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[۱]

اصغر مینایی*

علی دلاور**

محمدرصا فلسفی‌نژاد***

علیرضا کیامنش****

یحیی مهاجر*****

چکیده

زمینه: هدف سنجشهایی مانند تیمز آموزش و بهبود عملکرد دانش‌آموز است. در راستای دستیابی به این هدف متخصصان حوزه سنجش آموزشی، رویکرد سنجش تشخیصی شناختی(CDA) را پرورش داده‌اند. در پژوهش حاضر تلاش شد تا یکی از مدلهای پرکاربرد در حوزه سنجش شناختی را در مورد داده‌های دانش‌آموزان پایه هشتم ایران در آزمون ریاضی تیمز ۲۰۰۷ بکار گرفته شود و نقاط ضعف و قوت دانش‌آموزان ایران را در مهارتها و صفات شناختی زیربنایی آزمون ریاضی (مدل شناختی کورتر و تاتسوکا، ۲۰۰۲) شناسایی نماید و عملکرد دختران و پسران را در این صفات شناختی مورد مقایسه قرار دهد. **یافته‌ها:** نتایج حاصل از پژوهش نشان داد که مدل شناختی کورتر و تاتسوکا با داده‌های دانش‌آموزان پایه هشتم ایران در آزمون ریاضی تیمز ۲۰۰۷ از برازش مطلوبی برخوردار است. یافته‌های مربوط به وضعیت تسلطی نشان می‌دهد که دانش‌آموزان ایرانی، بیشترین ضعف را در صفات P1(فرمولبندی معادلات)، P2(هندسه)، P4(به کارگیری قواعد جبر)، و P5(تفکر قیاسی) که صفات مربوط به پردازش

---

۱. این مقاله از رساله دکتری نویسنده اول استخراج شده است

* استادیار دانشگاه علامه طباطبائی(نویسنده مسئول) (sghar.minaei@gmail.com).

** استاد دانشگاه علامه طباطبائی.

*** دانشیار دانشگاه علامه طباطبائی.

**** استاد دانشگاه آزاد اسلامی واحد علوم تحقیقات

***** استادیار دانشگاه علامه طباطبائی

تاریخ دریافت: ۱۳۹۲/۷/۱۶ **تاریخ پذیرش: ۱۳۹۲/۱۲/۵**

فصلنامهٔ اندازه‌گیری تربیتی، سال چهارم، شمارهٔ شانزدهم، تابستان ۱۳۹۳

شناختی هستند دارند. همچنین در دو صفت S5(بررسی صحت گزینه‌ها) و S6(الگوها و روابط) که به مهارتهای تفکر استقرایی مربوط می‌شوند نیز از عملکرد ضعیفی برخوردارند. در واقع می‌توان گفت که صفت S6 دومین صفت دشوار بعد از صفت P2 در بین دانش‌آموزان پایه هشتم ایران است. یافته‌های مربوط به عملکرد دختران و پسران نشان داد که تفاوتهای جنسیتی در این صفات و مهارتهای شناختی بسیار کم است. دامنه شاخص d کوهن از ۰/۳۰- تا ۰/۱۵+ با میانگین ۰/۰۳ به دست آمد. **بحث و نتیجه‌گیری:** بطور کلی پژوهش حاضر نشان می‌دهد که دانش‌آموزان ایران در اکثر صفات که برخی از آنها به مهندسی و علوم فیزیکی مربوط می‌شوند از سطح تسلطی ضعیف برخوردارند. از اینرو به نظر می‌رسد که سیاستگذاران آموزشی بایستی تغییرات اساسی در آموزش هندسه به عمل آورند، زیرا هندسه، آموزش مهارتهای تفکر ریاضی مهم را که برای مهندسی و علوم فیزیکی لازم هستند ممکن می‌سازد. در ارتباط با تفاوت‌های جنسیتی در ریاضیات نیز با توجه به یافته‌های پژوهش حاضر و سایر پژوهش‌های صورت گرفته می‌توان این نتیجه را گرفت که جنسیت یکی از بیشمار متغیری است که عملکرد در ریاضیات را، هر چند بسیار اندک، تحت تاثیر قرار می‌دهد.

کلیدواژه‌ها: *سنجش تشخیصی شناختی، تیمز، ریاضیات، مدل.*

مقدمه

هدف سنجش‌هایی مانند مطالعه روندها در ریاضیات و علوم[1] (TIMSS) "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 آموزشگاهی" است (مولیس[2] و همکاران، ۲۰۰۴، ص ۱۳).

اگر چه هدف اصلی از اجرای مطالعات TIMSS رتبه‌بندی کشورها نیست (کیامنش، ۱۳۸۸)، لکن 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[3]، ۱۹۹۸). با این حال، این گزارش‌ها و رتبه‌بندیها که بر اساس نمره کل صورت می‌گیرند اطلاعات کافی در زمینه نقاط ضعف دانش‌آموزان فراهم نمی‌آورند.

در راستای کمک به استفاده از سنجش‌ها برای بهبود و ارتقاء یادگیری، پژوهشگران و

1. Trends in Mathematics and Science Study
2. Mullis
3. Martin & Kelly

متخصصان حوزه سنجشِ آموزشی با اتکاء به پژوهش‌های صورت گرفته در حوزه روان‌شناسی یادگیری و با بهره‌گیری از مدل‌های آماری، رویکرد سنجش تشخیصی شناختی[1] (CDA) را به منظور انجام استنباط‌هایی راجع به ساختارها و فرایندهایی که زیربنای عملکرد در حیطه مورد نظر را تشکیل می‌دهند پرورش دادند. در سنجش تشخیصی شناختی، علاوه بر برآورد توانایی فرد در سازه یا سازه‌های مورد اندازه‌گیری، نیمرخی برای هر فرد فراهم می‌آید که تسلط یا عدم تسلط وی را در مجموعه‌ای از مهارت‌های اساسی یا صفات شناختی خُرد[2] از پیش تعیین‌شده 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 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.

کشور ایران به منظور ارزیابی و بهبود نظام آموزشی خود از سال ۱۹۹۱ میلادی رسما همکاری خود را با *انجمن بین‌المللی ارزشیابی پیشرفت تحصیلی*[3](IEA) آغاز کرد و تا کنون در ۶ مطالعه در فاصله سال‌های ۱۹۹۵، ۱۹۹۹، ۲۰۰۳، ۲۰۰۷، تیمز پیشرفته ۲۰۰۸، و ۲۰۱۱ شرکت کرده است. 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 و ۲۰۰۷ به ترتیب ۳۷، ۳۳، ۳۴، و ۳۴ بوده است. این در حالی است که تعداد کشورهای شرکت کننده در مطالعه، به ترتیب ۴۱، ۳۸، ۴۶، و ۵۹ بوده است.

1. Cognitive Diagnostic Assessment
2. fine-grained
3. International Association for the Evaluation of Educational Achievement

بنابراین، با توجه به 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؛ ۲- نشان دادن چگونکی کاربرد یکی از پرکاربردترین مدل‌های سنجش تشخیصی شناختی، یعنی *نسخه غیرجبرانی و کاهش‌یافته مدل یکپارچه با پارامترپردازی مجدد*[1] (*Reduced NC-RUM*) 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- مقایسه عملکرد دو گروه دانش‌آموزان دختر و پسر ایران در صفات و مهارتهای شناختی زیربنایی آزمون ریاضیات پایه هشتم تیمز ۲۰۰۷.

مدل یکپارچه با پارامترپردازی مجدد(RUM)

مدل‌های تشخیصی شناختی(CDM) زیادی ارائه شده است. برای مثال، فو و لی[2] (۲۰۰۷) تعداد ۶۲ مدل را که می‌توان برای اهداف تشخیصی شناختی استفاده نمود مورد بررسی قرار داده‌اند. یکی از مدل‌های پرکاربرد، مدل یکپارچه با پارامترسازی مجدد[3](RUM) است. این مدل که مدل التقاطی[4](FM) نیز نامیده می‌شود توسط هارتز(۲۰۰۲) و با الهام از ایده‌های مدل یکپارچه[5](UM؛ دی‌بلو، استوت، و روسوس[6]، ۱۹۹۵) ارائه گردید. تابع ریاضی مدل UM به گونه زیر است:

$$P(x_{ij}=1|\underline{\alpha}_j,\eta_j)=d_i\prod_{k=1}^{k}\pi_{ik}^{\alpha_{ik}\cdot q_{ik}}r_{ik}^{(1-\alpha_{jk})\cdot q_{ik}}P_{ci}(\eta_j)+(1-d_i)P_{bi}(\eta_j)$$

در فرمول فوق، π_{ik} بیانگر احتمال کاربرد صحیح مهارت k در سوال i توسط آزمودنی است مشروط بر آنکه آزمودنی در مهارت k به تسلط رسیده باشد؛ r_{ik} بیانگر احتمال کاربرد صحیح مهارت k در سوال i توسط آزمودنی است مشروط بر آن که آزمودنی در مهارت k به تسلط نرسیده باشد؛ P_{bi} نشان‌دهنده احتمال انتخاب راهبرد مبتنی بر ماتریس Q نسبت به سایر راهبردها است؛ P_{ci} که بر اساس تابع مدل راش تعریف می‌شود، احتمال پاسخ درست

1. Reduced Non-compensatory Reparameterized Unified Model
2. Fu & Li
3. Reparameterized Unified Model
4. fusion Model
5. unified Model
6. DiBello, Stout & Roussos

به سوال را نشان می‌دهد؛ $\underline{\alpha}_j$ برداری است مرکب از صفر و ۱ که وضعیت تسلطی آزمودنی را در مهارت‌ها یا صفات مورد اندازه‌گیری نشان می‌دهد؛ پارامتر η_j که پارامتر پس‌مانده[1]، یا پارامتر مکمل[2] نیز نامیده می‌شود و با استفاده از تابع مدل ساده راش به دست می‌آید، برآورد تک‌بعدی سطوح تسلط مهارتی آزمودنی در مهارت‌های غیر Q است(نمونه‌ای از ماتریس Q در صفحه ارائه شده است)؛ و سرانجام پارامتر c_i که پارامتر کامل‌بودن نامیده می‌شود، به کامل بودن دامنه شمول ماتریس Q اشاره دارد. پارامتر d_i نیز بیانگر راهبردهای مختلفی است که دانش‌آموزان برای حل سوال به کار می‌برند.

اگرچه مدل UM یک مدل انعطاف‌پذیر، تفسیرپذیر و به لحاظ مفهومی جذاب و گیرا است، اما پارامترهای سوالها ناهمانند[3] هستند و از اینرو به لحاظ آماری قابل برآورد نیستند. وجود پارامترهای ناهمانند در مدل UM باعث شد تا هارتز(۲۰۰۲) آن را مجددا به گونه‌ای پارامترپردازی نماید که در عین حال که همانند باشد پارامترهای آن نیز تفسیرپذیر باشند. برای کاهش پیچیدگی فضای پارامتر و افزایش توانایی برآورد پارامترها، مقدار پارامتر d_i، برابر با ۱ قرار داده شد و بدین ترتیب از مدل حذف شد. مدل کاهش‌یافته‌ای که هارتز تولید کرد و از آن تحت عنوان *مدل یکپارچه با پارامترپردازی مجدد*(RUM) نام برده می‌شود در مقایسه با مدل UM که برای هر سوال $2k_i + 3$ پارامتر برآورد می‌کند، برای هر سوال $2 + k_i$ پارامتر برآورد می‌نماید. معادله زیر، IRF مدل RUM را نشان می‌دهد که متکی بر همان پارامترهای توانایی α_j و η_j آزمودنی است که در مدل UM استفاده می‌شود. تعریف عبارت $P_{ci}(\eta_j)$ نیز همانند مدل UM است.

$$P(X_{ij} = 1|\underline{\alpha}_j, \eta_j) = \pi_i^* \prod_{k=1}^{k} r_{ik}^{*(1-\alpha_{jk})\times q_{ik}} P_{ci}(\eta_j)$$

در معادله فوق، π_i^*، که در دامنه ۰ تا ۱ مدل‌پردازی می‌شود، احتمال کاربرد درست *تمام* مهارت‌های مورد نیاز سوال، مشروط بر آنکه فرد در *تمام* آن مهارتها به تسلط رسیده باشد را نشان می‌دهد. در واقع در مدل RUM احتمال توام اجرای موفقیت‌آمیز تمام مهارت‌های مورد نیاز یک سوال، به صورت حاصل‌ضرب احتمال اجرای موفقیت‌آمیز تک‌تک مهارت‌های انفرادی، مدل‌پردازی می‌شود که بیانگر استقلال شرطی اجرای

1. residual
2. Supplement
3. Not-Identified

مهارت‌ها است. در واقع، $\pi_i^* = \prod_{k=1}^{K} \pi_{ik}^{q_{ik}}$، است.

پارامتر r_{ik}^* نیز که مقدار آن بین صفر و ۱ است، به صورت پارامتر تشخیص سوال i برای صفت k یا جریمه عدم تسلط در صفت k تفسیر می‌شود. شکل ریاضی این پارامتر به گونه زیر است:

$$r_{ik}^* = \frac{P(Y_{ijk}|\alpha_{jk}=0)}{P(Y_{ijk}|\alpha_{jk}=1)} = \frac{r_{ik}}{\pi_{ik}}$$

پارامتر r_{ik}^* نشانگر قابلیت تشخیص سوال i برای صفت k است. هر چقدر دادن پاسخ درست به سوال وابستگی بیشتری به تسلط بر صفت k داشته باشد، مقدار پارامتر r_{ik}^* پایین‌تر خواهد بود. هر چقدر مقدار این پارامتر به صفر نزدیکتر باشد، سوال i از قدرت بیشتری برای تمایزگذاری بین افراد مسلط و غیرمسلط در آن صفت برخوردار خواهد بود. بنابراین برای ارزشیابی قابلیت تشخیص آزمون، پارامترهای r_{ik}^* از اهمیت خاصی برخوردارند.

پارامتر c_i نیز که مقدار آن بین صفر تا ۳ است، شاخص کامل‌بودن ماتریس Q برای سوال i است. هر چقدر مقدار این پارامتر از صفر بیشتر باشد، بیانگر کامل‌تر بودن ماتریس Q است که به معنای این است که دادن پاسخ درست به سوال به مهارت‌های دیگری که در ماتریس Q وارد نشده‌اند وابستگی کمتری دارد.

در مدل RUM پارامتر η_j، که با مدل راش برآورد می‌شود، بیش از اندازه بر IRF[1] تاثیر می‌گذارد(جانگ، ۲۰۰۹) و علاوه بر این تعداد مهارت‌های ماتریس Q تقریبا همیشه به اندازه کافی زیاد است که بتوانند بخش اعظم واریانس موجود در پاسخ‌های سوال‌ها را تبیین نمایند(روسوس، تمپلین و هنسون، ۲۰۰۷). بنابراین پارامتر η_j بندرت مورد استفاده قرار می‌گیرد. در واقع، مقدار پارامتر c_i برابر با ۱۰ قرار داده می‌شود که باعث می‌شود پارامتر η_j از مدل حذف شود و مدل ساده‌تری با نام *Reduced-RUM*[2] بوجود آید. این مدل در واقع نسخه کاهش‌یافته و ساده‌شده مدل RUM است که در آن پارامتر مربوط به توانایی پس‌مانده یا مکمل حذف شده است. معنای تلویحی حذف عبارت $P_{ci}(\eta_j)$ از مدل بدان معنا است که ماتریس Q یک بازنمایی کامل از مهارت‌های مورد نیاز آزمون است یا اینکه مهارت‌های غیر Q مهارت‌های مهمی نیستند(وانگ[3]، ۲۰۰۹). تعریف و تفسیر پارامترهای

1. Item Response Function
2. Roussos , Templin & Henson
3. Wang

مدل *Reduced-RUM* دقیقا مثل مدل RUM است. احتمال پاسخ درست به سوال در این مدل به شکل زیر مدل‌پردازی می‌شود:

$$P(X_{ij}=1|\underline{\alpha}_j)=\pi_i^*\prod_{k=1}^{k} r_{ik}^{*(1-\alpha_{jk})\times q_{ik}}$$

مدل *Reduced-RUM* یک مدل غیرجبرانی است. به همین دلیل در مقالات و ادبیات پژوهش، این مدل را به صورت *Reduced NC-RUM* نشان می‌دهند. مدل‌های غیرجبرانی، که مدل‌های عطفی[1] نیز نامیده می‌شوند از شهرت و کاربرد بیشتری برخوداردند(روسوس، تمپلین و هنسون، ۲۰۰۷). مدل‌های غیرجبرانی بر این مفروضه استوارند که برای اینکه پاسخ‌دهنده بتواند به سوالی پاسخ درست بدهد باید در تمام صفات لازمه آن سوال به حد تسلط رسیده باشد.

روش پژوهش

از آنجا که در این پژوهش، برای شناسایی و تعیین قوتها و ضعفهای دانش‌آموزان ایرانی در ریاضیات پایه هشتم، از داده‌های درس ریاضی تیمز که در سال ۲۰۰۷ اجرا شده است استفاده گردید، از اینرو می‌توان گفت که پژوهش حاضر جزو تحقیقات تحلیل ثانویه[2] محسوب می‌گردد. البته با توجه به اینکه علاوه بر شناسایی و تعیین قوتها و ضعفهای ریاضی دانش‌آموزان، عملکرد دو گروه دانش‌آموزان دختر و پسر نیز مورد مقایسه قرار گرفت، بنابراین می‌توان آن را پژوهش توصیفی- مقایسه‌ای نیز تلقی کرد. با در نظر گرفتن مطالب فوق می‌توان نتیجه گرفت که پژوهش حاضر، جزو تحقیقات توصیفی- مقایسه‌ای از نوع تحلیل ثانویه به حساب می‌آید.

جامعه آماری، گروه نمونه و روش نمونه‌گیری

جامعه آماری و گروه نمونه پژوهش حاضر همان جامعه آماری و گروه نمونه پایه هشتم(جمعیت ۲) در آزمون ریاضی تیمز ۲۰۰۷ است که در سال تحصیلی ۱۳۸۶-۱۳۸۵ به اجرا درآمد. در سال تحصیلی مورد نظر تعداد ۱/۴۷۵/۳۶۸ دانش‌آموز سوم راهنمایی(پایه هشتم) در ۲۹/۹۵۶ کلاس، مشغول به تحصیل بودند (السن، مارتین و مولیس[3]، ۲۰۰۸).

1. conjunctive
2. secondary analysis
3. Olson, Martin & Mullis

نمونه پژوهش حاضر کلیه دانش‌آموزان ایرانی سال سوم راهنمایی شرکت‌کننده در تیمز ۲۰۰۷ بودند که تعداد آنها برابر با ۳۹۸۱ دانش‌آموز از ۲۰۸ مدرسه بود.

ابزار تحقیق

ابزار مورد استفاده در این پژوهش، آزمون ریاضیات پایه هشتم(سوم راهنمایی) تیمز ۲۰۰۷ است. این آزمون دارای ۲۱۵ سوال است که در ۱۴ دفترچه که از لحاظ حیطه‌های محتوایی و شناختی، متوازن محسوب می‌شوند قرار گرفته‌اند. هر کدام از دفترچه‌ها بین ۳۶-۳۲ سوال دارند.

آزمون ریاضی پایه هشتم تیمز ۲۰۰۷، چهار حیطه محتوایی و سه حیطه شناختی را پوشش می‌دهد. حیطه‌های محتوایی شامل اعداد، جبر، هندسه، و داده‌ها و احتمال است که هر کدام به ترتیب دارای ۶۳، ۶۴، ۴۷، و ۴۱ سوال است. حیطه‌های شناختی نیز شامل دانستن، کاربست و استدلال است که به ترتیب، ۳۵٪، ۴۰٪ و ۲۵٪ سوال‌ها را شامل می‌شوند.

از بین ۲۱۵ سوال، تعداد ۱۱۷ سوال، دو یا چند گزینه‌ای است که به صورت دو ارزشی (۰ یا ۱) نمره‌گذاری می‌شوند. بقیه سوال‌ها نیز که تعداد آنها برابر با ۹۸ سوال است، سوال‌های باز پاسخ هستند که به صورت چندارزشی از ۰ تا حداکثر ۳ نمره‌گذاری می‌شوند.

در پژوهش حاضر به منظور ایجاد سهولت در تحلیل داده‌ها و به پیروی از لی، پارک و تایلان(۲۰۱۱) سوال‌های چندارزشی از طریق کدگذاری مجدد به سوال‌های دو ارزشی تبدیل گردید. برای این منظور، اگر فردی در یک سوال چند ارزشی بالاترین نمره را به دست آورده بود، نمره ۱ و چنانچه نمره کامل نگرفته بود نمره صفر تعلق گرفت. در مورد سوال‌هایی که پاسخ‌دهنده آنها را خالی گذاشته بود، که در چارچوب تیمز آنها را سوال‌های *بدون پاسخ*[1] می‌نامند و همچنین سوال‌هایی که به دلیل کمبود زمان فرصت پاسخدهی به آنها نبوده است، که در چارچوب تیمز آنها را سوال‌های *جامانده*[2] می‌نامند، از یک روش دو مرحله‌ای (لادلو، و اُلیری[3]، ۱۹۹۹) استفاده گردید.

1. Omitted
2. Not-reached
3. Ludlow & Oleary

در این روش، که در مطالعات تیمز نیز از آن استفاده می‌شود، در مرحله اول که هدف آن برآورد پارامترهای سوال است، سوال‌های بدون پاسخ، به صورت سوال‌های نادرست(یعنی نمره صفر) و سوال‌های جامانده به صورت سوال‌های اجرا نشده، کدگذاری می‌شوند، و پارامترهای سوال برآورد می‌گردد. در مرحله دوم، پارامترهای سوال که در مرحله اول به‌دست آمده‌اند، به عنوان پارامترهای ثابت و معلوم فرض می‌شوند و پارامتر توانایی افراد برآورد می‌گردد. در این مرحله که هدف آن برآورد توانایی افراد است، سوالهای بدون پاسخ و جامانده، به عنوان سوال‌های نادرست در نظر گرفته می‌شوند و نمره صفر به آنها تعلق می‌گیرد.

تحلیل داده‌ها

با توجه به اینکه در پژوهش حاضر مدل *Reduced NC-RUM* در مورد داده‌های حاصل از آزمون ریاضیات پایه هشتم تیمز ۲۰۰۷ که با مدل‌های تک بُعدی IRT پرورش یافته است به کار برده شد، بنابراین می‌توان گفت که پژوهش حاضر بر رویکرد بازبرازش[1] یا مهندسی معکوس استوار است. در آن دسته از مطالعات CDA که بر رویکرد بازبرازش استوارند، اجرای CDA در چهار گام صورت می‌گیرد. چگونگی اجرای این مراحل ۴ گانه در پژوهش حاضر در زیر توضیح داده می‌شود.

گام اول: تعریف مهارت‌ها/ صفات شناختی

سنجش تشخیصی شناختی (CDA) معمولا با شناسایی مجموعه‌ای از صفات مورد اندازه‌گیری آزمون و فرمول‌بندی روابط میان صفات و همچنین روابط بین صفات و سوال‌های آزمون آغاز می‌شود. برای تعریف صفات می‌توان از منابع مختلف مانند جدول مشخصات آزمون، نظریه‌های حیطه محتوایی، تحلیل محتوای سوال، تحلیل پروتکل تفکر با صدای بلند و نتایج سایر پژوهش‌های مرتبط استفاده نمود.

در پژوهش حاضر نیز همانند دیگر پژوهش‌ها بایرن‌بام، تاتسوکا و زین، ۲۰۰۵؛ بایرن‌بام، تاتسوکا و یامادا، ۲۰۰۴؛ تاتسوکا، کورتر، و تاتسوکا، ۲۰۰۴؛ توکر، ۲۰۱۰؛ چن، ۲۰۰۶؛ چن، گورین، تامپسون و تاتسوکا[2]، ۲۰۰۶؛ دوگان و تاتسوکا، ۲۰۰۸) از مجموعه صفات

1. retrofitting
2. Chen, Gorin, Thompson & Tatsuoka

شناسایی شده توسط کورتر و تاتسوکا(۲۰۰۲، نقل از تاتسوکا، کورتر و تاتسوکا، ۲۰۰۴) استفاده گردید.

کورتر و تاتسوکا(۲۰۰۲) با انجام یک پژوهش گسترده و تحلیل سوال‌های ریاضی تیمز ۱۹۹۹و پاسخهای دانش‌آموزان، ۲۷ صفت را شناسایی و تعیین کردند. آنها این ۲۷ صفت را به ۳ طبقه صفات محتوایی(۵ صفت)، صفات فرایندی(۹ صفت) و صفات مهارتی یا نوع سوال(۹ صفت) تقسیم کردند. *صفات محتوایی* به خصوصیات و مفاهیم پایه در اعداد صحیح، کسرها و اعشارها، جبر مقدماتی، هندسه دو بعدی، داده‌ها و آمار مقدماتی اشاره دارد. *صفات فرایندی*، شامل صفاتی مانند کاربردهای قضاوتی دانش در حساب و هندسه، به کارگیری قواعد جبر، استدلال منطقی، جستجوی مساله، تولید، تجسم و خواندن اشکال و نمودارها، مدیریت داده‌ها و روش‌ها است. *صفات مهارتی*، صفاتی مانند کاربرد خواص و روابط عددی(درک اعداد)، تقریب/تخمین زدن، حل سوال‌های بازپاسخ، را شامل می‌شود. فهرست کامل این صفات ۲۷ گانه که از تاتسوکا، کورتر و تاتسوکا(۲۰۰۴) اقتباس شده است در جدول ۱ ارائه شده است.

جدول ۱: صفات محتوایی، فرایندی و مهارتی زیربنای عملکرد در آزمون ریاضیات تیمز ۲۰۰۷

صفات محتوایی	
C1	مفاهیم پایه، خواص و عملیات مربوط به اعداد حسابی و صحیح
C2	مفاهیم پایه، خواص و عملیات مربوط به اعداد کسری و اعشاری
C3	مفاهیم پایه، خواص و عملیات مربوط به جبر مقدماتی
C4	مفاهیم پایه، خواص و عملیات مربوط به هندسه دو بُعدی
C5	داده‌ها، احتمالات و آمار مقدماتی
C6	استفاده از ابزارها برای اندازه‌گیری (یا تخمین) طول، زمان، زاویه و دما
صفات فرایندی	
P1	ترجمه/فرمول‌بندی معادلات و عبارات جبری برای حل یک مساله
P2	کاربردهای محاسباتی دانش موضوعی حساب و هندسه
P3	کاربردهای قضاوتی دانش حساب و هندسه
P4	بکارگیری قواعد جبر
P5	استدلال منطقی– شامل استدلال در موقعیت‌ها یا شرایط خاص، مهارتهای تفکر قیاسی، اگر–آنگاه، شرط لازم وکافی، مهارتهای تعمیم‌دهی
P6	جستجوی مسائل؛ تفکر تحلیلی، بازسازی مساله و تفکر استقرایی
P7	تولید، تجسم و خواندن اشکال و نمودارها
P8	به کارگیری و ارزشیابی صحت ریاضی
P9	مدیریت داده‌ها و روشها
P10	خواندن کمّی و منطقی
صفات مهارتی	

S1	تبدیل واحد
S2	کاربرد خواص و روابط اعداد؛ درک اعداد/ خط (محور) اعداد
S3	استفاده از اشکال، جداول و نمودارها
S4	تقریب/ تخمین‌زدن
S5	ارزیابی/ بررسی صحت گزینه‌ها
S6	الگوها و روابط (توانایی استفاده از مهارت‌های تفکر استقرایی)
S7	استفاده از استدلال نسبتی
S8	حل مسائل جدید یا ناآشنا
S9	مقایسه دو یا چند شیء
S10	سوال باز پاسخ که یک جواب/پاسخ ندارد
S11	استفاده از کلمات یا واژه‌ها برای درک و پاسخ به سوالات (مسائل کلامی)

گام دوم: ساخت ماتریس *Q*

در پژوهش حاضر برای تشکیل ماتریس *Q*، پس از ترجمه پروتکل یا کتابچه راهنمای کدگذاری سوال‌ها، نسخه‌ای از سوال‌های ریاضیات پایه هشتم تیمز ۲۰۰۷ همراه با فهرست صفات و پروتکل کدگذاری در اختیار ۲ معلم ریاضی پایه هشتم با مدرک تحصیلی فوق لیسانس که سابقه ۱۰-۵ سال آموزش ریاضی داشتند و در طراحی و تصحیح سوال‌های ریاضی تیمز ۲۰۰۷ نیز شرکت داشتند قرار داده شد و از آنها خواسته شد تا به صورت جداگانه و مستقل از یکدیگر، در یک ماتریس دو بعدی که در ستونهای آن مهارت‌ها و در سطرهای آن سوال‌ها قرار دارند با قرار دادن ۰ یا ۱ مشخص کنند که هر سوال کدام صفت یا صفات را اندازه‌گیری می‌کند.

بدین ترتیب ۲ ماتریس اولیه *Q* به دست آمد. سپس در یک جلسه مشترک که با حضور ۲ متخصص مذکور و پژوهشگر تشکیل گردید، ماتریس‌های *Q* مورد بحث و بررسی قرار گرفت و پس از رفع ابهامات، اشکالات و اختلافات، که عمدتا به صفات فرایندی مربوط می‌شد، و دست‌یابی به اجماع کلی، ماتریس *Q* اولیه شکل گرفت. این ماتریس اولیه همراه با داده‌های دانش‌آموزان پایه هشتم ایران در ریاضیات تیمز ۲۰۰۷ وارد تحلیل گردید و از طریق برنامه *Arpeggio* (دی‌بلو و استوت[1]، ۲۰۰۸) مورد تحلیل قرار گرفت. نتایج حاصل از اولین اجرای برنامه *Arpeggio* نشان داد که شاخص r_{ik}^{*} که بیانگر قدرت تشخیص سوال برای صفت مورد اندازه‌گیری است برای برخی ازسوال‌ها از مقدار ۰/۹۰ بالاتر(کیم، ۲۰۱۱) است. در گام بعد، درایه‌های مربوط به این گونه سوال‌ها از ماتریس *Q* حذف گردید. یعنی

1. DiBello & Stout

در ماتریس Q عدد ۱ مربوط به این گونه سوال‌ها به ۰ تبدیل گردید و ماتریس Q جدیدی شکل گرفت و برنامه *Arpeggio* مجددا اجرا شد. این فرایند یک بار دیگر تکرار گردید. یعنی بعد از اجرای مجدد برنامه *Arpeggio* *شاخص* r_{ik}^{*} سوال‌ها مورد بررسی قرار گرفت و درایه‌های مربوط به سوال‌هایی که مقدار r_{ik}^{*} آنها بالاتر از ۰/۹۰ بود از ۱ به ۰ تبدیل پیدا کرد و برنامه *Arpeggio* برای بار سوم اجرا گردید.

گام سوم: برآورد مدل

هدف اصلی و اولیه تحلیل داده‌ها با مدل‌های CDA برآورد نیمرخ تسلط صفتی برای فراگیران انفرادی بر اساس عملکرد آنها در آزمون است. برای رسیدن به این هدف، انتخاب یک مدل برای مدرج‌سازی آزمون‌ها بخش جدایی‌ناپذیر سنجش تشخیصی شناختی محسوب می‌شود.

در پژوهش حاضر از *نسخه غیر جبرانی و کاهش‌یافته مدل یکپارچه با پارامترسازی مجدد (Reduced NC-RUM)* که از طریق برنامه *Arpeggio* (دی‌بلو، و استوت، ۲۰۰۸) قابل اجرا است استفاده گردید.

گام چهارم: گزارش نمره/ ارائه بازخورد تشخیص

از آنجا که پژوهش حاضر بر اساس رویکرد بازبرازش صورت گرفت که آزمون مورد استفاده در سال ۲۰۰۷ برگزار شده است و در نتیجه ارائه بازخورد به دانش‌آموزان کمکی به بهبود وضعیت آنها نمی‌کند، لذا ضرورتی برای ارائه گزارش تشخیصی وجود ندارد.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

یافته‌ها

جهت تحلیل داده‌ها و پاسخگویی به سوال‌های پژوهش از روش‌های آماری تحلیل رگرسیون چندگانه، آزمون t برای گروههای مستقل، آزمون z برای مقایسه نسبت‌ها و روش‌شناسی مدل *Reduced NC-RUM* استفاده گردید. برای انجام تحلیل‌ها از نسخه ۳/۱ نرم‌افزار *Arprggio* (دی‌بلو، و استوت، ۲۰۱۰)، SPSS15، و برای رسم نمودارها از Excel استفاده گردید.

در ادامه، نتایج مربوط به برازش مدل *Reduced NC-RUM* با داده‌ها ارائه می‌شود و سپس یافته‌های مربوط به سوال‌های پژوهش ارائه می‌گردد.

الف- ارزشیابی برازش مدل Reduced-NC-RUM

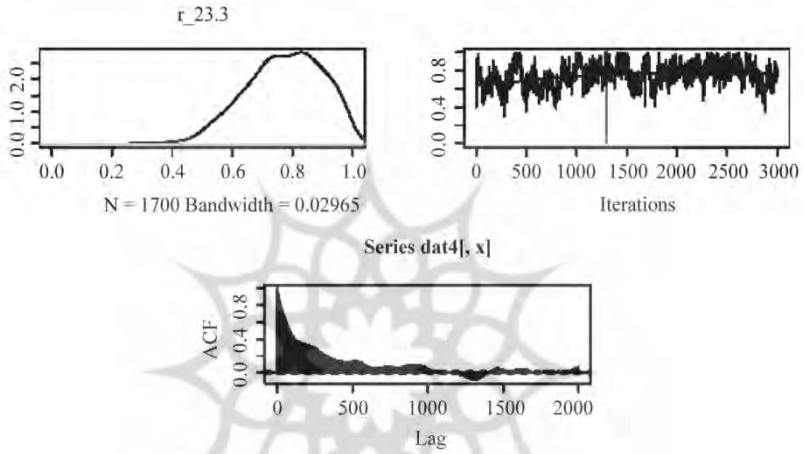
نرم‌افزار *Arpeggio* (دی‌بلو و استوت، ۲۰۰۸) برای برآورد مدل مورد نظر از الگوریتم MCMC در چارچوب بیزی استفاده می‌کند. در رویکرد MCMC برای برآورد باثبات پارامترها از زنجیره‌های مارکوفی طولانی(تکرارهای زیاد) استفاده می‌شود. به خاطر همگرایی نمونه‌گیری گیبس و عدم وابستگی به مقادیر اولیه معمولا تکرارهای اول نمونه‌ها کنار گذاشته می‌شوند. یعنی برای تولید یک نمونه Mتایی از هر پارامتر، $M+M_0$ مقدار تولید و M_0تای اول آن که تکرارهای سوخته[1] یا مرحله داغیدن نام دارند کنار گذاشته می‌شوند. سپس مقادیر مربوط به تکرارهای باقی‌مانده به عنوان توزیع پسین پارامتر، در نظر گرفته می‌شود و مقدار پارامتر و انحراف استاندارد آن بر اساس این توزیعهای پسین برآورد می‌گردد. روسوس و همکاران(۲۰۰۷) پیشنهاد می‌کنند که از دو زنجیره با طول حداقل ۵۰۰۰۰ تکرار که ۴۰۰۰۰ تکرار اول به عنوان تکرارهای سوخته در نظر گرفته شود استفاده گردد. در پژوهش حاضر از دو زنجیره مارکوفی با طول ۶۰۰۰۰ که ۴۰۰۰۰ تکرار اول آن به عنوان تکرارهای سوخته به کار رفت استفاده گردید.

همگرایی[2] یا تقارب الگوریتم MCMC از طریق وارسی چشمی نمودارهای زنجیره‌ای، توزیعهای پسین برآورد شده، خودهمبستگی‌های بین برآوردهای حاصل از زنجیره‌های مارکوفی، $\hat{R}$ گلمن و رابین (روسوس و همکاران، ۲۰۰۷) مورد بررسی قرار گرفت.

1. burn-in
2. Convergence

بررسی نمودارهای زنجیره‌ای، توزیعهای پسین برآورد شده، و خودهمبستگیها نشـان داد که تقارب یا همگرایی حاصل شده است. از آنجاییکه تعداد ۲۱۴ سوال وجود دارد(یکی از سوالها به دلیل آنکه هیچ دانش‌آموزی به آن پاسخ نداده بود و فاقد واریانس بـود از فراینـد تحلیل کنار گذاشته شد) و ارائه نمودارهای مربوط به تمام این سوالها در اینجا ممکن نیست لذا فقط نمودارهای مربوط به سوال M05_05 به عنوان نمونه ارائه شده است.

شکل ۱: نمودارهای چگالی، سریهای زمانی و خودهمبستگیها برای سوال M05_05



میانگین و انحراف استاندارد آماره $\hat{R}$ به ترتیب ۱/۰۲ و ۰/۰۳ به دست آمد. مقادیر $\hat{R}$کمتر از ۱/۲ بیانگر همگرایی است. میانگین و انحراف استاندارد خودهمبستگیها نیز به ترتیـب ۰/۱۴ و ۰/۱۶ به دست آمد.

بررسـی نمودارهـای زنجیـره‌ای، توزیعهـای پسـین و مقـادیر مربـوط بـه آمـاره $\hat{R}$ و خودهمبستگیها نشان داد که همگرایی حاصل شـده اسـت. بنـابراین تحلیـل‌هـای مربـوط بـه سوال‌های پژوهش صورت گرفت که در ادامه، یافته‌های مربـوط بـه هـر سـوال بـه صـورت جداگانه ارائه می‌گردد.

سوال اول پژوهش: آیا مجموعه مهارت‌ها و صفات شناختی ارائه شـده توسـط کورتر و تاتسوکا(۲۰۰۲) که در پژوهش‌های مختلف نیز بکـار بـرده شـده‌انـد، قابل تعمیم به عملکرد دانش‌آموزان پایه هشتم ایران است؟

مدل ارائه شـده توسـط تاتسـوکا و همکـارانش مرکـب از ۲۷ صـفت یـا مهـارت اسـت. در

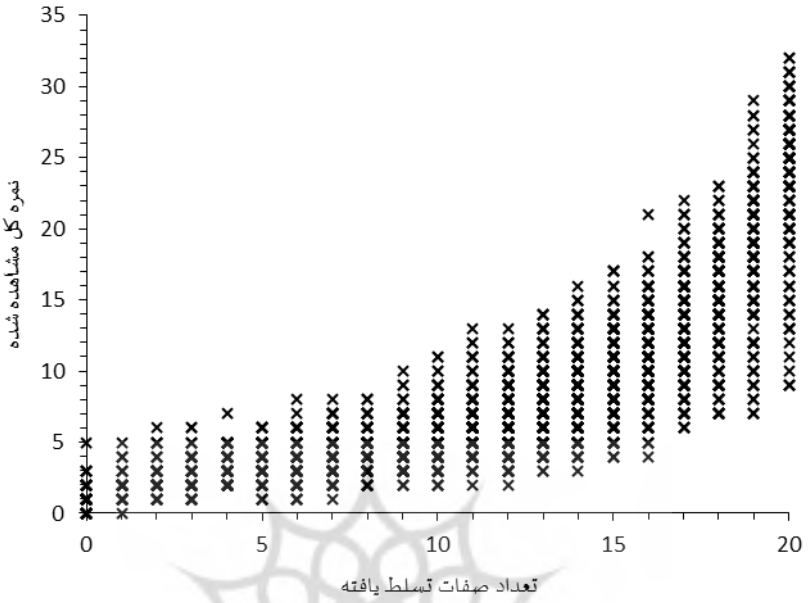
پژوهش حاضر پس از شکل‌گیری ماتریس اولیه Q بررسیهای مقدماتی این ماتریس اولیه نشان داد که از بین ۲۷ صفت مورد نظر، ۷ صفت C1، C6، P8، P9، S1، S2 و S4، یا به تمام سوال‌ها مربوط می‌شدند، یا به هیچ کدام از سوال‌ها مربوط نمی‌شدند و یا به کمتر از ۳ سوال مربوط می‌شدند. بنابراین، این ۷ صفت حذف گردیدند و ماتریس اولیه Q با ۲۰ صفت شکل گرفت. در پژوهش توکر(۲۰۱۰) نیز ۷ صفت P5، P6، S4، S5، S8، و S10 به دلیل اینکه به کمتر از ۳ سوال مربوط می‌شدند از ماتریس Q حذف شدند.

برای پاسخگویی به سوال اول پژوهش یا به عبارتی برای بررسی کفایت فهرست صفات و ماتریس Q مورد استفاده، از چندین رویکرد و تحلیل آماری بهره گرفته شد که نتایج حاصل در زیر ارائه می‌شود.

۱- برای بررسی برازش کلی مدل، رابطه بین تعداد صفات تسلط‌یافته با نمره کل مشاهده شده مورد بررسی قرار گرفت. فرض بر این است که وجود یک رابطه تکنوای افزاینده حاکی از برازش کلی مدل با داده‌ها است. در اکثر پژوهش‌های CDA که تا کنون صورت گرفته است برای تعیین تسلط یا عدم تسلط در صفات از نقطه برش ۰/۶۰(به عنوان مثال، کیم، ۲۰۱۱؛ جانگ، ۲۰۰۹) در مورد احتمال‌های پسین تسلط بر صفت(*ppm*) استفاده شده است. در پژوهش حاضر نیز از نقطه برش ۰/۶۰ برای تعیین وضعیت تسلطی دانش‌آموزان در مهارت‌ها استفاده گردید. اگر مقدار احتمال پسین تسلط(*ppm*) فردی در یک صفت، بالاتر از ۰/۶۰ بود به عنوان مسلط در آن صفت تلقی می‌شد. نمودار ۱ رابطه بین تعداد صفات تسلط‌یافته و نمره کل مشاهده شده را برای ۳۹۸۱ دانش‌آموز نشان می‌دهد. در این نمودار هر "x" نشان دهنده خوشه‌ای از افراد است.

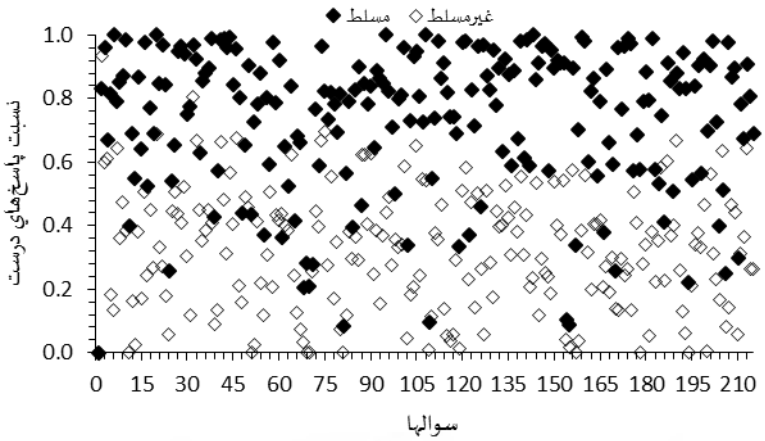
پرتال جامع علوم انسانی

نمودار ۱: رابطه بین تعداد صفات تسلط‌یافته و نمره کل مشاهده شده



۲- برای تعیین این که تا چه اندازه سوال‌های آزمون، اطلاعات تشخیصی فراهم می‌آورند، عملکرد دانش‌آموزانی که در مهارت‌های مورد نیاز یک سوال به تسلط رسیده‌اند با عملکرد دانش‌آموزان غیرمسلط مورد مقایسه قرار گرفت. آماره مورد استفاده برای این منظور در چارچوب مدل *Reduced NC-RUM*، آماره تسلط بر سوال(IMstats) نامیده می‌شود(هارتز و روسوس، ۲۰۰۵). در واقع، آماره IMstats درجه دشواری سوال‌ها را در بین دو گروه افراد مسلط و غیر مسلط در صفات مورد نیاز سوال مورد مقایسه قرار می‌دهد. نمودار۲ این آماره را برای ۲۱۴ سوال آزمون ریاضی تیمز ۲۰۰۷ نشان می‌دهد.

نمودار۲: مقایسه بین عملکرد افراد مسلط و غیر مسلط در سوالها



آماره IMstats بر این فرض استوار است که تفاوت در عملکرد افراد مسلط در صفات مورد نیاز سوال و عملکرد افراد غیرمسلط، شاهدی قوی بر روایی مدل شناختی و ظرفیت تشخیصی سوال‌ها فراهم می‌آورد. شکل فوق به وضوح نشان می‌دهد که عملکرد افراد مسلط در صفات، در اکثر سوال‌ها بهتر از عملکرد افراد غیرمسلط است. دامنه تفاوت بین درجه دشواری(P) سوال‌ها در بین افراد مسلط و غیر مسلط از ۰/۱۶ تا ۰/۸۷ و میانگین و انحراف استاندارد آن به ترتیب ۰/۴۱ و ۰/۲۰ است. تفاوت ۰/۴۰ بین عملکرد مسلط‌ها و غیر مسلط‌ها به عنوان ملاکی برای تعیین و شناسایی سوال‌های با ظرفیت و اطلاعات تشخیصی بالا به کار می‌رود. هرچه میزان این تفاوت بالاتر از ۰/۴۰ باشد، بیانگر این است که سوال از توانایی بالاتری در تمایزگذاری بین افراد مسلط و غیر مسلط برخوردار است. بررسی دقیق‌تر آماره IMstats نشان داد که تقریبا ۴۲٪ از سوال‌ها از آگاهی‌دهندگی تشخیصی[1] ضعیفی برخوردارند. به عبارات دقیق‌تر در ۴۲٪ از سوال‌ها تفاوت بین عملکرد مسلط‌ها و غیر مسلط‌ها کمتر از ۰/۴۰ بود.

۳- تحلیل دیگری که در زمینه ارزشیابی کفایت فهرست صفات و ماتریس Q صورت گرفت، تحلیل رگرسیون چندگانه بود. در این تحلیل درجه دشواری سوال‌ها در مدل راش، یعنی پارامتر b_i، به عنوان متغیر وابسته و ستونهای ماتریس Q نهایی (یعنی صفات دخیل در هر سوال) به عنوان متغیرهای مستقل وارد تحلیل شدند. مقدار R=۰/۷۵ به دست آمد که

1. diagnostic information

نشان می‌دهد ۵۶٪ از واریانس موجود در پارامتر b_i توسط ۲۰ صفت موجود در ماتریس Q تبیین می‌شود.

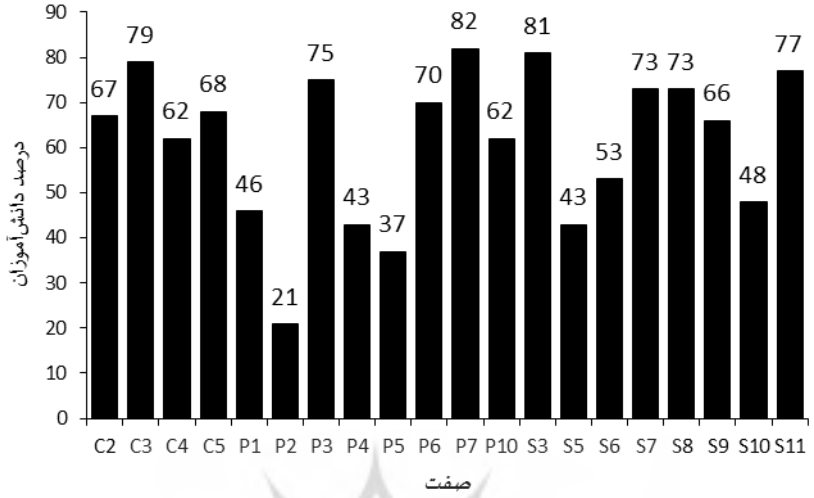
کفایت ماتریس Q در سطح دانش‌آموز نیز مورد بررسی قرار گرفت. برای این منظور، توان احتمال‌های تسلط بر صفت(ppm) در پیش‌بینی پارامتر دانش‌آموزان در مدل راش، با استفاده از تحلیل رگرسیون چندگانه مورد مطالعه قرار گرفت. در این تحلیل، احتمالهای تسلط بر صفت(ppm) به عنوان متغیرهای پیش‌بین و مقدار برآورد شده پارامتر توانایی( ) به عنوان متغیرهای وابسته بکار رفتند. نتایج حاصل نشان داد که تقریبا ۹۱٪، (R=۰/۹۵۴)، از واریانس پارامتر توانایی( ) دانش‌آموزان توسط ppm تبیین می‌شود.

با توجه به این که نتایج، حاکی از کفایت و برازش مطلوب مدل شناختی با داده‌های دانش‌آموزان ایران است، بنابر این گام مهم بعدی توصیف وضعیت تسلطی جامعه دانش‌آموزان در صفات مورد مطالعه و مقایسه بین دو گروه دانش‌آموزان دختر و پسر است.

سوال دو پژوهش: 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؟

نمودار۳ درصد دانش‌آموزان به حد تسلط رسیده در هر صفت را بر مبنای نقطه برش ۰/۶۰ نشان می‌دهد. با بررسی درصدهای ارائه شده در نمودار ۳ مشخص می‌شود که دانش‌آموزان پایه هشتم ایران در ۷ صفت P1(فرمولبندی معادلات)، P2(کاربرد دانش حساب و هندسه)، P4(به کارگیری قواعد جبر)، P5(تفکر استقرایی)، S5(بررسی صحت گزینه‌ها)، S6(الگوها و روابط)، و S10(سوال باز پاسخ) بر اساس نقطه برش ۰/۶۰ به حد تسلط نرسیده‌اند و در دو صفت C4(خواص هندسه دو بعدی) و P10(خواندن کمّی و منطقی) نیز وضعیت چندان مطلوبی ندارند. در واقع دانش‌آموزان ایران در ۵۰٪ از صفات شناختی مورد مطالعه در پژوهش حاضر از وضعیت ضعیف و نامطلوبی برخوردارند.

نمودار۳: درصد دانش‌آموزان مسلط در هر صفت براساس نقطه برش ۰/۶۰



سوال سوم پژوهش: آیا بین عملکرد دانش‌آموزان دختر و پسر پایه هشتم ایرانی در سوال‌های ریاضیات تیمز ۲۰۰۷ از لحاظ الگوی تسلط بر مهارت‌ها یا صفات شناختی تفاوت‌هایی وجود دارد؟

یک یافته مشترک و رایج در خصوص مقایسه پیشرفت ریاضی دختران و پسران این است که برتری پسران نسبت به دختران در ریاضیات در طول سالهای مقطع راهنمایی آغاز می‌شود و در سالهای مقطع متوسطه، در بسیاری از حوزه‌های ریاضی، پسران از دختران بسیار سبقت می‌گیرند(ما[1]، ۱۹۹۵؛ ناصر و بایرن بام[2]، ۲۰۰۵). بدون داشتن اطلاعات شناختی، مطالعه و پژوهش در زمینه تفاوتهای جنسیتی در پیشرفت ریاضی به نمرات کل و نمرات خرده مقیاس‌های آزمون‌های استاندارد، بسیار وابسته است. برای نمونه برخی از مطالعات تجربی نشان می‌دهد که عملکرد دانش‌آموزان پسر در موضوعاتی مانند اندازه‌گیری، تناسب و هندسه بالاتر از دختران است. از طرف دیگر دختران در سوال‌هایی که نیازمند مهارت‌های محاسبه، روابط نمادین و عملیات مجموعه‌ها است از عملکرد بالاتری نسبت به پسران برخوردارند. در حال حاضر هیچ یافته قطعی راجع به تفاوتهای

1. Ma
2. Naser & Birenbaum

جنسیتی در پیشرفت ریاضی وجود ندارد.

به عنوان جایگزینی برای نمره‌های کل و نمره‌های خرده مقیاسها مطالعه حاضر این امکان را داد تا مقایسه‌های بین دختران و پسران در سطح صفت صورت گیرد. برای انجام مقایسه بین دختران و پسران، میانگین *ppm* این دو گروه در صفات ۲۰ گانه با استفاده از آزمون t برای دو گروه مستقل مورد آزمون معناداری آماری قرار گرفت. به خاطر مقایسه‌های چندگانه‌ای که بین دختران و پسران صورت گرفت (به تعداد صفات) به منظور کنترل نرخ خطای نوع اول، سطوح معناداری آزمونt با استفاده از روش بنجامینی- هاچبرگ[1] (B-H؛ ۱۹۹۵) تعدیل گردید. نتایج حاصل از این تحلیل همراه با میانگین و انحراف استاندارد ppm برای دو گروه دختران و پسران در جدول ۱ ارائه شده است. در این جدول، در ستون متغیر *جنسیت*، اعداد داخل پرانتز، انحراف استاندارد و اعداد خارج پرانتز، میانگین‌های *ppm* مربوط به دختران و پسران است. در ستون مقدار آماره t نیز، اعداد داخل پرانتز، درجه آزادی و اعداد خارج پرانتز مقدار مشاهده شده آماره t است. سطوح معناداری که در ستون آخر ارائه شده است، با استفاده از روش بنجامینی- هاچبرگ اصلاح شده‌اند. بررسی سطوح معناداری نشان می‌دهد که از بین ۲۰ صفت مورد مقایسه، در ۶ صفت C2، C3، P1، P5، S5 و S7 بین میانگین‌های دو گروه دختران و پسران تفاوت آماری معنادار وجود دارد. بررسی بیشتر یافته‌ها نشان می‌دهد که در این ۶ صفت میانگین عملکرد پسران در ۴ صفت C2، P1، P5، و S7 بالاتر از دختران است و در دو صفت C3، و S5 دختران از میانگین بالاتری نسبت به پسران برخوردارند.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

1. Benjamini-Hochberg

جدول ۱: میانگین و انحراف استاندارد PPM به تفکیک جنسیت و نتایج مربوط به آزمون t

صفت	جنسیت		مقدار آماره t	سطح معناداری(P)	شاخص d کوهن	اندازه اثر r
	دختر(N=۱۷۸۶)	پسر(N=۲۱۹۴)				
C2	۰/۷۲(۰/۲۰)	۰/۷۴(۰/۲۰)	۲/۷۸(۳۹۷۸)	۰/۰۰۳	۰/۱۰	۰/۰۵
C3	۰/۷۶(۰/۱۸)	۰/۷۵(۰/۱۹)	۲/۸۴(۳۸۷۴)	۰/۰۰۲	-۰/۰۵	-۰/۰۳
C4	۰/۶۸(۰/۱۹)	۰/۶۸(۰/۲۰)	۰/۲۶(۳۸۶۸)	۰/۴۰	۰	۰
C5	۰/۷۱(۰/۱۶)	۰/۷۲(۰/۱۷)	۲/۲۷(۳۸۷۷)	۰/۱۱	۰/۰۶	۰/۰۳
P1	۰/۵۷(۰/۲۱)	۰/۶۰(۰/۲۲)	۳/۸۱(۳۹۷۸)	۰/۰۰۰۷	۰/۱۴	۰/۰۷
P2	۰/۲۶(۰/۳۴)	۰/۲۷(۰/۳۴)	۰/۲۷(۳۹۷۸)	۰/۳۹	۰/۰۳	۰/۰۱
P3	۰/۷۸(۰/۱۸)	۰/۷۷(۰/۱۹)	۰/۱۶(۳۹۷۸)	۰/۴۴	-۰/۰۳	-۰/۰۳
P4	۰/۵۰(۰/۳۴)	۰/۵۱(۰/۳۴)	۰/۶۷(۳۹۷۸)	۰/۲۵	۰/۰۳	۰/۰۱
P5	۰/۵۵(۰/۲۷)	۰/۵۸(۰/۲۷)	۳/۸۷(۳۸۶۷)	۰/۰۰۰۵	۰/۱۲	۰/۰۶
P6	۰/۷۰(۰/۲۱)	۰/۷۰(۰/۲۲)	۰/۰۳(۳۸۹۰)	۰/۴۹	۰	۰
P7	۰/۷۸(۰/۱۶)	۰/۷۹(۰/۱۶)	۱/۸۱(۳۹۷۸)	۰/۱۰	۰/۰۶	۰/۰۳
P10	۰/۷۰(۰/۲۰)	۰/۷۱(۰/۲۱)	۰/۴۶(۳۹۶۹)	۰/۳۲	۰/۰۵	۰/۰۳
S3	۰/۷۷(۰/۲۱)	۰/۷۷(۰/۲۱)	۰/۴۰(۳۹۷۸)	۰/۳۴	۰	۰
S5	۰/۵۸(۰/۳۳)	۰/۵۶(۰/۳۴)	۲/۵۲(۳۹۷۸)	۰/۰۰۵	-۰/۰۶	-۰/۰۳
S6	۰/۶۲(۰/۱۹)	۰/۶۲(۰/۲۰)	۰/۵۱(۳۹۷۸)	۰/۳۱	۰	۰
S7	۰/۷۴(۰/۱۶)	۰/۷۶(۰/۱۶)	۴/۵۲(۳۹۷۸)	۰/۰۰۰۳	۰/۱۳	۰/۰۶
S8	۰/۷۶(۰/۱۹)	۰/۷۵(۰/۱۹)	۱/۲۲(۳۹۷۸)	۰/۱۱	-۰/۰۵	-۰/۰۳
S9	۰/۷۳(۰/۱۶)	۰/۷۴(۰/۱۷)	۱(۳۸۷۲)	۰/۱۶	۰/۰۶	۰/۰۳
S10	۰/۶۶(۰/۲۹)	۰/۶۵(۰/۳۱)	۰/۷۵(۳۹۰۴)	۰/۲۳	-۰/۰۳	-۰/۰۲
S11	۰/۷۷(۰/۲۰)	۰/۷۷(۰/۲۰)	۰/۴۵(۳۹۷۸)	۰/۳۳	۰	۰
میانگین	–	–	–	–	۰/۰۳	۰/۰۲

بحث و نتیجه‌گیری

سنجش تشخیصی شناختی روشی سودمند برای مطالعه فرایندهای شناختی در سطح خُرد، مهارت‌های تفکر ریاضی و دانش است که می‌تواند ما را به کشف یافته‌های تازه رهنمون شود و بینشی را در مورد مشکلات امر آموزش فراهم نماید. در پژوهش حاضر تلاش شد تا پیشرفت ریاضی دانش‌آموزان پایه هشتم ایران در تیمز ۲۰۰۷ با استفاده از رویکرد سنجش تشخیصی شناختی(CDA) مورد بررسی قرار گیرد. برای توصیف وضعیت دانش‌آموزان ایران در مهارتها، صفات و تواناییهای شناختی که زیربنای عملکرد درسوال‌های ریاضی تیمز ۲۰۰۷ قرار می‌گیرند از مدل *Reduced NC-RUM*(هارتز، ۲۰۰۲) که از طریق برنامه نرم‌افزاری *Arpeggio*(دی‌بلو و استوت، ۲۰۰۸) قابل اجرا است استفاده گردید. در واقع، ۳

هدف اصلی پژوهش حاضر عبارت بود از: ۱- مطالعه برازش مجموعه صفات یا مدل شناختی کورتر و تاتسوکا(۲۰۰۲، نقل از تاتسوکا، کورتر و تاتسوکا، ۲۰۰۴) با عملکرد دانش‌آموزان پایه هشتم ایران در آزمون ریاضی تیمز ۲۰۰۷؛ ۲- توصیف وضعیت دانش‌آموزان پایه هشتم ایران در مهارتها یا صفات شناختی مطرح شده توسط کورتر و تاتسوکا بر اساس آزمون ریاضی تیمز ۲۰۰۷؛ و ۳- مقایسه عملکرد دانش‌آموزان دختر و پسر در مهارت‌ها یا صفات مورد نظر. در ادامه، یافته‌ها، تلویحات و نتیجه‌گیری‌های حاصل از پژوهش مورد بحث و بررسی قرار می‌گیرد.

برازش یا کفایت مدل شناختی

مدل شناختی مورد استفاده در پژوهش حاضر متشکل از فهرستی از صفات شناختی، همراه با یک ماتریس Q بود که رابطه صفات شناختی را با سوال‌ها توصیف می‌کند. این مدل شناختی که شامل ۲۷ صفت یا مهارت شناختی است توسط کورتر و تاتسوکا(۲۰۰۲، نقل از تاتسوکا، کورتر و تاتسوکا، ۲۰۰۴) پرورش یافت و تا کنون در مطالعات متعدد، از جمله بایرن‌بام، تاتسوکا و یامادا(۲۰۰۴) و توکر(۲۰۱۰) مورد استفاده قرار گرفته است. در پژوهش حاضر، برازش مدل شناختی مذکور با داده‌های حاصل از عملکرد دانش‌آموزان ایران مورد بررسی قرار گرفت. بررسی‌های مقدماتی نشان داد که از این صفات ۲۷ گانه تعداد ۷ صفت، که شامل C1(اعداد حسابی و صحیح)، C6(اندازه‌گیری)، P8(ارزشیابی صحت ریاضی)، P9(مدیریت داده‌ها)، S1(تبدیل واحد)، S2(درک عدد)، و S4(تخمین/تقریب) است یا به کمتر از ۳ سوال و یا به تمام سوال‌ها مربوط می‌شدند و یا به هیچ کدام از سوال‌ها مربوط نمی‌شدند و بنابراین از مدل شناختی و ماتریس Q حذف شدند. این یافته با نتایج حاصل از پژوهش تاتسوکا، کورتر و تاتسوکا(۲۰۰۴)، توکر(۲۰۱۰)، و چن(۲۰۰۶) از همخوانی خوبی برخوردار است. در پژوهش تاتسوکا و همکارانش(۲۰۰۴) چهار صفت C6(اندازه‌گیری)، S1(تبدیل واحد)، S9(مقایسه‌ها) و P8(ارزشیابی صحت ریاضی) و در پژوهش توکر(۲۰۱۰) که مدل شناختی مورد بحث را در مورد داده‌های حاصل از عملکرد دانش‌آموزان پایه هشتم ترکیه در آزمون ریاضی تیمز ۲۰۰۷ به کار برد، ۷ صفت P5(استدلال منطقی)، P6(تفکر قیاسی یا تحلیلی)، S4(تخمین/تقریب)، S5(ارزیابی/بررسی صحت گزینه‌ها)، S8(حل مسائل جدید) و

S10(سوال باز پاسخ) از مدل شناختی حذف شدند. در پژوهش چن(۲۰۰۶) نیز ۴ صفت C6(اندازه‌گیری)، S1(تبدیل واحد)، S9(مقایسه‌ها)، و P8(ارزشیابی صحت ریاضی) از ماتریس Q حذف شدند. دلیل حذف این صفات در ۳ پژوهش مذکور، همان دلایل ذکر شده برای حذف صفات در پژوهش حاضر بود.

در ارتباط با کفایت مدل شناختی مورد استفاده، ۳ تحلیل صورت گرفت که شامل رگرسیون چندگانه در سطح سوال و دانش‌آموز، رابطه بین تعداد صفات تسلط‌یافته با پارامتر توانایی( ) در مدل راش، و مقایسه عملکرد افراد مسلط و غیرمسلط در صفات در سوال‌ها(آماره IMstats؛ جانگ، ۲۰۰۹) بود.

نتایج حاصل از رگرسیون چندگانه نشان داد که ۵۶٪ از واریانس موجود در پارامتر b_i توسط ۲۰ صفت موجود در ماتریس Q تبیین می‌شود. این یافته با نتیجه حاصل از پژوهش تاتسوکا، کورتر و تاتسوکا(۲۰۰۴)، توکر(۲۰۱۰)، چن(۲۰۰۶) و دوگان و تاتسوکا(۲۰۰۸) هماهنگ است. پژوهش تاتسوکا، کورتر و تاتسوکا(۲۰۰۴) نشان داد که ۸۷٪ از واریانس موجود در درجه دشواری سوال‌ها توسط ۲۷ صفت موجود در ماتریس Q تبیین می‌شود. توکر(۲۰۱۰) نیز نشان داد که ۶۵٪ از واریانس موجود در دشواری سوالها توسط ۲۰ صفت مورد استفاده تبیین می‌شود. در پژوهش دوگان و تاتسوکا(۲۰۰۸) نیز که مدل شناختی کورتر و تاتسوکا(۲۰۰۲) را در مورد داده‌های حاصل از پاسخهای دانش‌آموزان ترکیه درتیمز ۱۹۹۹ بکار بردند، برای مطالعه کفایت مدل شناختی از تحلیل رگرسیون چندگانه استفاده گردید. نتیجه پژوهش آنها نشان داد که قریب به ۸۷٪ از واریانس موجود در دشواری سوال‌ها توسط صفات یا مدل شناختی تبیین می‌شود. همچنین چن(۲۰۰۶) نشان داد که ۸۷٪ از واریانس موجود در دشواری سوال‌ها از طریق ۲۳ صفت به کار رفته در ماتریس Q قابل تبیین است. همانطور که ملاحظه می‌شود درصد واریانس تبیین شده در پژوهش حاضر و پژوهش توکر(۲۰۱۰) با یکدیگر و درصد واریانس تبیین شده در ۳ پژوهش دیگر نیز با یکدیگر از همخوانی بیشتری برخوردارند. در عین حال، درصد واریانس تبیین شده در پژوهش حاضر و پژوهش توکر کمتر از ۳ پژوهش دیگر است. یک دلیل آن ممکن است این باشد که در پژوهش توکر نیز همانند پژوهش حاضر، از مدلی شناختی که بر اساس سوال‌های ریاضی تیمز ۱۹۹۹ شکل گرفته است استفاده گردید. با توجه به تغییرات اعمال شده در محتوای کتب درسی بر اساس یافته‌های تیمز و همچنین با

توجه به تفاوت‌های موجود بین چارچوب و محتوای آزمون ریاضی تیمز ۲۰۰۷ با تیمز ۱۹۹۹ این مقدار کاهش در تبیین واریانس دشواری سوال‌ها دور از انتظار نیست. دلیل دیگر کاهش در تبیین واریانس موجود در درجه دشواری سوال‌ها این است که در پژوهش توکر و پژوهش حاضر از ۲۰ صفت استفاده گردید در حالیکه در ۳ پژوهش دیگر از ۲۳ صفت استفاده شده است. در واقع در تمام این ۴ مطالعه مذکور و همچنین در پژوهش حاضر بیش از ۵۰٪ واریانس موجود در دشواری سوالها توسط مدل شناختی مورد استفاده تبیین می‌شود. بنابراین، می‌توان نتیجه گرفت که مدل پیشنهاد شده توسط کورتر و تاتسوکا(۲۰۰۲) برای سوال‌های ریاضی تیمز ۱۹۹۹ نه تنها عملکرد دانش‌آموزان ۲۰ کشور مورد مطالعه در پژوهش تاتسوکا، کورتر و تاتسوکا(۲۰۰۴) و عملکرد دانش‌آموزان تایوانی در مطالعه چن(۲۰۰۶) و عملکرد دانش‌آموزان ترکیه در پژوهش توکر(۲۰۱۰) را تبیین می‌کند بلکه قابل تعمیم به عملکرد دانش‌آموزان پایه هشتم ایران نیز است.

علاوه بر این با انجام یک تحلیل رگرسیون جداگانه کفایت مدل شناختی در پیش‌بینی و تبیین نمره توانایی در مدل راش مورد بررسی قرار گرفت(نمودار). نتایج حاصل نشان داد که ۹۵٪ از واریانس نمرات توانایی در مدل راش توسط احتمال‌های پسین تسلط برصفت(*ppm*) تبیین می‌شود. این یافته با نتیجه حاصل از پژوهش چن(۲۰۰۶) از همخوانی خوبی برخوردار است. چن نیز دریافت که ۹۲٪ از واریانس موجود در نمره‌های خام دانش‌آموزان بوسیله احتمال‌های تسلط بر صفت تبیین می‌شود.

شاهد دیگری که در ارتباط با کفایت مدل شناختی مورد توجه قرار گرفت، بررسی رابطه بین تعداد صفات تسلط‌یافته با نمره توانایی( ) در مدل راش بود. نتایج این تحلیل که در نمودار۱ ارائه شده است بیانگر وجود یک رابطه تکنوای افزاینده بین تعداد صفات تسلط‌یافته و عملکرد در آزمون است. این به آن معنا است که هر چه تعداد صفاتی که دانش‌آموز در آنها به حدتسلط دست یافته است افزایش می‌یابد عملکرد وی نیز در آزمون بهبود می‌یابد. به عبارت دیگر این یافته نشان می‌دهد که عملکرد در آزمون ریاضی پایه هشتم تحت تاثیر صفات یا مدل شناختی مورد استفاده در پژوهش است.

آخرین تحلیل در زمینه برازش مدل شناختی با داده‌ها، مقایسه بین عملکرد دانش‌آموزان مسلط و غیرمسلط در سوال‌ها بود. نتیجه این تحلیل که در چارچوب مدل *Reduced NC-RUM*، آماره IMstats نامیده می‌شود و در نمودار ۲ ارائه شده است نشان داد

که عملکرد دانش‌آموزان مسلط در صفات، در اکثر سوال‌ها بهتر از دانش‌آموزان غیرمسلط است. دامنه آماره IMstats از ۰/۱۶ تا ۰/۸۷ و میانگین و انحراف استاندارد آن به ترتیب ۰/۴۱ و ۰/۲۰ به دست آمد. بررسی دقیق‌تر آماره IMstats نشان داد که تقریبا ۴۲٪ از سوال‌ها از آگاهی‌دهندگی تشخیصی ضعیفی برخوردارند. به عبارات دقیق‌تر در ۴۲٪ از سوال‌ها تفاوت بین عملکرد مسلط‌ها و غیرمسلط‌ها کمتر از ۰/۴۰ بود. البته این یافته که ۴۲٪ از سوال‌های ریاضی تیمز ۲۰۰۷ فاقد قدرت تشخیص هستند دور از انتظار نیست، زیرا آزمون تیمز بر اساس رویکردهای سنجش تشخیصی تهیه نشده است. در واقع در پژوهش حاضر مدل *Reduced NC-RUM* که یک مدل تشخیصی است به آزمون ریاضی تیمز ۲۰۰۷ که یک آزمون غیر تشخیصی است، بازبرازش یافته است.

بررسی نتایج حاصل از تحلیل‌هایی که در خصوص برازش مدل شناختی صورت گرفت، نشان می‌دهد که مدل شناختی مورد مطالعه از برازش مطلوبی با داده‌های حاصل از پاسخهای دانش‌آموزان به سوال‌های آزمون ریاضی تیمز ۲۰۰۷ برخوردار است. بنابراین گام مهم بعدی توصیف وضعیت دانش‌آموزان ایران در این صفات ۲۰ گانه است. نتایج حاصل از این تحلیل‌ها زیر عنوان *وضعیت دانش‌آموزان ایران در صفات ۲۰ گانه* ارائه و مورد بحث و بررسی قرار می‌گیرد.

وضعیت دانش‌آموزان ایران در صفات ۲۰ گانه

در پژوهش حاضر نیز برای تعیین وضعیت تسلطی دانش‌آموزان ایرانی در هر کدام از صفات ۲۰ گانه مورد مطالعه از دو نقطه برش ۰/۶۰ در مورد *ppm*ها استفاده شد. بر اساس نقطه برش ۰/۶۰ اگر *ppm* دانش‌آموزی برای یک صفت، بالاتر از ۰/۶۰ بود به عنوان فرد مسلط و اگر مساوی یا پایین‌تر از ۰/۶۰ بود به عنوان فرد نامسلط طبقه‌بندی می‌شد.

نتایج حاصل از تحلیل‌ها (نمودار ۲) نشان می‌دهد که دانش‌آموزان ایرانی در ۷ صفت P1(فرمولبندی معادلات)، P2(هندسه)، P4(به کارگیری قواعد جبری)، P5(تفکر قیاسی یا تحلیلی)، S5(ارزیابی صحت گزینه‌ها)، S6(الگوها و روابط)، و S10(سوال بازپاسخ یا تفکر واگرا) به حد تسلط نرسیده‌اند.

بایرن‌بام، تاتسوکا و زین(۲۰۰۵) در مطالعه‌ای که با استفاده از مدل RSM[1] بر روی

1. Rule Space Model

داده‌های حاصل از پاسخهای دانش‌آموزان سنگاپور، اسرائیل و آمریکا در آزمون ریاضی تیمز ۱۹۹۹ انجام دادند به این نتیجه رسیدند که دانش‌آموزان سنگاپور در ۲۰ صفت از ۲۳ صفت، بالاترین میانگین احتمال‌های تسلط بر صفت را دارند که با رتبه یکم سنگاپور در مطالعات تیمز از همخوانی بسیار خوبی برخودار است. آنها به این نتیجه رسیدند که یک دانش‌آموز متوسط سنگاپوری در ۲۲ صفت به تسلط رسیده است و تنها در صفت S6(الگوها و روابط) به حد تسلط نرسیده است. در واقع، صفت S6 برای دانش‌آموزان سنگاپوری دشوارترین صفت بود.

بایرن‌بام، تاتسوکا و یامادا(۲۰۰۴) که با استفاده از مدل RSM و بر اساس مدل شناختی کورتر و تاتسوکا(۲۰۰۲) به مقایسه دانش‌آموزان ژاپنی، آمریکایی و اسرائیلی پرداختند، دریافتند که یک دانش‌آموز متوسط ژاپنی در ۲۲ صفت به تسلط رسیده است و تنها در صفت P4(به کارگیری قواعد جبر) به حد تسلط نرسیده است. پژوهش آنها همچنین نشان داد که یک دانش‌آموز متوسط آمریکایی و همتای اسرائیلی او در ۱۷ صفت به تسلط رسیده است.

بطور کلی یافته‌های مربوط به وضعیت تسلطی نشان می‌دهد که دانش‌آموزان ایرانی، بیشترین ضعف را در صفات P1(فرمولبندی معادلات)، P2(هندسه)، P4(به کارگیری قواعد جبر)، و P5(تفکر قیاسی) که صفات مربوط به پردازش شناختی هستند دارند. همچنین در دو صفت S5(بررسی صحت گزینه‌ها) و S6(الگوها و روابط) که به مهارتهای تفکر استقرایی مربوط می‌شوند نیز از عملکرد ضعیفی برخوردارند. در واقع می‌توان گفت که صفت S6 دومین صفت دشوار بعد از صفت P2 در بین دانش‌آموزان پایه هشتم ایران است. تاتسوکا و همکاران(۲۰۰۴) نیز دریافتند که صفت S6 دشوارترین صفت در میان دانش‌آموزان ۲۰ کشور مورد مطالعه است.

چن(۲۰۰۶) نیز نشان داد که صفت S6 دشوارترین صفت در میان دانش‌آموزان تایوانی است. در پژوهش توکر(۲۰۱۰) نیز صفت S6 چهارمین صفت دشوار در بین دانش‌آموزان ترکیه بود. دشوارترین و آسان‌ترین صفت برای دانش‌آموزان ترکیه به ترتیب، C1 و C4 بودند که هر دو در طبقه صفات محتوایی قرار دارند. یافته‌های پژوهش حاضر نشان می‌دهد که آسان‌ترین صفت برای دانش‌آموزان ایران، P7(تولید، تجسم و خواندن اشکال و نمودارها) و S3(استفاده از اشکال و نمودارها) و دشوارترین صفت، P2(هندسه) است.

تحلیل داده‌های حاصل از *پرسشنامه دانش‌آموز* نشان می‌دهد که تنها ۴۶٪ از دانش‌آموزان پایه سوم راهنمایی ایران در بیش از نیمی از درسها به مسائل مربوط به اشکال هندسی، خطوط و زوایا می‌پردازند. در کشورهای ژاپن، جمهوری چک، سنگاپور، کره و هنگ‌کنگ نیز وضعیت به همین شکل است. در این کشورها نیز در حدود ۵۰٪ دانش‌اموزان به مسائل مربوط به هندسه می‌پردازند. بنابراین دور از انتظار نیست که هندسه برای دانش‌آموزان اکثر کشورها دشوار باشد. به نظر می‌رسد که معلمان بایستی زمان بیشتری را صرف آموزش موضوعات مربوط به هندسه نمایند.

از آنجائیکه هندسه از همبستگی بالایی با مهارت‌های مهم ریاضی چون، استدلال نسبتی، کاربرد قضاوتی دانش، مفاهیم و نسبتها، و مهارت‌های پردازش و مدیریت داده‌ها برخوردار است، لذا هندسه را می‌توان مهارت کلیدی برای آموزش مهارت‌های تفکر ریاضی سطح بالاتر دانست. این امر ممکن است ناشی از این باشد که هندسه، حوزه‌ای است که به عنوان یک بافت یا زمینه کارآمد و موثر برای آموزش استدلال منطقی و مهارت‌های قضاوتی سطح بالاتر عمل می‌کند. این یافته بیانگر آن است که برنامه درسی در ایران بایستی تاکید بیشتری بر آموزش هندسه نماید، زیرا هندسه می‌تواند امکان آموزش مهارت‌های مهم تفکر ریاضی را که در علوم فیزیکی و مهندسی مورد نیاز هستند فراهم آورد. شواهد و یافته‌های پژوهشی (به عنوان مثال، هایبرت، و استیگلر، ۲۰۰۰؛ کاواناکا و استیگلر، ۱۹۹۹) نشان می‌دهد که تفاوت‌های موجود میان کشورها از لحاظ مهارت‌های تفکر ریاضی سطح بالاتر ناشی از تفاوت در شیوه‌های آموزش است. یافته‌های مربوط به *پرسشنامه دانش‌آموز* نشان می‌دهد که ۹۰٪ دانش‌آموزان ایران در بیش از نیمی از درس‌های ریاضی به معلم خود در حالی که سخنرانی می‌کند گوش می‌دهند. این رقم در به کشور سنگاپور ۵۳٪، در ژاپن ۵۸٪، در کره ۶۴٪، در هنگ‌کنگ ۷۳٪ و در جمهوری چک ۸۴٪ است. یافته‌های فوق نشان می‌دهد که در کشور سنگاپور و ژاپن که رتبه بالایی در مطالعات تیمز دارند آموزش به شیوه سخنرانی به میزان زیادی منسوخ شده است. در این کشورها معلمان به شیوه‌های مختلف تلاش می‌کنند تا دانش‌آموزان را در فرایند یادگیری و آموزش دخیل نمایند و دانش‌آموزان مشارکت موثری در فرایند آموزش داشته باشند. بنابراین، توصیه می‌شود که معلمان و در راس آنها وزارت آموزش و پرورش بایستی تلاش کنند تا با اتخاذ راهکارهای مناسب، شیوه سخنرانی در تدریس موضوعات مهمی چون

ریاضی را با روش‌های مناسب آموزش، مانند آموزش حل مساله و تفکر انتقادی(سیف، ۱۳۸۲) جایگزین نمایند.

همچنین یافته‌ها نشان می‌دهد یادگیری که در خارج از ساعات رسمی آموزش صورت می‌گیرد تاثیر به سزایی بر فرایندها و موفقیت یادگیری دانش‌آموزان دارد(چن و استیونسون، ۱۹۹۵). یافته‌های حاصل از *پرسشنامه دانش‌آموز* نشان می‌دهد که در کشور سنگاپور ۷۰٪ دانش‌آموزان این کشور گزارش کرده‌اند که معلم ریاضی آنها هر روز یا هفته‌ای ۳ الی ۴ بار تکلیف شب ریاضی به آنها می‌دهد. این رقم در هنگ‌کنگ ۷۳٪ و در ایران ۳۵٪ است. به نظر می‌رسد که معلمان ایران بایستی به دفعات بیشتری تکلیف شب ریاضی به دانش‌آموزان بدهند.

بررسی داده‌های حاصل از *پرسشنامه دبیر* نشان می‌دهد که سطح تحصیلات ۵۰٪ از معلمان ریاضی سوم راهنمایی ایران، پایین‌تر از لیسانس و ۵۰٪ دیگر نیز در سطح کارشناسی است. در کشورهای ژاپن، سنگاپور، کره، هنگ‌کنگ و جمهوری چک بطور متوسط بیش از ۷۰٪ معلمان دارای مدرک لیسانس و نزدیک به ۲۰٪ آنها نیز دارای مدرک تحصیلی فوق لیسانس و بالاتر هستند. بنابراین به نظر می‌رسد که وزارت آموزش و پرورش ایران باید تلاش کند تا از طریق فراهم آوردن تسهیلات ویژه و خاص، معلمان را تشویق نماید تا تحصیلات خود را در مقاطع بالاتر ادامه دهند. اگرچه صرف بالا بودن مدرک تحصیلی منجر به افزایش کارایی نمی‌شود، لکن داشتن مدرک تحصیلی بالا منجر به افزایش انتظارات فرد از خود و همچنین منجر به افزایش انتظارات مسولان مدارس و دانش‌آموزان از معلم می‌شود و در نتیجه باعث می‌شود معلم تلاش نماید تا کارایی خود را افزایش دهد.

داده‌ها نشان می‌دهد که در دوسال آخر منتهی به ۲۰۰۷، ۴۸٪ از معلمان ایرانی در دوره‌های آموزش ضمن خدمت ریاضی شرکت کرده‌اند، این در حالی است که در حدود ۶۵ تا ۶۹ درصد معلمان هنگ‌کنگ و سنگاپور در سالهای ۲۰۰۵ و ۲۰۰۶ در دوره‌های آموزشی ریاضی شرکت نموده‌اند. بنابراین می‌توان گفت که وزارت و ادارات کل آموزش وپرورش بایستی با برگزاری دوره‌های آموزش ریاضی و اجباری کردن شرکت معلمان در این دوره‌ها توانمندیها و دانش ریاضی معلمان ریاضی را افزایش دهد.

بطور کلی به نظر می‌رسد که برای آموزش مهارت‌های اثبات ریاضی، لازم است تا بر

آموزش هندسه تاکید و تمرکز بیشتری شود و به لحاظ مفهومی به شکل عمیق‌تری به این موضوع پرداخته شود. بنابراین، هندسه حوزه‌ای است که در آن می‌توان چنین مهارت‌های سطح بالاتر را به آسانی یاد گرفت و به کار برد. اگر این گفته صادق باشد، در این صورت مطلوب است که برنامه‌های درسی در ایران مورد تجدید نظر قرار گیرد تا زمان اختصاص یافته به آموزش هندسه افزایش یابد.

مقایسه عملکرد دختران و پسران

مساله تفاوتهای جنسیتی در ریاضیات توجه زیادی را در محافل علمی و غیر علمی به خود جلب نموده است. بخش عمده‌ای از این مجادلات و توجهات در این حوزه ناشی از این است که ریاضیات دروازه ورود به مشاغل و حرفه‌های پردرآمد و با پرستیژ، مانند علوم، تکنولوژی، مهندسی و استادی دانشگاه است(سِسی و ویلیامز، ۲۰۰۷، نقل از استوت و گی‌یری، ۲۰۱۲).

برای مقایسه میانگین *ppm* دختران و پسران در صفات یا مهارتهای ۲۰ گانه از آزمون t برای دو گروه مستقل استفاده گردید. نتایج نشان داد که از بین ۲۰ صفت مورد مقایسه، در ۶ صفت C2(اعداد کسری و اعشاری)، C3(جبر)، P1(فرمولبندی معادلات)، P5(تفکر قیاسی)، S5(ارزیابی صحت گزینه‌ها) و S7(استدلال نسبتی) بین میانگین‌های *ppm* دو گروه دختران و پسران تفاوت آماری معنادار وجود دارد. بررسی بیشتر یافته‌ها نشان می‌دهد که از بین این ۶ صفت در ۴ صفت C2، P1، P5، وS7 میانگین عملکرد پسران بالاتر از دختران است و در دو صفت C3، و S5 میانگین دختران از میانگین پسران بالاتر است.

در پژوهش چن(۲۰۰۶) تنها در دو صفت S5 و P5 بین میانگین‌های احتمال‌های تسلط بر صفت تفاوت آماری معنادار وجود داشت. در S5 میانگین دختران بالاتر از پسران بود در حالیکه پسران در P5(یعنی در استدلال منطقی) میانگین بالاتری داشتند. در پژوهش حاضر نیز در صفت S5 میانگین *ppm* دختران بالاتر از پسران بود در حالیکه در صفت p5 میانگین پسران بالاتر از میانگین گروه دختران بود که از همخوانی خوبی با یافته‌های پژوهش چن(۲۰۰۶) برخوردار است.

تاتسوکا و دودو(۲۰۰۰) در پژوهشی که با استفاده از آزمون‌های کمّی GRE در بین دانش‌آموزان آمریکایی و در سطح صفت انجام دادند دریافتند که بین عملکرد دختران و

پسران تفاوت‌های آماری معنادار وجود دارد. آنها به این نتیجه رسیدند که دانش‌آموزان پسر از دانش و آگاهی بیشتری برخوردارند و در حل مسائل کلامی، و استدلال ریاضی و مهارت‌های تست‌زنی از عملکرد یا مهارت بهتری برخوردارند.

اگرچه آزمونهای معناداری آماری در پژوهش حاضر و در پژوهش چن(۲۰۰۶) و تاتسوکا و دودو(۲۰۰۰) نشان می‌دهند که تفاوت آماری معنادار بین عملکرد دختران و پسران در آزمون‌های ریاضی وجود دارد لکن معنادار بودن این مقایسه‌ها ناشی ازبالا بودن حجم نمونه گروههای مورد مقایسه است. مقدار شاخص d کوهن(۱۹۸۸) که در جدول ۱ ارائه شده است برای اکثر صفات بسیار پایین‌تر از ملاک ۰/۲۰ است که توسط کوهن به عنوان اندازه اثر کم شناخته می‌شود. میانگین شاخص d کوهن برای ۲۰ صفت، ۰/۰۳ است که بسیار ناچیز و اندک محسوب می‌شود. در واقع بر اساس شاخص‌های اندازه اثر d و r می‌توان گفت که جنسیت ارتباط بسیار کمی با عملکرد در ریاضی دارد و بین دختران و پسران از لحاظ عملکرد در آزمونهای ریاضی و بویژه آزمون ریاضی پایه هشتم تیمز ۲۰۰۷ تفاوتی وجود ندارد. این یافته با نتایج حاصل از فراتحلیل هاید، فنما، و لامون(۱۹۹۰) همخوانی خوبی دارد. آنها با انجام فراتحلیل روی ۱۰۰ مقاله با ۲۵۴ اندازه اثر مستقل به این نتیجه رسیدند که متوسط شاخص اندازه اثر d=۰/۱۵ و به نفع گروه پسران است. هدجز و ناول(۱۹۹۵) مجموعه داده‌های حاصل از نمونه‌های بزرگی از نوجوانان آمریکا دریافتند که در مجموعه داده‌های مختلف، دامنه شاخص d کوهن از ۰/۰۳ تا ۰/۲۶ و به نفع پسران است. لیندبرگ، هایدی، پترسون و لین(۲۰۱۰) با انجام فراتحلیل روی نتایج حاصل از ۲۴۲ مقاله با نمونه‌ای بالغ بر ۱/۲۸۶/۳۵۰ آزمودنی بطور کلی به این نتیجه رسیدند که متوسط تفاوتهای جنسیتی در ریاضیات بر حسب شاخص d در مقطع ابتدایی نزدیک به صفر و در مقطع راهنمایی و متوسطه به ترتیب ۰/۲۳ و ۰/۱۸ است. بطور خلاصه تحقیقات پیشین نشان می‌دهد که تفاوتهای جنسیتی در عملکرد در ریاضیات بسیار کم و بسته به گروه نمونه و شاخص متغیر وابسته گاهی اوقات به نفع پسران و گاهی اوقات به نفع دختران است. در پژوهش حاضر نیز دامنه شاخص d از ۰/۰۳- تا ۰/۱۵+ با میانگین ۰/۰۳ به دست آمد. در برخی از صفات، عملکرد دختران بهتر از پسران و در برخی دیگر از صفات عملکرد پسران بهتر از دختران بود. البته همانطور که شاخص‌های اندازه اثر نیز نشان می‌دهد این تفاوت در عملکرد بسیار ناچیز و اندک است که از همسانی خوبی با نتایج تحقیقات صورت گرفته در

این حوزه(به عنوان مثال، ملیش و همکاران، ۲۰۰۸) برخوردار است. این پژوهشگران اندازه اثر مربوط به ۹ متغیر پیش‌بین را برای عملکرد در ریاضی در دانش‌آموزان ۱۰ ساله مقایسه نمودند. این ۹ متغیر شامل: وزن تولد، جنسیت، پایگاه اجتماعی- اقتصادی، تحصیلات پدر، تحصیلات مادر، درآمد خانواده، کیفیت محیط یادگیری خانواده، کارآمدی آموزشهای پیش‌دبستان، و کارآمدی آموزش‌های دبستان بود. یافته جالب پژوهش مذکور این بود که جنسیت ضعیف‌ترین پیش‌بینی کننده در میان این ۹ پیش‌بینی کننده بود(کمترین اندازه اثر را داشت). تحصیلات مادر، کیفیت محیط یادگیری خانه، و کارآمدی آموزشهای دبستان، قوی‌ترین پیش‌بینی کننده‌ها بودند. یافته پژوهش حاضر نیز با نتایج حاصل از پژوهش ملیش و همکاران(۲۰۰۸) از حیث اینکه جنسیت، پیش‌بینی کننده قوی عملکرد در ریاضیات نیست از همخوانی خوبی برخوردار است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

کتابنامه

Birenbaum, M., Tatsuoka, C., & Yamada, T. (2004). Diagnostic assessment in TIMSS-R: Betweencountries and within-country comparisons of eighth graders mathematics performance. *Studies in Educational Evaluation*, *30*(2), 151-173.

Chen, Y., Gorin, J., Thompson, M., & Tatsuoka, K. (2006, April). *Verification of cognitive attributes required to solve the TIMSS-1999 mathematics items for Taiwanese students.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, US.

DiBello, L. V., & Stout, W. (2008). *Arpeggio documentation and analyst manual* (Ver. 3.1.001) [Computer software]. St. Paul, MN: Assessment Systems Corporation.

DiBello, L. V., Stout, W. F., & Roussos, L. (1995). Unified cognitive psychometric assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-390). Hillsdale, NJ: Erlbaum.

Dogan, E., & Tatsuoka, K. K. (2008). An international comparison using a diagnostic testing model: Turkish students profile of mathematical skills on TIMSS-R. *Educational Studies in Mathematics*, *68*(3), 263-272.

Fu, J., & Li, Y. (2007). *An integrated review of cognitively diagnostic psychometric models*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL, April.

Kim, Y. H. (2011). Diagnosing EAP writing ability using the Reduced Reparameterized Unified Model. *Language Testing*, DOI: 10.1177/0265532211400860

Lee, Y-S., Park, Y-S., & Tayalan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and U.S national sample using TIMSS 2007. *International Journal of testing, 11*, 144-177.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J (2004). TIMSS 2003 international mathematics report. Chestnut Hill, MA: Boston College.

Mullis, I. V. S., Martin, M. O., Gonzales, E. J., Gregory, K. D., Garden, R. A., O Connor,K. M., Chrostowski, S. J., & Smith, T. A. (2000). *TIMSS 1999 International Mathematics Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade.* Chestnut Hill, MA: International Study Center, Boston College.

Olson, J. F., Martin,M. O., & Mullis, I. V. S. (2009). *TIMSS 2007 technical report*. Chestnut Hill, MA: IEA.

Roussos, L. A., DiBello, L. V., Stout, W. F., Hartz, S. M., Henson, R. A., &

Templin, J. L. (2007). The fusion model skills diagnosis system. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 275-318). Cambridge, UK: Cambridge University Press.

Tatsuoka, K. K., Corter, J., & Tatsuoka, C. (2004). Patterns of diagnosed mathematical content and process skills in TIMSS-R across a sample of 20 countries. *American Educational Research Journal, 41*(4), 901-926.

Toker, T. (2010). Cognitive diagnostic assessment of TIMSS-2007 mathematics achievement item for 8th grades in Turkey. (Master s thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 1482411)

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی