداند که هوش چیست؟ علی‌رغم تقریباً یک قرن مشاهده، تعریف و نظریه‌پردازی 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 کنیم. وجود یا فقدان و همچنین میزان آن را بایستی بطور غیرمستقیم استنباط کرد که حتی انجام این کار نیز چندان سهل و آسان نیست (براون، شربنو و جانسن^۱، ۱۹۹۷).

جانسن^۲ (۱۹۶۹) اظهار داشته که اندازه‌گیری هوش همانند اندازه‌گیری الکتريسته، آسان‌تر از تعریف آن است (ص ۵). اگر چه در ابتدا ممکن است این گفته جانسن، متناقض به نظر آید اما روان‌شناسان بدون اینکه تعریف کاملاً رضایت بخشی از سازه هوش ارائه دهند تعداد زیادی آزمون معتبر و روا برای اندازه‌گیری هوش کلی تهیه کرده‌اند. ویرایش نهم *سالنامه اندازه‌گیریهای روانی*^۳ (۱۹۸۵)، ۷۷ آزمون هوش را مورد بررسی قرار داده است. ۶۱ آزمون برای اندازه‌گیری هوش کلی طراحی شده‌اند و ۱۶ آزمون دیگر، تواناییهای ذهنی خاص را اندازه می‌گیرند. ۳۲ آزمونی که هوش کلی را اندازه می‌گیرند به صورت گروهی اجرا و توسط ماشین یا منشی، نمره گذاری می‌شوند و ۲۹ آزمون دیگر بایستی به صورت انفرادی و توسط آزمونگر آموزش دیده اجرا شود. (به نقل از مورفی و دیویدشوفر^۴، ۱۹۹۱).

علی‌رغم انتقادات زیاد، آزمونهای هوش ابزارهای سودمند و با ارزشی هستند. اگر آزمون از خصوصیات فنی مطلوب برخوردار باشد و اگر تفسیر نتایج بدست آمده توسط یک فرد متخصص و آگاه صورت گیرد، این آزمونها می‌توانند به نحو سودمندی در برآورد تواناییهای افراد جهت پیش‌بینی عملکرد در فعالیتهای خاص و کمک به فرایند تشخیص مورد استفاده قرار گیرند. متأسفانه، بسیاری از آزمونهای هوش برای سنجش هوش بخشهایی از افراد جامعه مناسب نیستند. برای نمونه، اکثر آزمونهای هوش از محرکات دیداری، مانند کارتهای تصویری، طراحی با مکعبها و غیره، استفاده می‌کنند. بدیهی

استاندارد سازی آزمون هوشی ویلیامز برای.../۱۱۷

است که این آزمونها برای سنجش هوش در افراد نابینا و افراد مبتلا به اختلالات دیداری مناسب نیستند.

آزمون افراد نابینا به دلیل ویژگیهایی که دارند همواره با مشکلاتی روبرو بوده است. آزمونهای شفاهی به سهولت قابل انطباق برای این گروه هستند، در حالی که آزمونهای عملی قابل استفاده نیستند. علاوه بر ارائه شفاهی آزمون توسط آزماینده از روشهای دیگری از قبیل ارائه آزمون با صفحه گرامافون یا نوار صوتی نیز استفاده شده است. بعضی از آزمونها نیز با خط بریل تهیه شده‌اند. کاربرد این روش تا حدودی محدود خواهد بود چون موادی که با خط بریل چاپ می‌شوند در مقایسه با مواد چاپی عادی حجیم هستند، خواندن مطالبی که با خط بریل چاپ شده‌اند کندتر صورت می‌گیرد، و تعداد کثیری از افراد نابینا مهارت کافی در خواندن خط بریل ندارند. (آناستازی، ۱۹۷۶، ترجمه براهنی، ۱۳۶۴).

به منظور سنجش هوش افراد نابینا، هم آزمونهای ویژه و خاص تهیه شده، مانند مقیاس هوشی لمسی ویژه بزرگسالان نابینا^۴ و هم برخی از آزمونهای هوش کلی مورد بازسازی قرار گرفته است، مانند آزمونهای بینه و وکسلر. تعدادی از آزمونهای گروهی هوش نیز با تغییراتی برای استفاده در افراد مبتلا به مشکلات دیداری آماده شده‌اند و فرمهایی از آنها با خط درشت یا خط بریل در دست است که از آن جمله می‌توان به آزمونهای توانش مدرسه و دانشگاه^۵ (SCAT)، آزمون استعداد تحصیلی شورای امتحانات ورودی دانشگاهها^۶ (CBSAT)، آزمون ورودی دوره‌های فوق لیسانس^۸ (GRE) اشاره کرد. (همان منبع).

یکی از آزمونهایی که برای سنجش هوش کلی افراد نابینا و نیمه بینا پرورش یافته است، آزمون هوشی ویلیامز (۱۹۵۶) است. این آزمون توسط به‌پژوه و پرند* (۱۳۸۰) بر روی ۱۳۴ دانش‌آموز نابینا و نیمه بینای ۱۱/۵ تا ۱۴/۵ سال شهر تهران اجرا و استاندارد گردید. آنها اعتبار آزمون را به روش دو نیمه کردن و با استفاده از فرمول اسپیرمن - براون به تفکیک گروه سنی بدست آوردند. دامنه ضرایب اعتبار در پژوهش آنها از ۰/۶۴ تا ۰/۸۷ است. آنها همچنین رابطه بهره هوشی با آخرین معدل پیشرفت تحصیلی را به

* در اینجا از این بزرگواران که نسخه اصلی و ترجمه شده آزمون را در اختیار ما قرار دادند و در طول انجام پژوهش نیز همکاری صمیمانه‌ای داشتند کمال تشکر و سپاس را داریم.

عنوان معیاری برای روایی آزمون مورد بررسی قرار دادند. آنها دریافتند که همبستگی بین بهره هوشی و آخرین معدل پیشرفت تحصیلی، ۰/۷۳ است که در سطح ۰/۰۱ معنادار است.

با توجه به اهمیت سنجش هوش کلی افراد مبتلا به آسیبهای بینایی و با عنایت به فقدان آزمون استاندارد به زبان فارسی، و با توجه به اینکه به پژوه و پرند (۱۳۸۰) این آزمون را صرفاً برای کودکان ۱۱/۵ تا ۱۴/۵ سال هنجاریابی کرده‌اند لذا آزمون هوشی ویلیامز جهت استانداردسازی و هنجاریابی انتخاب گردید.

روش

جامعه آماری و نمونه مورد مطالعه

پژوهش حاضر در زمره پژوهشهای توصیفی از نوع پیمایشی است. جامعه آماری شامل کلیه دانش‌آموزان نابینا و نیمه بینای ۵:۰ تا ۱۵:۶، است که در سال تحصیلی ۱۳۸۱-۸۲ در مدارس دخترانه و پسرانه ویژه دانش‌آموزان نابینا وابسته به سازمان آموزش و پرورش استثنایی و مدارس عادی (مدارس تلفیقی) استان تهران مشغول به تحصیل بودند. به علت اندک بودن تعداد دانش‌آموزان در مقاطع سنی ۵ تا ۶ سال، تعداد ۴۰ دانش‌آموز ۵ تا ۶ سال از مدرسه تقوا و امید شهر مشهد نیز مورد آزمون قرار گرفتند. ویژگیهای جمعیت شناختی گروه مورد مطالعه در جدول ۱، ارائه گردیده است.

جدول ۱، ویژگیهای جمعیت شناختی گروه هنجاری (N=۴۳۹)

| ویژگیها | فراوانی | درصد |
|--------------|---------|------|
| جنسیت | | |
| پسر | ۲۲۱ | ۵۱ |
| دختر | ۲۱۸ | ۴۹ |
| وضعیت بینایی | | |
| نابینا | ۲۳۱ | ۵۳ |
| نیمه‌بینا | ۲۰۸ | ۴۷ |
| سن | | |
| ۶ | ۲۲ | ۵ |
| ۷ | ۳۰ | ۷ |
| ۸ | ۵۲ | ۱۲ |
| ۹ | ۴۴ | ۱۰ |
| ۱۰ | ۳۸ | ۸ |
| ۱۱ | ۴۵ | ۱۰ |
| ۱۲ | ۴۹ | ۱۱ |
| ۱۳ | ۵۶ | ۱۳ |
| ۱۴ | ۵۱ | ۱۲ |
| ۱۵ | ۵۲ | ۱۲ |

ابزار

ابزار مورد استفاده که هدف پژوهش، استانداردسازی آن است آزمون هوشی ویلیامز برای افراد با آسیبهای بینایی^۹، است. این آزمون نخستین بار در انگلستان و در سال ۱۹۵۶ توسط مای فانوی ویلیامز^{۱۰} برای افراد نابینا و نیمه بینای ۳/۵ تا ۱۶ سال تهیه گردید (ویلیامز، ۱۹۵۶، ص ۷). آزمون مورد نظر دارای یک فرم و ۱۰۰ سؤال است و چهار حوزه از تواناییهای شناختی شامل: الف- استدلال کلامی^{۱۱}، ب- استدلال کمی^{۱۲}، ج- استدلال انتزاعی^{۱۳} و د- حافظه کوتاه مدت^{۱۴} را می‌سنجد (نیولند^{۱۵}، ۱۹۶۵).

این آزمون به صورت شفاهی و انفرادی ارائه می‌گردد و نیازی به خواندن و نوشتن نیست. سؤالات این آزمون عمدتاً از نسخه ۱۹۳۷ آزمون استنفرد- بینه (فرم M و L) اقتباس گردیده است. تعدادی از سؤالات نیز از آزمونهای هوشی والنتین برای کودکان^{۱۶}، آزمونهای استدلال برت^{۱۷} و برخی از آزمونهای گروهی هوش اقتباس شده است. آزمون ویلیامز، یک خرده آزمون واژگان نیز دارد. این خرده آزمون که دارای ۴۰ واژه است و نقش آزمون رهنمون را ایفاء می‌کند، از مقیاس وکسلر برای کودکان^{۱۸} استخراج شده است. نتیجه این خرده آزمون همراه با سن تقویمی فرد، جهت تعیین نقطه شروع آزمون اصلی بکار می‌رود.

شیوه اجرا و گردآوری داده‌ها

در این بخش شیوه اجرا و گردآوری داده‌ها توضیح داده می‌شود.

الف- تطبیق متن ترجمه با متن اصلی و انجام تغییرات لازم: همانطور که گفته شد این آزمون توسط به‌پژوه و پرند در سال ۱۳۸۰ به فارسی ترجمه گردید. به دلیل برخی نواقص و ضعفهای موجود، متن ترجمه با متن اصلی بطور دقیق و کامل تطبیق داده شد و اشکالات موجود رفع گردید.

فرم اولیه آزمون بر روی ۱۰۰ نفر (۵۰ دختر و ۵۰ پسر) در گروه سنی ۵؛۰ تا ۱۵؛۶ اجرا گردید. این نمونه به صورت تصادفی از بین دانش‌آموزان مجتمع‌های آموزشی شهید محبی، نرجس و خزائلی شهر تهران که ویژه نابینایان و نیمه بینایان است انتخاب گردید. اجرای این مرحله، پس از آموزش آزمونگران پژوهش که حداقل دارای مدرک لیسانس روان‌شناسی بودند انجام گرفت. پس از استخراج نتایج، تجزیه و تحلیل داده‌ها

براساس مدل کلاسیک اعتبار صورت گرفت. با استفاده از این مدل برای هر سؤال، ۲ پارامتر درجه دشواری^{۱۹} و قدرت تشخیص^{۲۰} برآورد گردید. نتایج حاصل از این امر نشان داد که کلیه سؤالات آزمون از ویژگیهای روان‌سنجی مطلوبی برخوردارند و نیاز به حذف و جایگزینی ندارند. سؤالات آزمون براساس درجه دشواری مرتب گردید. به این صورت فرم نهایی آزمون شکل گرفت.

ب- اجرای نهایی: برای جمع‌آوری اطلاعات جهت تشکیل جداول هنجاری، کلیه دانش‌آموزان دامنه سنی ۵;۰ (۵ سال تمام) تا ۱۵;۶ (پانزده سال و شش‌ماه) مجتمع‌های نابینایان شهید محبی، نجس، خزائی، حضرت عبدالعظیم (شهرری) امام خمینی (ورامین)، و پویا (کرج) مورد آزمون قرار گرفتند. افزون بر این، تعدادی از دانش‌آموزان نابینا و نیمه بینا که در مدارس تلفیقی شهر تهران مشغول به تحصیل بودند در نمونه مورد مطالعه گنجانده شدند. با عنایت به این که تعداد آزمودنیهای سنین ۵ و ۶ سال بسیار اندک بود، تعدادی از کودکان نابینا و نیمه بینای آموزشگاههای امید و تقوا در مشهد نیز مورد آزمون قرار گرفتند. نهایتاً آزمون بر روی ۳۴۹ نفر اجرا گردید. با توجه به این که فاصله زمانی بین دو مرحله اجرای تجربی و نهایی زیاد نبود، و چهارچوب نمونه‌گیری نیز در هر دو مرحله یکسان بود و علاوه بر این تغییری در محتوای آزمون ایجاد نشده بود، بنابراین جهت تشکیل جداول هنجاری و انجام سایر تحلیل‌های روان‌سنجی داده‌های حاصل از مرحله تجربی و مرحله نهایی ادغام گردید. در ادامه، نتایج حاصل از تحلیل‌ها ارائه می‌گردد.

نتایج

۱- اعتبار نتایج آزمون

مفهوم اعتبار^{۲۱} آزمون به این معناست که تا چه اندازه تفاوت‌های فردی در نمرات آزمون به تفاوت‌های واقعی در خصوصیات مورد مطالعه و تا چه اندازه به خطاهای تصادفی قابل استناد است. (آناستازی^{۲۲}، ۱۹۹۰، ص ۱۰۹).

از آن جهت که هدف آزمون، اندازه‌گیری ویژگی، توانایی و یا محتوای خاصی است. لذا هر چقدر سؤالات آزمون ارتباط بیشتری با یکدیگر داشته باشند خطای آزمون کمتر خواهد بود. (نیوکامر و هامیل^{۲۳}، ۱۹۹۷، ص ۵۵). همسانی درونی^{۲۴} آزمون با استفاده از

فرمول ۲۰ کودر- ریچاردسون در ۱۱ گروه سنی با استفاده از داده‌های حاصل از کل گروه هنجاری برآورد گردید. ضریب همسانی درونی برای کل گروه نمونه از طریق میانگین‌گیری ضرایب همسانی درونی تمام گروه‌های سنی با استفاده از روش تبدیل Z فیشر^{۲۵} بدست آمد. نتایج در جدول ۲ به تفکیک آزمون اصلی و واژگان گزارش شده است.

جدول ۲. ضرایب همسانی درونی آزمون اصلی و واژگان به تفکیک گروه‌های سنی (اعشار حذف شده است)

| میانگین ضرایب | سن به سال | | | | | | | | | | |
|------------------|-----------|----|----|----|----|----|----|----|----|----|-------|
| | ۱۵ | ۱۴ | ۱۳ | ۱۲ | ۱۱ | ۱۰ | ۹ | ۸ | ۷ | ۶ | آزمون |
| اصلی | ۹۶ | ۹۸ | ۹۶ | ۹۷ | ۹۷ | ۹۸ | ۹۸ | ۹۶ | ۹۶ | ۹۵ | ۹۴ |
| واژگان | ۹۱ | ۹۳ | ۹۳ | ۹۲ | ۹۳ | ۹۴ | ۹۲ | ۹۱ | ۸۷ | ۹۱ | ۸۷ |

ثبات زمانی آزمون با استفاده از روش آزمون- بازآزمون مورد مطالعه قرار گرفت. برای این منظور از بین کل نمونه هنجاری، تعداد ۷۰ نفر به طور تصادفی انتخاب و مجدداً توسط همان آزمونگران، آزمون شدند. فاصله بین دو اجرای آزمون ۳ تا ۴ ماه بود. نتایج حاصل در جدول ۳، ارائه گردیده است. در این جدول، میانگین و انحراف استاندارد هر بار اجرای آزمون و همچنین ضرایب همبستگی (r) که نشان دهنده میزان رابطه بین دو بار اجرای آزمون است ارائه گردیده است. اندازه ضرایب نشان دهنده این است که آزمون از ثبات زمانی قابل قبولی برخوردار است و خطای نمونه‌برداری زمانی آن اندک است.

جدول ۳. نتایج اعتبار بازآزمایی

| آزمون | اجرای اول | | اجرای دوم | | r |
|--------|-----------|-------|-----------|-------|------|
| | SD | M | SD | M | |
| اصلی | ۲۰/۲۵ | ۵۲/۸ | ۲۱/۷۰ | ۵۲/۸ | ۰/۸۸ |
| واژگان | ۹/۹۴ | ۲۰/۱۰ | ۹/۳۲ | ۲۰/۱۰ | ۰/۸۷ |

۲- روایی نتایج آزمون

مقصود از روایی^{۲۶} آزمون پاسخ دادن به این سؤال است که آزمون چه چیزی را اندازه‌گیری می‌کند و تا چه اندازه از این لحاظ کارآیی دارد. اغلب متخصصینی که با

اندازه‌گیری آموزشی و روان‌شناختی سروکار دارند توصیه می‌کنند افرادی که دست‌اندرکار ساخت آزمونها هستند بایستی شواهدی در زمینه سه نوع روایی ارائه دهند: روایی محتوا^{۲۷}، روایی ملاک مرجع^{۲۸} و روایی سازه^{۲۹} (آیکن^{۳۰} ۱۹۹۴؛ آناستازی، ۱۹۹۰؛ سالویا و یسلدایک^{۳۱}، ۱۹۹۵؛ به نقل از براون، شربنو و جانسن، ۱۹۹۷، ص ۸۷).
روایی محتوا: روایی محتوا بطور کلی متضمن بررسی منظم محتوای آزمون به منظور پاسخ دادن به این سؤال است که آیا آزمون نمونه‌های معرفی از حوزه رفتاری مورد اندازه‌گیری را در بر می‌گیرد یا نه. (آناستازی، ۱۹۹۰، ص ۱۴۰).

در خصوص روایی محتوایی آزمون مورد بحث دو نوع شاهد مورد بررسی قرار گرفت. اولاً مبنای منطقی انتخاب سؤالات و قالب آزمون بررسی گردید. ثانیاً شواهد کمی روایی محتوایی سؤالات که شامل نتایج حاصل از روشهای کلاسیک تحلیل سؤال است مورد بررسی قرار گرفت.

الف - مبنای منطقی انتخاب سؤالات و قالب آزمون: رواسازی محتوا به چند شیوه انجام می‌گیرد. متداولترین این روشها عبارتند از: ساخت یا پرورش سؤالات براساس مطالب مواد درسی و راهنماهای آموزشی، بررسی ادبیات پژوهش، بررسی محتوا و قالب ابزارهایی که هدفی مشابه با هدف آزمون دارند، و مشورت با متخصصان.

در نسخه انگلیسی آزمون مورد بحث، بعد کمی روایی محتوا مورد بحث قرار نگرفته اما شواهد کیفی آن ارائه شده است. ویلیامز (۱۹۵۶) اظهار می‌دارد، "به منظور جلوگیری از اتلاف وقت ناشی از اجرای آزمایشی سؤالات، تصمیم گرفتیم از سؤالات و مواد آزمونهایی که قبلاً بر روی گروههای بزرگی از کودکان بینا استاندارد شده بودند استفاده کنیم. بنابراین چندین آزمون هوش را مطالعه کردیم و سؤالاتی را که برای کودکان سنین ۳/۵ تا ۱۶ سال مناسب بودند، انتخاب نمودیم. سعی کردیم تا اکثراً سؤالات کلامی انتخاب شوند زیرا تحقیقات پیشین نشان داده بود که سؤالات عملی از تورش بسیار زیادی به نفع کودکان بینا برخوردارند." (ص ۷).

ویلیامز (۱۹۵۶) می‌نویسد، "سؤالاتی که در آزمون گنجانده شدند شامل تعداد قابل توجهی سؤال از نسخه تجدید نظر شده ۱۹۳۷ ترمن-مریل، عمدتاً از فرم M، تعدادی از سؤالات آزمون هوشی والنتین برای کودکان و برخی از سؤالات آزمونهای استدلال برت، تعداد کمی سؤال از آزمونهای گروهی هوش و خرده آزمون واژگان مقیاس وکسلر

برای کودکان بودند. در سؤالاتی که از آزمونهای گروهی اقتباس شدند، تغییراتی ایجاد گردید. (ص ۷).

نیولند (۱۹۶۵، ص ۶- ۸۴۵) آزمون ویلیامز را مورد بررسی و ارزشیابی قرار داده است. او می‌نویسد، "در ساخت این آزمون ۱۰۰ سؤالی از نسخه اصلاح شده آزمون هوشی/استنفرد - بینه به عنوان الگو استفاده شده است. حداقل ۶۴ سؤال این آزمون از مقیاس هوشی/استنفرد - بینه اقتباس شده که از این ۶۴ سؤال، ۴۵ سؤال از فرم M و ۷ سؤال از فرم L اقتباس گردیده و ۱۲ سؤال نیز در هر دو فرم مشترک است. در ۱۲ تا از این سؤالات تغییراتی داده شده است تا برای افراد مبتلا به مشکلات بینایی قابل استفاده باشند؛ ۳ تا از سؤالات از لحاظ فرهنگی برای کودکان بریتانیایی انطباق داده شده‌اند؛ و در ۳ تا از سؤالات نیز، برای اینکه برای دامنه وسیع‌تری قابل کاربرد باشند، اصلاحاتی بعمل آمده است. بقیه سؤالات از آزمون هوشی والننتین برای کودکان، آزمونهای استدلال برت و برخی از آزمونهای گروهی هوش اقتباس شده‌اند. آزمون واژگان که شامل ۴۰ واژه است، و صرفاً جهت تعیین نقطه شروع آزمون بکار می‌رود، از مقیاس هوشی وکسلر برای کودکان اقتباس گردیده است... هیچ اطلاعی در خصوص اعتبار و روایی آزمون گزارش نشده است... [اما] براساس آنچه که درباره ارزش آزمون بینه در پیش‌بینی یادگیری آموزشی در کودکان بینا مشخص شده است و با در نظر گرفتن ماهیت فرایندهای روان شناختی و محصولات یادگیری که در این آزمون، نمونه‌گیری شده است، می‌توان فرض کرد که نتایج حاصل از این آزمون از لحاظ نظری باید یک پایه نسبتاً خوب و مناسبی برای پیش‌بینی رفتار یادگیری کودکان مبتلا به مشکلات بینایی فراهم آورد.... بطور کلی این آزمون را می‌توان یک ابزار بالقوه سودمند در نظر گرفت."

علاوه بر مطالبی که از ویلیامز و نیولند در خصوص ساختار محتوایی آزمون گفته شد باید بگوییم که این آزمون به گونه‌ای طراحی شده است که برای یک موقعیت خاص و منحصر به فرد، مناسب باشد. هدف این بوده که آزمونی ساخته شود که تحت تأثیر بینایی قرار نگیرد و بنابراین برای استفاده در افرادی که نابینا و کم بینا هستند مناسب باشد. این آزمون کاملاً نوابسته به بینایی است. این هدفها باعث گردیده تا هم محتوا و هم قالب آزمون محدود گردد. ویلیامز (۱۹۵۶، ص ۷) اظهار می‌دارد که «به علت تفاوتی که در توانایی خواندن و نوشتن به خط بریل در کودکان یک گروه سنی وجود دارد و در

شکل‌گیری این توانایی، عوامل بسیاری به غیر از هوش نقش دارند از این رو معقول به نظر نمی‌رسد که سؤالات آزمون مستلزم دانستن خط بریل باشند. بنابراین لازم بود که سنجش به صورت انفرادی و شفاهی صورت گیرد».

سؤالات آزمون عمدتاً بر استدلال کلامی، استدلال کمی، استدلال انتزاعی و حافظه کوتاه مدت که فرض می‌شود از مؤلفه‌های اصلی هوش کلی هستند، تأکید دارند. شواهد تجربی زیادی وجود دارد که نشان می‌دهند این تواناییها، هم در موقعیتهای واقعی زندگی و هم در موقعیتهای آزمون، هوش کلی را به خوبی پیش‌بینی می‌کنند. در تهیه نسخه فارسی آزمون، جهت اطمینان از روایی محتوا نسخه اولیه آزمون در اختیار تعدادی از روانشناسان، روان‌سنجها و معلمانی که در مدارس نابینایان تدریس می‌کردند قرار گرفت و از آنها خواسته شد تا اصلاحات لازم را از لحاظ فرهنگی بعمل آورند و سؤالاتی را که احتمالاً مبهم هستند، مشخص کنند. متخصصان مذکور تغییرات لازم را در سؤالات بعمل آوردند.

ب- تحلیل کلاسیک سؤالات: شیوه‌های کلاسیک انتخاب سؤال بر قدرت تشخیص و درجه دشواری سؤال تأکید دارند. برای تعیین قدرت تشخیص سؤال که گاهی اوقات روایی سؤال^{۳۲} نیز نامیده می‌شود، روش همبستگی که در آن همبستگی هر سؤال با نمره کل محاسبه می‌شود، مورد استفاده قرار گرفت. گارت (۱۹۶۵)؛ به نقل از نیوکامر و هامیل، ۱۹۹۷، (ص ۷۲) عنوان کرده‌است که سؤالات دارای ضریب همبستگی ۰/۲۰ یا ۰/۳۰ می‌توانند قابل قبول تلقی گردند.

درجه دشواری سؤال، یعنی درصد آزمودنیهایی که به سؤال پاسخ درست داده‌اند، جهت تشخیص سؤالات بسیار آسان یا بسیار دشوار و همچنین به منظور تعیین ترتیب سؤالات در فرم نهایی محاسبه گردید. آناستازی (۱۹۹۰) بیان می‌کند که دشواری متوسط سؤالات باید تقریباً ۰/۵۰ باشد و همچنین باید پراکندگی نسبتاً زیادی در دشواری سؤالات وجود داشته باشد. سؤالات با درجه دشواری بین ۰/۱۵ تا ۰/۸۵ عموماً قابل قبول در نظر گرفته می‌شوند.

برای نشان دادن خصوصیات سؤالات نسخه نهایی، با استفاده از کل نمونه هنجاریابی یک تحلیل سؤال صورت گرفت. میانه ضرایب قدرت تشخیص و درجه دشواری بدست آمده برای سؤالات در جدول ۴، ارائه شده است. نتایج حاکی از آن است که سؤالات این

آزمون، شرایط در پیش توصیف شده را برآورده می‌کنند که این خود دلیلی بر روایی محتوایی آزمون است.

جدول ۴ - میانه درجه دشواری و قدرت تشخیص سؤالات آزمون اصلی به تفکیک گروه‌های سنی (اعشار حذف شده است)

| میانگین ضرایب | سن به سال | | | | | | | | | | آماره |
|------------------|-----------|----|----|----|----|----|----|----|----|----|-------------|
| | ۱۵ | ۱۴ | ۱۳ | ۱۲ | ۱۱ | ۱۰ | ۹ | ۸ | ۷ | ۶ | |
| ۵۳ | ۶۰ | ۶۰ | ۵۸ | ۵۸ | ۵۵ | ۴۸ | ۴۹ | ۴۸ | ۵۱ | ۴۸ | درجه دشواری |
| ۴۳ | ۴۹ | ۴۹ | ۴۷ | ۴۸ | ۴۷ | ۴۲ | ۴۵ | ۴۳ | ۴۰ | ۳۳ | ضرایب تشخیص |

درجه دشواری و قدرت تشخیص سؤالات آزمون واژگان به ترتیب دارای دامنه‌ای از ۰/۱۷ تا ۰/۸۷ با میانه ۰/۵۵ و ۰/۲۷ تا ۰/۷۶ با میانه ۰/۵۴ بود.

روایی وابسته به ملاک: آناستازی (۱۹۹۰، ص ۱۴۵) روایی وابسته به ملاک را به عنوان "کارآمدی آزمون در پیش‌بینی رفتار یک فرد در فعالیتهای خاص" توصیف می‌کند. وی اظهار می‌دارد که عملکرد فرد در آزمون با یک ملاک مقایسه می‌شود که خود نوعی شاخص مستقیم یا غیرمستقیم از آن چیزی است که آزمون جهت پیش‌بینی آن طراحی شده است. بنابراین، اگر آزمونی مانند آزمون ویلیامز که فرض می‌شود هوش کلی را اندازه‌گیری می‌کند دارای روایی باشد بایستی با سایر آزمونهایی که تصور می‌رود همان توانایی را اندازه می‌گیرند، همبستگی خوبی داشته باشد. روایی وابسته به ملاک آزمون مورد نظر به دو صورت مورد بررسی قرار گرفت:

الف - رابطه نمرات آزمون با هوشبهر کلامی و خرده آزمونهای مقیاس تجدید نظر شده هوشی وکسلر برای کودکان (ویسک- آر): جهت بررسی همبستگی این آزمون با خرده آزمونهای ویسک- آر، تعداد ۱۱۵ نفر از افراد نمونه نهایی بطور تصادفی انتخاب و خرده آزمونهای اطلاعات، شباهتها، حافظه، واژه‌ها و ادراک بر روی آنها اجرا گردید. جهت جلوگیری از طولانی شدن جلسه آزمون که ممکن بود باعث خستگی گشته و عملکرد افراد را تحت تأثیر قرار دهد، خرده آزمونهای ویسک- آر، با یک روز فاصله از آزمون اصلی اجرا گردید. مقیاس ویسک- آر، توسط سیما شهیم برای جامعه کودکان ۶ تا ۱۳ ساله شهر شیراز استاندارد شده است. ضرایب اعتبار خرده آزمونهای این مقیاس به غیر از خرده آزمون تطبیق علایم و حساب از ۰/۴۴ تا ۰/۹۴ با میانه ۰/۷۳ گزارش شده است

(شهیم، ۱۳۷۳، ص ۹). در جدول ۵، ضرایب همبستگی بین نمرات ۵ خرده آزمون کلامی و هوشبهر کلامی ویسک-آر با آزمون اصلی و واژگان ارائه شده است. ضرایب داخل پرانتز، بیانگر همبستگی بین نمرات خام آزمون اصلی و واژگان با نمرات خام خرده آزمونها و ضرایب خارج پرانتز، بیانگر همبستگی بین نمرات خام آزمون اصلی و واژگان با نمرات استاندارد خرده آزمونها هستند.

اگرچه ضرایب همبستگی بدست آمده از لحاظ آماری تماماً در سطح $P < 0.01$ معنی دار هستند، که این خود شاهی بر روایی وابسته به ملاک آزمون اصلی و واژگان است، اما با عنایت به اینکه مقیاس وکسلر برای کودکان شیراز استاندارد شده و نه کودکان تهران، لذا تفسیر ضرایب همبستگی باید با احتیاط صورت گیرد.

جدول ۵. ضرایب همبستگی بین آزمون واژگان و اصلی با هوشبهر کلامی و خرده آزمونها کلامی وکسلر برای کودکان (اعشار حذف شده است)

| هوشبهر کلامی | خرده آزمون کلامی وکسلر | | | | | آزمون اصلی |
|--------------|------------------------|--------|--------|---------|---------|------------|
| | حافظه | ادراک | واژگان | شباهتها | اطلاعات | |
| ۶۳ | ۷۸(۵۷) | ۸۰(۵۷) | ۷۶(۵۳) | ۸۲(۴۵) | ۸۳(۴۲) | اصلی |
| ۶۱ | ۶۶(۵۴) | ۷۵(۵۹) | ۶۲(۵۱) | ۶۵(۳۲) | ۷۴(۵۲) | واژگان |

ب- رابطه نمرات آزمون با نمرات دروس مدرسه و معدل ترم: هوش با پیشرفت تحصیلی رابطه مستقیم دارد (سالتر، ۱۹۸۲؛ به نقل از شیما، ۱۳۷۳). هرنشتاین و مورای (۱۹۹۴)؛ به نقل از کول^{۳۳}، (۱۹۹۵) معتقدند که هوش بخش مهمی از پراکندگی موجود در پیشرفت تحصیلی را تبیین می کند. آنها ادعا می کنند که هوشبهر یک ویژگی اساسی است که تعیین می کند چرا عملکرد تحصیلی بعضی از کودکان بهتر از دیگران است.

در پژوهش حاضر از همبستگی بین نمرات دروس و معدل تحصیلی پایان ترم به عنوان معیاری برای پیشرفت تحصیلی و رواسازی آزمون استفاده گردید. نمونه مورد نظر، ۱۲۵ کودک کلاس اول ابتدایی تا سوم دبیرستان بودند که بطور تصادفی از بین کل نمونه هنجاریابی استخراج گردید و پس از پایان ترم تحصیلی، نمرات دروس و معدل ترم از کارنامه تحصیلی دانش آموزان استخراج گردید. فاصله زمانی بین اجرای آزمون و امتحانات دانش آموزان که داده های مورد نیاز را به دست می داد، یک ترم تحصیلی بود. ضرایب همبستگی حاصل در جدول ۶، ارائه شده است.

جدول ۶. همبستگی بین آزمون اصلی و واژگان با نمرات و معدل دروس تحصیلی (اعشار حذف شده است)

| آزمون | | | آزمون | | |
|--------|------|---------------|--------|------|--------------|
| واژگان | اصلی | درس | واژگان | اصلی | درس |
| ۲۱ | ۲۵ | ریاضیات | ۲۱ | ۲۵ | دینی |
| ۲۲ | ۳۰ | علوم | ۳۱ | ۳۵ | انشا |
| ۲۷ | ۳۸ | عربی | ۲۲ | ۲۹ | املا |
| ۳۱ | ۳۶ | حرفه‌وفن | ۲۰ | ۲۲ | خواندن |
| ۳۰ | ۳۷ | گرامر زبان | ۲۱ | ۲۳ | علوم اجتماعی |
| ۳۱ | ۳۸ | املاي انگلیسی | ۲۳ | ۲۹ | جغرافی |
| ۲۵ | ۳۹ | معدل دروس | ۲۲ | ۲۵ | تاریخ |

بررسی جدول فوق نشان می‌دهد گرچه تمام ضرایب همبستگی، از لحاظ آماری (دست کم در سطح $P = 0/05$) معنی‌دار هستند اما ضرایب بدست آمده همگی در سطح متوسط قرار دارند. دو دلیل احتمالی برای این مسأله وجود دارد. اولاً اکثر رفتارها دلایل چندگانه دارند و بنابراین مشکل است که یک متغیر به تنهایی بتواند مقدار زیادی از واریانس [متغیر وابسته] را تبیین کند. (استیونس^{۳۴}، ۲۰۰۲، ص ۱۰). ثانیاً اعتبار نمرات دروس مختلف و در نتیجه اعتبار معدل پیشرفت تحصیلی دانش‌آموزان، پایین و حتی نامشخص است. پایین بودن اعتبار باعث کاهش همبستگی بین متغیرها می‌گردد.

روایی سازه: روایی سازه با این مسأله سر و کار دارد که یک آزمون خاص تا چه اندازه، سازه یا صفت بخصوصی را اندازه می‌گیرد. گرونلند ولین^{۳۵} (۱۹۹۰) برای اثبات این نوع روایی یک روش سه مرحله‌ای را پیشنهاد می‌کنند. ابتدا چند سازه 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 یا تجربی مورد بررسی قرار می‌گیرد. سازه‌های اساسی که فرض می‌شود زیر بنای عملکرد در این آزمون را تشکیل می‌دهند به صورت فرضیه‌هایی به شکل زیر مورد بررسی قرار گرفتند.

- ۱- از آنجا که هوش ماهیتاً وابسته به رشد است لذا عملکرد در این آزمون باید با سن تقویمی قویاً همبستگی داشته باشد. (تمایز گذاری سنی).
- ۲- با توجه به اینکه سؤالات آزمون، ویژگی یکسان و مشابهی را اندازه می‌گیرند لذا سؤالات باید همبستگی بالایی با نمره کل آزمون داشته باشند. (روائی سوال).

تمایزگذاری سنی: در جدول ۷، میانگین و انحراف استاندارد آزمون اصلی و واژگان در ۱۱ گروه سنی و همچنین ضرایب همبستگی بین نمرات آزمون اصلی و واژگان با سن ارائه شده است. محتوای جدول نشان می‌دهد که هر دو آزمون با سن رابطه دارند زیرا که با افزایش سن، میانگین آنها نیز افزایش می‌یابد. این نتیجه‌گیری با این واقعیت که ضرایب همبستگی ارائه شده در سمت چپ جدول به لحاظ آماری معنی‌دار هستند مورد تایید قرار می‌گیرد. این یافته‌ها فرضیه اول ما را درباره رابطه نمرات با سن تایید می‌کند.

جدول ۷، میانگین و انحراف استاندارد آزمونها در پنج گروه سنی و همبستگی آنها با سن

| همبستگی با سن | سن به سال | | | | | | | | | | آزمون |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|-----------------|--------|
| | ۱۵ | ۱۴ | ۱۳ | ۱۲ | ۱۱ | ۱۰ | ۹ | ۸ | ۷ | ۶ | |
| ۰/۷۲ | ۷۳/۶۹ (۲۶/۴۹) | ۷۳/۵۷ (۱۵/۰۵) | ۷۱/۷۸ (۱۷/۶۲) | ۶۳/۶۶ (۱۲/۰۷) | ۶۳/۶۵ (۱۹/۸۵) | ۴۷/۴۲ (۱۶/۳۰) | ۳۹/۶۴ (۱۲/۴۸) | ۳۲/۱۱ (۱۳/۱۹) | ۳۰/۱۰ (۱۰/۷۱) | ۲۱/۸۰ (۹/۲۲) | اصلی |
| ۰/۵۷ | ۲۶/۴۹ (۸/۴۱) | ۲۵/۲۷ (۷/۶۴) | ۲۳/۶۵ (۸/۸۸) | ۲۱/۴۱ (۷/۷۶) | ۱۹/۹۲ (۸/۵۲) | ۱۶/۱۶ (۷/۱۸) | ۱۴/۹۷ (۷/۴۴) | ۱۴/۴۵ (۵/۴۷) | ۱۳/۸۰ (۵/۹۸) | ۱۳/۰۰ (۶/۶۳) | واژگان |

*اعداد بدون پرانتز، میانگین و اعداد داخل پرانتز، انحراف استاندارد است.

روایی سؤال: گیلفورد و فروچتر^{۳۶} (۱۹۷۸) اظهار می‌دارند که از طریق همبسته نمودن عملکرد افراد در سؤالات با نمره کل آزمون می‌توان اطلاعاتی در زمینه روایی سازه بدست آورد. ضرایب همبستگی (ضرایب قدرت تشخیص) که در جدول ۴ ارائه شده است قویاً دلالت بر روایی سازه دارد.

قدرت تمایزگذاری: آخرین شاهد بر روایی که مورد بررسی قرار دادیم قدرت تمایزگذاری^{۳۷} 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 بدست می‌آید که تمام افراد نمره یکسان دریافت کنند. شاخص قدرت تمایزگذاری آزمون، دلتای فرگسون است که مقدار آن بین ۰ تا ۱ است (فرگسون، ۱۹۴۹؛ به نقل از کلاین^{۳۸}، ۲۰۰۰). مقدار این شاخص در توزیعهای نرمال ۰/۹۳ است. بطور کلی آزمون خوب، آزمونی است که قدرت تمایزگذاری آن بالاتر از ۰/۹۰ باشد. مقدار این شاخص برای هر دو آزمون، اصلی و واژگان، ۰/۹۹ است.

تبدل نمرات خام به نمرات استاندارد

جداول هنجاری برای نمرات خام آزمون اصلی و واژگان در ۲۰ گروه سنی به فاصله ۶ ماه از یکدیگر (از ۵;۷ تا ۱۵;۶) تهیه گردید.

جهت بدست آوردن نمرات استاندارد، ابتدا توزیع فراوانی نمرات خام شکل گرفت. آنگاه فراوانی تراکمی و رتبه‌های درصدی متناظر با نمرات خام محاسبه گردید. در گام بعد، مقادیر Z متناظر با رتبه‌های درصدی نمرات خام با استفاده از جدول توزیع نرمال استاندارد بدست آمد. در گام آخر، نمرات Z بدست آمده برای آزمون با استفاده از یک تبدیل خطی به توزیعی با میانگین ۱۰۰ و انحراف استاندارد ۱۵ تبدیل گردید.

دلیل اینکه توزیع نمرات 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 نمره‌های 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، ۱۹۶۶، ترجمه براهنی، ۱۳۷۰). علاوه براین، نرمال سازی توزیعهای چوله نمرات خام برای هدفهای مقایسه، مطلوب است (کوهن و اسوردلیک^{۳۹}، ۲۰۰۲).

نرمال سازی داده‌های خام یک توزیع زمانی قابل توجیه است که میزان انحراف توزیع نمرات خام از حالت نرمال، از لحاظ آماری معنادار نباشد. بنابراین تصمیم گرفتیم که فرضیه نرمال بودن توزیع داده‌های حاصل از آزمون اصلی را مورد بررسی قرار دهیم. هدف ما از این بررسی این بود که اگر مشخص شود، خصیصه مورد سنجش (هوش کلی) در جامعه‌ای که گروه نمونه از آن استخراج شده است توزیع نرمال دارد، در این صورت هم به نتایج حاصل از اجرای آزمون می‌توان اطمینان داشت و هم پایه‌ای منطقی برای نرمال سازی توزیع نمرات خام بدست می‌آید.

آزمونهای نموداری و غیر نموداری زیادی برای ارزیابی نرمال بودن توزیع وجود دارد. در مواردی که حجم نمونه مورد مطالعه، اندک تا متوسط است به خاطر خطای نمونه‌گیری زیاد، تشخیص اینکه آیا انحراف از حالت نرمال، واقعی است یا ظاهری،

مشکل است. به همین دلیل در این شرایط بهتر است که از آزمونهای غیر نموداری (آماری) استفاده کرد (استیونس، ۲۰۰۲).

از بین آزمونهای آماری که جهت بررسی نرمال بودن توزیع بکار می‌رود می‌توان به آزمون χ^2 خوبی برازندگی^{۴۰}، آزمون کلموگروف- اسمیرنوف^{۴۱}، آزمون شاپیرو- ویلک^{۴۲} و ضرایب چولگی^{۴۳} و کشیدگی^{۴۴} اشاره کرد. در مطالعه‌ای که ویلک، شاپیرو و چن^{۴۵} (۱۹۶۸). انجام دادند به این نتیجه رسیدند که ترکیبی از ضرایب چولگی و کشیدگی و آزمون شاپیرو - ویلک، بهترین شیوه تشخیص انحرافات از نرمال بودن است (به نقل از استیونس، ۲۰۰۲). جهت ارزیابی معنادار بودن ضرایب چولگی و کشیدگی، زمانی که حجم نمونه، اندک تا متوسط است، از آلفای سطح ۰/۰۱ یا ۰/۰۰۱ استفاده می‌گردد (تاباکنایک و فیدل^{۴۶}، ۲۰۰۱).

بر این اساس، برای بررسی درستی یا نادرستی این مسأله که آیا توزیع فراوانی نمره‌های هر گروه سنی با مدل احتمالات نرمال تطبیق می‌کند یا نه، توزیع فراوانی مشاهده شده نمرات خام با استفاده از آزمونهای چولگی، کشیدگی و شاپیرو- ویلک مورد آزمون قرار گرفت. نتایج بدست آمده در جدول ۷ خلاصه شده است. مقدار بحرانی برای آلفای یکصدم دوسویه برابر با ۲/۵۷ است. بررسی نتایج آزمونهای انجام شده نشان می‌دهد که بطور کلی می‌توان نتیجه گرفت که ویژگی مورد اندازه‌گیری این آزمون (هوش کلی) در جامعه مورد مطالعه توزیع نرمال دارد و نرمال سازی توزیع‌های نمرات خام گروههای سنی، منطقی و قابل توجیه است.

جدول ۸. نتایج حاصل از آزمونهای نرمال بودن توزیعیهای نمرات خام

| مقدار P | آزمون شاپیرو- ویلک | مقدار Z مشاهده شده | ضریب کشیدگی | مقدار Z مشاهده شد | ضریب چولگی | N | گروه سنی |
|---------|--------------------|--------------------|-------------|-------------------|------------|----|----------|
| ۰/۷۰ | ۰/۹۳ | ۰/۴۸ | ۰/۴۰ | -۱/۲۰ | -۰/۴۰ | ۲۲ | ۶ |
| ۰/۵۷ | ۰/۹۶ | ۰/۴۴ | ۰/۳۹ | -۱/۳۲ | -۰/۵۹ | ۲۰ | ۷ |
| ۰/۱۰ | ۰/۹۵ | ۰/۶۵ | ۰/۴۴ | -۱/۹۴ | -۰/۶۶ | ۵۲ | ۸ |
| ۰/۰۵ | ۰/۹۲ | ۱/۸۸ | ۱/۳۹ | ۲/۴۴ | ۰/۹۰ | ۴۴ | ۹ |
| ۰/۶۸ | ۰/۹۷ | ۱/۰۳ | ۰/۸۲ | ۰/۵۵ | ۰/۲۲ | ۳۸ | ۱۰ |
| ۰/۳۶ | ۰/۹۶ | -۰/۰۷ | -۰/۰۵ | ۱/۴۸ | ۰/۵۴ | ۴۵ | ۱۱ |
| ۰/۱۱ | ۰/۹۵ | ۰/۱۰ | ۰/۰۷ | ۱/۸۹ | ۰/۶۶ | ۴۹ | ۱۲ |
| ۰/۰۱ | ۰/۹۱ | ۰/۹۳ | ۰/۶۱ | -۲/۴۴ | -۰/۸۰ | ۵۶ | ۱۳ |
| ۰/۰۱ | ۰/۹۰ | -۱/۰۳ | -۰/۷۱ | ۲/۰۰ | ۰/۶۹ | ۵۱ | ۱۴ |
| ۰/۰۱ | ۰/۸۳ | ۳/۱۵ | ۲/۱۴ | -۴/۵۶ | -۱/۵۵ | ۵۲ | ۱۵ |

با توجه به این واقعیت که تفاوت بین نمرات گروه دختران و پسران در آزمون، برای هیچ یک از گروههای سنی مورد مطالعه از لحاظ آماری معنادار نبود گروههای دختر و پسر با هم ترکیب و نمرات مربوط به گروه کل آزمودنیها در هر یک از سطوح سنی مبنای محاسبه جداول توزیع فراوانی و نمره‌های گروههای مختلف قرار گرفت. نتایج حاصل از مقایسه میانگین گروههای دختر و پسر به تفکیک گروههای سنی در جدول ۸ نشان داده شده است.

جدول ۹- نتایج حاصل از مقایسه میانگین گروههای دختر و پسر

| | گروه سنی | | | | | | | | | |
|---------|----------|------|------|------|------|------|------|------|-------|------|
| | ۱۵ | ۱۴ | ۱۳ | ۱۲ | ۱۱ | ۱۰ | ۹ | ۸ | ۷ | ۶ |
| مقدار t | ۰/۲۷ | ۱/۸۰ | ۰/۲۶ | ۰/۹۰ | ۰/۸۵ | ۰/۶۵ | ۰/۳۸ | ۰/۰۶ | -۰/۵۸ | ۰/۶۸ |
| df | ۴۳ | ۳۵ | ۴۴ | ۳۹ | ۳۵ | ۲۹ | ۳۴ | ۴۲ | ۱۸ | ۸ |
| مقدار P | ۰/۷۹ | ۰/۱۲ | ۰/۸۰ | ۰/۲۲ | ۰/۲۰ | ۰/۵۲ | ۰/۷۱ | ۰/۹۵ | ۰/۵۷ | ۰/۵۵ |

معادلهای سنی: معادلهای سنی^{۴۷} متناظر با نمرات خام تهیه گردید. جهت تعیین معادلهای سنی، میانگین نمرات خام هر گروه سنی محاسبه و با استفاده از بسته نرم‌افزاری STATGRAPH نمودار پراکندگی که در محور Xهای آن سن بر حسب ماه و در محور Yهای آن میانگین نمرات خام قرار داشتند، رسم گردید. آنگاه نمودار بدست آمده هموار گردید و در مرحله بعد جهت تعیین نمرات خام متناظر با گروههای سنی خارج از محدوده گروه هنجاریابی و سایر گروههای سنی موجود در محدوده گروه هنجاری که مستقیماً مورد آزمون قرارنگرفته بودند از عمل برون‌یابی^{۴۸} و درون‌یابی^{۴۹} استفاده گردید.

بحث و نتیجه گیری

این پژوهش با هدف انطباق و استاندارد سازی آزمون هوشی ویلیامز برای افراد با آسیبهای بینایی به اجرا درآمد. نتایج حاصل از مطالعه ویژگیهای روانسنجی آزمون نشان می‌دهد که آزمون مورد نظر، برای تمام گروههای سنی مورد مطالعه از اعتبار و روایی بالایی برخوردار است و کاربران می‌توانند به عنوان یک ابزار معتبر و روا برای سنجش هوش کلی دانش آموزان نابینا و نیمه بینای ۵-۱۵ سال، به منظور اهداف نظری (پژوهشی) و کاربردی از آن استفاده نمایند.

البته تعمیم نتایج تحقیق حاضر به سراسر ایران و به جوامع آماری غیردانش‌آموزی باید با احتیاط صورت گیرد. بر عهده دانش‌پژوهان علاقمند به این زمینه است که با اجرای پژوهش در زمینه اعتبار و روایی این آزمون راه را برای شناخت بیشتر نقاط قوت و ضعف آن باز نمایند و زمینه پیشرفت بیشتر در این زمینه را فراهم نمایند. پژوهش حاضر نیز مانند بسیاری از پژوهشها دارای مشکلاتی بوده است. مهمترین این مشکلات عبارت است از: الف) کم بودن حجم نمونه در گروههای سنی مختلف، و ب) نامناسب بودن شرایط محیطی و فیزیکی مدارس که باعث افزایش خطا و کاهش اعتبار و روایی نتایج آزمون می‌گردد.

کم بودن حجم نمونه در گروههای سنی مختلف مانع از این شد که ساختار درونی آزمون مورد تحلیل قرار گیرد. همانطور که گفته شد، نیولند (۱۹۶۵) بیان می‌کند که این آزمون چهار حوزه از تواناییهای شناختی، الف: استدلال کلامی، ب: استدلال کمی، ج: استدلال انتزاعی، و د: حافظه کوتاه مدت را می‌سنجد. مناسبترین راه شناخت عاملهای مورد اندازه‌گیری آزمون و بررسی میزان انطباق ساختار درونی آن با تعاریف نظری سازه مورد اندازه‌گیری، روش تحلیل عوامل است. با عنایت به اینکه در این آزمون، گروههای سنی مختلف، به مجموعه نسبتاً متفاوتی از سوالات پاسخ می‌دهند، بنابر این لازم است که برای هر گروه سنی تحلیل عوامل جداگانه‌ای صورت گیرد، لکن با عنایت به کمی حجم نمونه در گروههای سنی انجام چنین تحلیلی مفید نبود.

یادداشت‌ها

- 1) Brown , Sherbenou & Johnson
- 2) Jensen, A
- 3) The Ninth Mental Measurement Yearbook
- 4) Murphy & Davidshofer
- 5) Haptic Intelligence Scale for Adult Blind
- 6) School and College Ability Tests (SCAT)
- 7) College Board Scholastic Aptitude Test (CBSAT)
- 8) Graduate Record Examination (GRE)
- 9) Williams Intelligence Test for Children with Defective Vision
- 10) Myfanway Williams
- 11) Verbal reasoning
- 12) Quantitative reasoning
- 13) Abstract reasoning
- 14) Short - term memory
- 15) Newland
- 16) Valentine's Intelligence Tests for Children
- 17) Burt's Reasoning Tests
- 18) Wechsler's Scale for Children
- 19) Difficulty
- 20) Discrimination

- | | |
|---------------------------------|-----------------------------------|
| 21) Reliability | 22) Anastasi,A |
| 23) Newcomer & Hammill | 24) Internal consistency |
| 25) Fisher's Z transformation | 26) Validity |
| 27) Content validity | 28) Criterion – related validity |
| 29) Construct validity | 30) Aiken, Lewis.R |
| 31) Salvia & Ysseldyke | 32) Item validity |
| 33) Cole, P | 34) Stevens, J |
| 35) Gronlund & Linn | 36) Guilford, & Fruchter |
| 37) Discrimination power | 38) Kline, P |
| 39) Cohen & Swerdlik | 40) χ^2 goodness of fit test |
| 41) Kolmogorov – Smirnov | 42) Shapiro – Wilk Test |
| 43) Skewness | 44) Kurtosis |
| 45) Wilk , Shapiro & Chen | 46) Tabachnik & Fidell |
| 47) Age equivalent | 48) Interpolation |
| 49) Extrapolation | |

منابع

- آناستازی ، آنا (۱۳۶۴). روان‌آزمایی. ترجمه محمد نقی‌براهنی. تهران انتشارات دانشگاه تهران. (تاریخ انتشار به زبان اصلی، ۱۹۷۶).
- به‌پژوه، احمد و پرنده، اکرم (۱۳۸۰). هنجاریابی آزمون هوش ویلیامز برای کودکان با نارسایی بینایی. مجله روانشناسی، ۱۹، صص ۲۳۰-۲۱۷.
- شهیم ، سیما (۱۳۷۳). انطباق و هنجاریابی مقیاس تجدید نظر شده هوشی وکسلر برای کودکان. شیراز. انتشارات دانشگاه شیراز .
- مگنو سون، دیوید (۱۳۷۰). مبانی نظری آزمونهای روانی. (ترجمه محمد نقی‌براهنی). تهران. انتشارات دانشگاه تهران. (تاریخ انتشار به زبان اصلی، ۱۹۶۶).

Aiken,L.R (1994). *Psychological Testing and Assesment* . Needham Heigths ,MA: Allyn & Bacon.

Anastasi, A.(1990). *Psychological Testing* (6th ed). New York: Macmillan

Brown,L.,Sherbenou,R.J., & Johnsen , S.K. (1997).*Test of Nonverbal Intelligence* (3rd ed).Austin , TX: PRO – ED.

Cohen, R. J. & Swerdlik, M. R. (2002). *Psychological Testing and Assessment: An Introduction to Test and Measurement* (5th ed). New York: McGraw-Hill.

- Cole, P.C. (1995). The Bell Curve: Should Intelligence be Used as The Pivotal Exploratory Concept of Student Achievement? *Issues in Educational Research*, 5 (1), 11-22.
- Gronlund, N. E., & Linn. R. L. (1990). *Measurement and Evaluation in Teaching* (6th ed). New York: Macmillan.
- Guilford, J.P., & Fruchter, B. (1978). *Fundamental Statistics in Psychology and Education*. New York: McGraw-Hill.
- Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39, 1-123.
- Kline, P. (2000). *The Handbook of Psychological Testing* (2nd ed). London: Routledge.
- Murphy, K.R. & Davidshofer, C.O. (1991). *Psychological Testing: Principles & Applications* (2nd ed). Englewood Cliff, NJ: Prentice-Hall.
- Newcomer, P.L., & Hammill, D.D. (1997). *Test of Language Development – Primary*. (3rd ed). Austin, TX: PRO-ED.
- Newland, T.E. (1965). Williams Intelligence Test for Children with Defective Vision. In O. K. Buros (Ed). *The Sixth Mental Measurement Yearbook*. Highland, NJ: Gryphon Press.
- Salvia, J., & Ysseldke, J., E. (1991). *Assessment* (6th ed). Boston: Houghton Mifflin.
- Stevens, J. P. (2002). *Applied Multivariate Statistics for The Social Sciences*. (4th ed). Mahwah, NJ: Lawrence Erlbaum Associations.
- Tabachnick, B. G. & Fidell, L. S. (2001). *Using Multivariate Statistics* (5th ed). Needham Heights, MA: Allyn & Bacon.
- Williams, M. (1956). *Williams Intelligence Test for Children with Defective Vision: A Test for Blind and Partially Sighted Children*. Birmingham: University of Birmingham Institute of Education.