



دانشگاه علامه طباطبایی
دانشکده روانشناسی و علوم تربیتی

رساله دکتری

مدل‌پردازی تشخیصی شناختی (CDM) سوالهای ریاضیات تیمز ۲۰۰۷ در دانش‌آموزان پایه هشتم
ایران با استفاده از مدل یکپارچه با پارامترپردازی مجدد (RUM) و مقایسه مهارت‌های ریاضی
دانش‌آموزان دختر و پسر

استاد راهنما

دکتر علی دلاور

دکتر محمد رضا فلسفی‌نژاد

استاد مشاور

دکتر علیرضا کیامنش

دکتر یحیی مهاجر

دانشجو

اصغر مینائی

۱۳۹۱

تقدیر و تشکر

پس از حمد و ثنای خداوند سبحان که تمامی تلاشها بدون عنایت او بی ثمر است بر خود لازم می دانم از اساتید خودم جناب آقای دکتر علی دلاور و جناب آقای دکتر محمدرضا فلسفی نژاد که از همان مراحل اولیه انتخاب موضوع، تهیه و تدوین پروپوزال و در به سرانجام رساندن رساله از هیچ کمک و همفکری دریغ نکردند صمیمانه تشکر نمایم.

از اساتید بزرگوارم جناب آقای دکتر علیرضا کیامنش و جناب آقای دکتر یحیی مهاجر که آنها نیز از همان مراحل ابتدایی شروع پژوهش نقطه نظرات و مشاوره‌های سودمندی را داشتند از صمیم قلب تشکر و قدردانی می‌کنم.

از جناب آقای دکتر ولی‌اله فرزاد و جناب آقای دکتر نورعلی فرخی نیز که زحمت داوری رساله را برعهده گرفتند و نظرات ارزشمندی را ارائه نمودند صمیمانه سپاسگزارم.

در پایان از همسر و فرزندانم (ارشیا و اشکان) که با صبر و حوصله، تمام مشقات و مشکلات دوران تحصیل وهمچنین مشکلات شغلی و حرفه‌ای اینجانب را تحمل کرده و می‌کنند از صمیم قلب سپاسگزارم و موفقیت و سلامتی آنها آرزوی قلبی من است.

**این رساله با حمایت مالی پژوهشگاه آموزش
و پرورش به اجرا در آمده است**

فهرست مطالب

| صفحه | مطالب |
|------|---|
| | فصل اول: کلیات پژوهش |
| ۸ | مقدمه |
| ۱۲ | مساله پژوهش |
| ۱۴ | ضرورت و اهمیت پژوهش |
| ۱۸ | اهداف پژوهش |
| ۲۰ | سوالات پژوهش |
| ۲۱ | تعریف مفاهیم |
| | فصل دوم: مبانی نظری و ادبیات پژوهش |
| ۲۴ | مقدمه |
| ۲۵ | رویکردهای سنتی روان‌سنجی |
| ۲۶ | نظریه کلاسیک آزمون |
| ۲۸ | نظریه سوال پاسخ |
| ۲۹ | سنجش تشخیصی شناختی |
| ۳۱ | روان‌شناسی شناختی و روایی سازه |
| ۳۳ | روان‌شناسی شناختی و تولید سوال |
| ۳۳ | روان‌شناسی شناختی و تفسیر تشخیصی نمره‌های آزمون |
| ۳۴ | مدلهای آماری سنجش تشخیصی شناختی |
| ۳۶ | مرور بر چسبهای موجود |
| ۳۸ | مروری بر مدل‌های CDM |
| ۴۵ | مدلهای تشخیصی شناختی مبتنی بر IRT |
| ۴۷ | مدلهای غیر جبرانی |
| ۵۳ | مدلهای جبرانی |
| ۵۳ | مروری بر کاربردهای مدل‌های CDA |

| | |
|-----|--|
| ۵۴ | سنجش آموزشی |
| ۶۲ | سنجش روانشناختی |
| ۶۳ | خلاصه و نتیجه گیری |
| | فصل سوم: روش شناسی پژوهش |
| ۶۷ | روش پژوهش |
| ۶۷ | جامعه آماری، گروه نمونه و روش نمونه گیری |
| ۶۸ | ابزار تحقیق |
| ۷۱ | شیوه تحلیل داده ها |
| ۷۱ | مجموعه تحلیل های اول |
| ۷۱ | مجموعه تحلیل های دوم |
| | فصل چهارم: یافته های پژوهش |
| ۷۸ | یافته های توصیفی |
| ۸۴ | ارزشیابی برازش مدل |
| ۹۵ | یافته های مربوط به سوالهای پژوهش |
| | فصل پنجم: بحث و نتیجه گیری |
| ۱۱۹ | برازش مدل یا کفایت مدل شناختی |
| ۱۲۴ | وضعیت دانش آموزان در صفات ۲۰ گانه |
| ۱۳۲ | مقایسه عملکرد دختران و پسران |
| ۱۳۶ | نتیجه گیری |
| ۱۴۰ | محدودیت ها |
| ۱۴۱ | پیشنهادات |
| ۱۴۴ | منابع |



فصل اول

کلیات پژوهش

مقدمه

نیاز و تقاضا برای آزمون‌ها از زمان الزامی شدن آموزش، یعنی اوایل قرن بیستم آغاز گشت. مشکل معلمان و مربیان در این دوره این بود که کودکانی که وارد مدرسه می‌شدند از لحاظ توانایی‌ها و زمینه‌های خانوادگی بسیار متنوع و گوناگون بودند و لذا معلمان، نیازمند روش‌هایی بودند که با استفاده از آنها بتوانند تعیین کنند کدام دسته از دانش‌آموزان از آموزش یکدست و یکپارچه‌ای که اساسا برای اکثریت افراد طراحی شده بود قادر خواهند بود بیشترین سود را ببرند. تلاش برای حل این معضل، باعث شد تا آزمون‌های گزینش تحصیلی به عنوان بخشی از راه‌حل، به کار گرفته شوند زیرا با استفاده از این آزمون‌ها می‌توانستند پیش‌بینی کنند که کدام گروه از افراد موفق خواهند بود، کدام یک به آموزش ویژه نیاز دارند و کدام افراد را بهتر است از رفتن به پایه بالاتر جلوگیری نمود (نیکولز^۱، ۱۹۹۴).

سنجش‌های سنتی که برای گزینش، پرورش یافت عموماً بر چیزی که اسنو و لهمان^۲ (۱۹۸۹) آن را رویکرد اندازه‌گیری روان‌سنجی آموزشی^۳ (EPM) می‌نامند استوار است. هدف نظریه‌های روان‌سنجی مبتنی بر مفهوم‌پردازی EPM، برآورد جایگاه شخص در متغیر مکنون زیربنایی است. در واقع، نظریه‌ها و روش‌های متکی بر رویکرد EPM بر مقیاس‌پردازی تک بُعدی آزمودنی‌ها در موضوعات اصلی و حوزه‌های شناختی مانند استدلال کلامی، توانایی ریاضی و غیره متمرکز است که غالباً با استفاده از مدل‌پردازی نظریه کلاسیک آزمون^۴ (CTT) یا مدل‌های تک بُعدی نظریه سوال پاسخ^۵ (IRT) و به ویژه مدل‌های لوجستیک یک، دو و سه پارامتری صورت می‌گیرد. سودمندی این مدل‌ها بر اساس کارایی آنها در درجه‌بندی یا جایابی^۶ افراد روی مقیاس متغیر مکنون، به منظور کمک به گزینش آنها، مورد قضاوت قرار می‌گیرد.

¹ - Nichols

² - Snow & Lohman

³ - Educational Psychometric Measurement

⁴ - Classical Test Theory

⁵ - Item Response Theory

⁶ - Locating

سنجش‌های مبتنی بر رویکرد EPM با نظریه‌های یادگیری اوایل قرن بیستم از هماهنگی و تناسب خوبی برخوردار است (نیکولز، ۱۹۹۴). برای مثال، آزمون‌های سنتی که با استفاده از رویکرد نمونه‌گیری ساخته می‌شود با مدل‌های یادگیری تداعی‌گرا^۱ که در آغاز قرن بیستم رواج داشتند از هماهنگی و تناسب خوبی برخوردار است (رسنیک و رسنیک، ۱۹۹۲، نقل از نیکولز، ۱۹۹۴). طراحی آنها متکی بر طبقه‌بندی منطقی و مشخصات محتوایی است و فاقد مدل‌های روان‌شناختی روشن و مدون در زمینه ساختارها و فرایندهای زیربنایی عملکرد در آزمون هستند (اسنو و ماندیناک^۲، ۱۹۹۱). نمره‌های این آزمون‌ها بر اساس حوزه‌های محتوایی تفسیر می‌شوند تا مکانیسم‌های شناختی.

متاسفانه، نمره‌های حاصل از مدل‌های سنتی روان‌سنجی، از جمله CTT و IRT، اطلاعات شناختی مستتر در آزمون را منعکس نمی‌سازند (اسنو و لهما، ۱۹۸۹). همانطور که امبرتسون (۱۹۸۳) اظهار می‌دارد "اگرچه مدل‌های IRT از مزایای زیادی نسبت به روش‌های پیشین برخوردارند، لکن ارتباط اندکی با موضوعات و مسائل مورد توجه در نظریه‌های شناختی درباره فرایندها، راهبردها، و ساختارهای دانش که زیربنای حل سوال قرار می‌گیرند، دارند" (صفحه ۱۲۵). در واقع، یکی از محدودیت‌های مدل‌های سنتی روان‌سنجی این است که عملکرد آزمون‌شوندگان در آزمون‌ها به صورت یک نمره منفرد، مانند نمره خام یا برآوردهای توانایی (تتا)، ارائه می‌شود که اطلاعاتی در خصوص اینکه دانش‌آموزان واقعا چه دانش یا مهارت‌هایی را دارا هستند فراهم نمی‌کنند. این محدودیت، اساسا ناشی از لحاظ نکردن اطلاعات شناختی در مدل‌های سنتی روان‌سنجی است (اسنو و لهما، ۱۹۸۹). این ضعف در فراهم آوردن اطلاعات شناختی به نوبه خود، سودمندی آزمون‌ها را به عنوان ابزاری برای دادن پسخوراند تشخیصی به معلمان کلاس درس، محدود می‌سازد. می‌توان گفت که در قرن بیستم از آزمون‌های پیشرفت تحصیلی عمدتا برای اهداف پذیرش، جایابی، اعطای بورس تحصیلی، ارتقاء پایه تحصیلی و همچنین رتبه‌بندی آزمون‌شوندگان در حوزه‌های شناختی مهم استفاده می‌شد تا تشخیص ضعف‌ها و

¹ - Associationist

² - Snow & Mandinach

قوت‌های آنان. به طور کلی، هدف اصلی آزمون‌های پیشرفت تحصیلی مبتنی بر مدل‌های سنتی روان‌سنجی این است که از طریق مقایسه عملکرد دانش‌آموز با یک گروه هنجار، درباره توانایی کلی وی دست به استنباط‌هایی بزنند (براون و هادسون^۱، ۲۰۰۲). چنین آزمون‌هایی به خاطر اینکه درباره ضعف‌ها و قوت‌های فراگیران در حیطه‌های تحصیلی مورد نظر، اطلاعات تشخیصی در اختیار آنها و معلمان قرار نمی‌دهند، مورد انتقاد قرار گرفته‌اند (نیکولز، ۱۹۹۴؛ اسنو و لهمان، ۱۹۸۹). با افزایش نارضایتی از این آزمون‌ها متخصصان و دست‌اندرکاران سنجش آموزشی، خواستار اطلاعات تشخیصی بیشتر برای هدایت یادگیری، بهبود آموزش و ارزشیابی پیشرفت تحصیلی دانش‌آموزان شدند.

در راستای کمک به استفاده از سنجش‌ها برای بهبود و ارتقاء یادگیری، پژوهشگران و متخصصان حوزه سنجش آموزشی با اتکاء به پژوهش‌های صورت گرفته در حوزه روان‌شناسی یادگیری و با بهره‌گیری از مدل‌های آماری، رویکرد سنجش تشخیصی شناختی^۲ (CDA) را به منظور انجام استنباط‌هایی راجع به ساختارها و فرایندهایی که زیربنای عملکرد در حیطه مورد نظر را تشکیل می‌دهند پرورش دادند. در واقع می‌توان گفت که از اواخر قرن ۲۰ و مخصوصاً با شروع قرن ۲۱ هدف‌های آموزش و پرورش تغییر پیدا کرد که این مساله به نوبه خود اهداف سنجش را نیز تغییر داده است (نیکولز، ۱۹۹۴). در گذشته در آموزش و پرورش بر گزینش افراد مستعد برای آموزش و نخبه‌پروری تاکید می‌شد. در قرن حاضر، وضعیت تغییر یافته است. امروزه، هدف نظام‌های آموزش و پرورش، تربیت عالی و خوب نسبت کوچکی از افراد جامعه نیست، بلکه بر کمک به همه افراد برای رسیدن به موفقیت تاکید می‌شود (فو و لی^۳، ۲۰۰۷). بجز مواردی از محرومیت از تحصیل و افت تحصیلی، همه کودکان و نوجوانان وارد مدرسه می‌شوند. کاملاً واضح است که ما در عصری قرار داریم که آموزش، یک کار کاملاً گزینشی یا انتخابی محسوب نمی‌شود بلکه بر پرورش جامعه‌ای با افراد تحصیلکرد تاکید می‌شود. در عصر

¹ -Brown & Hudson

² - Cognitive Diagnostic Assessment

³ - Fu & Li

حاضر بر گزینش افراد برای فرصت‌های آموزشی موجود تاکید کمتری می‌شود و عمدتاً تلاش بر این است که به افراد کمک شود تا در این فرصتها موفق شوند.

تغییر در اهداف سنجش، باعث شده است تا مدل‌های آماری جدیدی در حوزه روان‌سنجی مطرح شوند که با اهداف آموزش و پرورش به طور کلی و با اهداف سنجش آموزشی به طور خاص، هماهنگ و همخوان باشند. این دسته از مدل‌ها امروزه تحت عنوان مدل‌های سنجش تشخیصی شناختی مطرح هستند. احساس نیاز به این مدل‌ها، که در واقع نتیجه یا برآیند تلفیق پیشرفت‌های جدید علمی، سیاسی و آموزشی است دارای تاریخچه نسبتاً طولانی است (به عنوان مثال، کرونباخ^۱، ۱۹۷۵؛ اسنو و لهمان، ۱۹۸۹). دو فصل از کتاب *اندازه‌گیری‌های آموزشی* (لین^۲، ۱۹۸۹) در تشدید و گسترش علاقه و نیاز به سنجش تشخیصی شناختی، نقش به‌سزایی داشتند. در واقع، فصل ساموئل مسیک در مورد *روایی* و فصل ریچارد اسنو و دیوید لهمان با عنوان *تلویحات روانشناسی شناختی برای اندازه‌گیری‌های آموزشی*، به تحکیم رابطه روان‌شناسی شناختی با اندازه‌گیری‌های آموزشی کمک بسیاری نمودند. ایده‌های مطرح شده در این فصول، بسیاری از پژوهشگران را به سمت اندازه‌گیری‌های آموزشی سوق داد و آنان را واداشت تا امکان کاربرد روان‌شناسی شناختی در پرورش آزمون‌ها را مورد توجه قرار دهند. در واقع، سنجش آموزشی از مرحله درجه‌بندی پیشرفت تحصیلی دانش‌آموزان، فراتر رفته و به مرحله‌ای از تکامل رسیده است که می‌تواند در هر مرحله از آموزش اطلاعات تشخیصی سودمندی را فراهم آورد (بولت^۳، ۲۰۰۷). معلمان می‌توانند از این اطلاعات برای تعیین دانش و آموخته‌های پیشین دانش‌آموزان، تشخیص خطاها یا کج‌فهمی‌های آنان در طول فرایند آموزش و تعیین تاثیرات و پیامدهای آموزش، استفاده نمایند. بنابراین، بین آموزش و سنجش یک رابطه تنگاتنگ وجود دارد و سنجش بایستی نقش اساسی خود را در فرایند آموزش ایفا نماید.

^۱ - Cronbach

^۲ - Linn

^۳ - Bolt

مساله پژوهش

در طول چند دهه گذشته، پژوهش در زمینه سنجش آموزشی منجر به پیدایش مدل‌هایی شده است که عموماً از آنها تحت عنوان مدل‌های تشخیصی شناختی^۱ (CDM) نام برده می‌شود. این مدل‌ها به جای به دست دادن یک نمره کلی با بررسی فرایندهای شناختی مورد نیاز برای تکمیل موفقیت‌آمیز تکالیف، اطلاعات خاص و دقیقی را درباره مهارت یا صفتی* که دانش‌آموزان برای دادن پاسخ درست به سوال نیاز دارند فراهم می‌آورند (مک‌گلوهن^۲، ۲۰۰۴). روش‌های سنتی سنجش فقط تعیین می‌کنند که فرد چه چیزهایی را آموخته است، اما اطلاعاتی راجع به اینکه فرد ظرفیت و استعداد یادگیری چه چیزهایی را دارد ارائه نمی‌دهند (امبرتسون، ۱۹۸۳). یک سنجش تشخیصی خوب، پسخوراندی با کیفیت خوب فراهم می‌آورد و نشان می‌دهد که برنامه آموزشی و برنامه درسی تا چه اندازه کارآمد است.

در سنجش تشخیصی شناختی، علاوه بر برآورد توانایی فرد در سازه یا سازه‌های مورد اندازه‌گیری نیم‌رخه برای هر فرد فراهم می‌آید که تسلط یا عدم تسلط وی را در مجموعه‌ای از مهارت‌های اساسی یا صفات شناختی خرد^۳ از پیش تعیین شده که برای پاسخگویی به سوال‌های آزمون لازم و ضروری هستند، مشخص می‌سازد. به عبارت دقیق‌تر، در این مدل‌ها با توجه به تسلط یا عدم تسلط فرد بر مهارت‌ها و صفاتی که برای پاسخگویی درست به سوال، مورد نیاز هستند احتمال پاسخ درست فرد به سوال برآورد می‌شود. در این مدل‌ها می‌توان بر اساس وضعیت تسلطی افراد در هر یک از مهارت‌ها و صفات زیربنایی مورد نیاز سوال‌ها، آنها را به دو گروه مسلط و غیر مسلط و در صورت لزوم به ۳ گروه مسلط، غیرمسلط و نامعین تقسیم بندی نمود و از این طریق مشخص کرد که افراد یا دانش‌آموزان در کدام یک از مهارت‌ها و صفات زیر بنایی دارای مشکل هستند. با اجرای

¹ - Cognitive Diagnostic Models

* در این نوشته، و همچنین در اکثر منابع مربوط به سنجش تشخیصی شناختی، اصطلاحات "مهارت" و "صفت" و "صفت شناختی" به یک معنا به کار می‌روند.

² - McGlohen

³ - Fine-grained

این نوع سنجش و شناسایی افراد ضعیف می‌توان برای آنها برنامه‌های آموزشی جبرانی برگزار نمود و یا با انجام اصلاحات در کتابهای درسی و آموزشی، بر آن دسته از مفاهیم، مهارت‌ها و موضوعاتی که اکثریت دانش‌آموزان در آنها دارای مشکل هستند تاکید بیشتری نمود.

یکی از سنجش‌های آموزشی معتبر بین‌المللی، مطالعات تیمز (TIMSS)^۱ است که با فواصل ۴ ساله اجرا می‌شود. هدف سنجش‌هایی مانند تیمز "بهبود آموزش و یادگیری ریاضیات و علوم از طریق فراهم آوردن داده‌هایی درباره پیشرفت دانش‌آموزان در ارتباط با انواع مختلف برنامه‌های درسی، شیوه‌های آموزشی و محیط‌های تحصیلی یا آموزشی" است (مولیس، مارتین، گونزالس و کروس توسکی^۲، ۲۰۰۴، ص ۱۳).

یکی از کاربردهای جانبی یافته‌های این ارزیابی‌ها رتبه‌بندی کشورها بر اساس عملکرد آنها و شناسایی و تعیین عوامل موثر بر پیشرفت تحصیلی و یادگیری در این کشورها است (مارتین و کلی^۳، ۱۹۹۸). در تمام مطالعات تیمز، رتبه ایران در آزمون ریاضیات که براساس میانگین نمره کل تعیین می‌شود، بسیار پایین بوده است. رتبه ایران در ریاضیات سال سوم راهنمایی در سالهای ۱۹۹۵، ۱۹۹۹، ۲۰۰۳، ۲۰۰۷ و ۲۰۱۱ به ترتیب ۳۷، ۳۳، ۳۴، ۳۴ و ۳۲ بوده است. این در حالی است که تعداد کشورهای شرکت کننده در این مطالعات به ترتیب ۴۱، ۳۸، ۴۶، ۵۹ و ۴۲ بوده است (کریمی، بخشعلی‌زاده و کبیری، ۱۳۹۱). با این حال، این گزارش‌ها و رتبه‌بندیها که بر اساس نمره کل صورت می‌گیرند اطلاعات کافی در زمینه نقاط ضعف دانش‌آموزان فراهم نمی‌آورند. مولیس و همکاران (۲۰۰۰) که گزارش مربوط به آزمون ریاضیات سومین مطالعه بین‌المللی تیمز را تهیه کردند، تلاش نمودند تا اطلاعات بیشتری را در زمینه نقاط ضعف و قوت دانش‌آموزان فراهم نمایند. آنها اظهار می‌دارند "معلم، مربیان، برنامه‌ریزان و سیاست‌گذاران آموزشی بایستی نسبت به آنچه که دانش‌آموزان

¹ - Trends in Mathematics and Science Study

² - Mullis, Martin, Gonzalez, & Chrostowski

³ - Martin & Kelly

در ریاضیات می‌دانند و می‌توانند انجام دهند و همچنین نسبت به حوزه‌ها، مفاهیم و موضوعاتی که نیازمند تمرکز و تلاش بیشتر است آگاهی داشته باشند" (صفحه ۵۷).

کشور ایران به منظور ارزیابی و بهبود نظام آموزشی خود از سال ۱۳۷۰ برابر با ۱۹۹۱ میلادی رسماً همکاری خود را با/انجمن بین‌المللی ارزشیابی پیشرفت تحصیلی^۱ (IEA) آغاز کرد و تا کنون در ۶ مطالعه در فاصله سال‌های ۱۹۹۵، ۱۹۹۹، ۲۰۰۳، ۲۰۰۷، تیمز پیشرفته ۲۰۰۸ و ۲۰۱۱ شرکت کرده است.

با توجه به همکاری ایران با IEA از سال ۱۹۹۱ و مشارکت در TIMSS از سال ۱۹۹۵ و صرف هزینه‌های زیاد برای این مطالعات، و با توجه به عملکرد ضعیف دانش‌آموزان ایران در ریاضیات و رتبه پایین ایران در این مطالعات و همچنین با توجه به اینکه، تا جاییکه نویسنده اطلاع دارد، تا کنون پژوهش منسجمی در زمینه شناسایی و تعیین نقاط ضعف و قوت دانش‌آموزان ایران در صفات و مهارتهایی که زیربنایی عملکرد در ریاضیات را تشکیل می‌دهند صورت نگرفته است، از اینرو مساله اصلی پژوهش حاضر این است که نقاط ضعف و قوت دانش‌آموزان ایران در مهارت‌ها و صفات مورد اندازه‌گیری ریاضیات سال سوم راهنمایی بر اساس داده‌های تیمز ۲۰۰۷ چیست؟ و چه تفاوت‌هایی از این لحاظ بین دختران و پسران وجود دارد؟

ضرورت و اهمیت پژوهش

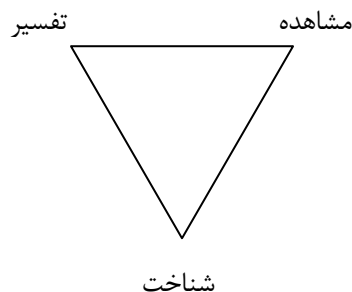
یک سنجش خوب بایستی وسیله‌ای برای اندازه‌گیری پیشرفت افراد (سنجش تراکمی^۲)، کمک به یادگیری (سنجش تکوینی^۳) و ارزشیابی برنامه‌ها فراهم آورد. انتخاب یک روش مناسب سنجش، برای دستیابی به اهداف آموزش از اهمیت اساسی برخوردار است و می‌تواند نگرش دانش‌آموزان و رویکردهای یادگیری آنان را تحت تاثیر قرار دهد و همچنین در بالا بردن انگیزه دانش‌آموزان و در نتیجه، یادگیری عمیق مطالب موثر باشد

^۱- International Association for the Evaluation of Educational Achievement

^۲ - Summative

^۳ - Formative

(سویلیکی، ۲۰۰۷، نقل از سانتیاگو رومان^۱، ۲۰۰۹). مثلث سنجش که توسط پلگرینو و همکاران (۲۰۰۱)، نقل از پلگرینو و چادوسکی^۲، ۲۰۰۳) پرورش یافت و بر پایه ایده سنجش به عنوان فرایند استدلال بر پایه شواهد (میسلوی، ۱۹۹۶) استوار است، چارچوب مناسب و سودمندی را برای تحلیل سنجش‌های فعلی و طراحی سنجش‌های آتی فراهم می‌آورد. مثلث سنجش شامل ۳ راس یا عنصر شناخت، مشاهده و تفسیر (شکل ۱-۱) است.



شکل ۱-۱: مثلث سنجش

در این چارچوب، مدل‌های اندازه‌گیری و به طور کلی سنجش، به صورت فرایند استدلال و انجام استنباط‌های معقول درباره دانسته‌های دانش‌آموزان بر پایه شواهد و مشاهده گفته‌ها و اعمال و رفتارهای آنان در موقعیت‌های معین تلقی می‌شود. برای داشتن یک سنجش کارآمد، این ۳ عنصر بایستی به صورت همزمان و هماهنگ با یکدیگر عمل نمایند. این عناصر ۳ گانه، تمام جنبه‌های طراحی و کاربردهای سنجش، از جمله محتوا، قالب، نمره‌گذاری و استفاده از نتایج را تحت تاثیر قرار می‌دهند (پلگرینو و چادوسکی، ۲۰۰۳). منظور از عنصر شناخت، یک نظریه یا مجموعه‌ای از عقاید درباره چگونگی پرورش و بروز و ظهور دانش، مهارت‌ها و شایستگی‌ها توسط

¹ - Santiago Roman

² - Pellegrino & Chudowsky

دانش‌آموزان در یک حیطه موضوعی خاص است. نظریه بایستی بر معتبرترین و دقیق‌ترین یافته‌های علمی در زمینه شیوه‌های بروز و ظهور دانش یادگیرندگان و پرورش مهارت‌ها در حیطه مورد نظر استوار باشد.

منظور از عنصر مشاهده نیز، توصیف یا تعیین مجموعه‌ای از مشخصات برای تکالیف سنجش است که پاسخ‌ها را از دانش‌آموزان فراخوانی خواهند کرد. این عنصر، محرک‌های ارائه شده به دانش‌آموزان و نتایج بدست آمده مانند پاسخهای شفاهی و کتبی آنها را توصیف می‌کند. پرورش و انتخاب تکالیف سنجش بایستی با در نظر گرفتن هدف سنجش صورت گیرد. سرانجام، عنصر تفسیر در برگرفته تمام روش‌ها و ابزارهای مورد استفاده برای استدلال بر پایه شواهد خطاپذیر^۱ است. این عنصر در واقع، شواهد حاصل از تکالیف شناختی را به دانش شناختی و صفات و مهارت‌های مورد سنجش، مرتبط می‌کند.

هدف سنجش تشخیصی شناختی (CDA) فراهم آوردن بازخورد تشخیصی و تکوینی از طریق گزارش دقیق نیمرخهای تسلط بر صفت یادگیرندگان است (امبرتسون، ۱۹۹۸). مشخصه بارز این رویکرد آن است که به منظور انجام استنباط درباره ضعف‌ها و قوت‌های فراگیران در مهارت‌ها و صفات مورد آزمون، نظریه‌های شناخت مرتبط با حیطه موضوعی مورد نظر را با مدل‌های آماری پیشرفته ترکیب می‌کند.

رویکرد CDA قصد دارد تا با فراهم آوردن اطلاعات مورد نیاز معلمان برای اصلاح آموزش و یادگیری در کلاس درس، سودمندی و نقش سنجش را در یادگیری ارتقاء بخشد. معلمان می‌توانند از اطلاعات تشخیصی تکوینی برای بازسازی و طراحی مجدد رویکردهای آموزشی، ارزشیابی منابع آموزشی و جبران ضعف‌های دانش‌آموزان استفاده کنند. این رویکرد می‌تواند با تشویق دانش‌آموزان به استفاده از سنجش به عنوان ابزار یادگیری، سهم و نقش آنها را در یادگیری افزایش دهد (جانگ، ۲۰۰۸). در واقع، از نتایج CDA می‌توان در اصلاح، تعدیل و متناسب کردن محتوا و روش‌های آموزش با نیازهای دانش‌آموزان و بر طرف ساختن ضعف‌ها و کمبودهای آنان استفاده نمود.

¹ - Fallible

در سند تحول بنیادین آموزش و پرورش (دبیرخانه شورای عالی انقلاب فرهنگی، وزارت آموزش و پرورش، و شورای عالی آموزش و پرورش، ۱۳۹۰)، که در ادامه مطالب به اختصار از آن تحت عنوان سند تحول نام خواهیم برد، بر متناسب‌سازی حجم و محتوای کتب درسی و روزهای آموزشی با توانمندی‌ها و ویژگی‌های دانش‌آموزان (ص ۱۸)، طراحی و تدوین برنامه تعلیم و تربیت انعطاف‌پذیر، متناسب با ویژگی‌های شخصیتی و محیطی استعدادها دانش‌آموزان به منظور شکوفایی استعدادهای خاص و افزایش کارآمدی و مفید بودن آنها (ص ۲۱) و بر استفاده از رویکرد ارزشیابی فرایندمحور (ص ۲۹) تاکید می‌شود.

با در نظر گرفتن مطالبی که در زمینه اهداف و کاربردهای CDA گفته شد، و همچنین با توجه به تاکیدات سند تحول بر ارزشیابی فرایندمحور، و به منظور دستیابی به اهداف این سند، و همچنین با عنایت به همکاری ایران با IEA از سال ۱۹۹۱ و مشارکت در TIMSS از سال ۱۹۹۵، و صرف هزینه‌های زیاد برای این مطالعات، و ضعف عملکرد دانش‌آموزان ایرانی در این مطالعات، ضروری است که با استفاده از مدل‌های تشخیصی شناختی، ضعف‌های دانش‌آموزان در مهارت‌ها و صفات مورد اندازه‌گیری ریاضیات تیمز ۲۰۰۷ و تفاوت‌های بین عملکرد دانش‌آموزان دختر و پسر در مهارت‌های مورد نظر را شناسایی کرد و رهنمودهای عملی و کاربردی در اختیار سیاست‌گذاران و برنامه‌ریزان آموزشی قرار داد تا بتوانند با اعمال تغییرات لازم و مناسب در محتوای درس و شیوه‌های آموزش ریاضیات، دانش و مهارت‌های دانش‌آموزان را در این حوزه مهم و اساسی ارتقاء بخشند.

بنابراین، اهمیت موضوع پژوهش حاضر، هم ناظر بر موضوع مورد مطالعه، یعنی شناسایی نقاط ضعف و قوت دانش‌آموزان ایران در مهارت‌های پایه‌ای ریاضیات بر اساس داده‌های تیمز ۲۰۰۷ و هم ناظر بر بخش روش‌شناسی تحقیق در حوزه سنجش‌های تربیتی و روانی است.

در بخش روش‌شناسی، تا جایکه نویسنده اطلاع دارد، اولین پژوهشی است که از مدل‌های تشخیصی شناختی^۱ (CDM) بر روی داده‌های تیمز در کشور استفاده می‌نماید. بنابراین، در نوع خود پیشرو است و زمینه

¹ - Cognitive Diagnostic Model

را برای استفاده از این مدل‌ها در پرورش و ارزشیابی آزمون‌های تشخیصی در حوزه‌های مختلف، و از جمله در حوزه مهم و اساسی ریاضیات، فراهم می‌نماید.

اهداف پژوهش

در چند دهه اخیر، تلاقی پیشرفت‌های جدید علمی، سیاسی و آموزشی باعث شده تا آزمون‌هایی با هدف‌های کاملاً متفاوت با آزمون‌های تراکمی تک بُعدی پرورش یابد. در واقع نیاز و تقاضای بسیار قوی و فوری به پرورش آزمون‌هایی وجود دارد که بتوانند تکه‌های^۱ خرد دانش را در موضوعات اصلی یا حوزه‌های مهم شناختی، به طور تکوینی مورد سنجش قرار دهند. منظور از تکوینی این است که نتایج حاصل از سنجش به طور مستقیم برای حمایت از آموزش و یادگیری مورد استفاده قرار می‌گیرد که با سنجش تراکمی که برای ارزشیابی دانش‌آموز در پایان دوره آموزش به کار می‌رود در تقابل است.

در واقع، امروزه اعتقاد بر این است که سنجش روانی و تربیتی فرایندی است برای استدلال از مشاهدات یا داده‌ها به استنباط‌هایی درباره افراد. پلگرینو^۲ و همکاران (۲۰۰۱، نقل از پلگرینو و چادوسکی، ۲۰۰۳) از طرح سنجش به عنوان فرایند استدلال از مشاهدات به استنباط‌ها، تحت عنوان مثلث سنجش نام می‌برند که شامل ۳ مولفه اصلی شناخت، مشاهده و تفسیر (شکل ۱-۱) است. معنای تلویحی این مثلث این است که فرایند پرورش سنجش نه تنها بایستی بر ملاحظات محتوایی و آموزشی استوار باشد بلکه بایستی به اندازه‌گیری یا سنجش فرایندهای شناختی دانش‌آموزان نیز توجه شود.

از اواسط دهه ۱۹۸۰ میلادی، اتحاد و پیوند روان‌شناسی شناختی و روانسنجی برای پژوهشگران و متخصصان بسیار جالب و گیرا بوده است. اتحاد این دو حوزه، به طور کلی و CDA به طور خاص، شیوه‌ای قانع‌کننده و

^۱ - Chunk

^۲ - Pellegrino

محکم برای اثبات روایی آزمون فراهم می‌آورد. این رویکرد با دیدگاه‌های ساموئل مسیک^۱ (۱۹۸۹، ۱۹۹۵) در مورد روایی سنجش، هماهنگ است. مسیک (۱۹۸۹) در مورد روایی بیان می‌دارد:

تقریباً هر نوع اطلاعات در مورد آزمون می‌تواند به روایی سازه آن کمک نماید. اما این کمک زمانی موثرتر خواهد بود که میزان برازش اطلاعات با منطق نظری زیربنایی تفسیر نمره، به روشنی مورد ارزیابی قرار گیرد. احتمالاً روشن‌گرترین اطلاعات، آنهایی هستند که فرایندهای زیربنایی پاسخهای آزمون را به طور مستقیم مورد کند و کاو و مدل‌پردازی قرار می‌دهند، رویکردی که با پیشرفت‌های مداوم در روان‌شناسی شناختی، روز به روز آسان‌تر، عملی‌تر، و نیرومندتر می‌شود (ص ۱۷).

بنابراین، با در نظر گرفتن مطالب فوق اهداف ۳ گانه زیر را برای پژوهش حاضر می‌توان در نظر گرفت. ۱- فراهم کردن شواهدی در مورد روایی سازه آزمون ریاضیات پایه هشتم تیمز ۲۰۰۷ بر اساس مدل‌های سنجش تشخیصی شناختی. پژوهش‌هایی که در زمینه سوال‌های تیمز صورت گرفته است منجر به شناسایی مجموعه‌ای از صفات و مهارت‌ها شده است که به نظر می‌آید عملکرد دانش‌آموزان برخی دیگر از کشورها را به اندازه کافی توصیف و تبیین می‌کند (تاتسوکا، کورتر، و تاتسوکا^۲، ۲۰۰۴). هدف پژوهش حاضر در این مرحله، رواسازی ماهیت آزمون ریاضیات پایه هشتم تیمز ۲۰۰۷ برای دانش‌آموزان ایرانی است؛ ۲- نشان دادن چگونگی کاربرد یکی از پرکاربردترین مدل‌های تشخیصی شناختی، یعنی نسخه غیرجبرانی و کاهش یافته مدل یکپارچه با پارامترپردازی مجدد^۳ (Reduced NC-RUM؛ هارتز^۴، ۲۰۰۲) برای بررسی صفات، فرایندها، و دانش مورد نیاز برای پاسخگویی به سوال‌های ریاضیات تیمز و شناسایی نقاط ضعف و قوت دانش‌آموزان ایرانی در این صفات با استفاده از داده‌های تیمز ۲۰۰۷ است؛ و ۳- همچنین می‌خواهیم بررسی کنیم که چه تفاوت‌هایی بین عملکرد دانش‌آموزان دختر و پسر ایرانی در صفات و مهارت‌های شناختی زیربنایی ریاضیات پایه هشتم تیمز ۲۰۰۷ وجود دارد.

¹ - Messick

² - Tatsuoka, Corter & Tatsuoka

³ - Reduced Non-compensatory Reparametrized Unified Model

⁴ - Hartz

سؤالات پژوهش

با عنایت به اینکه پیشرفت تحصیلی و به عبارت دقیق‌تر، عملکرد در آزمون‌های تیمز تحت تاثیر عوامل مختلف فردی، خانوادگی، نظام آموزشی، اجتماعی و حتی سیاسی قرار می‌گیرد، و با توجه به وجود تفاوت‌های زیاد بین ایران و سایر کشورها از لحاظ عوامل فوق و بویژه از لحاظ روش‌ها و محتوای نظام آموزشی، بنابراین امکان پیش‌بینی عملکرد دانش‌آموزان ایرانی در مهارت‌های مورد سنجش بر اساس پژوهش‌های صورت گرفته در سایر کشورها وجود ندارد. از طرف دیگر از آنجائیکه پژوهش حاضر نخستین پژوهشی است که به دنبال کاربرد مدل‌های CDM بر روی داده‌های دانش‌آموزان ایرانی در مطالعات TIMSS است، لذا امکان بیان فرضیه‌های دقیق و مشخص وجود ندارد. از اینرو در پژوهش حاضر به جای طرح فرضیه و آزمون آن، به دنبال پاسخگویی به سؤال‌های زیر هستیم.

۱- آیا مجموعه مهارت‌ها و صفات شناختی ارائه شده توسط کورتر و تاتسوکا (۲۰۰۲)، نقل از تاتسوکا، کورتر و تاتسوکا، (۲۰۰۴) که در پژوهش‌های مختلف نیز بکار رفته است قابل تعمیم به عملکرد دانش‌آموزان پایه هشتم ایران است؟

۲- عملکرد دانش‌آموزان ایران در این مهارت‌ها یا صفات شناختی چگونه است؟ به عبارت دیگر، دانش‌آموزان ایرانی در چه مهارت‌هایی به تسلط رسیده‌اند و در چه مهارت‌هایی به تسلط نرسیده‌اند؟

۳- آیا بین عملکرد دانش‌آموزان دختر و پسر پایه هشتم ایران در مهارت‌ها یا صفات شناختی زیربنایی ریاضیات تیمز ۲۰۰۷ تفاوت‌هایی وجود دارد؟