پژوهش در حیطه کودکان استثنایی ۱| سال اول، شماره ۱| پاییز ۱۳۸۰| ۵۱-۳۵
Research on Exceptional Children, Fall 2001, Vol.1, No. 1, 35-51

انطباق و هنجاریابی آزمون رشد زبان TOLD-P:3 برای کودکان فارسی زبان تهرانی*(بخش اول)

سعید حسن زاده، اصغر مینایی****

چکیده:

پژوهش حاضر با هدف انطباق، اعتباریابی، رواسازی وهنجاریابی آزمون رشد زبان، روی نمونه‌ای با حجم ۱۲۳۵ نفر (۶۲۶ پسر و ۶۰۹ دختر) از کودکان ۰-۴ (سال تمام) تا ۱۱-۸ (۸ سال و ۱۱ ماه) شهر تهران که با روش نمونه برداری چند مرحله‌ای انتخاب شدند، اجرا گردید. این آزمون مشتمل بر ۹ خرده آزمون (۶ خرده آزمون اصلی و ۳ خرده آزمون تکمیلی) است. علاوه بر این، از ترکیب نمرات خرده آزمون‌های اصلی، ۶ نمره مرکب نیز به‌دست می‌آید. برای انطباق و هنجاریابی‌این آزمون ابتدا دستورالعمل اجرایی و گویه‌های آزمون به زبان فارسی ترجمه شد. آنگاه با توجه به ویژگی‌های نحوی، معنی شناختی و واج شناختی زبان فارسی و با مد نظر قرار دادن مبانی نظری آزمون، بسیاری از گویه‌ها مجدداً طراحی گردید. با اجرای مراحل پیش تجربی و تجربی و تجزیه و تحلیل داده‌ها بر پایه مدل کلاسیک اعتبار، فرم نهایی آزمون شکل گرفت. ضریب همسانی درونی خرده آزمون‌ها با استفاده از فرمول آلفای کرونباخ و ثبات زمانی خرده آزمون‌ها وترکیب‌ها با استفاده از روش آزمون و بازآزمون به‌دست آمد. ضرایب آلفا برای ترکیب‌ها(نمرات مرکب) با استفاده از فرمول گیلفورد که برای برآورد آلفای نمرات مرکب طراحی شده است، به‌دست آمد. روایی محتوایی، ملاکی و سازه خرده آزمون‌ها و ترکیب‌ها با استفاده از روش‌های مناسب مورد بررسی قرار گرفت. نتایج به‌دست آمده در زمینه اعتبار و روایی آزمون، اطمینان کافی برای استفاده از این ابزار را برای ارزیابی مهارت‌های زبانی کودکان فارسی زبان تهران به‌دست می‌دهد.

کلید واژه‌ها: *هنجاریابی، زبان دریافتی، زبان بیانی، اعتباریابی، رواسازی*

□ □ □

* این تحقیق در پژوهشکده کودکان استثنایی انجام شده است.

** عضو هیأت علمی پژوهشکده کودکان استثنایی

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

مقدمه

نسخه اولیه آزمون رشد زبان *TOLD*[1] در سال ۱۹۷۷ میلادی توسط *نیوکامر* و *هامیل* برای ارزیابی رشد زبان کودکان انگلیسی زبان ۰-۴ (۴ سال تمام) تا ۱۱-۸ (۸ سال و ۱۱ ماه) در آمریکا تنظیم شد. این نسخه شامل ۵ خرده آزمون اصلی: واژگان تصویری، واژگان شفاهی، درک دستوری، تقلید جمله و تکمیل دستوری و ۲ خرده آزمون تکمیلی:تمایز گذاری کلمه و تولید کلمه بود. با جمع نتایج خرده آزمون‌های اصلی بهره زبانی محاسبه می‌شد. در سال ۱۹۸۲ میلادی نام آزمون به *TOLD-P*[2] تغییر یافت. علت این تغییر برای آن بود که این آزمون از همتای دیگر خود یعنی *TOLD-I*[3] که برای ارزیابی مهارت‌های زبانی کودکان ۶-۸ تا ۱۱-۱۲ ساله تهیه شده بود، مجزا گردد. علاوه بر تغییر نام، چند تغییر قابل ملاحظه دیگر نیز در این نسخه از آزمون صورت گرفته بود. اول آنکه آزمونگر می‌توانست از طریق ترکیب نتایج حاصله از خرده آزمون‌ها، *نمرات مرکب*[4] که شامل نمره‌های گوش کردن، صحبت کردن، نحو و معنی‌شناسی می‌شدند را به‌دست آورد و دوم آنکه در دستورالعمل اجرایی آزمون توجه ویژه‌ای به مبانی نظری آزمون و همچنین نحوه تفسیر نتایج شده بود و بالاخره آنکه جمعیت نمونه هنجاری به ۱۰۱۴ نفر افزایش پیدا کرده بود. همچنین شواهدی دال بر روایی محتوایی ارائه شده بود ولی با این حال محتوای اصلی و گویه‌های این نسخه با نسخه قبلی یکسان بود. در سال ۱۹۸۸ میلادی ویرایش دوم این آزمون با عنوان *TOLD-P:2* منتشر شد. در این نسخه خرده آزمون‌های تمایز گذاری کلمه و تولید کلمه مجدداً طراحی شدند. از ترکیب نتیجه این دو خرده آزمون، نمره ترکیبی آواشناسی به‌دست می‌آمد. نمونه هنجاری این نسخه از آزمون به ۲۴۳۶ نفر رسید. برای افزایش اعتبار آزمون تعداد گویه‌های برخی از خرده آزمون‌ها نیز افزایش پیدا کرد. آخرین نسخه آزمون مورد نظر در سال ۱۹۹۶ هنجاریابی و در سال ۱۹۹۷ انتشار یافت.

در مطالعه حاضر، آخرین نسخه آزمون برای انطباق و هنجاریابی به زبان فارسی مورد استفاده قرار گرفته است. در این نسخه نسبت به نسخه‌های قبلی علاوه بر آنکه مجدداً هنجاریابی بر روی گروهی ۱۰۰۰ نفری از ۲۸ ایالت آمریکا انجام شده است، شواهدی مبنی بر عدم تورش ناشی از جنس و نژاد ارائه شده است و ضریب اعتبار برای زیر گروه‌های نمونه

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

هنجاری نیز محاسبه شده است. گذشته از این مطالعات بسیاری در زمینه روایی آزمون انجام شده است تا نشان دهد آزمون علاوه بر آنکه برای جمعیت کل روایی دارد در ارتباط با زیر گروه‌ها نیز از روایی کافی برخوردار است. دو خرده آزمون تحلیل واجی و واژگان ربطی به مجموعه آزمون اضافه شده و سه خرده آزمون تمایزگذاری کلمه، تحلیل واجی و تولید کلمه نیز به‌عنوان خرده آزمون‌های تکمیلی معرفی شده‌اند. مبنای نظری آزمون با توجه به نظریات جاری زبانشناختی روزآمد شده و کلیهٔ تصاویر خرده آزمون‌های مصور رنگی شده‌اند.

چارچوب نظری آزمون رشد زبان

اکثر آزمون‌های استاندارد شده خوش ساخت براساس یک الگو یا چارچوب که توجیه کننده منطق به کار رفته در ساخت آزمون و گویه‌های خرده آزمون‌ها است، تهیه شده‌اند. هنگامی که اولین نسخه انگلیسی آزمون رشد زبان توسط سازندگان اصلی تهیه شد، آنها مصمم بودند که از یک الگوی زبانشناختی به عنوان مبنای نظری استفاده کنند و البته این تصمیم را در ویرایش‌های بعدی آزمون نیز حفظ کردند. برای تحقق این امر یک دیدگاه نظری خاص را دنبال نکردند، بلکه برای تهیه الگوی نظری خود از مساعدت‌های بسیاری از زبانشناسان معتبر و معروف از جمله: *بلوم*، ۱۹۸۳؛ *بلوم و لیهی*، ۱۹۷۸؛ *بلوم*، *تکف و لیهی*، ۱۹۸۴؛ *آر. براون*، ۱۹۷۳؛ *چامسکی*، ۱۹۵۷، ۱۹۶۵، ۱۹۶۸، ۱۹۶۹، ۱۹۸۱؛ *دویتو*، ۱۹۷۰؛ *جاکوبسون*، *فانت و هال*، ۱۹۶۳؛ *لئونارد لوئب*، ۱۹۸۸؛ *شلزینگر*، ۱۹۷۷ و *ویگوتسکی*، ۱۹۷۷ بهره بردند (نیوکامر و هامیل، ۱۹۹۷).

الگوی اصلی این آزمون کارهای جدید و رایج نظریه‌پردازان زبانشناسی اجتماعی، از قبیل *اونز* (۱۹۹۵)، را در بر نمی‌گیرد. نظریه‌های زبانشناسی اجتماعی اساساً به مطالعه کاربردها (استفاده از زبان برای برقراری ارتباط) می‌پردازند. زبانشناسان اجتماعی بیشترین اهمیت را به درک گوینده از موقعیتی که در طی آن ارتباط صورت می‌گیرد، می‌دهند. براساس این دیدگاه، کاربرد زبان به طور گسترده تحت تأثیر افرادی است که کودک با آنها تعامل دارد و متأثر از موقعیتی است که تعامل در آن صورت می‌گیرد(نینو و اسنو، ۱۹۸۸). این توجه به کاربرد که از سالهای ۱۹۷۰ با کارهای *باتس* (۱۹۷۶)، بلوم(۱۹۷۴) و *ریز*(۱۹۷۸) شروع شد، در حال حاضر به مطالعه چگونگی درک معنای گفتار توسط کودکان با توجه به رشد آنها پرداخته

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

است. کاری که توسط ویگ (۱۹۸۹) و نیپولد (۱۹۸۸) در مورد مهارت‌های فرا زبانشناختی نوجوانان انجام شده، نشان دهنده این تأکید و توجه است. اگر چه ارزش این نوع مطالعات و نظریات زبانشناختی مشخص است، ولی این اطلاعات برای تهیه آزمون‌های هنجار مرجع استاندارد شده مناسب نیستند. بهتر است از این نوع تحقیقات زبانی، همراه با دیگر رویکردهای غیررسمی ارزیابی- آموزشی استفاده شود (نیو کامر و هامیل، ۱۹۹۷).

در اینجا هر یک از مختصات زبانشناختی و نیز هر یک از نظام‌های زبانشناختی مختصراً توضیح داده می‌شود تا خواننده بتواند منطق حاکم بر خرده‌آزمون‌ها را بهتر درک کند. البته خواننده بایستی به خاطر داشته باشد که اگر چه این مختصات به صورت جداگانه و مستقل توضیح داده می‌شوند، اما در واقع یکپارچه و مرتبط با هم هستند. در ادامه به بررسی نظام‌ها و مختصات زبانی و رابطه آنها با الگویی که برای مفهوم بخشیدن به این آزمون به کار گرفته شده است، می‌پردازیم.

اشخاصی که از قواعد نحوی آگاهی ضمنی ندارند، نمی‌توانند جملات دستوری قابل قبولی را بیان کنند. با این حال به دلایل مختلف، بسیاری از افرادی که آگاهی مبنایی از نحو دارند هنوز هم در دستور زبان دچار مشکلاتی هستند. این کاربرد دستوری نامنظم یا استاندارد نشده کمتر مورد توجه زبانشناسان می‌باشد، زیرا آنها عمدتاً به درک این موضوع علاقمند هستند که چرا متکلّمان یک زبان دارای توانایی و بیان یک سری جملات نامحدود هستند (نیوکامر و هامیل، ۱۹۹۷).

معناشناسی: اصطلاح معناشناسی به مطالعه معنا در زبان اطلاق می‌شود (یعنی رابطه میان زبان و تفکر). دانش‌معنایی مستلزم درک نحوه ارتباط معنا با اشکال تکواژ شناختی یا واج شناختی است و همچنین قوانینی که چگونگی انتقال معنا را از طریق جملات نشان می‌دهند. به عبارت دیگر، لغات بخشی از معنا شناسی محسوب می‌شوند، زیرا به توانایی درک و کاربرد معنای کلمات مستقلی مانند "توپ" مربوط هستند. به علاوه، کاربرد کلمات مستقل مانند "توپ" در یک جمله معنی‌دار به آگاهی معناشناسی ویژه‌ای نیاز دارد. برای نمونه، جمله "توپ، پسر را زد" از نظر نحوی، تکواژشناسی و واج شناسی صحیح است اما از نظر معناشناسی غیرممکن است. همچنین معناشناسی فراتر از سطح جمله بوده و به بافتی وسیع‌تر مرتبط می‌شود.

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

نظام‌های زبانشناختی

علاوه بر تقسیم زبان به مختصات آن، بسیاری از نظریه‌پردازان زبانشناسی مانند *بلوم* (۱۹۷۴)، *چپمن* (۱۹۷۴) و *اینگرام* (۱۹۷۴) زبان را براساس نظام‌هایی که گوش‌کردن(دریافتی)، سازماندهی (*ترکیبی ـ واسطه‌ای*)[5] و صحبت کردن(بیانی) نامیده می‌شوند، طبقه‌بندی می‌کنند.به‌طورکلی باید گفت این نظام‌ها نشان‌دهنده چگونگی کاربرد و درک زبان هستند. به‌طور دقیق‌تر، گوش کردن شامل عملیات رمزگشایی است که توسط آن معنی نمادین و انتزاعی به گفتار شنیده شده اختصاص می‌یابد. صحبت کردن، توصیف کننده عملیات رمزگذاری است که در طی آن گفتار معنی‌دار فرمول‌بندی و تولید می‌شود.ظاهراً بسیاری از فعالیت‌های زبانی، این دو نظام را در بردارند. با این وجود، از آنجا که نمی‌توان تمام تکالیف زبانی را به آسانی به دو شکل صحبت کردن و یا گوش کردن تقسیم کرد، محققان علاقمند به الگوهای پردازش اطلاعات در حوزه رشد زبانی همانند *مارکمن* (۱۹۸۱) و *رابینوویتز* و *گلاسر* (۱۹۸۵)،نظام‌های زبانی را گسترده‌تر کرده‌اند تا یک نظام سازماندهی را نیز به آنها اضافه کنند. این نظام به ظرفیت مغز در تفکیک یا سازماندهی اطلاعات زبانشناختی وارده به وسیله طبقه‌بندی کردن اشاره دارد. *کاس* (۱۹۷۸) ثابت کرد که به طور نظری ظرفیت حافظه(یعنی توانایی فراخوانی اطلاعاتی که قبلاً فراگرفته شده‌اند) به کارآیی پردازش سازمانی بستگی دارد. دو راهکار اساسی سازماندهی عبارتند از *واسطه‌گری*[6] و *تداعی‌گری*[7] که در آن یک نماد به نماد دیگر متصل می‌شود، همانگونه که "گرم و سرد" یا "پدر و مادر" به هم مرتبط شده‌اند.

الگوی دوبعدی

چارچوب نظری زبانشناختی که در قسمت قبل به آن پرداخته شد را می‌توان به صورت یک الگوی دو بعدی که شامل مختصات زبانشناختی و نظام‌های زبانشناختی است، نشان داد (به جدول ۱ نگاه کنید). این جدول، ارتباط میان مفاهیم موجود در الگو و خرده آزمون‌های آزمون رشد زبان اولیه را نیز نشان می‌دهد.

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

جدول ۱- الگوی دو بعدی ساختار زبانی، که در ساختن خرده‌آزمون‌های آزمون رشد زبان اولیه از آن استفاده شده است

نظام‌های زبانشناختی / مختصات زبانشناختی	گوش‌کردن (مهارت‌های دریافتی)	سازماندهی (مهارت‌های یکپارچه‌سازی و واسطه‌گری)	صحبت‌کردن (مهارت‌های بیانی)
معناشناسی	واژگان تصویری	واژگان ربطی	واژگان شفاهی
نحو	درک دستوری	تقلید جمله	تکمیل دستوری
واج‌شناسی	تمایزگذاری کلمه	تحلیل واجی	تولید کلمه

مختصات زبانشناختی

مختصه معناشناسی به وسیله سه خرده آزمون ارائه شده است. خرده آزمون واژگان تصویری که نظام گوش کردن را دربر می‌گیرد؛ خرده‌آزمون واژگان ربطی که شامل نظام سازمانی است و خرده‌آزمون واژگان شفاهی که نظام صحبت کردن را در بر دارد. همان‌گونه که از اسامی این سه خرده‌آزمون برداشت می‌شود آنها خزانه لغات را می‌سنجند. اگر چه خزانه لغات تنها جنبه محدودی از معناشناسی را نشان می‌دهد، اما این جنبه بسیار حائز اهمیت است و برای آموزگاران و والدین نیز آشنا و ملموس می‌باشد.

مختصه نحوی نیز به وسیلهٔ سه خرده آزمون ارائه شده است. خرده‌آزمون درک دستوری، عمدتاً نظام گوش کردن، خرده آزمون تقلید جمله، نظام سازماندهی و خرده آزمون تکمیل دستوری نیز نظام صحبت کردن را دربر می‌گیرند. خرده‌آزمون درک دستوری به اندازه‌گیری توانایی فهم نظم و ترتیب کلمات می‌پردازد. خرده آزمون تکمیل دستوری، آگاهی و اطلاعات کودکان را از تکواژشناسی (یعنی توانایی آنها را در به کارگیری تکواژهای مناسب در بافت جمله) می‌سنجد.

مختصه واج‌شناسی نیز به وسیله سه خرده آزمون اندازه‌گیری می‌شود که هر یک از آنها نشان دهنده یک نظام خاص زبانشناختی است. خرده آزمون تمایزگذاری کلمه، نظام گوش کردن، خرده آزمون تحلیل واجی، نظام سازماندهی و خرده‌آزمون تولید کلمه نیز نظام صحبت کردن را در برمی‌گیرند. محتوای واقعی خرده‌آزمون‌های تمایزگذاری کلمه و تولید کلمه بیشتر به اندازه‌گیری گفتار می‌پردازد تا زبان (نیوکامر و هامیل، ۱۹۹۷).

محتوای خرده‌آزمون تحلیل واجی به مهارت منحصر به فردی احتیاج دارد که پیش‌بینی

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

کنندهٔ بسیار خوبی از توانایی خواندن است (لایبرمن، شنکویلر و لایبرمن، ۱۹۸۹). در کل، تمامی این سه خرده آزمون جنبه‌های مهمی از رشد زبانی کودکان خردسال را می‌سنجند. عموماً، کودکان بزرگتر از ۶ یا ۷ ساله، به‌طور موفقیت‌آمیزی اکثر توانایی‌های واج‌شناسی را در نظام‌های زبانی خود گنجانده‌اند.

نظام‌های زبانشناسی

خرده آزمون‌هایی که نظام گوش کردن را می‌سنجند عبارتند از: تمایزگذاری کلمه(واج‌شناسی)، واژگان تصویری(معناشناسی) و درک دستوری(نحو) که جنبه‌هایی از عملیات رمزگشایی موجود در درک گفتار را می‌سنجند.

نظام سازماندهی توسط خرده‌آزمون‌های تحلیل واجی(واج‌شناسی)، واژگان ربطی(معناشناسی) و تقلید جمله(نحو) سنجیده می‌شود. این خرده آزمون‌ها جنبه‌هایی از راهکارهای تداعی‌گری و واسطه‌گری را که برای سازماندهی زبان گفتاری استفاده می‌شوند، اندازه‌گیری کرده و آنها را به نمادهای دیگر متصل می‌کنند.

نظام صحبت کردن توسط سه خرده آزمون تولید کلمه(واج شناسی)، واژگان شفاهی(معناشناسی) و تکمیل دستوری(نحو) سنجیده می‌شود. تمامی این سه خرده آزمون، جنبه‌های رمزگذاری را که برای تولید گفتار معنی‌دار استفاده می‌شوند، ارزیابی می‌کنند.

توصیف خرده‌آزمون‌ها و ترکیب‌ها

الگوی دو بعدی که مورد بررسی قرار گرفت اساس نظری تهیه ۹ خرده آزمون است. خرده آزمون‌های مربوط به معناشناسی و نحو، خرده‌آزمون‌های اصلی و سه خرده آزمون مربوط به واج‌شناسی جزء خرده آزمون‌های تکمیلی هستند.

خرده آزمون‌های اصلی

خرده‌آزمون‌های اصلی شامل شش خرده آزمون هستند که نتایج آنها ترکیب شده و نمرات ترکیبی (یعنی زبان گفتاری، گوش کردن، سازماندهی، صحبت کردن، معناشناسی و نحو) را شکل می‌دهند. این خرده آزمون‌ها برای سنجش معنا شناسی و نحو به کار می‌روند. خرده

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

آزمون‌هایی که واج‌شناسی را اندازه‌گیری می‌کنند، اجزای تکمیلی آزمون رشد زبان اولیه هستند. جداسازی نمرات واج شناسی از نمرات ترکیبی گوش کردن، سازماندهی، صحبت کردن و زبان گفتاری، باعث تمایزی واضح و آشکار میان توانش گفتاری (یعنی مهارت‌های ارتباطی اصلی که توسط خرده‌آزمون‌های واج شناسی اندازه‌گیری می‌شوند) و توانش زبانی شده و موجب گردیده است تا اختلالات ویژه در هر یک از این حیطه‌ها، به خوبی و با کارآیی بالا مشخص شوند.

خرده آزمون ۱- واژگان تصویری: این خرده آزمون معناشناختی دارای ۳۰ گویه است و میزان درک کودک را از معانی مرتبط با کلمات فارسی اندازه‌گیری می‌کند.

خرده آزمون ۲- واژگان ربطی: این خرده آزمون معناشناختی که ۳۰ گویه دارد به سنجش توانایی کودک در درک و بیان شفاهی ارتباط میان دو کلمه می‌پردازد.

خرده‌آزمون ۳- واژگان شفاهی:این خرده آزمون معنا شناختی، ۲۸ گویه دارد و به اندازه‌گیری توانایی کودک در ارائه تعاریف شفاهی از کلمات رایج فارسی می‌پردازد.

خرده‌آزمون ۴- درک دستوری: این خرده آزمون نحوی، ۲۵ گویه دارد و توانایی کودک را در درک و فهم معنای جمله‌ها می‌سنجد.

خرده‌آزمون ۵- تقلید جمله: این خرده آزمون نحوی که به صورت ۳۰ گویه‌ای طراحی شده است، به اندازه‌گیری توانایی کودکان در تولید صحیح جمله‌های فارسی می‌پردازد.

خرده‌آزمون ۶- تکمیل دستوری: این خرده‌آزمون نحوی، ۲۸ گویه دارد و به سنجش توانایی کودکان در شناختن، فهمیدن و به کارگیری اشکال تکواژ شناختی رایج زبان فارسی می‌پردازد.

خرده‌آزمون‌های تکمیلی

این گروه از خرده آزمون‌ها شامل ۳ خرده آزمون زیر است:

خرده‌آزمون ۷- تمایزگذاری کلمه: این خرده‌آزمون واج شناختی با داشتن ۲۰ گویه به سنجش توانایی کودکان در تشخیص تفاوت‌های موجود در صداهای گفتاری مشخص می‌پردازد.

خرده‌آزمون ۸- تحلیل واجی: این خرده‌آزمون ۱۴ گویه دارد و یک تکلیف سازماندهی است که به بررسی توانش کودک در مورد نظام واج شناختی زبان فارسی می‌پردازد. این خرده‌آزمون

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

یک جنبه از مهارت پردازش شنیداری، به ویژه توانایی تجزیه کلمات به واحدهای واجی کوچکتر را می‌سنجد.

خرده‌آزمون ۹- تولید کلمه: این خرده‌آزمون واج شناختی، ۲۰ گویه دارد و توانایی کودک را در تولید صداهای گفتاری مهم زبان‌فارسی اندازه‌گیری می‌کند.

ترکیب‌ها

همان‌گونه که قبلاً گفتیم ، اساس تهیه خرده‌آزمون‌ها یک الگوی دو بعدی بوده است. هر خرده‌آزمون، یک نظام زبانی(گوش‌کردن، سازماندهی، صحبت کردن) و یک مختصه زبانی(معناشناسی، نحو، واج‌شناسی) را شامل می‌شود. بنابراین، خرده‌آزمون‌ها با توجه به نظام‌ها و یا مختصه‌های مشترک می‌توانند گروه‌بندی شوند و ترکیب‌های زیر را به وجود آورند:

گوش کردن(واژگان تصویری + درک دستوری)

سازماندهی (واژگان ربطی + تقلید جمله)

صحبت کردن(واژگان شفاهی + تکمیل دستوری)

معنی‌شناسی (واژگان تصویری + واژگان ربطی + واژگان شفاهی)

نحو (درک دستوری + تقلید جمله + تکمیل دستوری)

زبان گفتاری (واژگان تصویری + واژگان ربطی + واژگان شفاهی + درک دستوری+ تقلید جمله+ تکمیل دستوری)

مراحل هنجاریابی و نمونه مورد مطالعه

۱- مرحله پیش تجربی[۸]: برای بررسی کیفی گویه‌ها و آگاهی نسبت به مسائل و مشکلات احتمالی اجرایی، آزمون ابتدا بر روی ۹۶ نفر (۴۹ پسر و ۴۷ دختر) از کودکان ۹-۴ ساله مراکز پیش‌دبستانی و پایه‌های اول تا سوم مدارس ابتدایی اجرا گردید و براساس نتایج حاصله تغییراتی در برخی از گویه‌ها داده شد و همچنین تعدادی از گویه‌های مبهم حذف و گویه‌های مناسب‌تری جایگزین آنها شد.

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

۲- مرحله تجربی[۹]: در این مرحله، آزمون بر روی ۳۸۰ کودک(۱۹۰ دختر و ۱۹۰ پسر) در گروه سنی ۰ - ۴ تا ۱۱ - ۸ ساله اجرا گردید. نمونه فوق به صورت تصادفی‌از بین کودکان مراکز پیش‌دبستانی و دبستان‌های مناطق ۱۹گانه شهرتهران انتخاب گردید. برای انتخاب نمونه از بین کودکانی که به مرکز پیش دبستانی و یا مدرسه نمی‌رفتند از نمونه‌های در دسترس استفاده شد.

اجرای این مرحله، پس از آموزش آزمونگران طرح که حداقل دارای مدرک لیسانس روانشناسی بودند انجام گرفت. پس از استخراج نتایج، تجزیه و تحلیل داده‌ها بر اساس مدل کلاسیک اعتبار صورت گرفت. با استفاده از این مدل برای هر گویه، ۲ پارامتر درجه دشواری و قدرت تشخیص برآورد گردید. براساس نتایج حاصل از این مرحله تعدادی از گویه‌ها حذف گردید و فرم نهایی آزمون براساس گویه‌های باقی‌مانده شکل گرفت.

۳- اجرای نهایی: در این مرحله برای انتخاب نمونه برای تشکیل جداول هنجاری از روش نمونه‌گیری چند مرحله‌ای استفاده گردید. به این صورت که ابتدا شهر تهران به سه منطقه جغرافیایی شمال، مرکز و جنوب تقسیم گردید. آنگاه با بهره‌گیری از اطلاعات منتشر شده توسط مرکز آمار ایران، نمونه لازم برای هر منطقه جغرافیایی براساس متغیرهای جنسیت و سن کودکان و همچنین وضعیت اقتصادی - اجتماعی والدین تعیین گردید. در گام بعدی با توجه به حجم جمعیت در هر منطقه جغرافیایی، تعدادی مدرسه دخترانه و پسرانه و نیز مرکز پیش‌دبستانی به‌طور تصادفی انتخاب شد که شامل ۱۰۰ مدرسه دخترانه و پسرانه و ۱۰۰ مرکز پیش دبستانی بودند. سپس از بین دانش‌آموزان مدارس و مراکز پیش دبستانی، نمونه لازم استخراج و آزمون بر روی آنها اجرا شد. برای انتخاب نمونه از بین کودکانی که به مرکز پیش‌دبستانی و یا مدرسه نمی‌رفتند از نمونه‌های در دسترس استفاده شد. در نمونه انتخاب شده تعدادی کودک کم توان ذهنی، دارای اختلالات گفتاری، ناتوانی یادگیری و سایر معلولیت‌ها نیز گنجانده شد. در نهایت آزمون بر روی ۱۲۳۵ کودک اجرا و جداول هنجاری براساس نتایج حاصل از این تعداد به دست آمد. آزمونگران این مرحله، همان آزمونگران مرحله تجربی بودند که در اجرای آزمون تجربه کافی داشتند. ویژگی‌های نمونه نهایی برای تشکیل جداول هنجاری به تفکیک متغیرهای جمعیت شناختی در جدول ۲ آمده است.

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

جدول ۲- ویژگی‌های جمعیت شناختی گروه هنجاری (N = ۱۲۳۵)

ویژگی‌ها	فراوانی	درصد
جنسیت		
مذکر	۶۲۶	٪۵۱
مؤنث	۶۰۹	٪۴۹
ناحیه جغرافیایی تهران		
شمال	۳۱۷	٪۲۶
مرکز	۴۸۶	٪۳۹
جنوب	۴۳۲	٪۳۵
وضعیت روانی		
عادی	۱۱۳۷	٪۹۲
دارای ناتوانی یادگیری	۴۷	٪۳/۸
دارای اختلال گفتاری	۲۶	٪۲/۱
کم توان ذهنی	۱۷	٪۱/۴
سایر معلولیت‌ها	۸	٪۰۰۷
سطح تحصیلات مادر		
فوق لیسانس و بالاتر	۲۵	٪۰۲
دیپلم تا لیسانس	۶۹۲	٪۵۶
کمتر از دیپلم	۵۱۹	٪۴۲
سطح تحصیلات پدر		
فوق لیسانس و بالاتر	۷۷	٪۰۶
دیپلم تا لیسانس	۵۹۳	٪۴۸
کمتر از دیپلم	۵۶۹	٪۴۶
سن		
۴ ساله	۱۹۷	٪۱۶
۵ ساله	۲۳۷	٪۲۰
۶ ساله	۲۷۸	٪۲۲
۷ ساله	۲۴۵	٪۲۰
۸ ساله	۲۷۸	٪۲۲

نمرات هنجاری

نمرات استاندارد برای خرده‌آزمون‌ها و ترکیب‌ها

جداول هنجاری برای نمرات خام هر یک از خرده‌آزمون‌ها در ۱۰ گروه سنی به فاصله ۶ ماه از یکدیگر (از ۰-۴ تا ۱۱-۸) تهیه گردید. با توجه به این‌که در آزمون‌هایی که به صورت انفرادی اجرا می‌شوند، عوامل زمان و هزینه، تعداد افراد مورد نیاز برای هنجاریابی را محدود

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

می‌کنند، لذا نوعاً از نمونه‌هایی با حجم ۱۰۰ الی ۲۰۰ نفر برای هر گروه سنی یا کلاسی برای هدف‌های هنجاری استفاده می‌گردد. نوسانات نمونه‌گیری ناشی از حجم نسبتاً اندک گروه هنجاری، مشکلاتی را در پرورش هنجارها به‌وجود می‌آورد. بنابراین با توجه به کمی حجم نمونه در هر گروه سنی (تقریبا ۱۰۰ نفر) و به منظور کاهش مشکلات ناشی از این موضوع، از روشی که *انگاف* و *رابرتسون* (۱۹۸۷) ارائه داده‌اند، استفاده شد. در این روش، پس از تعدیل تفاوت‌های موجود در میانگین‌ها و انحراف استانداردهای گروه‌های سنی متوالی، داده‌های گروه‌های سنی مختلف به یک توزیع تبدیل می‌گردد. کاربرد این روش منجر به هنجارهایی می‌گردد که در آن مقایسه گروه‌های سنی مختلف بهتر صورت می‌گیرد.

پس از تشکیل توزیع کلی، نمرات خام به نمرات استاندارد هنجار تبدیل شدند. نتایج به‌دست آمده برای هر خرده آزمون با استفاده از یک تبدیل خطی به نمرات استاندارد با میانگین ۱۰ و انحراف استاندارد ۳، تبدیل گردید.

نمره استاندارد متناظر با آزمون‌های ترکیبی که بهره نامیده می‌شود با استفاده از روش *گیلفورد* و فروچر (۱۹۷۸) برای ادغام واریانس‌ها محاسبه شد. این فرمول برای ادغام نمرات استاندارد خرده آزمون‌هایی که آزمون‌های ترکیبی را تشکیل می‌دهند، بکار برده شد. نمرات استاندارد آزمون‌های ترکیبی دارای میانگین ۱۰۰ و انحراف استاندارد ۱۵ است.

معادل‌های سنی[۱۰]

معادل‌های سنی خرده آزمون‌ها نیز تهیه گردید. این ارزش‌ها، بیانگر سطح یا سن توانایی زبانی متناظر با نمره خام آزمودنی است. برای تعیین معادل‌های سنی، میانگین نمرات خام هر خرده آزمون برای هر گروه سنی محاسبه و با استفاده از بسته نرم‌افزاری استات گراف، نمودار پراکندگی که در محور x های آن سن بر حسب ماه و در محور y های آن میانگین نمرات خام قرار داشتند رسم گردید. آنگاه نمودار به‌دست آمده هموار گردید و در مرحله بعد برای تعیین نمرات خام متناظر با گروه‌های سنی خارج از محدودهٔ گروه هنجاریابی و سایر گروه‌های سنی موجود در محدودهٔ گروه هنجاریابی که به طور مستقیم مورد آزمون قرار نگرفته بودند از عمل برون‌یابی[۱۱] و درون‌یابی[۱۲] استفاده گردید.

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

نتایج

۱- اعتبار[۱۳] آزمون

مفهوم اعتبار 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، ۱۹۸۸، ص ۱۰۹). به همین دلیل سازندگان یک آزمون، به ویژه آزمون‌های تشخیصی مانند آزمون مورد بحث باید شواهد قابل توجهی مبنی بر این که آزمونشان از اعتبار کافی برخوردار است و می‌تواند با اطمینان برای هدف‌های عنوان شده مورد استفاده قرار گیرد، ارائه دهند. مطالعهٔ اعتبار یک آزمون بر برآورد مقدار واریانس خطای مرتبط با نمرات آن تأکید دارد. این خطا می‌تواند ناشی از نمونه‌برداری محتوایی، نمونه‌برداری زمانی و تفاوت‌های نمره‌گذاران باشد. صرف نظر از این که کدام منبع واریانس خطا مورد بررسی قرار گرفته باشد، نتایج معمولا برحسب ضریب اعتبار که کاربرد ویژه‌ای از ضریب همبستگی متداول است گزارش می‌شوند. برای این که آزمونی به‌عنوان یک آزمون معتبر مورد توجه قرار گیرد، ضرایب اعتبار نمرات آن باید نزدیک به ۰/۸ و یا بیشتر از آن باشد. ضریب اعتبار ۰/۸ توسط بیشتر صاحبنظران برای تمیز بین آزمون‌های معتبر و غیرمعتبر پذیرفته شده است (آناستازی، ۱۹۸۸). در اینجا وضعیت خرده‌آزمون‌های اصلی و تکمیلی و ترکیب‌ها در ارتباط با دو منبع خطای محتوایی و زمانی مورد بحث قرار می‌گیرد.

۱-۱- نمونه‌برداری محتوایی

خطای ناشی از نمونه‌برداری محتوایی، به میزان قابل توجهی به درجهٔ همگنی بین مواد یک آزمون یا خرده‌آزمون ارتباط دارد. از آن جهت که هدف آزمون، اندازه‌گیری ویژگی، توانایی و یا محتوای خاصی است، لذا هر چقدر مواد آزمون ارتباط بیشتری با یکدیگر داشته باشند، خطای آزمون کمتر خواهد بود. هرگاه مواد آزمون، با یکدیگر ارتباط نداشته باشند، ویژگی‌های متفاوتی را اندازه می‌گیرند و در نتیجه مقدار خطای ناشی از نمونه برداری محتوایی فزونی خواهد گرفت (نیوکامر و هامیل، ۱۹۹۷، ص۵۵).

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

جدول ۳- ضرایب آلفا در پنج گروه سنی*

	سن به سال					
	۴	۵	۶	۷	۸	میانگین ضرایب
خرده‌آزمون‌ها						
واژگان‌تصویری	۸۱	۷۹	۷۰	۷۸	۷۴	۷۶
واژگان‌ربطی	۸۷	۸۹	۹۰	۹۱	۸۹	۸۹
واژگان‌شفاهی	۸۶	۸۸	۹۰	۹۲	۸۹	۸۹
درک‌دستوری	۷۳	۷۴	۷۸	۷۴	۷۲	۷۴
تقلیدجمله	۸۷	۹۰	۹۱	۹۲	۹۲	۹۰
تکمیل‌دستوری	۷۹	۷۷	۸۴	۸۴	۸۲	۸۱
تمایزگذاری‌کلمه	۹۱	۸۸	۸۴	۹۲	۹۳	۹۰
تحلیل‌واجی	۹۳	۹۶	۹۴	۹۲	۹۴	۹۴
تولیدکلمه	۸۳	۸۴	۸۴	۷۸	۸۱	۸۲
ترکیب‌ها						
گوش‌کردن	۸۵	۸۱	۷۸	۸۴	۸۳	۸۲
سازماندهی	۸۹	۹۱	۹۳	۹۴	۹۴	۹۲
صحبت‌کردن	۸۸	۸۹	۹۱	۹۳	۹۱	۹۰
معنی‌شناسی	۹۳	۹۳	۹۴	۹۵	۹۴	۹۳
نحو	۸۹	۹۱	۹۱	۹۴	۹۴	۹۱
زبان‌گفتاری	۹۶	۹۶	۹۶	۹۷	۹۷	۹۶

* (اعشار حذف شده است)

خطای نمونه‌برداری محتوایی (اعتبار همسانی درونی) با استفاده از روش ضریب آلفای کرونباخ مورد بررسی قرار گرفت. ضرایب آلفا برای خرده‌آزمون‌ها و ترکیب‌ها در پنج محدوده سنی با استفاده از داده‌های حاصل از کل نمونه هنجاریابی محاسبه شد. ضرایب آلفا برای ترکیب‌ها با استفاده از فرمول *گیلفورد* (۱۹۵۴، ص ۳۹۳) که برای برآورد آلفای نمرات مرکب طراحی شده است به‌دست آمد. این نتایج در جدول ۳، گزارش شده است. ضرایب آلفا برای کل نمونه از طریق میانگین‌گیری ضرایب آلفای سطوح سنی مختلف با استفاده از روش تبدیل z به‌دست آمد.

همیشه نمی‌توان گفت که چون یک آزمون برای کل جامعه معتبر است، پس برای هر یک از زیر گروه‌های درون آن جامعه نیز از اعتبار یکسانی برخوردار است. بنابراین سازندگان آزمون باید ثابت کنند که آزمونشان در واقع برای گروه‌های فرعی، به‌ویژه زیر گروه‌هایی که

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

احتمالاً مورد ارزیابی قرار می‌گیرند یا کسانی که به‌دلیل اختلاف نژادی، معلولیتی و زبانی ممکن است مورد مناسبی برای سنجش نباشند نیز معتبر است. در جدول ۴، ضرایب آلفا برای ۶ زیر گروه گزارش شده است.

زیر گروه‌های مورد مطالعه را پسران، دختران، کودکانی که مبتلا به نقص توجه / بیش فعالی، اختلال یادگیری، تأخیر در گفتار و زبان و کم‌توانی ذهنی تشخیص داده شده بودند، تشکیل می‌دهند.

جدول ۴- ضرایب اعتبار خرده‌آزمون‌ها و ترکیب‌های دو گروه اصلی و چهار گروه فرعی

	زیر گروه‌ها					
	پسران	دختران	نقص توجه بیش فعالی	ناتوانی یادگیری	تأخیر در گفتار و زبان	کم‌توانی ذهنی
خرده‌آزمون‌ها						
واژگان تصویری	۷۷	۷۹	۸۶	۸۴	۹۲	۸۴
واژگان ربطی	۹۱	۹۰	۷۰	۸۸	۷۵	۸۳
واژگان شفاهی	۹۱	۹۱	۸۱	۸۷	۸۴	۸۶
درک دستوری	۷۸	۷۸	۸۶	۸۳	۸۳	۸۷
تقلید جمله	۹۳	۹۲	۹۳	۹۶	۷۹	۸۶
تکمیل دستوری	۸۶	۸۸	۸۶	۸۰	۷۸	۸۵
تمایزگذاری کلمه	۹۲	۹۰	۹۸	۷۷	۸۲	۹۶
تحلیل واجی	۹۶	۹۶	۹۵	۸۳	۹۱	۹۸
تولید کلمه	۸۷	۷۷	۹۶	۸۰	۸۷	۸۴
ترکیب‌ها						
گوش کردن	۸۶	۸۷	۸۶	۸۵	۸۵	۸۲
سازمان‌دهی	۹۵	۹۴	۹۰	۹۶	۸۳	۸۷
صحبت کردن	۹۴	۹۴	۹۱	۹۰	۸۱	۹۳
معنی‌شناسی	۹۵	۹۵	۷۶	۹۴	۷۹	۷۹
نحو	۹۵	۹۵	۸۰	۹۳	۸۰	۸۶
زبان گفتاری	۹۷	۹۷	۸۶	۹۷	۸۷	۸۷

۲-۱- نمونه‌گیری زمانی

خطای ناشی از نمونه‌گیری زمانی به میزان ثبات عملکرد فرد در آزمون، در طول زمان

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

اشاره دارد. این خطا معمولا با روش آزمون - بازآزمون برآورد می‌گردد. ثبات زمانی آزمون حاضر نیز با استفاده از روش آزمون - باز آزمون مورد مطالعه قرار گرفت. تعداد افراد مورد مطالعه در این مرحله ۹۶ نفر بودند که به‌طور تصادفی از بین کل افراد مورد استفاده در هنجاریابی انتخاب گردیدند. فاصله بین دو اجرای آزمون ۳ ماه بود. نتایج حاصل در جدول ۵، ارائه گردیده است. در این جدول میانگین و انحراف استاندارد خرده‌آزمون‌ها و ترکیب‌ها در هر اجرای آزمون و همچنین ضرایب همبستگی (r) که نشان دهنده میزان رابطه بین دو اجرای آزمون است، ارائه گردیده است.

جدول ۵- نتایج اعتبار بازآزمایی

	اجرای اول		اجرای دوم		
	M	SD	M	SD	r
خرده‌آزمون‌ها					
واژگان‌تصویری	۲۲/۹	۴/۲	۲۳/۲	۴/۱	۰/۷۸
واژگان‌ریطی	۱۳	۶/۳۵	۱۵/۶	۶	۰/۷۸
واژگان‌شفاهی	۱۴/۱	۶	۱۱/۳	۶/۱	۰/۸۵
درک‌دستوری	۱۷/۳	۴/۴	۱۸/۳	۴/۳	۰/۸۲
تقلیدجمله	۱۵/۶	۷/۸	۱۷/۵	۷/۶	۰/۸۶
تکمیل‌دستوری	۱۴/۷	۶/۳	۱۶/۷	۵/۸	٪۸۵
تمایزگذاری‌کلمه	۲۲/۶	۵/۲	۲۴/۱	۳	۰/۸۶
تحلیل‌واجی	۱۰/۱۸	۵/۴	۱۰/۹	۴/۹	۰/۸۴
تولیدکلمه	۱۸/۳	۲/۶	۱۸/۸	۱/۶	۰/۸۲
ترکیب‌ها					
گوش‌کردن	۴۰/۳	۸/۳	۴۱/۵	۷/۸	۰/۸۸
سازمان‌دهی	۲۸/۶	۱۲/۵	۳۳	۱۲/۵	۰/۸۸
صحبت‌کردن	۲۵/۹	۱۱/۴	۲۸/۱	۱۱	۰/۸۵
معنی‌شناسی	۴۷/۲	۱۴/۸	۵۰/۱	۱۴/۹	۰/۸۴
نحو	۴۷/۷	۱۶/۸	۵۲/۶	۱۵/۸	۰/۸۷
زبان‌گفتاری	۹۴/۸	۲۹/۹	۹۴/۸	۳۰/۹	۰/۸۸

خلاصه‌ای از ضرایب اعتباردر جدول ۶، ارائه شده است. این جدول وضعیت آزمون را در ارتباط با دو منبع خطا یعنی خطاهای محتوایی و زمانی نشان می‌دهد. میانگین ضرایب با استفاده از روش تبدیل z فیشر (گیلفورد و فروچر، ۱۹۷۸، ص ۳۹۳) به‌دست آمده است.

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]

جدول ۶- خلاصه‌ای از اعتبار آزمون در ارتباط با دو منبع خطای آزمون*

	منابع خطای آزمون		
	نمونه‌برداری محتوایی	نمونه‌برداری زمانی	متوسط ضرایب
خرده‌آزمون‌ها			
واژگان تصویری	۷۶	۷۸	۷۷
واژگان ربطی	۸۹	۷۸	۸۴
واژگان شفاهی	۸۹	۸۵	۸۷
درک دستوری	۷۴	۸۲	۷۸
تقلید جمله	۹۰	۸۶	۸۸
تکمیل دستوری	۸۱	۸۵	۸۳
تمایزگذاری کلمه	۹۰	۸۶	۸۸
تحلیل واجی	۹۴	۸۴	۸۹
تولید کلمه	۸۴	۸۲	۸۳
ترکیب‌ها			
گوش کردن	۸۲	۸۸	۸۵
سازمان‌دهی	۹۲	۸۸	۹۰
صحبت کردن	۹۰	۸۵	۸۸
معنی‌شناسی	۹۳	۸۴	۸۸
نحو	۹۱	۸۷	۸۹
زبان گفتاری	۹۶	۸۸	۹۲

* اعشار حذف شده است.

□ □ □

یادداشت‌ها

1) *Test of Language Development*

2) *Test of Language Development- Primary*

3) *Test of Language Development - Intermediate*

4) *Composite score*

5) *Integrating- Mediating*

6) *Mediational*

7) *Associative*

8) *Pretryout*

9) *Tryout*

10) *Age equivalents*

11) *Extrapolation*

12) *Interpolation*

13) *Reliability*

□ □ □

[Downloaded from mail.joec.ir on 2026-10-21]

[DOR: 20.1001.1.16826612.1380.1.1.3.1]