

مقاله پژوهشی اصیل

ساخت و رواسازی آزمون تشخیصی حساب‌نارسایی برای کودکان پایه سوم دبستان

حجت‌الله فراهانی^۱
دکتر مختار ملک‌پور
دکتر شعله امیری
دکتر حسین مولوی
دکتر مهرداد کلانتری

هدف: هدف این پژوهش ساخت و رواسازی آزمون تشخیصی حساب‌نارسایی پایه سوم دبستان شهر اصفهان بود. **روش:** در این پژوهش پیمایشی، کتاب ریاضی پایه سوم دبستان تحلیل محتوا شد. ضریب کربیندورف مبنی بر توافق تحلیل محتوا ۰/۹۱ به دست آمد. پس از آن بر اساس مقوله‌های استخراج‌شده و اهداف رفتاری هر مقوله و با توجه به نظریه شناختی بلوم و حجم هر مقوله در کتاب ریاضی پایه سوم تعداد زیادی سؤال طرح و به‌طور مقدماتی اجرا گردید. سپس فرم نهایی روی یک نمونه ۳۰۰ نفری از دانش‌آموزان پایه سوم دبستان (۱۵۵ پسر و ۱۴۵ دختر) که در سال تحصیلی ۸۶-۱۳۸۵ به روش نمونه‌برداری خوشه‌ای چندمرحله‌ای از نواحی پنج‌گانه آموزشی شهر اصفهان انتخاب شده بودند، اجرا و پارامترهای سؤال آزمون بر اساس نظریه کلاسیک آزمون (CTT) و نظریه سؤال - پاسخ (IRT) محاسبه گردید. **یافته‌ها:** نتایج نشان داد که این آزمون دارای روایی مطلوب و نیز پایایی مناسب ($\alpha=0/93$)، ضریب بازآزمایی ($=0/87$) است و حساسیت و ویژگی (دقت) آن به ترتیب ۰/۹۸ و ۰/۹۶ به دست آمد. پارامتر تشخیصی $a(g)$ و دشواری $b(g)$ بر اساس نظریه سؤال - پاسخ نیز نشان داد که همه سؤال‌ها دارای توان تشخیص مناسب (بالای ۰/۵) و دشواری مطلوب‌اند و با مدل دو پارامتری برازش کامل دارند ($p>0/05$). منحنی تابع آگاهی آزمون نیز نشان داد که مناسب‌ترین کاربرد آن در سطوح توانایی ریاضی $R=-0/6$ تا $R=+0/6$ است. **نتیجه‌گیری:** این آزمون دارای ویژگی‌های روان‌سنجی مناسب برای تشخیص کودکان حساب‌نارسای پایه سوم می‌باشد.

کلید واژه‌ها: آزمون تشخیصی، حساب‌نارسایی، نظریه کلاسیک آزمون (CTT)، نظریه سؤال - پاسخ (IRT)

مقدمه

کودکان سبب انطباق مواد درسی و روش آموزش با سطح یادگیری دانش‌آموز می‌گردد. از طریق سنجش و اندازه‌گیری روایی و پایایی می‌توان برنامه آموزشی درست و جامعی تدوین کرد که به بهترین روش درک مفاهیم نظری ریاضی را با کاربرد آن در زندگی روزمره همراه سازد.

دشواری در فهم روابط عددی برای برخی از کودکان از همان سال‌های نخست زندگی که سایر کودکان در شمارش، درک واحد، جورکردن و مقایسه تجاربی به دست می‌آورند، آغاز می‌شود که می‌تواند کارکردی پایین‌تر از حدانتظار را در آزمون‌های پیشرفت

ریاضی یکی از درس‌هایی است که با وجود توجه زیادی که در مدارس به آن می‌شود، یادگیری آن برای بسیاری از دانش‌آموزان مشکل است. این در حالی است که معلمان می‌توانند با ارزشیابی درست توانایی‌های ریاضی دانش‌آموزان، بازده تدریس خود و در نتیجه موفقیت تحصیلی و پیشرفت ریاضی دانش‌آموزان را افزایش دهند. زیرا شناسایی نقاط ضعف و نارسایی‌های ریاضی

۱- نشانی تماس: اصفهان، خیابان هزارجریب، گروه روانشناسی، دانشگاه اصفهان.

Email: icphaf@yahoo.com

حجت‌الله فراهانی و همکاران

هنجاریایی گردید (محمد اسماعیل و هومن، ۱۳۸۱). آزمون دیگری که برای سنجش شایستگی ریاضی طراحی شده، آزمون شایستگی اوترخت برای ریاضیات کودکان است، که شامل هشت مقیاس و در مجموع ۱۲۰ سوال است (کرپرال^۴، ۱۹۹۵؛ اندوریخت^۵، ۱۹۹۲، به نقل از عریضی، کاووسیان و کدیور، ۱۳۸۳). این آزمون را عریضی و همکاران (۱۳۸۳) پایایی سنجی و رواسازی کرده‌اند. در پژوهش دیگری چین‌آوه، عرب‌زاده و منشی‌زاده (۱۳۸۳) به ساخت و هنجاریایی آزمون تشخیصی ریاضی برای دانش‌آموزان ۶ تا ۱۳ ساله شهر شیراز پرداختند. آنچه در این پژوهش مورد توجه قرار گرفت، این بود که با توجه به تنوع این اختلال و یک‌دست نبودن ویژگی‌های آن (احدی و کاکاوند، ۱۳۸۲)، به نظر می‌رسد استفاده از یک آزمون استاندارد ملی پژوهشگران را به نتیجه مطلوبی نرسانده و بر روایی محتوایی این آزمون‌ها تأکید خاص نشده است. از سوی دیگر بنا به نظر باترورث^{۱۶} (۲۰۰۳، ۲۰۰۵)، جوردن و مونتانی^{۱۷} (۱۹۹۷)، لندرل^{۱۸}، بوان^{۱۹} و باترورث (۲۰۰۴) و آشکرافت^{۲۰}، یاماشیتا^{۲۱} و آرام^{۲۲} (۱۹۹۲) اگر در مقوله شمارش زمان لحاظ نگردد، کودکان حساب‌نارسان نیز می‌توانند مانند سایر کودکان عادی به سؤال‌ها پاسخ درست بدهند، زیرا آنها از راهکارهای نادرست یعنی شمارش با انگشتان استفاده کرده و در مدت زمان طولانی‌تر پاسخ درست را ایجاد می‌کنند. بنابراین در این پژوهش، خرده‌آزمون شمارش زمان‌مند گردید و از سوی دیگر برای دستیابی به پارامترهای دقیق سؤال (ضرایب دشواری و تشخیصی)، جامعه آماری آزمودنی‌ها (تتا) و تعیین منحنی آگاهی آزمون علاوه بر نظریه کلاسیک آزمون^{۲۳} (CTT) از نظریه سؤال-پاسخ^{۲۴} (IRT) نیز استفاده شد. در ایران بر اساس نظریه سؤال-پاسخ (IRT) در زمینه ریاضیات هیچ آزمونی ساخته

تحصیلی دوران دبستان برای آنان پیش‌بینی کند (رمضانی، ۱۳۸۰). حساب‌نارسایی که به عنوان یک اختلال در سومین نسخه دستنامه تشخیصی و آماری اختلالات روانی^۱ (DSM-III) (یعنی از سال ۱۹۸۰ به بعد) مطرح گردید، عبارت است از ناتوانی در انجام مهارت‌های حساب با توجه به ظرفیت هوش و سطح آموزشی مورد انتظار از کودک که این مهارت‌ها می‌بایست به کمک آزمون‌های میزان‌شده فردی اندازه‌گیری شده باشد. بر اساس ویراست چهارم اصلاح‌شده دستنامه تشخیصی و آماری اختلالات روانی^۲ (DSM-IV-TR) این کودکان در چهار گروه از مهارت‌های زبانی، ادراکی، ریاضی و توجهی مرتبط با ریاضیات مشکل دارند.

با وجود اینکه این اختلال مشکل جدیدی تلقی نمی‌شود، اما ماهیت آن تا حد زیادی مبهم مانده است (فوکس^۳ و فوکس، ۲۰۰۵). به همین دلیل پژوهشگران از این اختلال تقسیم‌بندی‌های متعددی ارائه داده‌اند (جوردن^۴، هانیک^۵ و کاپلان^۶، ۲۰۰۳؛ روسلی^۷ و آردیلا^۸، ۱۹۸۹). فوکس و فوکس (۲۰۰۵) همه‌گیری اختلال ریاضی را در دبستان پنج درصد، رمضانی (۱۳۸۲) و حمید (۱۳۸۵) در شهر تهران به ترتیب ۱/۶۳ و ۳/۶ درصد برآورد کرده‌اند. در سبب‌شناسی این اختلال فرضیه‌هایی مطرح شده که در مجموع می‌توان به ترکیبی از تأثیرات محیط و ژنتیک اشاره کرد (روسولی^۹، نوئل^{۱۰}، ۲۰۰۷). بنا به تأکید انجمن روان‌پزشکی آمریکا تشخیص قطعی فقط پس از اجرای آزمون استاندارد شده انفرادی در ریاضی و کسب نمره‌ای بسیار پایین‌تر از سطح مورد انتظار (بر حسب میزان تحصیلات و ظرفیت هوشی کودک) امکان‌پذیر است (کاپلان و سادوک^{۱۱}، ۱۳۸۳). بنابراین تشخیص صحیح و دقیق آن نیازمند اجرای یک آزمون تشخیصی استاندارد می‌باشد. برای سنجش این اختلال روش‌های رسمی و غیررسمی متعددی طراحی شده است که به عنوان نمونه می‌توان به آزمون ریاضی کی‌مت^{۱۲} اشاره نمود که کنولی^{۱۳} (۱۹۸۸)، به نقل از رمضانی، ۱۳۸۲) آن را در سال ۱۹۸۸ هنجاریایی کرد. این آزمون که به منظور تعیین نقاط قوت و ضعف دانش‌آموزان در حوزه‌های مختلف ریاضی به کار می‌رود، شامل ۱۳ خرده‌آزمون و سه بخش مفاهیم اساسی، عملیات و کاربردها است که از اهمیت آموزشی تقریباً یکسانی برخوردارند. این آزمون در ایران برای دانش‌آموزان

1- Diagnostic and Statistical Manual of Mental Disorders-Third Edition
 2- Diagnostic and Statistical Manual of Mental Disorders-Text Revision
 3- Fuchs
 4- Jordan
 5- Hanich
 6- Kaplan
 7- Rosselli
 8- Ardila
 9- Rousselle
 10- Noel
 11- Sadock
 12- Key Math
 13- Conolly
 14- Korporal
 15- Andoricht
 16- Butterworth
 17- Montani
 18- Landrel
 19- Bevan
 20- Ashcraft
 21- Yamashita
 22- Aram
 23- Classic Test Theory
 24- Item-Response Theory

نشده و این اولین پژوهش در این راستا است. هدف این رویکرد آن است که علاوه بر برآورد دقیق پارامترهای آزمون و سؤال، نارسایی‌های نظریه کلاسیک آزمون، از جمله وابستگی پارامتر سؤال به نمونه خاص را از میان بردارد (هومن، ۱۳۷۱). بنابراین، مطالعه حاضر به منظور طراحی آزمون تشخیصی حساب‌نارسایی برای کودکان پایه سوم دبستان شهر اصفهان و بررسی ویژگی‌های روان‌سنجی و هنجاریابی آن برای این کودکان انجام شد.

روش

جامعه، نمونه آماری و روش نمونه‌گیری

جامعه آماری این پژوهش کلیه کودکان دختر و پسر دبستانی پایه سوم شهر اصفهان بودند که در سال تحصیلی ۸۶-۱۳۸۵ در مدارس ابتدایی این شهر تحصیل می‌کردند (جدول ۱). از این تعداد با استفاده از فرمول موزر^۱ و کالتون^۲ و با در نظر گرفتن دو درصد برآورد نسبت همه‌گیرشناسی این اختلال در جامعه و یک درصد خطای مورد پذیرش، ۳۰۰ دانش‌آموز (۱۵۵ پسر و ۱۴۵ دختر) از نواحی پنج‌گانه به‌عنوان نمونه انتخاب شدند (رمضانی، ۱۳۸۲).

در جدول ۱ تعداد افراد جامعه مورد بررسی به تفکیک ناحیه آموزشی و جنسیت آمده است. بدین صورت که با مراجعه به اداره کل آموزش و پرورش اصفهان فهرست مدارس ابتدایی هر پنج ناحیه تهیه شد و سپس با توجه به جمعیت دانش‌آموزان پایه سوم، از هر ناحیه چند مدرسه و از میان آنها یک یا چند کلاس به‌صورت تصادفی انتخاب گردیدند.

ابزار پژوهش و مراحل ساخت آن

برای تهیه پرسش‌های این آزمون ابتدا کتاب ریاضی پایه سوم به‌وسیله تعدادی از اساتید روان‌شناسی و علوم تربیتی، متخصصان

روان‌سنجی، پژوهشگران و سرگروه‌های آموزش ریاضی پایه سوم تحلیل محتوا شد و ضریب توافق کرپندورف^۱ ۰/۹۱ به‌دست آمد که نشان‌دهنده توافق افراد در تحلیل محتوای این کتاب است. سپس عناوین مجدداً به تأیید سرگروه‌های آموزشی رسید و سپس سؤال‌ها با توجه به نظریه شناختی بلوم^۲ (۱۹۶۵) که همسو با نظریه گانیه^۳ (۱۹۷۰)، به نقل از چین‌آوه و همکاران، (۱۳۸۳) است، طرح گردید. تعداد سؤال‌های هر مقوله، با توجه به حجم آن در کتاب ریاضی تعیین شد. این مقولات شامل نه حیطه اعداد، جمع و تفریق، ارزش مکانی، اندازه‌گیری، ضرب و تقسیم، هندسه، کسر و اعشار، نسبت و درصد، تخمین و حل مسأله بود. سپس این سؤال‌ها مورد بررسی مقدماتی قرار گرفتند و بعد روی یکصد دانش‌آموز (۵۰ دختر و ۵۰ پسر) اجرا شدند. پرسش‌ها به‌صورت انفرادی و تشریحی اجرا گردید و برای هر پاسخ درست نمره یک و برای هر پاسخ غلط نمره صفر منظور شد. نتایج تحلیل و پرسش‌ها از نظر سطح دشواری (p)، ضریب تشخیص (d)، همسانی درونی (α) و روایی، همزمان با آزمون‌کی‌مت ریاضی ایران (محمد اسماعیل و هومن، ۱۳۸۱) و نمره حساب آزمون و وکسلر کودکان (وکسلر، ۱۳۷۳) محاسبه گردید. سپس سؤال‌های دارای ضریب بالا (α) (بالای ۰/۸) و ضرایب تشخیص معنادار (بالای ۰/۴) انتخاب شدند. برای کمک به استقلال موضعی سؤال‌ها (شرط لازم برای تحلیل سؤال - پاسخ) سؤال‌ها بر اساس درجه دشواری به‌طور تصادفی در هر خرده‌آزمون قرار داده شدند.

در پایان ۵۰ سؤال باقی‌ماند و از آنجا که ریاضی یک مجموعه به‌هم پیوسته تلقی می‌شود، آزمون تشخیصی حساب‌نارسایی پایه دوم (۳۰ سوال) نیز که بر اساس اصولی مشابه این آزمون ساخته شده بود،* به سؤال‌ها اضافه و در مجموع ۸۰ سؤال به‌عنوان آزمون تشخیصی حساب‌نارسایی پایه پنجم اجرا گردید. برای آگاهی از وضعیت هوشی و ادراکی - دیداری شرکت‌کنندگان علاوه بر اجرای آزمون هوش کودکان

جدول ۱- تعداد افراد جامعه مورد بررسی به تفکیک ناحیه آموزشی و جنسیت

ناحیه	۱	۲	۳	۴	۵
پسر	۹۹۳	۱۵۵۸	۲۷۵۳	۳۶۰۰	۳۳۰۸
دختر	۱۰۰۵	۱۵۳۳	۲۵۲۸	۳۳۲۱	۳۱۰۳

۱- Mozer
 ۲- Kalton
 ۳- Krippendorff
 ۴- Bloom
 ۵- Ganie
 * این مقاله قسمتی از پایان‌نامه دکترای روان‌شناسی با تأکید بر روان‌سنجی بود که ضمن آن برای هر پنج پایه دبستان، یک آزمون تشخیصی حساب‌نارسایی ساخت و رواسازی شد.

حجت‌الله فراهانی و همکاران

دو نیمه‌سازی محاسبه و نقاط برش، حساسیت و ویژگی (دقت) سؤال‌ها از طریق تحلیل تمیزی^۳ تعیین شد. منظور از حساسیت یک آزمون امکان تشخیص افراد حساب‌نارسانا از افراد عادی و منظور از ویژگی (دقت) یک آزمون، توان یک آزمون در تشخیص موارد سالم است. پارامترهای سؤال در این پژوهش براساس مدل دو پارامتری سؤال- پاسخ (IRT) برآورد شد. ضرایب تشخیص سؤال‌ها $a(g)$ ، پارامتر دشواری $b(g)$ و عامل حدس $c(g)$ بود که به دلیل ماهیت تشریحی این آزمون عامل حدس در پاسخگویی تأثیر نداشت و صفر در نظر گرفته شد. بنابراین از مدل دو پارامتری استفاده شد. در این مدل، پارامتر تشخیص سؤال به قدرت سؤال‌های یک آزمون در تشخیص و جداسازی افراد قوی و ضعیف اشاره دارد (آلن^۴ و ین^۵، ۱۳۷۴). معمولاً دامنه $a(g)$ در عمل بین ۰/۵ تا ۲/۵ تغییر می‌کند که مقادیر کمتر از ۰/۵ نشان‌دهنده توان تشخیص پایین سؤال‌ها و مقادیر بزرگتر نیز (که به ندرت به دست می‌آید) (هومن، ۱۳۷۱)، پارامتر دشواری سؤال $b(g)$ است که معمولاً بین ۲/۵- تا ۲/۵+ مقیاس‌پردازی می‌شود. پارامتر دشواری سؤال نشان‌دهنده نقطه عطف منحنی شاخص سؤال است که ۵۰ درصد آزمودنی‌هایی که در آن سطح قرار دارند جواب سؤال را می‌دانند یا می‌توانند جواب آن را پیدا کنند (ستاری، ۱۳۸۲).

برای تعیین برآزش سؤال‌ها با مدل دو پارامتری سؤال - پاسخ مقادیر مجذور کای محاسبه و سطح معناداری آنها تعیین شد. بررسی داده‌ها نشان داد که برای احراز دو پیش فرض تحلیل IRT یعنی استقلال موضعی^۶ و تک‌بعدی بودن^۷ آزمون، بیش از ۲۰ درصد واریانس آزمون در یک عامل قرار می‌گیرد که این نشان‌دهنده برقراری فرض تک‌بعدی بودن است که اگر رعایت شود استقلال نیز قدری رعایت می‌شود (باقی، ۱۹۹۵)، به نقل از ستاری، ۱۳۸۲). برای کمک به برقراری این فرض سؤال‌هایی با درجه دشواری‌های تصادفی قرار داده شدند. در تمام مراحل از نرم‌افزارهای SPSS، Testfact و Bilog استفاده شد.

(وکسلر، ۱۳۷۳) و آزمون ادراکی- دیداری آندره‌ری (میرهاشمی، ۱۳۷۱) که از روایی و پایایی مطلوبی برخوردار بودند، نمرات روخوانی فارسی، انشا و دیکته نیز از آنها پرسیده شد. همچنین طی یک مصاحبه بالینی از آنها خواسته شد متن مشخصی را بخوانند و آن را برای آزمونگر بازگو کنند و در مواردی که شرکت‌کننده دچار مشکل ادراکی، بالینی یا هوشی بود از نمونه خارج و به جای آن فرد دیگری به‌طور تصادفی انتخاب و جایگزین می‌شد. برای تعیین زمان مطلوب در خرده‌آزمون شمارش، مدت زمانی که ۷۰ درصد شرکت‌کنندگان (آناستازی^۱، ۱۳۷۳؛ ثرن‌دایک^۲، ۱۳۷۵؛ سیف، ۱۳۷۵) قادر به پاسخ‌گویی سؤال‌ها بودند، به‌عنوان زمان مطلوب در نظر گرفته شد. برای تعیین نمونه دارای حساب‌نارسایی بر اساس روش غیرآزمون، به‌منظور تعیین نقطه برش، حساسیت و ویژگی (دقت) و روایی تشخیصی سؤال‌های آزمون تشخیصی نهایی، از آموزگار پایه سوم کلاس مربوطه خواسته شد بر مبنای شناخت فردی، دانش آموزشی را که فقط در درس ریاضی با مشکل روبه‌رو بودند، معرفی کند. نمره کمتر از ۱۲ (۱۰ درصد پایین توزیع) به‌عنوان اولین ضابطه تعیین گردید و سپس از آموزگار خواسته شد که به فهرست واریسی تشخیصی DSM-IV-TR برای هر مورد پاسخ دهد و پس از پاسخ‌دهی، ماده‌هایی که با بلی پاسخ داده می‌شدند جمع می‌شدند و در صورتی که آزمودنی حداقل چهار ماده (از ۱۱ ماده فهرست واریسی تشخیصی) را با «بلی» پاسخ داده بود به‌عنوان حساب‌نارسایی تلقی می‌شد. بدین ترتیب ۲۵ نفر تعیین شدند.

این آزمون به‌وسیله پژوهشگر و دانشجویان کارشناسی ارشد و کارشناسی که در مورد ایجاد رابطه و برقراری ارتباط مؤثر و همدلانه با کودک و اجرای آزمون تجربه داشتند و آموزش کامل دیده بودند، در هفته‌های پایانی سال تحصیلی ۸۶-۱۳۸۵ اجرا گردید.

پارامترهای سؤال در این پژوهش در نظریه کلاسیک آزمون (CTT)، ضریب دشواری (یعنی نسبت افرادی که به سؤال‌ها پاسخ درست دادند، و ضریب تشخیص (روایی) سؤال بود که از طریق همبستگی دو رشته‌ای نقطه‌ای سؤال با نمره کل به‌دست آمد. پایایی از طریق آلفای کرونباخ، بازآزمایی و

1- Anastasia
3- discrimination analysis
5- Yen
7- unidimensionality

2- Thorndike
4- Allen
6- local independence

يافته‌ها

كه حاكي از روايي همزمان مطلوب خرده‌آزمون‌ها و نمره كل اين آزمون است.

همچنين همبستگي بين آزمون تشخيصي حساب‌نارسايي پايه سوم و خرده‌آزمون حساب در آزمون هوش ويسك-آر نيز معنادار بود كه نشان‌دهنده روايي همزمان و مطلوب اين آزمون است ($r=0/79$ ، $p<0/001$).

براي تعيين روايي تشخيصي سؤال‌هاي آزمون تشخيصي حساب‌نارسايي، اين آزمون روي افراد عادي و گروه حساب‌نارسا كه بر اساس فهرست وارسى DSM-IV-TR و با نمره کمتر از ۱۲ در رياضي انتخاب شده بودند، اجرا و تحليل شد. تحليل MANOVA سؤال‌ها نشان داد كه بر اساس همه سؤال‌ها مي‌توان افراد حساب‌نارسا و عادي را به‌طور معناداري ($p<0/05$) تفكيك كرد. اگر آزموني اين ويژگي را داشته باشد، داراي روايي سازه است (سرمد، بازرگان و حجازي، ۱۳۷۶).

پايابي آزمون تشخيصي حساب‌نارسايي: براي تعيين پايابي اين آزمون از سه روش آلفاي كرونباخ، دونيمه‌سازي (همساني دروني) و بازآزمائي (ثبات) استفاده شد. ضرايب آلفاي كرونباخ در خرده‌آزمون‌ها در محدوده ۰/۷۴-۰/۹۲ و در كل ۰/۹۳، در روش دونيمه‌سازي در محدوده ۰/۷۱-۰/۹۱ و در كل ۰/۸۹ و در بازآزمائي در محدوده ۰/۸۲-۰/۹۲ و در كل ۰/۹۲ به‌دست آمد. براي تعيين ضريب بازآزمائي، شركت‌كنندگان پس از چهار هفته مجدداً آزمون شدند.

نقطه برش، حساسيت و ويژگي (دقت) آزمون تشخيصي حساب‌نارسايي: نقطه برش در خرده‌آزمون‌ها در محدوده ۱۰-۲ و در كل ۵۱، حساسيت خرده‌آزمون‌ها در محدوده ۹۸-۸۸ درصد و در كل ۹۸ درصد و ويژگي (دقت) خرده‌آزمون‌ها در محدوده ۹۶-۸۴ درصد و در كل ۹۶ درصد به‌دست آمد.

حساسيت و ويژگي (دقت) آزمون در حد بسيار مطلوبي بود، يعني با اين آزمون مي‌توان ۹۹ درصد موارد حساب‌نارسا و ۹۶ درصد موارد سالم را به درستي تشخيص داد.

ضرايب تشخيصي و دشواري سؤال‌هاي آزمون براساس نظريه كلاسيك آزمون (CTT) و سؤال - پاسخ (IRT): ضرايب تشخيصي و دشواري سؤال‌هاي آزمون براساس نظريه كلاسيك آزمون (CTT) و سؤال-پاسخ (IRT) در جدول ۳ آورده شده است.

روايي صوري و محتوايي آزمون: براي تعيين روايي صوري و محتوايي اين آزمون مراحل زير انجام شد:

(الف) كتاب رياضي پايه سوم دبستان تحليل محتوا و مقوله‌هاي مختلف استخراج گرديد.

(ب) ضريب توافق كريندورف مبني بر توافق تحليل محتوا محاسبه شد كه ۰/۹۱ به‌دست آمد.

(ج) اهداف آموزشي و رفتاري بر اساس مقوله‌هاي استخراج شده تعريف و تدوين شدند.

(د) بر اساس بند «ج» مهارت‌هايي كه نشان‌دهنده محتواي مورد نظر بودند، تعيين شدند.

(ه) از ميان سؤال‌هاي امتحان‌هاي رياضي دو نوبت (صبح و عصر) پايه سوم دبستان‌هاي نواحي پنج‌گانه آموزش و پرورش شهر اصفهان خزانه‌اي از سؤال‌هاي پنج سال اخير تهيه و بررسي شد.

(و) بر اساس مراحل «الف» تا «ه» مجموعه‌اي از سؤال‌ها طراحي و تنظيم شدند.

(ز) سؤال‌هاي نهايي به‌وسيله اساتيد گروه علوم تربيتي و روان‌شناسي، متخصصان روان‌سنجي، اساتيد راهنما و مشاور و سرگروه‌هاي آموزش رياضي پايه سوم دبستان مورد بازنگري نهايي قرار گرفت و نسبت روايي محتوايي (CVR_i) بر اساس توافق در تعيين روايي محتوايي سؤال‌ها محاسبه و دامنه‌اي از ۰/۷۹-۰/۸۲ به‌دست آمد كه معنادار و حاكي از روايي محتوايي بالا و مطلوب اين آزمون بود.

(ح) پس از اجراي مقدماتي، سؤال‌هاي مبهم (از نظر دانش‌آموزان) شناسايي، مجدداً عبارت‌بندي و در صورت لزوم در محتواي آنها تغيير داده شد.

روايي آزمون تشخيصي حساب‌نارسايي پايه پنجم: براي تعيين روايي همزمان اين آزمون از خرده‌آزمون‌هاي مرتبط در آزمون رياضي ايران كي‌مت (محمد اسماعيل و هومن، ۱۳۸۱) و خرده‌آزمون حساب در آزمون هوش ويسك-آر (وكسلر، ۱۳۷۳) استفاده شد. نتايج نشان داد كه همبستگي بين همه خرده‌آزمون‌ها و نمره كل اين آزمون و خرده‌آزمون‌هاي مرتبط در آزمون رياضي ايران كي‌مت در محدوده ۰/۸۷-۰/۸۹ و معنادار بودند ($p<0/001$).

جدول ۲- ضرایب تشخیصی و دشواری سؤالات، مقادیر کای دو و سطح معناداری

ضریب گویه							ضریب گویه							
df	P	χ^2	b(g)	a(g)	r_{pbis}	pi	df	P	χ^2	b(g)	a(g)	r_{pbis}	pi	
۸	۰/۲۳۸	۱۱/۷۱	۱/۵۲۹	۱/۴۸۳	۰/۷۵	۰/۶۷	سوال ۴۱	۸	۰/۱۲۱	۱۳/۲۱	-۰/۴۶۲	۱/۸۴۳	۰/۷۴	۰/۵۷
۸	۰/۴۵۱	۹/۳۱	۰/۴۷۲	۲/۴۷	۰/۵۴	۰/۵۵	سوال ۴۲	۹	۰/۲۲۱	۱۲/۲۴	۰/۷۲۹	۱/۸۲۹	۰/۷	۰/۴۱
۸	۰/۱۳۰	۱۲/۸	۱/۰۲۶	۰/۷۵۷	۰/۵۳	۰/۶۳	سوال ۴۳	۸	۰/۴۳۹	۱۱/۷۱	۰/۹۷	۱/۸۸۸	۰/۷۶	۰/۴۰
۸	۰/۲۸۱	۱۰/۳۲	۱/۹۴۵	۰/۷۸	۰/۴۵	۰/۵۸	سوال ۴۴	۹	۰/۴۳۷	۱۱/۸۴	۰/۹۸۴	۱/۸۴۳	۰/۷۲	۰/۳۸
۹	۰/۱۰۴	۱۴/۵	۰/۷۱۶	۰/۵۴	۰/۶۲	۰/۴۳	سوال ۴۵	۸	۰/۲۱۰	۱۲/۷۲	۰/۷۱۶	۱/۸۷۵	۰/۷۵	۰/۴۱
۹	۰/۱۷۱	۱۵/۳۲	۱/۳۸۵	۰/۸۹۴	۰/۴۶	۰/۴۹	سوال ۴۶	۸	۰/۱۰۱	۱۴/۳۱	۰/۴۶۷	۱/۸۴	۰/۷۱	۰/۵۳
۹	۰/۰۴	۱۴/۷	۰/۸۳۵	۰/۷۷۱	۰/۶۸	۰/۵۹	سوال ۴۷	۸	۰/۲۸۱	۱۲/۵۱	۰/۴۴۹	۱/۹۰۳	۰/۷۷	۰/۵۲
۹	۰/۰۸۱	۱۵/۸	۱/۸۸۳	۰/۵۰۹	۰/۸۲	۰/۶۱	سوال ۴۸	۸	۰/۳۸۱	۱۱/۲	۰/۴۸۱	۱/۹۳۵	۰/۷۹	۰/۵۳
۹	۰/۱۵۲	۱۱/۳۱	۱/۷۲۱	۰/۶۸۴	۰/۷۳	۰/۶۲	سوال ۴۹	۹	۰/۴۴۹	۹/۳۱	۰/۵۱۲	۲/۳۴	۰/۸۲	۰/۵۴
۹	۰/۱۶۷	۱۲/۴۱	۱/۴۶۶	۰/۷۸۳	۰/۸۲	۰/۶۳	سوال ۵۰	۹	۰/۲۱۹	۱۲/۳۲	۰/۵۱۲	۱/۹۶۹	۰/۷۹	۰/۵۷
۹	۰/۱۰۵	۱۵/۷	۱/۰۶	۱/۴۷۴	۰/۸۱	۰/۶۰	سوال ۵۱	۸	۰/۱۴۱	۱۱/۲۱	۱/۹۴۰	۱/۵۱۷	۰/۷۸	۰/۴۶
۹	۰/۳۵۱	۱۰/۸	۱/۳۵۶	۰/۶۷۵	۰/۴۵	۰/۴۷	سوال ۵۲	۸	۰/۱۳۲	۱۱/۹	۲/۴۲	۰/۷۱۷	۰/۴۵	۰/۳۸
۹	۰/۱۱۷	۱۲/۸	۳/۰۰۰	۱/۴۶۲	۰/۴۳	۰/۵۹	سوال ۵۳	۸	۰/۰۲۱	۱۷/۱	۱/۶۴۵	۱/۴۹	۰/۷۸	۰/۴۶
۶	۰/۲۰۲	۸/۷۱	۱/۵۹۹	۱/۴۸	۰/۴۳	۰/۶۷	سوال ۵۴	۹	۰/۱۰۸	۱۳/۸	۰/۸۴۹	۱/۲۶۱	۰/۷۴	۰/۴۷
۷	۰/۲۱۴	۸/۹۲	۰/۲۶۲	۰/۷۷۵	۰/۴۷	۰/۴۱	سوال ۵۵	۹	۰/۱۳۸	۱۱/۲	۰/۱۰۶	۱/۴۴	۰/۴۲	۰/۳۱
۸	۰/۹۵۵	۱۱/۷	۲/۲۵۹	۰/۵۱۳	۰/۴۴	۰/۶۲	سوال ۵۶	۸	۰/۳۲۱	۱/۹۲	۱/۳۶۶	۱/۵۰۶	۰/۴۵	۰/۷۹
۸	۰/۰۶۷	۱۴/۶	۱/۷۹۸	۰/۵۱۶	۰/۴۷	۰/۳۹	سوال ۵۷	۸	۰/۱۵۳	۱۰/۷	۰/۶۴۵	۰/۵۶۶	۰/۴۸	۰/۴۲
۶	۰/۲۰۷	۸/۴	۲/۵۹	۰/۵۲۴	۰/۵۶	۰/۶۸	سوال ۵۸	۸	۰/۱۲۸	۱۱/۷۹	۰/۴۲۴	۰/۵۲۷	۰/۵۳	۰/۶۳
۶	۰/۲۳۱	۸/۱۷	-۰/۵۸۱	۱/۵۰۱	۰/۵۹	۰/۵۲	سوال ۵۹	۸	۰/۱۱۱	۱۲/۸	۱/۸۲۰	۱/۵۴۲	۰/۵۴	۰/۶۴
۶	۰/۹۵۴	۱/۶	۱/۰۶۴	۰/۶۰۹	۰/۴۹	۰/۴۸	سوال ۶۰	۹	۰/۱۲۸	۱۱/۶۳	۱/۳۵۸	۰/۵۱۰	۰/۵۲	۰/۶۵
۹	۰/۲۱۱	۸/۷۳	۱/۰۱۱	۰/۹۶۸	۰/۶۲	۰/۶۱	سوال ۶۱	۹	۰/۱۶۱	۱۰/۲۱	۰/۵۲۱	۱/۴۵۷	۰/۵۱	۰/۵۶
۶	۰/۱۱۷	۱۱/۹۳	۱/۶۲۲	۲/۵۸	۰/۸۶	۰/۶۰	سوال ۶۲	۸	۰/۲۳۱	۱۰/۸۱	۰/۴۷۲	۲/۴۵۸	۰/۴۴	۰/۵۷
۷	۰/۴۷۹	۶/۵	۱/۶۰۲	۲/۵۵	۰/۴۴	۰/۶۲	سوال ۶۳	۹	۰/۱۲۸	۱۱/۸۳	۰/۴۶۶	۱/۸۴۳	۰/۴۲	۰/۴۸
۷	۰/۰۹۳	۱۳/۵۲	۱/۷۴۵	۰/۷۴۶	۰/۶۸	۰/۵۷	سوال ۶۴	۸	۰/۵۳۲	۸/۷۱	۰/۴۷۲	۲/۵	۰/۵۴	۰/۶۷
۷	۰/۱۲۱	۱۱/۸۱	۰/۳۹۲	۱/۴۹۷	۰/۵۱	۰/۶۸	سوال ۶۵	۸	۰/۴۶۱	۹/۳۲	۰/۴۷۲	۱/۹۷	۰/۴۲	۰/۲۸
۷	۰/۱۱۵	۱۲/۸	-۰/۹	۰/۸۶۴	۰/۴۷	۰/۵۳	سوال ۶۶	۷	۰/۱۳۱	۱۱/۸۲	۰/۵۲۴	۰/۷۸۶	۰/۴۳	۰/۵۳
۶	۰/۲۳۱	۸/۲۰	۱/۳۰۸	۰/۷۶۷	۰/۵۲	۰/۸۱	سوال ۶۷	۹	۰/۱۵۳	۱۰/۷	۰/۴۴۵	۰/۵۶۶	۰/۷۳	۰/۵۷
۷	۰/۲۲۷	۹/۳۷	-۰/۶۲۲	۰/۵۸	۰/۴۴	۰/۵۵	سوال ۶۸	۹	۰/۱۱۲	۱۲/۴	۱/۴۸۵	۱/۵۰۸	۰/۸۲	۰/۶۱
۷	۰/۱۷۳	۱۰/۷	۱/۳۸۵	۰/۷۹۴	۰/۴۵	۰/۶۶	سوال ۶۹	۹	۰/۲۲۱	۱۱/۳	۲/۳۳۳	۰/۵۱۳	۰/۴۷	۰/۴۸
۷	۰/۲۲۷	۱۱/۲۱	۰/۴۳۹	۰/۸۸۹	۰/۶۸	۰/۵۶	سوال ۷۰	۸	۰/۰۸۳	۱۵/۲	۰/۴۱۲	۰/۶۴۹	۰/۴۵	۰/۴۳
۷	۰/۳۷۲	۸/۵۱	۱/۴۲۰	۰/۹۱	۰/۴۷	۰/۸۳	سوال ۷۱	۷	۰/۱۳۷	۱۱/۷۲	۱/۱۶۴	۰/۵۲۳	۰/۴۳	۰/۶۲
۷	۰/۱۱۵	۱۲/۹	۰/۲۴۶	۰/۵۰۶	۰/۸۵	۰/۶۱	سوال ۷۲	۷	۰/۱۲۰	۱۰/۷	۰/۶۴۵	۱/۵۶۶	۰/۶۰	۰/۵۵
۶	۰/۲۳۱	۸/۲۰	۰/۴۴۳	۰/۴۹۸	۰/۴۲	۰/۶۳	سوال ۷۳	۷	۰/۸۵۷	۱۰/۲	۰/۵۶۷	۰/۵۶۶	۰/۴۷	۰/۴۵
۷	۰/۲۲۷	۹/۳۷	۰/۷۱	۱/۴۹	۰/۴۶	۰/۸۲	سوال ۷۴	۷	۰/۱۶۲	۱۰/۸۲	۱/۲۱	۰/۵۱۹	۰/۴۰	۰/۶۳
۷	۰/۱۷۳	۱۰/۷	-۰/۶۵۹	۰/۴۷۸	۰/۴۲	۰/۵۲	سوال ۷۵	۷	۰/۱۶۴	۱۰/۳۱	۱/۲۹۱	۰/۶۰۱	۰/۴۰	۰/۵۶
۹	۰/۳۱۷	۹/۵۳	۰/۱۶۴	۱/۵۲۳	۰/۶۲	۰/۶۲	سوال ۷۶	۷	۰/۱۱۸	۱۱/۸۵	۱/۴۶۲	۰/۵۲۲	۰/۴۱	۰/۴۹
۸	۰/۲۹۴	۱۰/۷۳۱	۱/۶۲۲	۱/۵۸۰	۰/۸۳	۰/۶۳	سوال ۷۷	۸	۰/۲۵۴	۱۰/۲	۱/۹۲۵	۰/۵۳۹	۰/۴۹	۰/۵۴
۸	۰/۱۴۱	۱۳/۸۳	۰/۴۲۶	۲/۴۴۷	۰/۵۲	۰/۵۸	سوال ۷۸	۷	۰/۲۶۵	۸/۸	۰/۸۰۴	۰/۷۱۶	۰/۴۷	۰/۴۸
۸	۰/۱۴۱	۹/۸۴	۱/۶۲۲	۰/۵۸۳	۰/۷۸	۰/۵۹	سوال ۷۹	۷	۰/۱۴۳	۷/۶۱	۳/۰۰	۰/۵۴۲	۰/۵۴	۰/۲۸
۸	۰/۳۲۶	۹/۸۳	۱/۶۲۳	۱/۵۸	۰/۷۵	۰/۴۲	سوال ۸۰	۹	۰/۲۴۱	۱۰/۳۱	۱/۱۷۹	۰/۵۵۵	۰/۶۲	۰/۳۴

Pi: ضریب دشواری هر سؤال در نظریه کلاسیک آزمون (CTT)؛ r_{pbis} : همبستگی دو رشته‌ای نقطه‌ای سؤال با نمره کل (ضریب تشخیص) در نظریه کلاسیک؛ $a(g)$: ضریب تشخیص هر سؤال در نظریه سؤال پاسخ (ITR)؛ $b(g)$: ضریب دشواری در نظریه سؤال پاسخ (ITR)؛ P: سطح معناداری؛ df: درجه آزادی

جدول ۳- رتبه‌های درصدی نمرات در هر یک از خرده‌آزمون‌های آزمون تشخیصی حساب‌نارسایی و کل

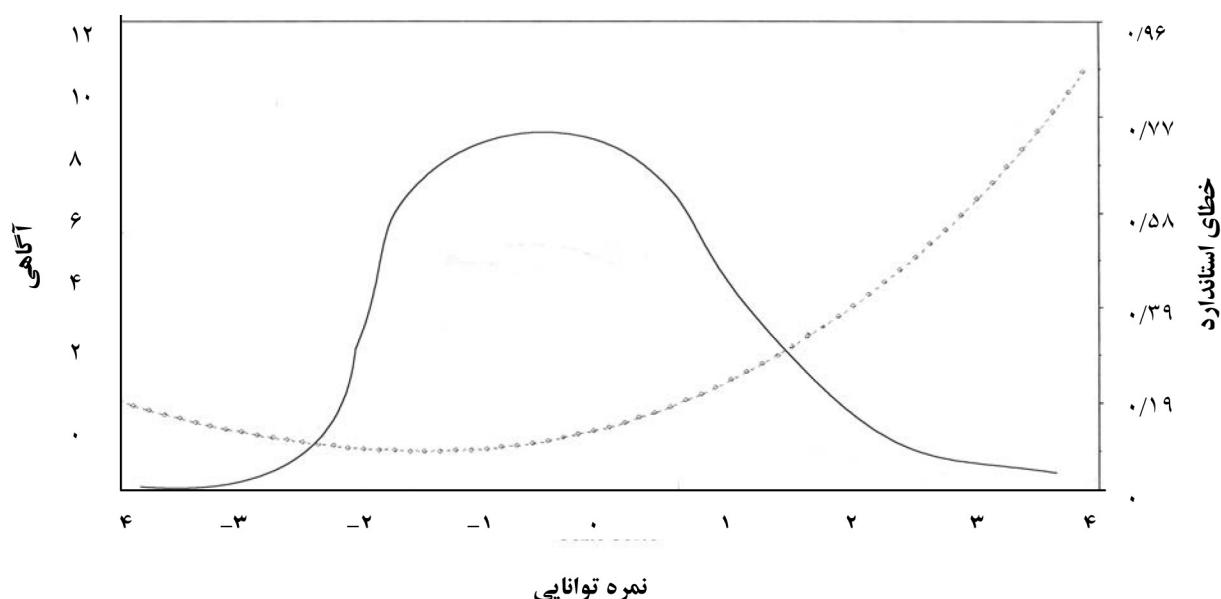
رتبه درصدی	خرده‌آزمون‌ها						
	تعداد	هندسه	جمع و تفریق	تقسیم	کسرها	اندازه‌گیری	حل مسئله
۵	۸	۲	۸	۷	۲	۸	۷
۲۵	۱۰	۲	۹	۸	۳	۹	۸
۵۰	۱۲	۳	۱۱	۱۱	۴	۱۲	۹
۷۵	۱۴	۳	۱۳	۱۳	۴	۱۳	۱۱
۹۵	۱۵	۴	۱۵	۱۴	۵	۱۵	۱۲

رتبه‌های درصدی نمرات در هر یک از خرده‌آزمون‌های آزمون تشخیصی حساب‌نارسایی: همان‌گونه که در جدول ۴ ملاحظه می‌گردد، نمرات پنج درصد افراد در خرده‌آزمون شمارش کمتر از هشت، در خرده‌آزمون هندسه کمتر از دو، در خرده‌آزمون جمع و تفریق کمتر از هشت، در خرده‌آزمون ضرب و تقسیم کمتر از هفت، در خرده‌آزمون کسر و اعشار کمتر از دو، در خرده‌آزمون اندازه‌گیری کمتر از هشت، در خرده‌آزمون حل مسئله کمتر از هفت و در کل کمتر از ۴۷ است.

منحنی تابع آگاهی آزمون (TIF): همان‌گونه که در شکل ۱ ملاحظه می‌شود، بیشترین میزان آگاهی‌دهندگی این آزمون در فاصلهٔ نتای ۰/۶- تا ۰/۶+ قرار گرفته که نشان‌دهندهٔ آن است که این آزمون برای افرادی که دارای این سطوح توانایی هستند، بیشترین و مناسب‌ترین کاربرد را دارد. میزان آگاهی‌دهندگی این

همان‌گونه که در جدول ۳ ملاحظه می‌گردد، ضرایب براساس نظریه کلاسیک (r_{pbis}) بین ۰/۴ تا ۰/۸۵ نوسان دارد که نشان می‌دهد هیچ یک از ضرایب، منفی و پایین نیست. همچنین ضرایب دشواری سؤال‌ها (P_i) بین ۰/۲۸ تا ۰/۸۸، ضرایب تشخیص $a(g)$ براساس نظریه سؤال- پاسخ (IRT) بین ۲/۵ تا ۰/۵۲۳ با میانگین ۱/۱۹۸ و انحراف معیار ۰/۱۶۰ است. هیچ یک از ضرایب کمتر از ۰/۵ نیست. ضرایب دشواری $b(g)$ نیز بین ۳/۰۰۰ تا ۰/۹- با میانگین ۱/۷۲ و انحراف معیار ۰/۷۱۵ است. سطح معناداری مقادیر کای دو سؤال‌ها نیز نشان می‌دهد که همگی با مدل دو پارامتری (2PLM) برازش دارند ($p > ۰/۰۵$).

بررسی پارامتر توانایی (θ) آزمودنی‌ها در ریاضیات نیز نشان داد که آزمودنی‌ها دارای پارامتر توانایی (θ) بین ۲/۱۳۹- تا ۲/۹۵ با میانگین ۱/۴۳۵ و انحراف معیار ۰/۰۲۱ هستند.



شکل ۱- منحنی تابع آگاهی آزمون (TIF) حساب‌نارسایی پایه سوم

رواسازی آزمون تشخیصی دقیق برای سنجش حساب‌نارسانی پایه سوم به‌عنوان هدف این آزمون در نظر گرفته شد.

بررسی همه جانبه، تأکید بر تحلیل و مقوله‌بندی محتوای کتاب ریاضی پایه سوم، طراحی سؤال‌های مختلف و نظرخواهی از اساتید و صاحب نظران و سرگروه‌های آموزشی ریاضی پایه سوم، محاسبه و شاخص‌های کمی، توافق در تحلیل محتوا و تعیین روایی صوری و محتوایی این آزمون، همگی مطلوب بودن روایی آزمون را نشان می‌دهند. همچنین همبستگی خرده‌آزمون‌ها و نمره کل این آزمون با آزمون ریاضیات ایران کی‌مت و خرده‌آزمون حساب در آزمون هوش ویسک-آر (وکسلر، ۱۳۷۳) هم حاکی از روایی همزمان و هم نشان‌دهنده روایی همگرایی (سازه) این آزمون است که امکان تفکیک معنادار افراد حساب‌نارسا و عادی را براساس سؤال‌های این آزمون فراهم می‌آورد. وجود نقاط برش گواه دیگری برای روایی سازه این آزمون است، زیرا اگر آزمودنی بتواند بین دو گروه تمایز قائل شود سازه دارای روایی است (سرمد و همکاران، ۱۳۷۶). بررسی پایایی خرده‌آزمون‌های این آزمون نشان داد که همسانی درونی این آزمون از طریق آلفای کرونباخ، دونیمه‌سازی و ثبات خرده‌آزمون‌ها، و پایایی نمره کل آن از طریق بازآزمایی مطلوب و قابل قبول است. مطابق با اعتقاد نانالی و برنشتاین (۱۹۹۴)، به نقل از مستامورن، (۱۳۸۵) ضریب پایایی بالای ۰/۷ مطلوب تلقی می‌شود، بنابراین ضرایب پایایی خرده‌آزمون‌ها و کل این آزمون مطلوب و پذیرفتنی است.

حساسیت این آزمون نشان داد که در ۹۸ درصد موارد با این آزمون می‌توان افراد حساب‌نارسا را از سالم تفکیک کرد. بررسی ویژگی (دقت) این آزمون نیز نشان داد که با استفاده از این آزمون می‌توان در ۹۶ درصد موارد، افراد سالم را به درستی تشخیص داد. بررسی ضرایب همبستگی دو رشته‌ای نقطه‌ای سؤال‌های آزمون (ضرایب تشخیص) در نظریه کلاسیک نشان داد که هیچ یک از این ضرایب منفی و کمتر از ۰/۲ نیستند. حداقل ضرایب تشخیصی در نظریه کلاسیک آزمون (CTT) ۰/۲ می‌باشد (متسامورن، ۱۳۸۵). از سوی دیگر ضرایب دشواری (P_i) در این آزمون نشان داد که ضرایب دشواری سؤال‌ها در محدوده ۰/۲۸ تا ۰/۸۸ است، بنابراین سؤال ۷۵ این آزمون ساده‌ترین و سؤال‌های ۳۰ و ۵۹ دشوارترین

آزمون در سطوح توانایی بسیار پایین (۴-) و بسیار بالا (۴+) بسیار اندک است. میزان آگاهی‌دهندگی این آزمون در این سطوح به کمتر از ۲ واحد می‌رسد.

مدت زمان لازم برای اجرای خرده‌آزمون شمارش در آزمون تشخیصی: زمان لازم برای اجرای خرده‌آزمون شمارش در این آزمون، مدت زمانی در نظر گرفته شد که ۷۰ درصد آزمودنی‌ها در آن دوره زمانی توانستند به همه سؤال‌ها پاسخ بدهند. این زمان در خرده‌آزمون شمارش ۶ دقیقه و ۵۰ ثانیه به دست آمد.

بحث

به نظر استیونس (۱۹۸۸)، به نقل از چین آوه و همکاران، (۱۳۸۳) تعداد روش‌های ارزشیابی که طراحی آنها به گونه‌ای باشد که مستقیماً به آموزش‌های جبرانی منجر شود اندک است، بنابراین هدف اصلی این پژوهش ساخت و رواسازی آزمونی دقیق با حداکثر قدرت تشخیص موارد حساب‌نارسا و عادی بود. بررسی آزمون‌های مختلفی که در حیطه ریاضیات تدوین شده‌اند نشان می‌دهد که در هر یک از این آزمون‌ها اگرچه نقاط مثبت هم وجود دارد، اما ضعف‌های جدی نیز به چشم می‌خورد. مثلاً در آزمون تشخیصی ریاضی فقط به ارزشیابی حل مسأله پرداخته شده (روچ و همکاران، ۱۹۲۱)، به نقل از چین آوه و همکاران، (۱۳۸۳) و یا پرسشنامه ریاضی پیش‌دبستانی کرانر (۱۹۷۴) برای ارزیابی توانایی کاربرد مفاهیم کمی کودکان تهیه شده است. از سوی دیگر برخی از این آزمون‌ها فقط برای پایه‌های خاصی تهیه شده و در آنها به پیوستاری بودن مفاهیم و مهارت‌های ریاضی توجهی نشده است؛ مثلاً در آزمون تشخیصی ریاضی کالیفرنیا مخصوص پایه‌های سوم تا هشتم، پرسشنامه ریاضی پیش‌دبستانی کرانر برای کودکان پیش‌دبستانی و پایه اول دبستان (به نقل از باس و وگن، ۲۰۰۰)، و نیز در برخی ابزارهای روایی محتوایی که صرفاً هنجاریابی شده‌اند محتوای دروس ریاضی پایه‌های مختلف و حجم این مطالب در کتاب‌های ریاضی ایرانی کمتر مورد توجه قرار گرفته است. همچنین هیچ یک از این آزمون‌ها زمان‌مند نبوده و در نتیجه قادر به تفکیک افراد حساب‌نارسا و عادی نیستند (باترورث، ۲۰۰۳، لندل و همکاران، ۲۰۰۴؛ جوردن و مونیتانی ۱۹۹۷؛ آشکرافت و همکاران، ۱۹۹۲). بنابراین ساخت و

(۲۰۰۳) هماهنگ است. براساس یافته‌های این پژوهشگران اگر در آزمون‌های تشخیصی ریاضی (خرده‌آزمون‌های شمارش) زمان لحاظ نشود، کودکان حساب‌نارسان نیز مانند کودکان عادی عمل خواهند کرد. در نتیجه زمانی این آزمون‌های تشخیصی دارای حساسیت کافی خواهند بود که زمانمند باشند.

بررسی منحنی آگاهی آزمون نشان داد که بیشترین میزان آگاهی‌دهندگی آن در فاصله سطح توانایی $0/6$ تا $0/6$ است که نشان‌دهنده آن است که این آزمون برای افرادی که در این سطح توانایی قرار دارند، بهترین و بیشترین کاربرد را خواهد داشت. میزان آگاهی‌دهی این آزمون در این سطوح توانایی حدود ۹ و پایین‌ترین میزان در سطوح توانایی بسیار پایین ۳- و بسیار بالا ۴+ است که در این سطوح این میزان به کمتر از یک واحد می‌رسد. با توجه به متوسط تابع آگاهی این آزمون که به وسیله رایانه $4/87$ به دست آمده، منحنی این آزمون یک تابع آگاهی متوسط به بالا را نشان می‌دهد.

بنابراین این آزمون دارای قدرت لازم جهت سنجش متغیر زیربنایی (خصیصه مکنون ریاضی) است و بر اساس آن می‌توان به ارزشیابی نسبتاً دقیق وضعیت پیشرفت ریاضی و تفکیک افراد دارای ناتوانی ریاضی از افرادی عادی پرداخت. موارد استفاده از آن با کاربرد آزمون ایرانی کی‌مت ریاضیات (محمداسماعیل و هومن، ۱۳۸۱) و آزمون پیشرفت ریاضی (چین آوه و همکاران، ۱۳۸۳) مشابه است. از این آزمون می‌توان در تعیین نقاط ضعف و قوت دانش آموزان پایه پنجم در حوزه‌های مختلف ریاضی و نیز تفکیک دقیق افراد حساب‌نارسان و عادی سود جست. به پژوهشگران نیز پیشنهاد می‌شود که برای تشخیص دیگر مشکلات یادگیری کودکان، مانند اختلال در دیکته‌نویسی و خواندن، با استفاده از نظریه جدید روان‌سنجی سؤال - پاسخ (IRT) آزمون‌هایی طراحی و ویژگی‌های دقیق روان‌سنجی آنها را تعیین کنند.

بودند. بهترین ضریب دشواری آزمون‌های تشریحی، ضرایب دشواری نزدیک به $0/5$ می‌باشند (عباس‌زادگان و فتوت، ۱۳۸۴؛ شریفی و نجفی‌زند، ۱۳۷۸). بررسی ضرایب دشواری این سؤال‌ها نشان داد که بیشتر آنها دارای ضریب دشواری بهینه یعنی نزدیک به $0/5$ هستند.

برای تفسیر ضرایب دشواری و تشخیصی سؤال‌ها در نظریه کلاسیک، باید این دو در ارتباط با یکدیگر در نظر گرفته شوند؛ سؤال‌هایی که ضرایب تشخیصی بالا و دشواری متوسط (در سؤال‌های تشریحی $0/5$) دارند، در تفکیک افراد قوی از ضعیف مطلوب تلقی می‌شوند (متسامورنن، ۱۳۸۵)، بنابراین همه سؤال‌ها براساس نظریه کلاسیک دارای ضرایب تشخیصی مطلوب و ضرایب دشواری نزدیک به $0/5$ هستند و در نتیجه می‌توان بر اساس آنها افراد دارای توانمندی ریاضی بالا و پایین را تفکیک کرد.

بررسی ضرایب تشخیص $a(g)$ و دشواری $b(g)$ سؤال‌ها براساس نظریه سؤال-پاسخ (IRT) نشان داد که سؤال‌های این آزمون دارای ضرایب تشخیصی $a(g)$ مطلوب (بالای $0/5$) هستند و براساس آنها می‌توان افراد ضعیف و قوی را تفکیک کرد. همچنین میانگین کل ضرایب تشخیصی نیز نشان می‌دهد که در مجموع قدرت تمیز آزمون نیز مطلوب است. ضرایب دشواری آزمون نیز نشان داد که سؤال‌ها در طیف وسیعی از بسیار ساده تا بسیار دشوار در نوسان‌اند ($-0/09$ تا $+3$) و میانگین ضرایب دشواری نیز نشان می‌دهد که متوسط دشواری این آزمون بیشتر از متوسط توانمندی ریاضی افراد است. به دلیل ماهیت تشریحی سؤال‌ها عامل شانس صفر بود $(c(g)=0)$ و احتمال پاسخ‌دهی شانسی وجود نداشت. بررسی سطح معنادار مقادیر χ^2 دو نیز نشان داد که همه سؤال‌ها با مدل دو آماره‌ای برازش کامل دارند.

در این آزمون زمان لازم برای خرده‌آزمون شمارش شش دقیقه و ۵۰ ثانیه به دست آمد که مدت زمانی بود که ۷۰ درصد آزمودنی‌ها در آن فاصله زمانی توانستند به سؤال‌های این خرده‌آزمون پاسخ دهند. این نکته با یافته‌های جوردن و مونتسیانی (۱۹۹۷)، لندرل و همکاران (۲۰۰۴) و باترورث

منابع

- احدی، ح.، و کاکاوند، ع. ر. (۱۳۸۲). *اختلال‌های یادگیری (از نظریه تا عمل)*. تهران: انتشارات ارسباران.
- آلن، م. جی.، وین، و. ام. (۱۳۷۴). *مقدمه‌های بر نظریه‌های اندازه‌گیری (روان‌سنجی)* (ترجمه ع. دلاور). تهران: انتشارات سمت.
- آناستازی، ا. (۱۳۷۳). *روان‌آزمایی* (ترجمه م. ن. براهنی). تهران: انتشارات دانشگاه تهران.
- چین آوه، م.، عرب‌زاده، ز.، و منشی‌زاده، ف. (۱۳۸۳). ساخت و هنجاریابی آزمون تشخیص ریاضی برای دانش‌آموزان ۶ تا ۱۳ ساله شهر شیراز. *مجله دانش و پژوهش در علوم تربیتی، دانشگاه آزاد واحد اسلامی خوراسگان اصفهان*، ۳، ۳۲-۱۵.
- حمید، ن. (۱۳۸۵). بررسی اختلال یادگیری ریاضی در دانش‌آموزان دختر و پسر دوره ابتدایی ناحیه یک شهر تهران و اثر آموزش کاربردی، تقویت ژتونی و آرامش عضلانی در کاهش اختلال یادگیری ریاضی در آنان. *مجله علوم تربیتی و روان‌شناسی دانشگاه شهید چمران اهواز*، ۲، ۱۳۶-۱۱۹.
- رمضانی، م. (۱۳۸۰). بررسی میزان شیوع حساب‌نارسایی در دانش‌آموزان پایه‌های چهارم و پنجم مدارس ابتدایی شهر اصفهان. تهران: پژوهشکده کودکان استثنایی.
- رمضانی، م. (۱۳۸۲). تحلیل خطاهای محاسباتی دانش‌آموزان حساب‌ناراسا. *مجله پژوهش در حیطه کودکان استثنایی*، ۴، ۲۴۴-۲۲۳.
- ژندایک، ا. (۱۳۷۵). *روان‌سنجی کاربردی* (ترجمه ح. ع. هومن). تهران: انتشارات دانشگاه تهران.
- ستاری، ب. (۱۳۸۲). *روان‌سنجی پیشرفته کاربردی*. مشهد: انتشارات آستان قدس رضوی.
- سرمد، ز.، بازرگان، ع.، و حجازی، ا. (۱۳۷۶). *روش‌های تحقیق در علوم رفتاری*. تهران: انتشارات آگاه.
- سیف، ع. ا. (۱۳۷۵). *روش‌های اندازه‌گیری و ارزشیابی آموزشی*. تهران: نشر دوران.
- شریفی، ح. پ.، و نجفی‌زند، ج. (۱۳۷۸). *اصول روان‌سنجی و روان‌آزمایی*. تهران: انتشارات رشد.
- عباس‌زادگان، م. ر.، و قوت، ا. ر. (۱۳۸۴). کاربرد پایایی و روایی در پژوهش. تهران: انتشارات کیهان.
- عریضی، ح. ر.، کاووسیان، ج.، و کدیور، پ. (۱۳۸۳). پایایی سنجی و اعتباریابی مقیاس شایستگی ریاضی کودکان اوت‌رخت در دانش‌آموزان دختر و پسر پیش‌دبستانی و دبستانی شهر اصفهان. *فصلنامه نوآوری‌های آموزشی*، ۹، ۶۵-۴۹.
- کاپلان، ه. ا.، و سادوک، ب. جی. (۱۳۸۳). خلاصه *روان‌پزشکی* (ترجمه ح. رفیعی). تهران: انتشارات ارجمند.
- محمد اسماعیل، ا.، و هومن، ح. ع. (۱۳۸۱). انطباق و هنجاریابی آزمون ریاضیات ایران کی‌مت. *مجله پژوهش در حیطه کودکان استثنایی*، ۴، ۳۳۲-۳۲۳.
- مستامورن، ج. (۱۳۸۵). *مبانی نظری آزمون و آزمون‌سازی* (ترجمه ش. کامکار و ا. اسرای). تهران: انتشارات بهینه.
- میرهاشمی، م. (۱۳۷۱). *هنجاریابی آزمون تصاویر در هم‌اندره ری در دانش‌آموزان ۷ تا ۱۵ سال و بزرگسالان شهر تهران*. پایان‌نامه کارشناسی ارشد دانشگاه آزاد کرج.
- هومن، ح. ع. (۱۳۷۱). *اندازه‌گیری‌های روانی و تربیتی (فن تهیه تست)*. تهران: انتشارات درسا.
- و کسلر، د. (۱۳۷۳). مقیاس تجدید نظر شده هوش و کسلر برای کودکان، *انطباق و هنجاریابی* (ترجمه س. شهیم). شیراز: انتشارات دانشگاه شیراز.

Ashcraft, M. H., Yamashita, T. S., & Aram, D. M. (1992). Mathematics performance in left and right brain-lesioned children and adolescents. *Brain and Cognition*, 19, 208-252.

Bloom, B. S. (1965). *Taxonomy of educational objectives; the classification of educational goals, by a committee of college and university examiners*. New York: Longmans, Green.

Butterworth, B. (2003). *Dyscalculia screener. Highlighting pupils with specific learning difficulties in maths*. London, UK: Nelson Publishing Company.

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 46(1), 3-18.

Fuchs, D., & Fuchs, L. D. (2005). Enhancing mathematical problem solving for student with disabilities. *Journal of Special Education*, 39, 45-57.

Jordan, N. C., & Montani, T. O. (1997). Cognitive arithmetic and problem solving: A comparison of children with specific and general mathematics difficulties. *Journal of Learning Disabilities*, 30, 624-634.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development*, 74, 834- 850.

Korporal, A. (1995). *The utrecht early mathematical competences scales. Analysis from classical and modern test theory*. Utrecht University, The Netherlands.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8-9-year-old student. *Cognition*, 93, 99-125.

Rosselli, M., & Ardila, A. (1989). Calculation deficits in patients with right and left hemisphere damage. *Neuropsychologia*, 27, 607-617.

Rousselle, L., & Noel, M-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition*, 102, 361-365.