

# سنجش کیفیت یادگیری ریاضیات در دانش‌آموزان پایه ششم شهر تهران

■ مسعود کبیری\* ■ شهرناز بخشعلی‌زاده\*\*

## چکیده:

مشاهده کیفیت خدمات آموزشی برای حصول اطمینان از کیفیت یادگیری دانش‌آموزان از وظایف عمده نظام‌های آموزشی به‌شمار می‌رود تا جایی که یکی از شاخص‌های کیفیت آموزش و پرورش در نظر گرفته شده است. سنجش یادگیری دانش‌آموزان در درس ریاضی، که یکی از دروس اصلی است، نیز اهمیت بسیاری دارد. در این مطالعه برای سنجش کیفیت یادگیری درس ریاضی در پایه آخر دوره ابتدایی، تعداد ۹۷۲۰ نفر از دانش‌آموزان پایه ششم شهر تهران با استفاده از روش نمونه‌گیری طبقه‌ای خوشه‌ای دومرحله‌ای انتخاب شدند. تعداد ۲۶۴ سؤال ریاضی با استفاده از نمونه‌گیری ماتریسی میان دانش‌آموزان توزیع شد. نتایج نشان داد که میانگین عملکرد دانش‌آموزان در پایه ششم برابر با ۴۶۳ بود که پایین‌تر از میانگین مقیاس (۵۰۰) است. همچنین دختران در مقایسه با پسران عملکرد بهتری را در ریاضی نشان دادند. مقایسه عملکرد انواع مدارس نیز نشان داد که دانش‌آموزان مدارس دولتی عملکرد پایین‌تری از سایر مدارس داشتند، اما تفاوت در میان دانش‌آموزان انواع دیگر مدارس معنادار نبود. علاوه بر این، تفاوت قابل توجهی میان میانگین مناطق آموزشی وجود داشت به طوری که بهترین عملکرد در مناطق ۳، ۲، ۵ و ۶ و ضعیف‌ترین عملکرد در مناطق ۱۹، ۱۵ و ۱۸ دیده شد. تفاوت‌های ملاحظه‌شده حتی در مقایسه میان مدارس دولتی و هیئت‌امانی نیز حفظ شد.

پیشرفت تحصیلی در ریاضی، مطالعه برکات، آموزش ریاضی، دوره ابتدایی

کلید واژه‌ها:

□ تاریخ پذیرش مقاله: ۱۴۰۰/۵/۹

□ تاریخ شروع بررسی: ۹۹/۷/۴

□ تاریخ دریافت مقاله: ۹۹/۹/۴

\* نویسنده مسئول) استادیار مرکز ملی مطالعات تیمز و پرلز، پژوهشگاه مطالعات آموزش و پرورش. E-mail: maskabiri@yahoo.com

\*\* کارشناس مرکز ملی مطالعات تیمز و پرلز، پژوهشگاه مطالعات آموزش و پرورش. E-mail: sbakhshalizadeh@yahoo.com

## مقدمه

مرور تحولات آموزش و پرورش در کشورهای گوناگون دنیا نشان می‌دهد که در نظام‌های آموزشی مرسوم است با استناد به داده‌هایی چون میزان ثبت‌نام و سایر آمار ورودی‌های نظام بر گسترش آموزش و پرورش از جنبه کمی تأکید شود (کی‌پرش<sup>۱</sup> و همکاران، ۲۰۱۳)، به طوری که حتی در شعار اهداف توسعه هزار<sup>۲</sup> با عنوان «آموزش برای همه»<sup>۳</sup> افزایش میزان ثبت‌نام هدف اصلی کشورها دانسته شده است. با این حال، تأملات بعدی نشان داد که این شاخص نمی‌تواند برای بازنمایی وضعیت و نحوه توسعه کشورها در امر آموزش و پرورش کامل باشد و لازم است به شایستگی‌ها، عدالت در کیفیت آموزشی و کارایی آموزش‌ها توجهی ویژه شود. با این نگاه، تمرکز گفتمان آموزشی از تأکید بر تعداد ورودی‌های آموزش و پرورش به کیفیت خروجی‌های آن تغییر جهت داد. در رویکرد جدید، کیفیت آموزشی به‌منزله آنچه دانش‌آموزان پس از گذراندن زمانی معین کسب می‌کنند، توانمند به اجرای آن هستند و نحوه ارتباط این خروجی‌ها با انتظارات جامعه و برنامه‌های درسی تعریف می‌شود (ویجمیکر<sup>۴</sup>، ۲۰۱۴).

با این حال، در دسترس نبودن اطلاعات منظم در زمینه میزان اکتسابات آموزشی دانش‌آموزان مشکل تصمیم‌گیران نظام آموزشی در غالب کشورها بود زیرا بیشتر اطلاعات گردآوری شده حول آمار ورودی‌های آموزش و پرورش از قبیل تعداد ثبت‌نام‌کنندگان، معلمان، و بودجه بود. نبود توازن اطلاعاتی در ابعاد گوناگون باعث شد اندیشه سیاست‌مداران بر ورودی‌های آموزش و پرورش متمرکز شود و حتی برخی از آن‌ها بخواهند که کیفیت ضعیف آموزش و پرورش را نادیده بگیرند یا بپوشانند و توجه به یادگیری را از داده‌ها، گزارش‌های رسمی مدیریتی و دستور کار سیاست‌گذاران خارج کنند (بانک جهانی<sup>۵</sup>، ۲۰۱۸). لذا یکی از تمهیدات اصلی کشورها برای رفع کمبودهای اطلاعاتی در دهه‌های اخیر بازنگری در نظام‌های سنجش به‌منظور سنجش منظم و دوره‌ای یادگیری دانش‌آموزان است. این تغییر دیدگاه در برنامه‌های اصلی نظام‌های سنجش، شامل امتحانات و به‌خصوص مطالعات کلان‌مقیاس بین‌المللی، ملی و محلی، به‌منزله ابزار اصلی اندازه‌گیری و بهبود کیفیت آموزشی دیده می‌شود.

در کنار این موارد، گسترش مطالبه پاسخ‌گویی از دولت‌ها و آموزش و پرورش نیز یکی دیگر از دلایل گسترش سنجش‌های متمرکز بر کیفیت یادگیری است. نتایج مطالعات نشان می‌دهد میان استفاده از سنجش‌ها برای پاسخ‌گو نگه‌داشتن مدارس و کیفیت بهتر خروجی‌های مدرسه ارتباطی آشکار وجود دارد (یونسکو، ۲۰۱۷). پاسخ‌گویی زمانی محقق می‌شود که اطلاعات موثقی درباره کیفیت آموزش در اختیار والدین و جامعه قرار گیرد. در غیر این صورت شاخص‌های دیگری به‌عنوان کیفیت نظام آموزشی جایگزین خواهند شد یا افراد جامعه درباره وضعیت آموزش و پرورش خود تصورات اشتباهی خواهند داشت. در مثالی از عدم مطالبه جدی می‌توان به موردی از اوگاندا اشاره کرد. تقریباً ۷۵ درصد از والدین دانش‌آموزان این کشور درباره کیفیت آموزش و پرورش کشورشان احساس رضایت کرده بودند، ولی تنها

۲۵ درصد از دانش‌آموزان پایه چهارم این کشور توانسته بودند در آزمون ریاضی با محتوای متناسب با پایه دوم قبول شوند (بانک جهانی، ۲۰۱۸).

برای آگاهی از کیفیت یادگیری تأکید فراوانی بر دروس اصلی<sup>۶</sup> به‌منزله نشانگر<sup>۷</sup> می‌شود. این دروس در دوره ابتدایی، در شکل سنتی، شامل خواندن و نوشتن زبان مادری، ریاضی و علوم است. بر اساس مطالعه روی سنجش‌های ملی کشورهای متفاوت مشخص شد که همه کشورهای دو حوزه مهارت‌های زبانی و عددی (در قالب درس ریاضی) را در سنجش‌های ملی خود می‌سنجیدند که این مسئله اهمیت این دو موضوع را منعکس می‌کند (گرنی و کلاگان<sup>۸</sup>، ۲۰۰۸). علاوه بر این، ارتباط تنگاتنگ ریاضی با موفقیت اقتصادی کشورها دلیل دیگر توجه خاص به این درس است (لوهمان<sup>۹</sup>، ۲۰۰۶).

مطالعات انجام‌شده در زمینه یادگیری ریاضی در ایران نشان می‌دهد که عملکرد دانش‌آموزان در این درس وضعیت مطلوبی ندارد. بر اساس نتایج مطالعه بین‌المللی روندهای آموزش علوم و ریاضی (تیمز)، که از سال ۱۹۹۵ تا ۲۰۱۹ هر چهار سال یک بار در بیش از ۵۰ کشور از جمله ایران اجرا شده است، عملکرد ریاضی دانش‌آموزان پایه چهارم ایران همیشه پایین‌تر از میانگین بین‌المللی بوده است. نتیجه آخرین دوره مطالعه در سال ۲۰۱۵ نشان داد که میانگین ریاضی دانش‌آموزان ایران در پایه چهارم برابر با ۴۳۱ است که در قیاس با دوره قبل از آن (۲۰۱۱) تفاوتی نشان نمی‌دهد. علاوه بر این، ۳۵ درصد از دانش‌آموزان پایه چهارم به حداقل سطح شایستگی موردانتظار در ریاضی نرسیدند و تنها ۱ درصد از آنان به سطح بالاترین معیار عملکردی رسیدند (کبیری و همکاران، ۱۳۹۵). مطالعه «سنجش صلاحیت‌های پایه: ارزشیابی درون‌داده‌ها و برون‌داده‌های آموزشی در ایران» که در پایه پنجم ابتدایی انجام شد نیز نشان می‌دهد که عملکرد دانش‌آموزان وضعیت مطلوبی ندارد (کیامنش و خیریه، ۱۳۷۹).

طبق مرور پژوهش‌ها، در عملکرد ریاضی میان دختران و پسران در تهران و کل کشور الگوهای گوناگون و متفاوتی دیده می‌شود. نتایج پنج دوره اجرای مطالعه تیمز تا سال ۲۰۱۵ نشان می‌دهد که با وجود تغییرات روندی میان عملکرد دختران و پسران در پایه چهارم، تفاوت معناداری میان عملکرد آن‌ها وجود ندارد (کبیری و همکاران، ۱۳۹۵) اما نتایج پژوهش‌های انجام‌شده در شهر تهران تفاوت‌هایی جزئی، هرچند باز هم غیرمعنادار، را به نفع دختران نشان می‌دهد. در مطالعه‌ای که روی ۳۰۰ دانش‌آموز پایه اول دبیرستان از منطقه ۶ (کیامنش و پوراصغر، ۱۳۸۸) و ۳۹۰ دانش‌آموز پایه سوم راهنمایی (کرامتی و شهرآرای، ۱۳۸۳) اجرا شد، باینکه عملکرد دختران نسبت به پسران برتری داشت اما تفاوت‌های مشاهده‌شده معنادار نبود. همچنین در مقایسه عملکرد دختران و پسران روی ۸۰۰ دانش‌آموز پایه سوم دبیرستان (قاضی طباطبایی و همکاران، ۱۳۹۳) و ۵۰۹ دانش‌آموز پایه اول دبیرستان (افضلی و همکاران، ۱۳۹۳)، باینکه معناداری تفاوت‌ها بررسی نشده بودند، عملکرد ریاضی دختران بهتر از پسران گزارش شد؛ البته تفاوت‌ها نشان می‌دهد که در صورت بررسی مقایسه، تفاوت معناداری به‌دست نمی‌آید. باوجود این، مقایسه بین ۳۶۶ دانش‌آموز دختر و پسر در پایه سوم راهنمایی برتری معنادار عملکرد ریاضی

دختران را از پسران نشان داد (کبیری، ۱۳۸۲). باین‌همه به نظر می‌رسد که حجم نمونه نسبتاً محدود این پژوهش‌ها دلیلی بر نبود معناداری عملکرد بهتر دختران باشد.

علاوه بر تأثیر جنسیت در عملکرد دانش‌آموزان، مباحث بسیاری چه در مجامع علمی و چه در عموم جامعه در مورد تأثیر نوع مدرسه در عملکرد تحصیلی دانش‌آموزان ایران مطرح است. تفاوت میان عملکرد مدارس هنگامی بیشتر ملموس شد که در برخی مطالعات از روش‌های تحلیل چندسطحی برای تحلیل داده‌های پیشرفت تحصیلی استفاده شد. برای مثال، در مطالعه‌ای برای تحلیل رابطه میان مهارت‌های تدریس معلمان با پیشرفت تحصیلی دانش‌آموزان در درس‌های محاسباتی همبستگی درون‌طبقه‌ای پیشرفت تحصیلی به میزان ۰/۴۱ به دست آمد (مدنی و همکاران، ۱۳۹۸). مشابه چنین مقداری (۰/۴۷) در مطالعه‌ای گزارش شد که روی معدل تحصیلی دانش‌آموزان پایه سوم متوسطه در رشته علوم انسانی انجام شد (حجازی و همکاران، ۱۳۹۷). این یافته‌ها نشان دادند که نزدیک به نیمی از پراکندگی عملکرد تحصیلی دانش‌آموزان قابل انتساب به مدرسه‌ای است که در آن مشغول به تحصیل‌اند. باین‌حال، نوع مدرسه نیز در عملکرد دانش‌آموزان تأثیرگذار تشخیص داده شد. زکی (۱۳۸۴) در تحلیلی که روی ۲۰۰ دانش‌آموز دولتی و غیردولتی انجام داد تأثیر نوع مدرسه در عملکرد تحصیلی را بیش از جنسیت و رضایت اجتماعی دانش‌آموزان تعیین کرد. در مقایسه میان مدارس مختلط و تک‌جنسیتی دولتی، پیشرفت تحصیلی مدارس تک‌جنسیتی نیز بیشتر گزارش شده است (خانی و محمدی، ۱۳۹۴). با وجود این، در مطالعه نوغانی و همکاران (۱۳۹۰)، که روی ۴۰۰ دانش‌آموز پیش‌دانشگاهی در مدارس دولتی و غیردولتی انجام شد، متغیر نوع مدرسه در میان چهار متغیر مؤثر در الگو قرار نگرفت. در این مطالعه تأثیر ۱۰ متغیر در میزان موفقیت در آزمون سراسری ورود به دانشگاه بررسی شد.

با وجود مطالعات ارزش‌مندی که انجام شده است، شواهد معتبری در زمینه میزان یادگیری دانش‌آموزان به تفکیک استان‌ها موجود نیست. باین‌حال، به نظر می‌رسد که با وجود برنامه‌های درسی مشترک، عملکرد دانش‌آموزان استان‌های مختلف متفاوت باشند که در این صورت نمی‌توان از نتایج کل کشور (که در مطالعاتی همچون تیمز منتشر می‌شود) برای برنامه‌ریزی‌ها و سیاست‌گذاری‌های استانی استفاده کرد. همچنین، آگاهی از کیفیت یادگیری دانش‌آموزان به دلیل اجرای سیاست ارزشیابی توصیفی و تأکید بر نظر معلمان و استفاده نکردن از آزمون‌های استاندارد شده در دوره ابتدایی اهمیت دوچندانی دارد. به عبارت دیگر، مشخص نیست که دانش‌آموزان در درسی مانند ریاضی تا چه اندازه در پایه آخر دوره ابتدایی به مهارت‌های لازم دست یافته‌اند. بنابراین نیاز به داشتن چنین اطلاعاتی در پایه ششم، که پایه آخر دوره ابتدایی به‌شمار می‌رود، بیشتر به چشم می‌خورد. جدای از این، اطلاعات دقیقی در مورد تفاوت‌های جنسیتی در میان انواع مدارس، به‌خصوص در شهر تهران، در دسترس نیست.

مطالعه حاضر در راستای اطلاع از کیفیت یادگیری دانش‌آموزان دوره ابتدایی تهیه شده است. در این مطالعه، یادگیری ریاضی دانش‌آموزان پایه ششم شهر تهران با توجه به اهداف ریاضی و همچنین به مدد

نمونه معرفی بررسی شد. این مطالعه بخشی از «برنامه رصد کیفیت آموزشی شهر تهران» (برکات) است که اداره کل آموزش و پرورش شهر تهران آن را با هدف پایش کیفیت آموزشی و فراهم کردن داده‌های باکیفیت برای پشتیبانی از توسعه سیاست‌گذاری آموزشی طراحی کرده است. این برنامه اولین مطالعه کلان‌مقیاس سنجش کیفیت یادگیری در سطح استانی است که در کشور برگزار می‌شود (کبیری، ۲۰۱۹ و ۱۳۹۹). دوره اول برنامه برکات با دو موضوع ریاضی (پایه ششم) و مهارت حل مسئله (پایه یازدهم) شروع شد که این مطالعه نتایج سنجش ریاضی آن را تشریح می‌کند. در این مقاله، در پی آن‌ايم که عملکرد ریاضی دانش‌آموزان پایه ششم شهر تهران را بررسی کنیم و سپس عملکرد ریاضی دانش‌آموزان میان گروه‌های اصلی نمونه شامل گروه‌های جنسیتی، انواع مدرسه و مناطق آموزشی را باهم مقایسه کنیم. مقایسه عملکرد ریاضی میان گروه‌های گوناگون دو هدف را دنبال می‌کند: اول، حصول اطمینان از کیفیت یادگیری در گروه‌های متفاوت جامعه که در رویکرد عدالت آموزشی نیز ملاحظه شود. دوم، امکان برنامه‌ریزی بهینه برای متوازن نگه‌داشتن کیفیت یادگیری در گروه‌های متفاوت جامعه. گروه‌هایی که معمولاً در شهر تهران به آن‌ها توجه می‌شود در دسته‌بندی‌های جنسیتی، نوع مدرسه و مناطق آموزشی قرار می‌گیرند

## روش پژوهش

به منظور تعیین عملکرد ریاضی دانش‌آموزان پایه ششم شهر تهران، نمونه معرفی با استفاده از نمونه‌گیری طبقه‌ای خوشه‌ای دومرحله‌ای<sup>۱۰</sup> انتخاب شد. در مرحله اول نمونه‌گیری، مدارس بر اساس طبقات تعیین‌شده انتخاب شدند. در مرحله دوم و پس از تعیین هریک از مدارس هدف، فهرست کلاس‌های ششم هر مدرسه استخراج شد. از میان کلاس‌های ششم آن مدرسه یکی به‌طور تصادفی انتخاب شد و همه دانش‌آموزان آن کلاس برای جامعه نمونه این مطالعه انتخاب شدند. برای اجرای طبقه‌بندی مدارس از دو نوع طبقه‌بندی صریح و ضمنی (لاروشه<sup>۱۱</sup> و همکاران، ۲۰۱۶) استفاده شد. مناطق آموزشی طبقه‌بندی صریح نمونه‌گیری<sup>۱۲</sup> برای تشکیل چارچوب‌های جداگانه تحلیل در نظر گرفته شدند و دو متغیر جنسیت و نوع مدرسه به‌منزله متغیرهای طبقه‌بندی ضمنی<sup>۱۳</sup> به کار رفتند که سهم اصلی آن‌ها مرتب کردن مدارس در درون هریک از مناطق بود. علاوه بر این برای انتخاب مدارس در هریک از مناطق، از روش تصادفی منظم احتمالات متناسب با حجم<sup>۱۴</sup> استفاده شد.

پس از گردآوری و ثبت نحوه مشارکت دانش‌آموزان هر مدرسه در مطالعه و همچنین به دلیل استفاده از نمونه‌گیری طبقه‌ای، انتخاب هریک از اعضای نمونه احتمالات نامساوی داشت. برای رفع این مشکل از وزن‌های نمونه‌گیری<sup>۱۵</sup> (روتکوفسکی<sup>۱۶</sup> و همکاران، ۲۰۱۰) استفاده شد. به عبارت دیگر، وزن‌های نمونه‌گیری برای جبران احتمالات نامساوی نمونه‌گیری در هر منطقه و همچنین جبران موارد ناشی از مشارکت نکردن دانش‌آموزان در محاسبات به کار گرفته شدند. این وزن‌ها از طریق محاسبات مربوط به

تعیین احتمال هریک از اعضا در میزان مشارکت آنان به‌دست آمد. در تحلیل‌های این مطالعه از وزن کلی دانش‌آموز<sup>۱۷</sup> استفاده شد.

از هر منطقه آموزشی ۱۷ تا ۱۸ مدرسه انتخاب شدند. در نتیجه، ۹۷۲۰ دانش‌آموز پایه ششم از ۳۲۵ مدرسه در این مطالعه مشارکت داشتند که از هر منطقه ۴۵۱ تا ۵۶۶ دانش‌آموز در مطالعه مشارکت داشتند. همچنین، از کل جامعه نمونه ۴۷۳۴ نفر (۴۸/۷ درصد) دختر و ۴۹۸۶ نفر (۵۱/۳ درصد) پسر بودند.

### ■ ابزار پژوهش

برای طراحی آزمون و ساخت سؤال‌های ریاضی، ابتدا محدوده دانش و مهارت‌های موردنیاز سنجش ریاضی بررسی شدند و روی آن‌ها توافق صورت گرفت. به علت اینکه درس ریاضی در دوره ابتدایی هنوز فاقد برنامه درسی مصوب است، برای تعیین چارچوب سنجش و شناسایی مفاهیم و مهارت‌های مورد سنجش، کتاب‌های درسی اصلی‌ترین منبع رسمی در نظر گرفته شدند. از طریق واکاوی مفاهیم مندرج در آن‌ها و تعیین سلسله‌مراتب ارائه آن‌ها، نقشه‌های مفهومی تهیه شد. این کار از طریق تشکیل گروهی از متخصصان آموزش ریاضی طی جلسات متعدد حاصل شد. در این بررسی، پنج حیطه محتوایی اعداد؛ نسبت، تناسب و درصد؛ جبر؛ اندازه‌گیری و هندسه؛ و کار با داده‌ها به همراه سه حیطه شناختی دانستن؛ به‌کاربردن؛ و استدلال کردن با نسبت‌های نامساوی در نظر گرفته شدند.

با مشخص شدن چارچوب سنجش، گروهی یازده‌نفره متشکل از متخصصان آموزش ریاضی و آموزش ابتدایی و معلمان پایه ششم برای هریک از مباحث آموزشی سؤال طراحی کردند. این سؤال‌ها پس از تأیید گروه دیگری که کار تأیید سؤال‌ها را به عهده داشتند به خزانه سؤال‌ها وارد می‌شدند. مجموعه سؤال‌های طراحی‌شده و تأییدشده شامل ۲۳۴ سؤال بود. علاوه‌براین، با دو هدف پوشش دادن به سطوح پایین عملکرد دانش‌آموزان و همچنین پیوند دادن مقیاس نمره‌های این مطالعه به مقیاس مورد استفاده در مطالعه تیمز، ۵۳ سؤال ریاضی از سؤال‌های منتشرشده مطالعه تیمز در دو سال ۲۰۱۵ و ۲۰۱۱ انتخاب و به مجموع سؤال‌ها اضافه شدند. بدین ترتیب، ۲۸۷ سؤال به‌عنوان سؤال‌های اصلی مطالعه در نظر گرفته شد. با توجه به اینکه چنین حجمی از سؤال‌ها را هیچ‌یک از دانش‌آموزان نمی‌توانستند پاسخ دهند، با استفاده از نمونه‌گیری ماتریسی<sup>۱۸</sup> سؤال‌ها در ۲۰ دفترچه جداگانه توزیع شد. با اجرای چنین شیوه‌ای هر دانش‌آموز فقط به بخشی از سؤال‌هایی پاسخ می‌داد که تصادفی انتخاب شده و در اختیار او گذاشته شده بود. توزیع سؤال‌ها در بین دفترچه‌ها با در نظر گرفتن پراکندگی سطوح حیطه‌های محتوایی و شناختی و نوع سؤال‌ها انجام شد. پراکندگی سؤال‌ها بین دفترچه‌های بیست‌گانه به‌گونه‌ای انجام شد که به هر سؤال ۱۹۰۰ تا ۲۰۰۰ نفر پاسخ دهند تا برای محاسبات مربوط به تحلیل سؤال کافی باشد.

به‌منظور تعیین کیفیت سؤال‌های نهایی، پس از اجرای آزمون، هریک از سؤال‌ها در معرض تحلیل‌های متعدد قرار گرفت تا کیفیت سؤال‌های نهایی مشخص شود. شاخص‌های مورد بررسی از مجموعه تحلیل‌های کلاسیک آزمون‌سازی شامل ضریب تمیز، درصد پاسخ‌گویی به سؤال و ضریب دشواری، درصد پاسخ‌گویی به هریک از گزینه‌ها (سؤال‌های چندگزینه‌ای) یا کدهای نمره‌گذاری (سؤال‌های بازپاسخ)، همبستگی‌های دورشته‌ای نقطه‌ای برای هریک از گزینه‌ها و کدهای نمره‌گذاری و شاخص‌ها بودند و از مجموعه تحلیل‌های نظریهٔ پرسش‌پاسخ شامل برآورد جایگاه از مجموعه تحلیل‌های الگوی راش و تعیین ضرایب جایگاه، شیب و حدس به همراه میزان خطا و آگاهی هر سؤال بودند. مضاف‌براین، تحلیل‌های دیگری از جمله بُعدیت سؤال‌ها، تحلیل کارکرد افتراقی سؤال‌ها و برازش سؤال‌ها نیز انجام شد. اجرای این تحلیل‌ها به شناسایی اشکالات برخی از سؤال‌ها منجر شد. به‌منظور مشخص کردن سؤال‌های با کیفیت مناسب برای تحلیل‌های نهایی، از ترکیب سؤال‌های چندبخشی، تبدیل سؤال‌های دونمره‌ای به یک‌نمره‌ای و حذف سؤال‌های نامطلوب برای رفع اشکالات شناسایی شده استفاده شد. در نتیجهٔ این تحلیل، ۲۶۴ سؤال برای تحلیل نهایی انتخاب شدند و ۲۳ سؤال حذف شدند. درنهایت، اعتبار سؤال‌های هر دفترچه با روش آلفای کرونباخ نیز محاسبه شد. طبق این بررسی، میانگین اعتبار دفترچه‌ها برابر با ۰/۸۴۵ شد که از ۰/۶۰۵ تا ۰/۹۰۳ در نوسان بود.

از ۲۸۷ سؤال آزمون ریاضی، ۱۳۳ سؤال (۴۶/۳ درصد) را سؤال‌های چندگزینه‌ای و ۱۵۴ سؤال (۵۳/۷ درصد) را سؤال‌های بازپاسخ تشکیل می‌دادند. نمره‌دهی به سؤال‌های بازپاسخ مستلزم استفاده از نمره‌گذاران آموزش‌دیده بود. برای اینکه نمرات داده‌شده به سؤال‌های بازپاسخ دقت و اعتبار کافی داشته باشند به چند کار اقدام شد. ابتدا، راهنمای نمره‌گذاری برای هریک از سؤال‌های بازپاسخ تهیه شد که نحوهٔ نمره‌گذاری به هریک از سؤال‌ها را تشریح می‌کرد. این راهنما به‌گونه‌ای تدوین شد که علاوه‌بر مشخص شدن درستی یا نادرستی پاسخ (یا ناقص بودن پاسخ در سؤال‌های دونمره‌ای) تنوع و انواع پاسخ‌های درست یا نادرست (یا ناقص در سؤال‌های دونمره‌ای) نیز مشخص شود. اقدام دوم شامل آموزش نمره‌گذاران بود که در مواقعی با تمرین و ارائهٔ بخشی از سؤال‌ها به‌صورت آزمایشی و با نظارت بر نحوهٔ نمره‌دهی نمره‌گذاران همراه بود. اقدام سوم، انتخاب تصادفی ۲۰ درصد از دفترچه‌ها به‌منزلهٔ دفترچه‌های نمره‌گذاری مجدد بود. بدین صورت که این دفترچه‌ها را دو نمره‌گذار متفاوت نمره‌گذاری می‌کردند و میزان توافق آنان از طریق اعتبار میان ارزیاب<sup>۱۹</sup> محاسبه می‌شد. این کار برای پی‌بردن به این موضوع بود که از یک‌طرف تا چه اندازه راهنمای نمره‌گذاری هر سؤال با شفافیت تدوین شده است و از طرف دیگر کیفیت کار نمره‌گذاران چگونه است. ضرایب اعتبار میان ارزیاب بین ۰/۷۱۵ تا ۰/۹۹۹ در نوسان بود. سؤال‌هایی که اعتبار بین ارزیاب آنان کمتر از ۰/۸۵ بودند (شامل ۲۳ سؤال) بررسی شدند و در صورت وجود مشکل دیگر، سؤال نامزد برای حذف شدن از مجموع سؤال‌ها در نظر گرفته می‌شدند

## روش تجزیه و تحلیل

برای محاسبه برآوردهای دقیق، برای هر یک از دانش‌آموزان، از روش‌شناسی مقادیر محتمل<sup>۲۰</sup> (ون‌داویر<sup>۲۱</sup> و همکاران، ۲۰۰۹؛ وو<sup>۲۲</sup>، ۲۰۰۵) استفاده شد. در این شیوه، از تلفیق داده‌های پیشینه‌ای دانش‌آموزان (که عمدتاً از پرسش‌نامه‌های پیشینه‌ای دانش‌آموزان به دست می‌آید و با کدگذاری تصنعی به داده‌های دویخشی تبدیل شده است) با داده‌های آزمون نمره‌هایی از دانش‌آموزان (نمرات پسین) تولید می‌شوند که علاوه بر امکان بررسی خطاهای بیشتر، نزدیک‌ترین حالت را با برآوردهای جامعه داشته باشند. بدین منظور برای هر دانش‌آموز، پنج نمره تولید خواهد شد که تقریبی از میانگین نمره‌های عملکرد هر فرد را معین خواهد کرد و با آشکار کردن پراکندگی میان هر یک از پنج برآورد، دقت برآورد هر فرد را نیز مشخص خواهد کرد. در کل، این شیوه هم برآوردهای دقیق‌تری در جامعه حاصل می‌کند و هم بخشی از خطا را استخراج می‌کند که با شیوه‌های دیگر قابل اندازه‌گیری نیست و در مجموع اعتبار بیشتری را حاصل خواهد کرد.

به‌منظور افزایش قابلیت مقایسه نتایج مطالعه با مطالعات قبلی در زمینه ریاضی، علاوه بر روش‌شناسی مقادیر محتمل، از برآوردهای سؤال‌های مشترک میان این مطالعه و مطالعه تیمز برای پیونددهی میان آن‌ها استفاده شد. این کار باعث می‌شود نتایج این مطالعه با مطالعه تیمز در پایه چهارم مقایسه‌پذیر باشد و به عبارت دیگر نتایج این مطالعه در همان مقیاس تیمز ۲۰۱۵ در پایه چهارم ارائه شود. این فرایند از طریق به‌کارگیری روش انتقال خطی<sup>۲۳</sup> از مجموعه فنون پیونددهی<sup>۲۴</sup> و هم‌ترازسازی آزمون (فوی و ین<sup>۲۵</sup>، ۲۰۱۶) انجام شد. با توجه به اینکه مقیاس لوجیت<sup>۲۶</sup> برای محاسبه برآورد توانایی به کار می‌رود، برای پرهیز از ارائه نمرات به صورت منفی یا اعشاری از مقیاسی با میانگین ۵۰۰ و انحراف استاندارد ۱۰۰ استفاده شد. در این مقیاس، نمره ۵۰۰ معادل با میانگین نمرات مقیاس در مطالعه و نمرات کمتر یا بیشتر از ۵۰۰ به معنای نمرات کمتر یا بیشتر از میانگین مقیاسی است که میزان کمتر یا بیشتر شدن آن را می‌توان با استناد به انحراف استاندارد ۱۰۰ قیاس کرد.

## یافته‌های پژوهش

با استفاده از روش‌شناسی مقادیر احتمالاتی و همچنین پیوند دادن نتایج به مقیاس مورد استفاده در درس ریاضی مطالعه تیمز مشخص شد که میانگین ریاضی دانش‌آموزان پایه ششم شهر تهران برابر با ۴۶۳ است. این عدد بدین معناست که میانگین پیشرفت تحصیلی دانش‌آموزان در درس ریاضی با حدود ۰/۳۷ انحراف استاندارد (معادل با ۳۷ نمره در مقیاس با میانگین ۵۰۰ و انحراف استاندارد ۱۰۰) به طور معناداری از نمره متوسط مقیاس پایین‌تر است؛ به عبارت دیگر، دانش‌آموزان پایه ششم شهر تهران نتوانسته‌اند به حد وسط انتظارات یادگیری در درس ریاضی



برسند. نقطه درصدی ۵ و ۹۵ به ترتیب برابر با ۳۰۱ و ۶۰۰ بود؛ به عبارت دیگر، پنج درصد از دانش‌آموزان نمره‌های کمتر از ۳۰۱ و ۵ درصد از آنان نمره‌های بالاتر از ۶۰۰ گرفته بودند. چارک‌های اول و سوم (برابر با نقطه درصدی ۲۵ و ۷۵) نیز به ترتیب برابر با ۴۰۱ و ۵۳۲ بودند؛ به عبارت دیگر، یک‌چهارم دانش‌آموزان پایه ششم شهر تهران نمراتی کمتر از ۴۰۱ دارند و از طرف دیگر یک‌چهارم قوی‌تر دانش‌آموزان نمراتی بالاتر از ۵۳۲ کسب کرده‌اند. وضعیت عملکرد دانش‌آموزان در درس ریاضی در نمایه ۱ نشان داده شده است.

نمایه ۱. توزیع عملکرد دانش‌آموزان پایه ششم شهر تهران در درس ریاضی



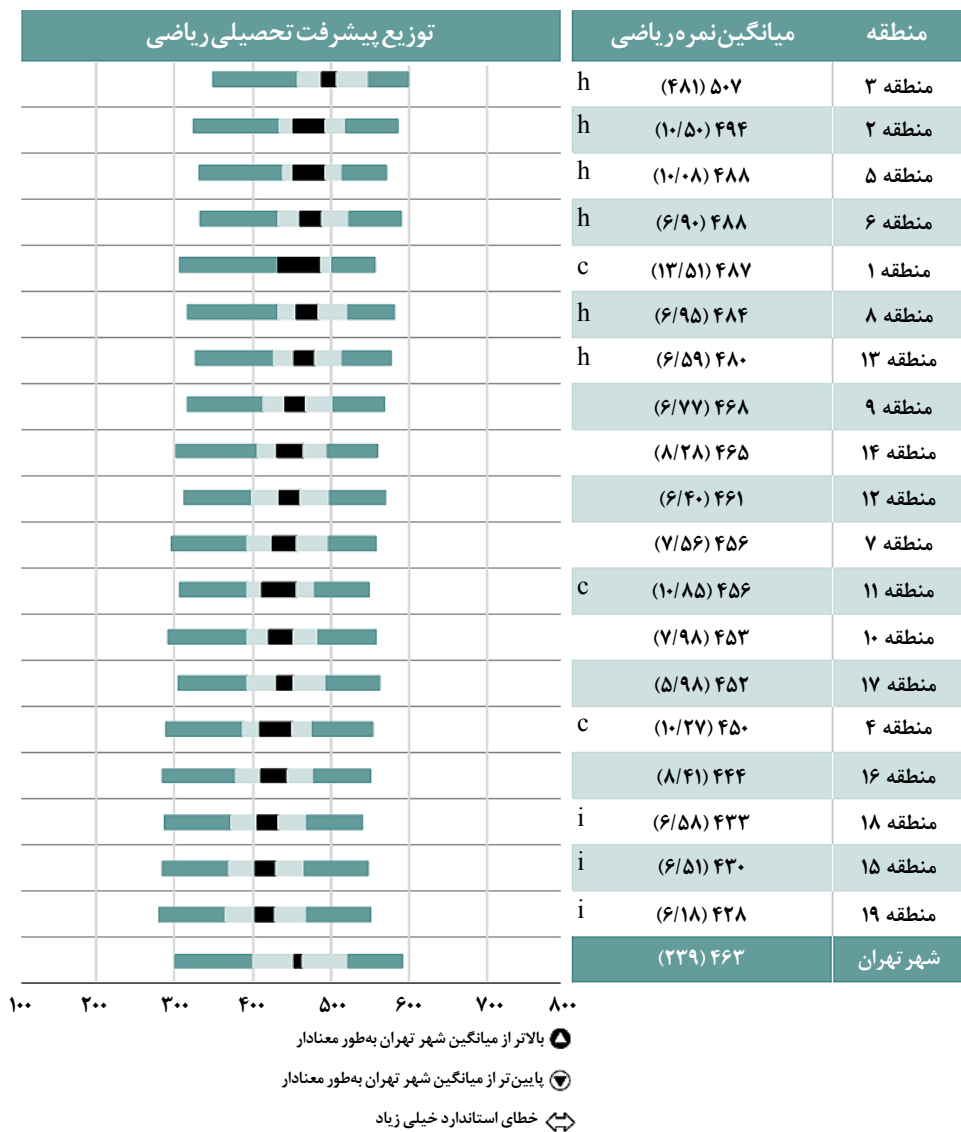
تفاوت عملکرد دانش‌آموزان دختر و پسر در درس ریاضی یکی دیگر از موضوعات موردعلاقه در این مطالعه بود. توزیع عملکرد هریک از دو جنسیت در نمایه ۲ نشان داده شده است.

نمایه ۲. مقایسه توزیع پیشرفت تحصیلی در درس ریاضی در بین دو جنسیت



این نمایه نشان می‌دهد که دختران با اختلاف ۱۹ نمره نسبت به پسران عملکرد بهتری از خود در ریاضی نشان دادند. میانگین نمره دختران برابر با ۴۷۳ برآورد می‌شود در حالی که میانگین نمره پسران ۴۵۴ است که به طور معناداری از میانگین دختران پایین‌ترند. توزیع پیشرفت تحصیلی دانش‌آموزان پسر نیز نشان می‌دهد که عملکرد ضعیف‌تری از خود نشان داده‌اند. نمره ۵ درصد از دانش‌آموزان پسر از ۲۹۰ کمتر است که نمره متناظر با این درصد در دختران برابر با ۳۱۷ است. با این حال، تفاوت چندانی میان دانش‌آموزان دختر و پسر در منطقه عملکرد بالای توزیع مشاهده نشد، به طوری که نمره ۵ درصد بالایی دانش‌آموزان پسر برابر با ۶۰۰ و دانش‌آموزان دختر برابر با ۶۰۲ بود. تفاوت میان عملکرد ریاضی مناطق آموزشی نیز از جمله تحلیل‌های این مطالعه بود. نتایج در نمایه ۳ نشان داده شده است.

جدول ۳. مقایسه توزیع پیشرفت تحصیلی در درس ریاضی در مناطق آموزشی



این نمایه نشان می‌دهد که بالاترین پیشرفت تحصیلی در درس ریاضی در منطقه ۳ تهران به‌دست آمده است. میانگین عملکرد دانش‌آموزان در این منطقه ۵۰۷ است. پس از این منطقه، به ترتیب دانش‌آموزان مناطق ۲، ۵، ۶، ۱، ۸ و ۱۳ قرار دارند که فاصله آنان با میانگین شهر تهران، به‌جز منطقه ۱، معنادار است. از طرف دیگر، سه منطقه ۱۹، ۱۵ و ۱۸ به ترتیب پایین‌ترین پیشرفت تحصیلی در ریاضی را دارند که هر سه به‌طور معنادار

از میانگین شهر تهران پایین‌ترند. فاصله بین بالاترین عملکرد (منطقه ۳ با میانگین ۵۰۷) تا پایین‌ترین عملکرد (منطقه ۱۹ با میانگین ۴۲۸) ۷۹ نمره برآورد می‌شود. همچنین، دقت در توزیع پیشرفت تحصیلی نشان می‌دهد درحالی‌که ۵ درصد دانش‌آموزان ضعیف منطقه ۳ نمره کمتر از ۳۴۹ گرفته‌اند، در منطقه ۱۹ همین نسبت از دانش‌آموزان نمره کمتر از ۲۸۱ کسب کرده‌اند. از طرف دیگر با اینکه ۵ درصد بالای دانش‌آموزان در منطقه ۱۸ نمرات بالاتر از ۵۶۷ کسب کرده بودند، در منطقه ۲ نمره این نسبت از دانش‌آموزان بالاتر از ۶۲۷ بود. نمایه فوق‌بالیکنه معناداری عملکرد ریاضی هر منطقه را از میانگین شهر تهران نشان می‌دهد، ولی تفاوت‌های معنادار را میان مناطق روشن نمی‌کند. برای آگاهی از معناداری فاصله میان مناطق نمایه ۴ را ببینید.

نمایه ۴. مقایسه پیشرفت تحصیلی در درس ریاضی میان مناطق آموزشی شهر تهران

منطقه	مقیاس نمره مقیاسی	منطقه ۳	منطقه ۲	منطقه ۵	منطقه ۶	منطقه ۱	منطقه ۸	منطقه ۱۳	منطقه ۹	منطقه ۱۴	منطقه ۷	منطقه ۱۱	منطقه ۱۰	منطقه ۱۷	منطقه ۴	منطقه ۱۶	منطقه ۱۸	منطقه ۱۵	منطقه ۱۹
منطقه ۱۹	۵۰۷ (۴۸۱)																		
منطقه ۱۵	۴۹۴ (۱۰/۵۰)																		
منطقه ۱۸	۴۸۸ (۱۰/۰۸)																		
منطقه ۱۶	۴۸۸ (۶/۹۰)																		
منطقه ۴	۴۸۷ (۱۳/۵۱)																		
منطقه ۱۷	۴۸۴ (۶/۹۵)																		
منطقه ۱۰	۴۸۰ (۶/۵۹)																		
منطقه ۱۱	۴۶۸ (۶/۷۷)																		
منطقه ۷	۴۶۵ (۸/۲۸)																		
منطقه ۱۲	۴۶۱ (۶/۴۰)																		
منطقه ۵	۴۵۶ (۷/۵۶)																		
منطقه ۲	۴۵۶ (۱۰/۸۵)																		
منطقه ۱۳	۴۵۲ (۷/۹۸)																		
منطقه ۸	۴۵۲ (۵/۹۸)																		
منطقه ۱	۴۵۰ (۱۰/۲۷)																		
منطقه ۶	۴۴۴ (۸/۴۱)																		
منطقه ۱۸	۴۳۳ (۶/۵۸)																		
منطقه ۱۵	۴۳۰ (۶/۵۱)																		
منطقه ۱۹	۴۲۸ (۶/۱۸)																		

نمودار فوق معناداری فاصله‌ها میان مناطق را نشان می‌دهد. علاوه بر مقایسه میان جنسیت و منطقه آموزشی، مقایسه عملکرد ریاضی میان انواع مدرسه نیز در نمایه ۵ ارائه شده است.

**نمایه ۵. مقایسه توزیع پیشرفت تحصیلی در درس ریاضی در انواع مدرسه**

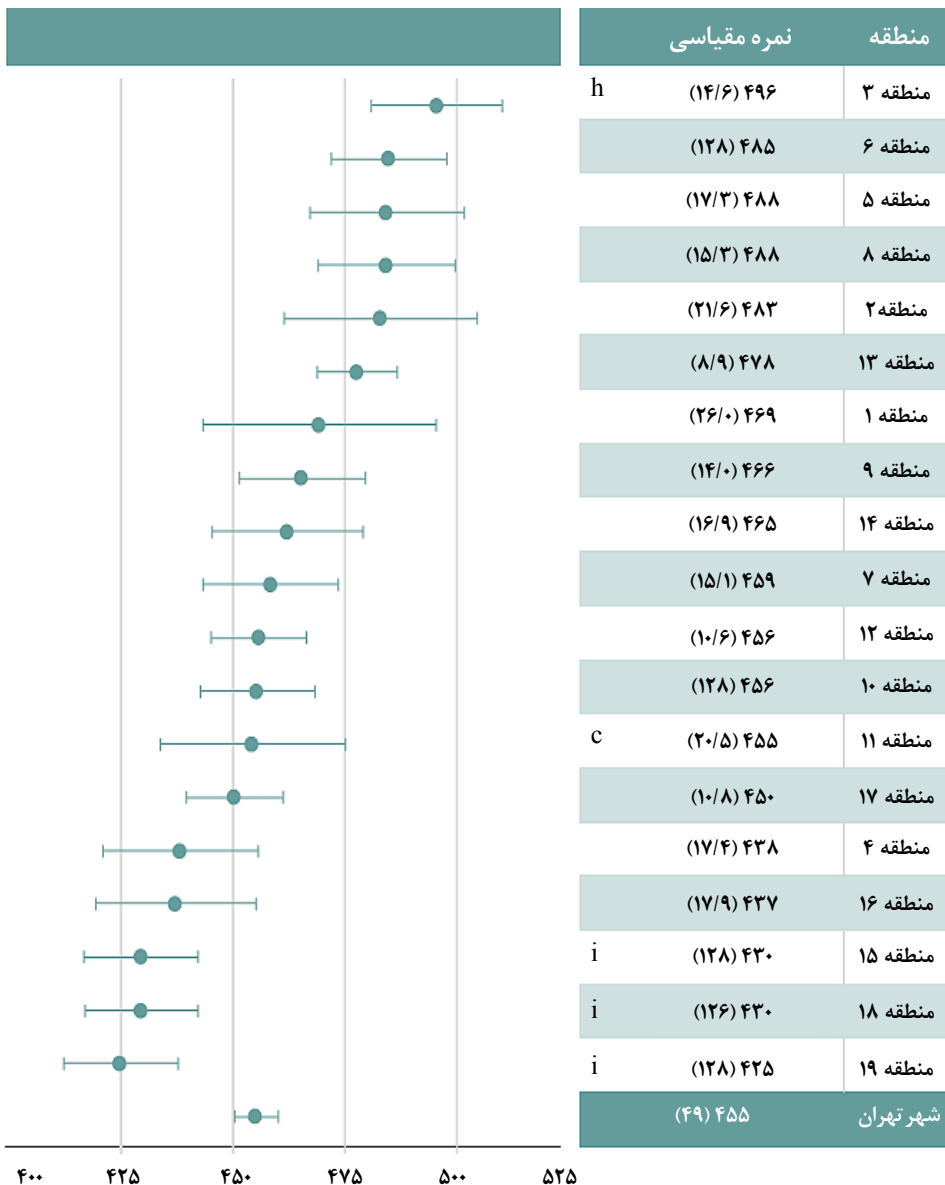


بر اساس نمایه ۵، بیشترین عملکرد تحصیلی در مدارس شاهد و غیردولتی مشاهده شد که از میانگین شهر تهران به‌طور معناداری بالاترند. میان این دو نوع مدرسه تفاوت معناداری دیده نمی‌شود. عملکرد مشاهده‌شده در دانش‌آموزان مدارس شاهد همگون‌تر است، به طوری که ۵ درصد بالایی و پایینی در این مدارس در قیاس با مدارس غیردولتی به میانگین نزدیک‌تر است. پس از این مدارس، مدارس هیئت‌امنایی و دولتی قرار دارند.

بهبود عملکرد مدارس شاهد و غیردولتی ایده مقایسه مناطق را بر اساس مدارس دولتی و هیئت‌امنایی مطرح کرد. منطق پشت این مقایسه آن است که در مناطق آموزشی گوناگون، میان مدرسی که همگی از منبع بودجه دولتی تأمین مالی می‌شوند چه تفاوتی وجود دارد. به عبارت دیگر، این سؤال پیش آمد که میان عملکرد مناطق آموزشی، بدون مدرسی که خانواده‌ها تأمین‌کننده هزینه‌های آن‌ها هستند، چه تفاوتی وجود دارد. نمایه ۶ این مقایسه را نشان می‌دهد.

این نمایه نشان می‌دهد که در کل کیفیت آموزشی مناطق پس از حذف مدارس غیردولتی و شاهد تفاوت چندانی نکرده است. در اینجا نیز مدارس منطقه ۳ بهترین عملکرد را نشان دادند. منطقه ۶ از رتبه قبلی خود (چهارمی) به رتبه دوم ارتقا پیدا کرده است که نشان می‌دهد عملکرد مدارس دولتی و هیئت‌امنایی در این منطقه بهتر از سایر مناطق است. در عوض، منطقه ۲، که عملکرد کلی‌اش در نمایه ۳ رتبه دوم را داشت، در نمایه مربوط به مدارس دولتی و هیئت‌امنایی به جایگاه پنجم تنزل یافته است

نمایه ۶. مقایسه عملکرد ریاضی مناطق آموزشی در مدارس دولتی و هیئت‌امنائی



▲ بالاتر از میانگین شهر تهران به‌طور معنادار

▼ پایین‌تر از میانگین شهر تهران به‌طور معنادار

↔ خطای استاندارد خیلی زیاد

## ■ بحث و نتیجه‌گیری ■

این مطالعه با هدف سنجش کیفیت یادگیری ریاضی دانش‌آموزان پایه ششم شهر تهران اجرا شد. بدین منظور، ابتدا گستره مفاهیم و موضوعات درسی از طریق بررسی مجدد کتاب‌های ریاضی ابتدایی احصا شد و سپس برای هر یک از مباحث درسی به تعداد متناسبی سؤال ریاضی در سطوح مختلف شناختی طراحی شد. مجموع سؤال‌های طراحی شده و سؤال‌های مطالعه تیمز، که برای اجرای پیونددهی استفاده شده بودند (جمعاً ۲۶۴ سؤال شامل سؤال‌های چندگزین‌های و بازپاسخ)، در ۲۰ دفترچه جداگانه توزیع شدند. هر دفترچه به یکی از ۹۷۲۰ دانش‌آموزی داده شد که تصادفی به‌عنوان نمونه معرف دانش‌آموزان پایه ششم شهر تهران انتخاب شده بودند. نمره‌های ریاضی دانش‌آموزان با استفاده از روش‌شناسی مقادیر محتمل به مقیاسی تبدیل شد که میانگین آن ۵۰۰ و انحراف استاندارد آن ۱۰۰ است.

نتایج نشان داد که عملکرد ریاضی دانش‌آموزان شهر تهران در پایه ششم از میانگین مورد انتظار از آنان پایین‌تر است، به طوری که میانگین آنان در مقیاس اشاره‌شده، در فوق، ۴۶۳ بود. کاهش میانگین عملکرد نسبت به میانگین مورد انتظار به میزان ۰/۳۷ انحراف استاندارد مقیاس در نظر گرفته می‌شود. با وجود این توزیع عملکرد ریاضی دانش‌آموز بیشتر به سمت دامنه پایین توزیع گسترده شده است، به طوری که نقطه ۵ درصدی برابر با ۳۰۱ و نقطه ۹۵ درصدی معادل ۶۰۰ برآورد شد که نشان می‌دهد توزیع به سمت دامنه پایین گرایش دارد.

همچنین، مقایسه عملکرد دختران و پسران نشان داد که عملکرد دختران به‌طور معناداری بهتر از پسران است. با این حال، بررسی توزیع عملکرد ریاضی روی هر دو جنسیت نشان داد که کاهش عملکرد پسران در مقایسه با دختران در دامنه پایین توزیع قرار دارد. به عبارت دیگر، باینکه دانش‌آموزان ضعیف پسر عملکرد پایین‌تری در مقابل دانش‌آموزان ضعیف دختر دارند، ولی تفاوتی میان دانش‌آموزان دختر و پسر قوی وجود ندارد. علاوه بر این، مقایسه عملکرد انواع مدارس نیز حاکی از پایین‌تر بودن عملکرد دانش‌آموزان مدارس دولتی در مقایسه با سایر مدارس است. با این حال، تفاوت معناداری میان عملکرد دانش‌آموزان مدارس شاهد، غیردولتی، و هیئت‌امنائی ملاحظه نشد.

مقایسه عملکرد ریاضی در مناطق آموزشی مختلف شهر تهران تفاوت قابل توجهی را از کیفیت یادگیری در میان آن‌ها نشان داد. بهترین عملکرد در منطقه ۳ مشاهده شد که میانگین آن در این منطقه اندکی بالاتر از میانگین شهر تهران (۵۰۷) بود. منطقه‌های ۲، ۵، ۶ و ۱ هم پس از منطقه ۳ بهترین عملکرد را در میان مناطق شهر تهران حاصل کرده‌اند. از طرف دیگر، سه منطقه ۱۹، ۱۵ و ۱۸ به ترتیب پایین‌ترین پیشرفت تحصیلی در ریاضی را دارند که هر سه به‌طور معنادار از میانگین شهر تهران پایین‌ترند. تفاوت ۷۹ نمره‌ای میان بالاترین و پایین‌ترین عملکرد منطقه نشان‌دهنده فاصله قابل توجه و توزیع نامتناسب مشاهده‌شده در کیفیت یادگیری

در میان مناطق آموزشی شهر تهران است. برخلاف انتظار، تفاوت میان مناطق آموزشی به علت نسبت مدارس غیردولتی و شاهد نبود، به طوری که مقایسه عملکرد مناطق فقط در مدارس دولتی و هیئت‌امانی نیز فاصله نسبتاً مشابهی میان بالاترین و پایین‌ترین منطقه (۷۱ نمره) نشان داد. به علاوه باینکه تفاوت‌های محدودی در جایگاه قبلی و بعدی دیده شد (برای مثال، کاهش سه‌رتبه‌ای برای منطقه ۲ و افزایش دورتبه‌ای برای مناطق ۶ و ۸)، ولی جایگاه تقریباً یکسانی برای مناطق با کنار گذاشتن مدارس غیردولتی و شاهد ملاحظه شد.

مقایسه میان نمره‌های حاصل از این مطالعه و مطالعه تیمز در سال ۲۰۱۵ نشان می‌دهد که اگرچه افزایش حدود ۳۲ نمره‌ای میان این دو مطالعه وجود دارد، ابهامی در مورد انتساب تفاوت به پایه‌های سنجش شده یا نمونه‌ها (شهر تهران در مقابل ایران) وجود دارد. باین حال، با احتمال بیشتر تفاوت‌های مشاهده‌شده تفاوت عملکرد در استان‌های مختلف است زیرا بر اساس مطالعه پیونددهی که میان نتایج مطالعه حاضر و مطالعه تیمز ۲۰۱۵ ایجاد شد، نتایج مطالعه حاضر در همان مقیاس تیمز ۲۰۱۵ در پایه چهارم قرار می‌گیرد و بنابراین نتایج مقایسه‌پذیر با یکدیگرند. پایین‌تر بودن عملکرد نسبت به نقطه وسط مقیاس (نمره ۵۰۰) موضوع جدیدی نیست و در همه سنجش‌های بین‌المللی این موضوع دیده می‌شود، اما نکته قابل توجه این است که این سنجش در شهر تهران، و نه در سطح کشور، برگزار شده است؛ به این معنا که در این نمونه دیگر از مدارس روستایی و چندپایه خبری نیست، که معمولاً فرض می‌شود در کاهش میانگین کشور مؤثرند. بنابراین اینکه میانگین پیشرفت تحصیلی در درس ریاضی دانش‌آموزان پایه ششم در شهر تهران هنوز به سطح انتظار نرسیده است، نشان می‌دهد که نقصان چشمگیری در آموزش عمومی کشور وجود دارد و هنوز باید کار بیشتری در این زمینه صورت پذیرد.

برتری عملکرد دختران شهر تهران در مقایسه با هم‌تایان پسر خود با رویه‌های منتشرشده در مطالعه تیمز همخوانی ندارد. بر اساس نتایج مطالعه تیمز اگرچه عملکرد دختران ایرانی در پایه چهارم ۱۰ نمره بیش از پسران بود، این تفاوت از لحاظ آماری معنادار نبود. در مطالعه حاضر، عملکرد دختران ۱۹ نمره بیش از پسران است و این افزایش معنادار به‌شمار می‌رود. این نتیجه ممکن است مؤید این نکته باشد که الگوی برابری عملکردی در بین دو جنسیت دانش‌آموزان در دوره ابتدایی، که از ابتدا در ایران وجود داشته است، لزوماً در تهران مصداق ندارد. مطالعات دیگری که عملکرد ریاضی دانش‌آموزان را میان دو جنسیت مقایسه کرده‌اند نیز کمابیش همین موضوع را نشان می‌دهند. طبق برخی از این مطالعات، عملکرد بالاتر دختران معنادار بوده است (کبیری، ۱۳۸۲) و در برخی دیگر، با وجود بیشتر بودن عملکرد دختران، تفاوت معناداری دیده نشده است (کیامنش و پوراصغر، ۱۳۸۸؛ کرامتی و شهرآرای، ۱۳۸۳). مشاهده‌نشدن تفاوت معنادار میان دو جنسیت تا اندازه بسیاری ناشی از حجم نمونه نسبتاً محدودی است که در این دو مطالعات دیده شده است، به طوری که در حجم‌های حدوداً ۳۰۰ نفری تفاوت‌های معنادار به‌شمار می‌آید ملاحظه می‌شود. در هر صورت، آنچه می‌توان از این مقایسه استنباط کرد آن است که دختران تهرانی

نسبت به پسران عملکرد بهتری دارند اما این برتری در کل کشور دیده نمی‌شود. در این مطالعه با بهره‌گیری از حجم نمونه‌ای چشمگیر تأثیر نوع مدرسه در عملکرد دانش‌آموزان بررسی شد. این مقایسه، برخلاف تحلیل‌های قبلی، همه انواع مدارس را در دوره ابتدایی (دولتی، غیردولتی، هیئت‌امنائی و شاهد) شامل می‌شد. باین حال، باید توجه داشت که تعداد برخی مدارس، همانند مدارس شاهد، چندان زیاد نیست که تعداد فراوانی از مدارس نمونه را شامل شود. باین همه، نتایج نشان داد که عملکرد مدارس دولتی از سایر مدارس ضعیف‌تر است، ولی تفاوت معناداری میان عملکرد سایر مدارس مشاهده نشد. باین حال، از لحاظ میانگین مشاهده شده عملکرد مدارس شاهد بهتر از همه بود و پس از آن مدارس غیردولتی و هیئت‌امنائی بهتر بودند. تفاوت غیرمعنادار میان عملکرد مدارس شاهد و هیئت‌امنائی تا اندازه‌ای ناشی از تعداد کم مدارس شاهد است. علاوه بر این باید توجه داشت که عملکرد مدارس شاهد تقریباً معادل میانگین مدارس غیردولتی ولی همگن‌تر است. این یافته ناشی از تفاوتی چشمگیر در کیفیت خدمات آموزشی مدارس غیردولتی، به خصوص در دوره ابتدایی، است. با وجود اینکه عملکردهای قابل توجهی در برخی از مدارس غیردولتی دیده می‌شود، برخی از آن‌ها نیز عملکرد قابل دفاعی ندارند. میزان ۵ درصد از دانش‌آموزان ضعیف مدارس غیردولتی نمره کمتری از ۳۲۷ دریافت کرده‌اند در حالی که درصد مشابه در مدارس شاهد و هیئت‌امنائی به ترتیب برابر با ۳۳۸ و ۳۱۸ بود. این نکته نشان می‌دهد که کیفیت آموزشی در مدارس ضعیف غیردولتی چندان مطلوب نیست. البته باید به این نکته توجه داشت که این یافته مربوط به دوره ابتدایی است و احتمالاً تعمیم‌پذیر به دوره متوسطه نیست.

تفاوت‌های عملکرد مناطق آموزشی شهر تهران گویای شکاف عملکردی قابل توجه میان مناطق آموزشی است. باین حال، زمانی که صرفاً مدارس دولتی و هیئت‌امنائی باهم مقایسه شدند نیز جایگاه مناطق تفاوت جدی نمی‌کرد. این نتیجه نشان می‌دهد که تفاوت‌های میان مناطق آموزشی و شرایط اقتصادی و اجتماعی ساکنان آن مناطق است و این موارد نمی‌تواند توجیه مناسبی برای کیفیت نامناسب آموزشی در برخی از مناطق باشد. باین همه به علت اینکه داده‌های قبلی برای این مطالعه موجود نبود نمی‌توان قضاوت کرد که تفاوت میان کیفیت آموزشی در مناطق مختلف در مقایسه با قبل بیشتر شده است یا کمتر. تکرار مطالعاتی از این جنس می‌تواند پاسخ مناسبی برای این سؤال مهم باشد.

با توجه به اینکه داده‌های این مطالعه اطلاعات لازم برای برنامه‌ریزی و سیاست‌گذاری مبتنی بر شواهد را در اختیار مدیران، کارشناسان و برنامه‌ریزان آموزش و پرورش، به خصوص در شهر تهران، قرار می‌دهد، انتظار می‌رود که راهبری آموزشی بر اساس این اطلاعات قرار گیرد. این فرایند باعث می‌شود اهداف عملیاتی سنجش‌پذیری همچون ارتقای هدف‌گذاری شده یادگیری دانش‌آموز و کاهش فاصله عملکردی میان مناطق آموزشی به‌منزله برنامه آموزش و پرورش در نظر گرفته شود. تکرار مطالعاتی از این دست، موفقیت در دستیابی به این اهداف را نشان خواهد داد.



منابع  
REFERENCES

- افضل، افشین، دلاور، علی، فلسفی نژاد، محمدرضا، فرخی، نورعلی و برجعلی، احمد. (۱۳۹۳). کاربرد مدل‌های تشخیصی شناختی در تعیین ماهیت تفاوت در عملکرد دانش‌آموزان دختر و پسر پایه اول دبیرستان در ریاضیات. *دست‌آوردهای روان‌شناختی*، ۴ (۲)، ۸۹-۱۰۴.
- زکی، محمدعلی. (۱۳۸۴). بررسی جامعه‌شناسی نقش عوامل اجتماعی، رضایت اجتماعی و عزت‌نفس بر موفقیت تحصیلی. *مجله علوم تربیتی و روان‌شناسی دانشگاه سیستان و بلوچستان*، ۲، ۴۱-۲۱.
- حجازی، الهه، عباسی، فهیمه و مسلحی، حامد. (۱۳۹۷). تحلیل چندسطحی عوامل سطح دانش آموز و مدرسه بر پیشرفت تحصیلی دانش‌آموزان. *فصلنامه نوآوری‌های آموزشی*، ۶۷، ۹۴-۷۱.
- خانی، محمدحسین و محمدی، محسن (۱۳۹۴). نقش تفکیک جنسیتی بر پیشرفت تحصیلی و مقبولیت اجتماعی در مدارس ابتدایی. *پژوهش‌های تربیتی*، ۳۰، ۱۱۹-۱۰۵.
- قاضی طباطبایی، محمود، حاتمی، معصومه و نقش، زهرا. (۱۳۹۳). رابطه خودکارآمدی و نگرش ریاضی با پیشرفت ریاضی. *مجله روان‌شناسی*، ۱۸ (۲)، ۱۶۰-۱۴۶.
- کرامتی، هادی و شهرآرای، مهرناز. (۱۳۸۳). بررسی نقش خودکارآمدی ادراک‌شده در عملکرد ریاضی. *فصلنامه نوآوری‌های آموزشی*، ۱۰ (۳)، ۱۱۵-۱۰۳.
- کبیری، مسعود. (۱۳۸۲). نقش خودکارآمدی ریاضی به همراه متغیرهای شخصی دانش‌آموزان دختر و پسر سال سوم راهنمایی (پایان‌نامه کارشناسی ارشد). دانشگاه تربیت معلم تهران.
- کبیری، مسعود. (۱۳۹۹). برنامه رصد کیفیت آموزشی شهر تهران (برکات): تعیین کیفیت یادگیری ریاضیات در دانش‌آموزان پایه ششم شهر تهران (جلد اول، طرح پژوهشی منتشر نشده). اداره کل آموزش و پرورش شهر تهران.
- کبیری، مسعود، کریمی، عبدالعظیم و بخشعلی‌زاده، شهرناز (۱۳۹۵). یافته‌های ملی تیمز ۲۰۱۵: روند ۲۰ ساله آموزش علوم و ریاضیات ایران در چشم‌انداز بین‌المللی. انتشارات مدرسه.
- کیامنش، علیرضا و پوراصغر، نصیبه. (۱۳۸۸). بررسی تفاوت‌های دختر و پسر در متغیرهای مرتبط به عملکرد تحصیلی (خودپنداره ریاضی، انگیزش یادگیری ریاضی عملکرد قبلی ریاضی) و نقش آن بر پیشرفت ریاضی. *مجله روان‌شناسی دانشگاه تبریز*، ۱۳، ۱۹۲-۱۶۳.
- کیامنش، علیرضا و خیریه، مریم. (۱۳۷۹). سنجش صلاحیت‌های پایه: ارزشیابی درون‌دادها و بروندادهای آموزشی در ایران، پایه پنجم ابتدایی (گزارش منتشر نشده). پژوهشکده تعلیم و تربیت، وزارت آموزش و پرورش.
- مدنی، سیداحمد، پورسینا، داوود، امینی، محمد و یوسفی نژاد. (۱۳۹۸). تحلیل روابط چندگانه میان مهارت‌های تدریس معلمان با پیشرفت تحصیلی دانش‌آموزان در درس‌های محاسباتی. *فصلنامه نوآوری‌های آموزشی*، ۷۰، ۵۲-۳۱.
- نوغانی، محسن، آهنچیان، محمدرضا و رفیعی، محمدتقی. (۱۳۹۰). تأثیر سرمایه‌های اقتصادی، اجتماعی و فرهنگی بر موفقیت در آزمون سراسری ورود به دانشگاه. *جامعه‌شناسی آموزش و پرورش*، ۱، ۱۹۱-۲۱۸.

- Foy, P., & Yin, L. (2016). Scaling the TIMSS 2015 achievement data. In M. Martin, I. Mullis, & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015* (pp. 13.11-13.63). Boston: TIMSS & PIRLS International Study Center.
- Greaney, V., & Kellaghan, T. (2008). *Assessing national achievement levels in education* (Vol. 1). World Bank Publications.
- Kabiri, M. (2019). BARAKAAT: Program for monitoring educational quality in Tehran - Iran's first provincial large-scale assessment, UNESCO-NEQMAP website. <https://neqmap.bangkok.unesco.org/barakaat-program-for-monitoring-educational-quality-in-tehran-irans-first-provincial-large-scale-assessment/>
- Kirsch, I., Lennon, M., von Davier, M., Gonzalez, E., & Yamamoto, K. (2013). On the growing importance of international large-scale assessments. In M. von Davier, E. Gonzalez, I. Kirsch, & K. Yamamoto (Eds.), *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research* (pp. 1-11). Springer.
- LaRoche, S., Joncas, M., & Foy, P. (2016). Sample design in TIMSS 2015. In M. O. Martin, I. V. S. Mullis, & M.

- Hooper (Eds.), *Methods and Procedures in TIMSS 2015* (pp. 3.1-3.37). TIMSS & PIRLS International Study Center and International Association for the Evaluation of Educational Achievement (IEA).
- Lehmann, R. (2006). What should be measured in a cross-national study? In K. Ross & I. Genevois (Eds.), *Cross-national studies of the quality of education: Planning their design and managing their impact*. (pp. 135-154). International Institute for Educational Planning.
  - Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). *International large-scal assessment data*. *Educational Researcher*, 39(2), 142.
  - UNESCO. (2017). *Analyzing and utilizing assessment data for better learning outcomes*. UNESCO.
  - Von Davier, M., Gonzalez, E., & Mislevy, R. J. (2009). What is plausible values and why are they useful? In M. von Davier and D. Hastedt (eds.) *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments* (Vol. 2, pp. 9-36). IEA-ETS Research Institute.
  - Wagemaker, H. (2014). International large-scale assessments: From research to policy. In L. Rutkowski, M. v. Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 11-36). FL: CRC Press.
  - World Bank (2018). *World development report: Learning to realize education's promise*. World Bank.
  - Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, 31(2-3), 114-128.

## پی‌نوشت‌ها

1. Kirsch
2. Millennium Development Goals
3. Education for All
4. Wagemaker
5. World Bank
6. core subjects
7. منظور از نشانگر این است که برخی از دروس نشانه‌ای از کلیت کیفیت آموزش‌اند و به آن‌ها توجه می‌شود.
8. Greaney & Kellaghan
9. Lehmann
10. stratified two-stage cluster sample design
11. LaRoche
12. Implicit stratification
13. explicit stratification
14. Probabilities Proportional to their Size (PPS)
15. sampling weights
16. Rutkowski
17. total weight of students
18. matrix sampling
19. inter-rater reliability
20. plausible values
21. Von Davier
22. Wu
23. linear transformation
24. Linking
25. Foy & Yin
26. logit metric