فصلنامه مطالعات اندازه‌گیری و ارزشیابی آموزشی
سال نهم، شماره ۲۶، تابستان ۱۳۹۸، صص ۱۳۴-۱۰۹

Journal of Educational Measurement & Evaluation Studies
Vol. 9, No. 26, Summer 2019

سنجش صلاحیت‌های شناختی چندسطحی سواد ریاضی دانش‌آموزان پایه نهم: کاربردی از مدل پی‌جی‌دینا

مریم محسن‌پور*

تاریخ دریافت مقاله: ۹۶/۱۱/۳۰

تاریخ پذیرش مقاله: ۹۷/۰۵/۱۳

چکیده:

در چند دهه اخیر، چارچوب نظری جدیدی در زمینه اندازه‌گیری آموزشی با عنوان «سنجش شناختی تشخیصی» به وجود آمده که با پیوند دادن نظریه‌های شناختی با آموزش، بازخورد تشخیصی تکوینی از طریق گزارش نیمرخ تسلط آزمودنی در صلاحیت‌های شناختی مورد نیاز را برای پاسخگویی به سؤال‌ها ارائه می‌دهد. در این پژوهش با استفاده از یک مدل اندازه‌گیری جدید در سنجش شناختی تشخیصی به نام *پی‌جی‌دینا* (چن و دی‌لتوره، ۲۰۱۳) روی یک آزمون سواد ریاضی، به تعیین میزان اطلاعات تشخیصی که هر یک از سؤال‌های آزمون می‌توانند به دست دهند، اقدام شده است. طراحی آزمون سواد ریاضی بر مبنای چارچوب برنامه سنجش بین‌المللی دانش‌آموزان صورت گرفته است. صلاحیت‌های شناختی زیربنای این آزمون شامل ارتباطات، ریاضی‌وار کردن، بازنمایی، استدلال، طراحی راهبرد برای حل مسئله و استفاده از زبان و عملیات نمادین، رسمی و فنی، در قالب ماتریس کیو توسط تیم متخصص در حوزه آموزش ریاضی تدوین و داده‌ها با استفاده از اجرای آزمون روی ۷۰۰ دانش‌آموزان ۱۵ ساله به دست آمده است. شاخص‌های برازش مدل، نیمرخ کلاس‌های مکنون، پارامترهای سؤال‌ها و نیمرخ صلاحیتی آزمودنی‌ها بر مبنای مدل پی‌جی‌دینا، تحلیل شده‌اند. یافته‌ها نشان‌دهنده برازش مناسب این مدل با داده‌ها و ارائه اطلاعات تشخیصی مفید در سطح سؤال‌ها بر مبنای مدل مذکور است.

واژگان کلیدی: سنجش شناختی تشخیصی، مدل پی‌جی‌دینا، صلاحیت‌های چندسطحی، سواد ریاضی

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

*استادیار گروه روانشناسی تربیتی، دانشکده علوم تربیتی و روانشناسی، دانشگاه الزهرا، تهران، ایران (نویسنده مسئول: m.mohsenpour@alzahra.ac.ir)

مقدمه و بیان مسئله

در دهه‌های اخیر، نقش مهم نظریه‌های شناختی در اندازه‌گیری‌های آموزشی برای بسیاری از متخصصان آموزشی آشکار شده است؛ بر این اساس، پژوهشگران حوزه اندازه‌گیری آموزشی و روان‌شناسی شناختی، به دنبال طراحی آزمون‌هایی برای سنجش عملکرد دانش‌آموزان بوده‌اند تا با نتایج به دست آمده، بتوانند نقاط قوت و ضعف صلاحیت‌های شناختی دانش‌آموزان را استخراج کنند (لیتون، گیرل و هونکا[1]، ۲۰۰). از این‌رو، یکی از دیدگاه‌های نوین در حوزه اندازه‌گیری آموزشی، استفاده از سنجش‌های شناختی تشخیصی[2] جهت سنجش توانایی‌های آزمون‌شوندگان است، زیرا اطلاعات به دست آمده از این نوع سنجش‌ها که طراحی آنها بر مبنای نظریه‌های شناختی است، می‌تواند به وسیله کسانی که در آموزش حیطه‌های محتوایی مانند ریاضی صاحب‌نظر هستند، به‌طور معنی‌داری تفسیر شود. یک تعریف دقیق از سنجش شناختی تشخیصی در اندازه‌گیری آموزشی به معنای طراحی آزمونی برای سنجش و درک فرایندهای شناختی مستتر در پاسخ‌های دانش‌آموزان است (راپ، تمپلین و هنسون[3]، ۲۰۱۰). هدف رویکرد سنجش شناختی تشخیصی برخلاف *سنجش بازده‌های یادگیری*[4]، ارتقای *سنجش برای یادگیری و به‌عنوان فرایند یادگیری است*[5]. *این رویکرد* با فراهم کردن اطلاعات مورد نیاز برای اصلاح آموزش و یادگیری در کلاس درس و به وسیله معلم، به وجود آمده است (جانگ[6]، ۲۰۰۸). در سنجش شناختی تشخیصی، به جای لحاظ کردن یک پیوستار صلاحیت تک‌بعدی، چیرگی را به‌عنوان مجموعه‌ای از صلاحیت‌ها مجزا ولی مرتبط با هم می‌بینند. بنابراین اطلاعات به دست آمده برای آزمودنی‌ها در حد خُرد است (چن و دی‌لتوره[7]، ۲۰۱۳). به‌علاوه، در سنجش شناختی تشخیصی، اندازه خرد بودن یک صلاحیت، راه حلی[8] (تحلیلی)

---

1. Leighton, Gierl & Hunka
2. Cognitive Diagnostic Assessment
3. Rupp, Templin & Henson
4. Assessment of Learning Outcomes
5. Assessment Used for Learning and as Learning Process
6. Jang
7. Chen & de la Torre
8. Resolution

است که با آن پژوهشگر یک فرایند پاسخ شناختی را تشریح و مؤلفه‌های تشکیل‌دهنده آن را توصیف می‌کند. به‌عنوان مثال، در یک مطالعه شناختی، دی‌لتوره و داگلاس[۱] (۲۰۰٤) روی سازه *تفریق کسرها* کار کرده‌اند. در مطالعه دیگر کونینا، راپ و ویلهلم[۲] (۲۰۰۸) روی مهارت‌های پایه حساب سنجش شناختی انجام داده‌اند. در مطالعه اول تکلیف‌های تفریق کسر در مقایسه با مطالعه دوم که انجام عملیات حسابی صورت می‌گیرد، محدودتر هستند. بنابراین هرچه قلمرو تکالیف مورد تحلیل افزایش یابد، تعریف‌های اندازه خرد، گسترده‌تر می‌شوند (راپ و همکاران، ۲۰۱۰).

نکته مهم درباره صلاحیت‌ها این است که نوع تعامل صلاحیت‌های شناختی با هم (جمع یا ضرب)، مدل‌های شناختی تشخیصی را به دو دسته‌بندی مدل‌های جبرانی[۳] و غیر جبرانی[٤] تقسیم کرده است. در مدل‌های جبرانی، نقص در یک صلاحیت از طریق قوت در صلاحیت دیگر جبران می‌شود، در حالی که در مدل‌های غیر جبرانی لازم است تا همه صلاحیت‌ها وجود داشته باشند تا بتوان به سؤال، پاسخ درست داد. هنگامی‌که یک مدل آماری از رابطه جمع استفاده می‌کند، صلاحیت‌ها به یک شیوه جبرانی با هم ترکیب می‌شوند و هرگاه مدل از رابطه ضرب استفاده کند، صلاحیت‌ها به روشی غیر جبرانی با هم ترکیب می‌شوند. همچنین، صلاحیت‌های شناختی می‌توانند ماهیت دوارزشی[٥] یا چندارزشی[٦] داشته باشند[۷] (کارلیز[۸]، ۲۰۰٤ و ون داویر[۹]، ۲۰۰۵، به نقل از راپ و همکاران، ۲۰۱۰). در مدل‌های شناختی تشخیصی، تعداد سطوح صلاحیت‌های شناختی و تعریف هر سطح پیش از فرایند برازش داده‌ها، توسط متخصصان حوزه محتوایی، تعیین و ساختار صلاحیت‌های مورد نیاز برای هر سؤال در نیمرخ صلاحیتی مشخص می‌شوند. این نیمرخ صلاحیت‌ها در غالب ماتریس کیو ارائه می‌شوند؛ بنابراین ماتریس کیو، تعیین‌کننده کیفیت اطلاعات

---

1. De la Torre & Douglas
2. Kunina, Rupp & Wilhelm
3. Compensatory
4. Non Compensatory
5. Dichotomous
6. Polytomous
۷. در این نوشتار واژه‌های دوارزشی و چندارزشی به ترتیب معادل دوسطحی و چندسطحی به کار برده شده‌اند.
8. Karelitz
9. Von Davier

تشخیصی از ابزار سنجش است. هر سطر این ماتریس، یک فرضیه درباره صلاحیت‌های مورد نیاز برای کسب پاسخ درست به یک سؤال و دارای مرتبه $n \times k$ است که در آن n، تعداد سؤال‌ها و k تعداد صلاحیت‌های مورد نیاز برای سؤال‌ها است (راپ و همکاران، ۲۰۱۰). در حال حاضر، مدل‌های شناختی تشخیصی متنوعی برای سنجش صلاحیت‌های دوارزشی شامل مدل *دینا* (هارتل[1]، ۱۹۸۹؛ جانکر و سیجسما[2]، ۲۰۰۱)، مدل *دینودینو* (ین و هنسون، ۲۰۰۶)، مدل لجستیک خطی (ماریس[3]، ۱۹۹۹)، مدل *جی‌دینا* (دی لتوره، ۲۰۱۱)، مدل شناختی تشخیصی لگاریتم خطی (هنسون، تمپلین و ویلز[4]، ۲۰۰۹) وجود دارد، در حالی که تعداد کمی از آنها برای تحلیل صلاحیت‌های چندارزشی مانند *جی‌دی‌ام* (هابرمن، ون‌داویر و لی[5]، ۲۰۰۸ و ون‌داویر، ۲۰۰۵) و کدگذاری صلاحیت طبقه‌ای مرتب شده (کارلیتز، ۲۰۰٤) قابلیت کاربرد دارند (به نقل از چن و دی‌لتوره[6]، ۲۰۱۳).

مدل *پی‌جی‌دینا*[7] از مدل‌هایی است که برای تحلیل ساختار صلاحیت‌های چندارزشی تدوین شده و توسط متخصصان کاربرد دارد[8]. این مدل بر مبنای چارچوب کدگذاری صلاحیت طبقه‌ای مرتب شده[9] برای صلاحیت‌های چندارزشی است که شامل تنوعی از مدل‌های سنجش شناختی تشخیصی است. این مدل را نخستین بار چن و دی‌لتوره (۲۰۱۳) معرفی کردند و مبنای ساختاری آن برگرفته از مدل *جی‌دینا* است که در آن قابلیت انعطاف در زمینه روابط جبرانی و غیر جبرانی در بین صلاحیت‌های موجود در یک آزمون وجود دارد (دی‌لتوره، ۲۰۱۱).

مبانی نظری مدل *پی‌جی‌دینا*

---

1. Haertel
2. Junker & Sijtsma
3. Maris
4. Willse
5. Haberman, Von Davier & Lee
6. Chen & de la Torre
7. Polytomous Generalized Deterministic Inputs, Noisy, and Gate (PG-DINA)

۸. در این روش، صلاحیت‌ها و ساختار سطوح آنها پیش از فرایند برازش داده‌ها، توسط متخصصان محتوایی تدوین می‌شود.

9. Ordered-Category Attribute Coding

به‌طور کلی در مدل‌های آماری شناختی تشخیصی، پاسخگویی درست به سؤالی که نیازمند یک صلاحیت شناختی M سطحی است، می‌تواند آزمودنی‌ها را به M گروه تقسیم کند به‌گونه‌ای که هر گروه احتمال موفقیت مختص به خود را دارد. واضح است که در چنین شرایطی یک مدل شناختی تشخیصی اشباع شده[1] دارای پارامترهای زیادی است و با افزایش تعداد صلاحیت‌های سؤال، پیچیدگی مدل افزایش می‌یابد. به‌عنوان نمونه، سؤال‌هایی که نیازمند یک، دو و سه صلاحیت M سطحی هستند، تعداد پارامترهای هر سؤال به ترتیب به M، M^2 و M^3 می‌رسد. برای مدل‌های مبتنی بر صلاحیت‌های چندسطحی، به‌منظور مدیریت بهتر، فرض می‌شود هر سؤال بین دو نوع آزمودنی (دو گروه مکنون) تفاوت قائل شود: آزمودنی‌هایی که رو یا بالاتر از سطح ویژه صلاحیت هستند و آزمودنی‌هایی که پایین‌تر از این سطح هستند. آزمودنی‌هایی که رو یا بالای سطح خاص از صلاحیت هستند، فرض می‌شود که سطح تسلط مورد نیاز برای پاسخگویی به سؤال را دارند و برای افرادی که پایین سطح خاص هستند، چنین نیست. سؤال‌ها با چنین مفروضه‌ای به‌عنوان *سؤالات تسلط سطح ویژه صلاحیت*[2] تلقی می‌شوند. این کار اجازه می‌دهد سطوح متفاوت صلاحیت‌ها قابل اندازه‌گیری باشند. در ماتریس تعیین شده به وسیله سؤال‌ها (چنانچه فرض کنید q_{jk} درایه سطر j ام و ستون k ام از یک ماتریس کیو J*K شد)، q_{jk} می‌تواند مقادیر 0 تا M_k-1 را بگیرد (جایی که M_k تعداد سطوح صلاحیت kام است). اگر تعداد سطوح همه صلاحیت‌ها یکسان باشد، در این صورت $M_K=M$. همچنین، اگر برای پاسخگویی درست به سؤال j ام، نیاز به تسلط بر سطح m یا بالاتر در صلاحیت kام باشد، $q_{jk} = m$. بنابراین هرگاه در یک سؤال تسلط سطح ویژه صلاحیت، مقدار $q_{jk} = m$ باشد، تفاوتی بین سطوح تسلط m تا M-1 نیست و به‌طور مشابه تمایزی بین سطوح تسلط 0 تا m-1 نیست. نکته دیگر اینکه اگر سؤال j به هیچ سطحی از صلاحیت k ام نیاز نداشته باشد $(q_{jk} = 0)$، یعنی صلاحیت مورد نظر برای سؤال مذکور نیاز نیست (چن و دی‌لتوره، ۲۰۱۳).

در مدل اندازه‌گیری *پی‌جی‌دینا*، برای نشان دادن تعداد صلاحیت‌های مورد نیاز برای سؤال j ام از معادله زیر استفاده می‌شود که در آن $I(\cdot)$ تابع نشانگر است:

---

[1]. مدل اشباع شده مدلی است که بیشترین تعداد پارامتر برآورد شده را دارا است.

[2]. Specific Attribute Level Mastery (SALM) Items

$$k_j^* = \sum_{k=1}^{k} I(q_{jk} > 0)$$

فرض بر این است که k_j^* صلاحیت اول، صلاحیت‌های مورد نیاز برای سؤال j ام هستند. به‌علاوه، صلاحیت‌های مورد نیاز برای سؤال jام می‌تواند به وسیله بردار کاهش یافته $\alpha_{lj}^* = \left(\alpha_{l1}. \ldots . \alpha_{lk_j^*}\right)'$ نشان داده شود که در آن $l = 1. \ldots . M^{k_j^*}$ و $M^{k_j^*}$ نشان‌دهنده تعداد الگوهای صلاحیت‌های منحصربه‌فرد است. گفتنی است هر سطح صلاحیت کاهش یافته یا نیافته، می‌تواند مقادیر 0 تا m-1 را بگیرد. به این ترتیب، تعداد بردارهای صلاحیت‌هایی که باید برای سؤال j ام در نظر گرفته شوند از M^k به $M^{k_j^*}$ کاهش می‌یابند. با استفاده از *سؤال‌های تسلط سطح ویژه صلاحیت*، که تنها بی دو گروه مکنون برای هر صلاحیت مربوطه تمایز قائل می‌شوند، فرمول مدل شناختی تشخیصی می‌تواند ساده‌تر شود. با مشخص کردن $q_{jk} = \mathrm{m}$ بری سؤال jام، صلاحیت M سطحی α_{lk} می‌تواند به یک صلاحیت دوارزشی α_{lk}^{**} کاهش یابد که در آن:

$$\alpha_{lk}^{**} = \begin{cases} 0 & if \quad \alpha_{lk} < q_{jk} \\ 1 & otherwise \end{cases}$$

چن و دی لتوره (۲۰۱۳)، $\alpha_{lj}^{**} = \left(\alpha_{l1}^{**}. \ldots . \alpha_{lk_j^*}^{**}\right)'$ را به‌عنوان بردار صلاحیت کاهش یافته[1] نشان می‌دهند که در آن $l = 1. \ldots . 2^{k_j^*}$ است. با ساده‌سازی بیشتر، آنها پارامترها را در فضای α_{lj}^{**} برآورد می‌کنند. بنابراین برای سؤالی که به k_j^* تا صلاحیت نیاز دارد، تعداد الگوهای منحصربه‌فرد α_{lj}^{**} به $2^{k_j^*}$ کاهش می‌یابد. به‌عنوان نمونه، در مدل *پی‌جی‌دینا*، احتمال اینکه آزمودنی با بردار صلاحیت کاهش یافته α_{lj}^{**} پاسخ صحیحی به سؤال j ام دهد، به‌صورت زیر نشان داده می‌شود:

$$P(X_i = 1|\alpha_{li}^{**}) = P(\alpha_{li}^{**})$$

[1]. Collapsed Attribute Vector

فرم اشباع شده مدل *پی‌جی‌دینا* می‌تواند با استفاده از تابع سؤال پاسخ زیر بر اساس توزیع یکنواخت بیان شود:

$$P(\alpha_{lk}^{**}) = \delta_{j0} + \sum_{k=1}^{k_j^*} \delta_{jk}\alpha_{lk}^{**} + \sum_{\acute{k}>k}^{k_j^*}\sum_{k=1}^{k_j^*} \delta_{jk\acute{k}}\,\alpha_{lk}^{**}\alpha_{l\acute{k}}^{**} + \cdots + \delta_{j1\ldots..k_j^*}\prod_{k=1}^{k_j^*}\alpha_{lk_j^*}^{**}$$

در آن δ_{j0} عرض از مبدأ سؤال j، δ_{jk} اثر اصلی، α_{lk}^{**}، $\delta_{jk\acute{k}}$ اثر تعامل ناشی از α_{lk}^{**} و $\alpha_{l\acute{k}}^{**}$ و $\delta_{j1\ldots..k_j^*}$ اثر تعامل ناشی از $\alpha_{l1}^{**} \ldots . \alpha_{lk^*}^{**}$ است.

مرور پیشینه سنجش شناختی تشخیصی، نشان‌دهنده مطالعات اندک در حوزه صلاحیت‌های چندارزشی و کاربرد مدل *پی‌جی‌دینا* است. تنها در مطالعه چن و دی‌لتوره (۲۰۱۳) با عنوان «یک مدل شناختی تشخیصی کلی برای صلاحیت‌های چندارزشی تعیین شده توسط متخصصان» است که با معرفی مدل *پی‌جی‌دینا*، یک آزمون ۱۵ سؤالی چندگزینه‌ای مبتنی بر چهار صلاحیت (شامل دو صلاحیت دوسطحی و دو صلاحیت سه‌سطحی) طراحی و روی ۳۹۳ دانش‌آموز اجرا شده است. سپس با استفاده از مدل *پی‌جی‌دینا* و *جی‌دینا*، پارامترهای برآورد شده، مقایسه شد و با ایجاد دو ماتریس ساختاری دوارزشی و چندارزشی از صلاحیت‌های شناختی نشان داده شد که مدل *پی‌جی‌دینا* برازش بهتری در مقایسه با مدل *جی‌دینا* دارد.

با توجه به مطالب فوق، اگرچه ساختار مدل‌های شناختی چندسطحی نظیر *پی‌جی‌دینا*، پیچیده است، اما در صورت تدوین آزمون‌های شناختی و تحلیل نتایج آنها بر مبنای چنین مدل‌هایی می‌توان اطلاعات تشخیصی مفیدی از اشتباهات پردازشی آزمودنی‌ها به دست آورد و سپس به طراحی برنامه‌های آموزشی متناسب با صلاحیت‌های شناختی آنها اقدام کرد. یکی از مطالعات بین‌المللی که در راستای طراحی آزمون‌های شناختی چندسطحی در حوزه محتوایی ریاضی گام برمی‌دارد، «برنامه بین‌المللی سنجش دانش‌آموزان»[1] (پیزا) است. در پیزا، تأکید بر سنجش توانایی

[1] Program for International Students Assessment (PISA)

فرد در شناسایی و درک نقشی که ریاضیات در این دنیا ایفا می‌کند، با عنوان سواد ریاضی[۱] صورت می‌گیرد (سازمان همکاری و توسعه اقتصادی، ۲۰۰۹). یک دهه تجربه در تدوین سؤال‌های آزمون سواد ریاضی در پیزا و تحلیل روش‌های پاسخ‌گویی دانش‌آموزان به آنها، نشان داده که یک مجموعه صلاحیت‌های ریاضی[۲] بنیادی وجود دارد که سواد ریاضی را در عمل، حمایت می‌کند. نیس و همکاران دانمارکی او (نیس و جنسن[۳]، ۲۰۰۲؛ نیس، ۲۰۰۳؛ نیس و هوجگارد[۴]، ۲۰۱۱)، هشت صلاحیت ریاضی را شناسایی کرده‌اند که می‌توان آنها را برای شناخت رفتار ریاضی فرد، به کار گرفت (به نقل از سازمان همکاری و توسعه اقتصادی، ۲۰۱۳). در چارچوب مطالعه پیزا (۲۰۱۲)، صلاحیت‌ها با نام‌های الف- *ارتباطات*؛ ب- *ریاضی‌وار کردن*؛ ج- *بازنمایی*؛ د- *استدلال و بحث* ه- *طراحی راهبرد برای حل مسائل*؛ و- *استفاده از زبان و عملیات نمادین رسمی و فنی*؛ ز- *استفاده از ابزار و وسایل ریاضی*، معرفی شده‌اند[۵] (استیسی[۶]، ۲۰۱۲). از آنجا که تاکنون کشور ایران در مطالعات پیزا شرکت نکرده، طراحی آزمونی بر مبنای این چارچوب و سنجش سطوح صلاحیت‌های شناختی دانش‌آموزان در حوزه سواد ریاضی می‌تواند گامی مؤثر در تشخیص نقاط قوت و ضعف آزمون‌شوندگان داشته باشد. بنابراین، در پژوهش حاضر قصد بر این است که بر مبنای مدل *پی‌جی‌دینا*، نتایج آزمونی که برای سنجش سواد ریاضی دانش آموزان ایرانی ۱۵ ساله مبتنی بر چارچوب پیزا تهیه شده بود[۷]، تحلیل شود تا معلوم شود که

---

[1]. Mathematical Literacy

[۲].صلاحیت ریاضی؛ شامل دانش درک کردن، انجام دادن، به‌کاربردن و نیز داشتن یک طرز تلقی درباره ریاضی و فعالیت‌های ریاضی، در زمینه‌های مختلفی است که ریاضی می‌تواند در آنها، نقشی داشته باشد (نیس و هوجگارد، ۲۰۱۱).

[3]. Niss & Jensen

[4]. Niss & Højgaard

[۵]. ترنر؛ نیس و موگان (۲۰۱۵) در پروژه‌ای به تعریف صلاحیت‌ها و تعیین سطح هر صلاحیت پرداخته‌اند، که تعریف و سطوح صلاحیت‌های شناختی در پیوست مقاله ارائه شده است.

[6]. Stacey

[۷]. برای اطلاع بیشتر در خصوص نحوه شل‌گیری آزمون و محتوای آن به منبع محسن‌پور، گویا، شکوهی یکتا، کیامنش و بازرگان (۱۳۹۳) رجوع شود.

چه اندازه سؤال‌های آزمون شناختی سواد ریاضی می‌تواند اطلاعات مفید تشخیصی بر مبنای مدل *پی‌جی‌دینا* برای تفسیر نتایج آزمون فراهم کند.

روش پژوهش

داده‌ها

در این پژوهش از آزمون سواد ریاضی مبتنی بر چارچوب نظری مطالعات پیزا و متشکل از ۲۰ سؤال بر اساس سازگار کردن محتوای سؤال‌های آزمون با توانایی دانش‌آموزان ایرانی و ارزش‌های فرهنگی جامعه ایرانی استفاده شد. جامعه آماری، دانش‌آموزان پایه دهم شهر تهران در سال تحصیلی ۱۳۹۳-۱۳۹۲ در سه رشته «ریاضی-فیزیک»، «علوم تجربی» و «ادبیات و علوم انسانی» از شاخه نظری بودند. انتخاب پایه دهم به دو دلیل صورت گرفت؛ نخست آنکه محتوای آزمون سواد ریاضی پیزا، بر اساس سال خروج از دوره آموزش عمومی صورت گرفته است که به‌نوعی، معادل پایان پایه نهم در ایران است. دوم اینکه در ایران، طیف سنی دانش‌آموزان در زمان آزمون (شروع پایه دهم)، بین ۱۵/۲ سال تا ۱۶/۲ سال است که با جمعیت هدف مطالعات پیزا همخوانی داشت. رویکرد نمونه‌گیری در این پژوهش، رویکرد «احتمال متناسب با بزرگی[1]» بوده که با استفاده از این رویکرد، حجم احتمالی نمونه، ۷۰۰ نفر به دست آمد. آزمون طراحی شده روی افراد مذکور، اجرا شد و یک نمونه ۶۸۸ نفری به سؤال‌های آزمون پاسخ دادند که پس از حذف ۳۱ برگه مخدوش (سفید یا نامربوط)، تعداد نهایی به ۶۵۷ نفر رسید. داده‌ها با استفاده از نرم‌افزار آر[2] و پکیج سی‌دی‌ام نسخه ۳-۳-۳ (روبیتز، کیفر، جوروج و آنلو[3]، ۲۰۱۴) بر مبنای ساختار ماتریس کیو مذکور تحلیل شدند. برای برآورد پارامترها، بیشینه احتمال حاشیه‌ای با استفاده از الگوریتم بیشینه سازی-انتظار استفاده شده است.

تدوین ماتریس کیو

همان‌طور که پیش‌تر بحث شد، با شش صلاحیت شناختی، زیربنای سواد ریاضی شناسایی شد. بررسی ماتریس کیوی شش‌صلاحیتی نشان داد که دو صلاحیت

---

1. Probability Proportional to Size (PPS)
2. R Software
3. Robitzsch, Kiefer, George & Uenlue

ارتباطات و *استفاده از زبان و عملیات نمادین، رسمی و فنی* در بیشتر سؤال‌ها با هم ظاهر می‌شوند. همچنین، بررسی همبستگی صلاحیت‌ها نشان داد، بین ضرایب همبستگی پلی کوریک - همبستگی بین دو صلاحیت مذکور، بیشترین ضریب همبستگی وجود داشته است. از سویی، در ادبیات مربوط به مدل شناختی به کار برده شده در این پژوهش، بنا به پیشنهاد نیس و هوجگارد (۲۰۱۱)، دو صلاحیت *ارتباطات* و *استفاده از زبان و عملیات نمادین، رسمی و فنی*، در یک گروه مشترک قرار گرفتند. از این گذشته، پیشینه صلاحیت‌های شناختی نشان می‌داد که برای به‌کارگیری صلاحیت *ریاضی‌وار کردن*، استفاده از صلاحیت *بازنمایی* معمول است (نیس و هوجگارد، ۲۰۱۱). پس صلاحیت‌های *ارتباطات* و *استفاده از زبان و عملیات نمادین، رسمی و فنی* با هم و دو صلاحیت *ریاضی‌وار* کردن و *بازنمایی*، نیز با هم ترکیب شدند. در آخرین مرحله، ساختار بارگیری جدید در ماتریس کیو با چهار صلاحیت *ارتباطات* و *استفاده از زبان و عملیات نمادین، رسمی و فنی؛ ریاضی‌وار* کردن و *بازنمایی؛ استدلال و بحث و طراحی راهبرد برای حل مسئله* از طریق یک جلسه هم‌اندیشی انجام شد. با توجه به اینکه هدف پژوهش حاضر بررسی مدل *پی‌جی‌دینا* در ساختار آزمون شناختی سواد ریاضی بود، باید ساختار صلاحیت‌های شناختی مورد نیاز برای هر یک از سؤال‌های آزمون مشخص می‌شد، بنابراین برای رسیدن به توافق در زمینه سطوح صلاحیت‌های شناختی، یک جلسه هم‌اندیشی با هدایت یک متخصص آموزش ریاضی و چهار نفر شامل دو دانشجوی دکتری آموزش ریاضی که از دبیران ریاضی باتجربه در دوره متوسطه بودند، یک دبیر باتجربه ریاضی و یک دانشجوی دکتری سنجش آموزش، تشکیل شد. نخست، تک‌تک سؤال‌ها بررسی شد و انتساب صلاحیت‌های شناختی به هر سؤال انجام گرفت. در ادامه جلسه، درباره سطوح صلاحیت‌های هر سؤال آن‌قدر، بحث می‌شد تا جمع، درباره درجه‌بندی آن سؤال به اجماع می‌رسید. در جدول (۱) ماتریس کیو نهایی با چهار صلاحیت چهارسطحی گزارش شده است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

جدول (۱) ماتریس کیو چندسطحی حاصل از جلسه هم‌اندیشی

صلاحیت / سؤال	ارتباطات و استفاده از زبان و عملیات...	ریاضی‌وار کردن و بازنمایی	استدلال	طراحی راهبرد
۱	۱	۰	۰	۰
۲	۱	۱	۰	۰
۳	۲	۲	۰	۱
٤	۲	۲	۱	۱
٥	۳	۲	۱	۳
٦	۲	۱	۰	۰
۷	۳	۲	۰	۲
۸	۲	۱	۱	۱
۹	۰	۲	۱	۰
۱۰	۳	۲	۲	۳
۱۱	۱	۰	۰	۱
۱۲	۱	۰	۰	۰
۱۳	۲	۰	۰	۰
۱٤	۲	۲	۳	۲
۱٥	۲	۳	۲	۲
۱٦	۲	۰	۱	۰
۱۷	۲	۰	۱	۱
۱۸	۲	۲	۲	۲
۱۹	۳	۱	۱	۰
۲۰	۲	۳	۳	۲

بر اساس جدول، سه سؤال نیازمند یک صلاحیت شناختی، پنج سؤال نیازمند دو صلاحیت شناختی، چهار سؤال نیازمند سه صلاحیت شناختی و هشت سؤال نیازمند چهار صلاحیت شناختی هستند.

یافته‌ها

شاخص‌های برازش مدل

به‌طور کلی، اگر داده‌ها با مدل برازش مناسبی نداشته باشند، نتایج یک مدل سنجش شناختی تشخیصی معنی‌دار نیست. برازش مدل از دو راه امکان‌پذیر است: ۱- مقایسه شاخص‌های برازش مدل با داده‌ها (برازش مطلق)؛ ۲- مقایسه مدل با مدل‌های رقیب (برازش نسبی) (مایدیو[1]، ۲۰۱۳). در این پژوهش شاخص‌های مطلق *مدل پی‌جی‌دینا* با مقایسه آماره‌های پیش‌بینی شده و مشاهده شده در جدول (۲) ارائه شده است.

جدول (۲) آماره‌های برازش مدل *پی‌جی‌دینا*

شاخص برازش	برآوردها	معنی‌داری
maxX2	۸/۷٤۱	۰/۵۹۱
MADcor	۰/۰۳۰	-
۱۰۰*MADRESIDCOV	۰/۵۳۹	-
MADQ3	۰/۰۳۷	-
RMSEA	۰/۰٦٦	-
SRMSR	۰/۰۳۹	-

maxX2 آزمون برازش کلی مدل بر اساس متوسط تفاوت بین فراوانی‌های پاسخ مدل مشاهده شده و مدل پیش‌بینی شده است. چنانچه تفاوت بزرگ باشد نشان از وابستگی بین سؤال‌هاست، از آنجا که صلاحیت‌های شناختی زیربنایی یکسانی برای پاسخگویی به همه سؤال‌ها وجود دارد، این همبستگی دور از انتظار نیست. اما چنانچه مدل مورد نظر با داده‌ها برازش خوبی داشته باشد، انتظار می‌رود آماره آزمون X2 صفر باشد (راپ و همکاران، ۲۰۱۲، ۲٦۹). در این پژوهش عدم معناداری این شاخص حاکی از برازش مناسب مدل با داده‌ها است.

[1]. Maydeu

MADcor آماره تفاوت مطلق میانگین برای همبستگی‌های زوج-سؤال‌هاست که نشان‌دهنده تفاوت میان همبستگی‌های هر زوج از سؤال‌ها در مدل پیش‌بینی شده و مشاهده شده است. به‌طور کلی قاعده مشخصی برای اغلب شاخص‌های برازش مدل وجود ندارد. اما هرچقدر این شاخص‌ها به صفر نزدیک‌تر باشند، برازش مدل بهتر است. در این پژوهش این شاخص ۰/۰۳ به دست آمده که طبق نظر دی‌بلو، روسوس و استوات[1] (۲۰۰۷) حاکی از برازش مناسب است.

MADRESIDCOV میانگین کواریانس باقی‌مانده است که تفاوت میانگین بین ماتریس‌های همبستگی سؤال‌های *مشاهده* شده و دوباره تولید شده است. در این پژوهش مقدار آن ۵۳٪ و مناسب است.

MADQ3 از طریق تفاضل پاسخ‌های دو مدل پیش‌بینی شده و مشاهده شده در آزمودنی‌ها و محاسبه متوسط همبستگی زوجی این باقی‌مانده‌ها است. در داده‌های این پژوهش مقدار آن ۰/۰۳۷ به دست آمده که طبق نظر دی‌بلو و همکاران (۲۰۰۷) مناسب است.

RMSEA ریشه مجذور میانگین خطای برآورد برای پارامترهای سؤال است. در داده‌های این پژوهش مقدار آن ۰/۰٦٦ و مناسب است.

به‌طور کلی، نتایج تحلیل داده‌ها نشان‌دهنده آن بود که صلاحیت *استدلال* که ٤۰٪ آزمودنی‌ها به آن تسلط یافته بودند، ساده‌ترین صلاحیت بود. پس از آن *ریاضی‌وار کردن و بازنمایی* (۳۸٪)، *ارتباطات و استفاده از زبان و عملیات نمادین* (۳۳٪) و *طراحی راهبرد* (۱۸٪) قرار داشت[2]. همچنین با توجه به ساختار چهارسطحی بودن چهار صلاحیت شناختی در مدل *پی‌جی‌دینا*، تعداد کل کلاس‌های مکنون ۲٥٦ کلاس بود و امکان نمایش نیمرخ‌های صلاحیت‌های ممکن[3] برای هر کلاس مکنون در این

---

[1]. Di Bello, Roussos & Stout

[2]. در این بخش به‌منظور خلاصه‌نویسی گاهی به جای ذکر نام چهار صلاحیت ارتباطات و استفاده از زبان و عملیات نمادین؛ ریاضی‌وار کردن و بازنمایی؛ استدلال و طراحی راهبرد، به ترتیب از اسامی صلاحیت‌های اول تا چهارم استفاده می‌شود.

[3]. با توجه به اینکه زیربنای پاسخگویی به سؤالات نیازمند چهار صلاحیت شناختی است و هر صلاحیت دارای ساختار چهارسطحی (۰ تا ٤) است، تمام کلاس‌های مکنون ممکن برای تشکیل نیمرخ صلاحیت (٤*٤*٤*٤) یعنی ۲٥٦ مورد است.

نوشتار میسر نبود؛ بنابراین تنها اطلاعات مربوط به چهار کلاس مکنون اول و آخر در جدول (۳) ارائه شده است.

جدول (۳) احتمالات کلاسی

کلاس مکنون	نیمرخ صلاحیت	احتمال کلاسی	فراوانی مورد انتظار در هر کلاس
۱	۰۰۰۰	۰/۰۸٤	٥٤/۹۷
۲	۰۰۰۱	۰/۰۱۷	۱۱/۱٥
۳	۰۰۰۲	۰/۰۰۷۷	٥/۰۷
٤	۰۰۰۳	۰/۰۰۷۸	٥/۱۷
–	–	–	
۲٥۳	۳۳۳۰	۰/۰۰٤۳	۲/۸۳
۲٥٤	۳۳۳۱	۰/۰۰٤۳	۲/۸٥
۲٥٥	۳۳۳۲	۰/۰۰۹۷	٦/٤
۲٥٦	۳۳۳۳	۰/۰٤۹	۳۲/۳۹

همان‌گونه که در جدول ۳ دیده می‌شود نیمرخ صلاحیت اول بیشترین احتمال کلاسی را دارد؛ یعنی حدود ۸٪ از کل افراد مورد مطالعه (حدود ٥٥ نفر) به کلاس (۰۰۰۰) تعلق دارند پس انتظار می‌رود ۸٪ از کل افراد به هیچ‌یک از چهار صلاحیت شناختی تسلط ندارند. پس از آن نیمرخ صلاحیت آخرین کلاس، بیشترین احتمال کلاسی را دارد؛ به‌طوری که انتظار می‌رود حدود ٥٪ از کل افراد به همه چهار صلاحیت شناختی تسلط یافته‌اند. در ادامه به پارامترهای برآورد سؤال‌ها اشاره می‌شود. چون ارائه تعداد کل پارامترها از حوصله این نوشتار خارج است، تنها به پارامترهای دو سؤال اول آزمون در جدول (٤) اشاره شده است.

پرتال جامع علوم انسانی

جدول (٤) پارامترهای مدل *پی‌جی‌دینا*

شماره سؤال	صلاحیت‌های مورد نیاز	الگوی تسلط صلاحیت	احتمال موفقیت در سؤال
۱	۰۰۰۱	۰۰۰۰	۰/۹۲
۱	۰۰۰۱	۰۰۰۱	۰/۹۷
۲	۰۰۱۱	۰۰۰۰	۰/۳۸
۲	۰۰۱۱	۰۰۰۱	۰/۸۷
۲	۰۰۱۱	۰۰۱۰	۰/۸۱
۲	۰۰۱۱	۰۰۱۱	۰/۹۳

به‌طور کلی تعداد پارامترهای برآورد شده در هر سؤال تابع تعداد صلاحیت‌های مورد نیاز[1] برای هر سؤال است. با توجه به اینکه مدل *پی‌جی‌دینا* یک مدل شناختی تشخیصی اشباع شده است، تمام اثرات اصلی و اثرات تعاملی برای این مدل برآورد می‌شود. به‌عنوان نمونه، سؤال‌هایی که نیازمند یک صلاحیت برای پاسخگویی هستند (مانند سؤال ۱، ۱۲ و ۱۳)، دارای یک عرض از مبدأ و یک اثر اصلی هستند. در سؤال‌هایی که برای پاسخگویی نیازمند دو صلاحیت هستند (مانند سؤال ۲، ٦، ۹، ۱۱ و ١٦)، چهار پارامتر برآورد می‌شود که شامل یک عرض از مبدأ، دو اثر اصلی و یک اثر تعاملی است. همچنین در سؤال‌های سه صلاحیتی (مانند سؤال ۳، ۷، ۱۷ و ۱۹)، هشت پارامتر برآورد می‌شود که شامل یک عرض از مبدأ، سه اثر اصلی و چهار اثر تعاملی است. به‌علاوه، در سؤال‌های نیازمند چهار صلاحیت (مانند (٤، ٥، ۸، ۱۰، ١٤، ۱۵، ۱۸ و ۲۰) شانزده پارامتر برآورد می‌شود که شامل یک عرض از مبدأ، چهار اثر اصلی و یازده اثر تعاملی است. پارامتر عرض از مبدأ، احتمال پاسخگویی درست به سؤال را هنگام عدم تسلط بر هیچ‌یک از صلاحیت‌های مورد نیاز برای سؤال را نشان می‌دهد. پارامتر اثر اصلی، افزایش در احتمال پاسخگویی درست به سؤال برای هر

[1]. در هر سؤال متناسب با ساختار ماتریس کیو، صلاحیت‌های مورد نیاز برای هر سؤال مشخص می‌شود. مثلاً در سؤال یک فقط نیاز به یک صلاحیت ارتباطات و استفاده از زبان و عملیات نمادین است بنابراین الگوی (۰۰۰۱) معرف این موضوع است.

یک از صلاحیت‌های مورد نیاز سؤال را ارائه می‌کند و درنهایت پارامتر اثر تعاملی، نشان‌دهنده افزایش در احتمال پاسخگویی درست به سؤال هنگام تسلط بر یک ترکیب از صلاحیت‌های مورد نیاز سؤال است. بر اساس جدول (٤)، در مورد سؤال یک، ٪۹۲ احتمال پاسخگویی درست به این سؤال در صورت عدم تسلط به هیچ‌یک از صلاحیت‌های مورد نیاز برای این سؤال، وجود دارد (به‌طور شانسی پاسخ صحیح دادن). احتمال پاسخگویی درست به سؤال یک برای افراد مسلط به صلاحیت *ارتباطات و استفاده از زبان و عملیات نمادین* ۰/۹۷ بیشتر از افرادی است که به این صلاحیت تسلط نیافته‌اند. بنابراین تسلط داشتن بر صلاحیت *ارتباطات و استفاده از زبان و عملیات نمادین* با احتمال بسیار قوی پاسخگویی درست به سؤال همراه است. در مورد سؤال دوم، ٪۳۸ احتمال پاسخگویی درست به این سؤال در صورت عدم تسلط به هیچ‌یک از صلاحیت‌های مورد نیاز برای این سؤال، وجود دارد (به‌طور شانسی پاسخ صحیح دادن). احتمال پاسخگویی درست به سؤال دوم برای افراد مسلط به صلاحیت *ارتباطات و استفاده از زبان و عملیات نمادین* ۰/۸۷ بیشتر از افرادی است که به این صلاحیت تسلط نیافته‌اند. همچنین احتمال پاسخگویی درست به سؤال دوم برای افراد مسلط به صلاحیت *ریاضی‌وار کردن و بازنمایی* ۰/۸۱ بیشتر از افرادی است که به این صلاحیت تسلط نیافته‌اند. بنابراین تسلط بر *ارتباطات و استفاده از زبان و عملیات نمادین*، احتمال پاسخگویی بیشتری را در مقایسه با تسلط بر *ریاضی‌وار کردن و بازنمایی* (۰/۸۷ در مقابل ۰/۸۱) افزایش می‌دهد. همچنین احتمال پاسخگویی درست به سؤال دوم برای افراد مسلط به دو صلاحیت *ارتباطات و استفاده از زبان و عملیات نمادین* و *ریاضی‌وار کردن و بازنمایی* ۰/۹۳ بیشتر از افرادی است که به این دو صلاحیت تسلط نیافته‌اند.

از دیگر نتایج قابل توجه در خروجی‌های مدل *پی‌جی‌دینا* احتمال تسلط هر آزمودنی به هر یک از صلاحیت‌های شناختی مورد نیاز در یک آزمون با توجه به الگوی پاسخ آزمودنی[۱]، و نیمرخ صلاحیتی محتمل هر آزمودنی است. جدول (۵)،

---

[۱]. الگوی پاسخ، تصویری از نحوه پاسخگویی فرد به سؤال‌های آزمون است. در این آزمون ۲۰ سؤالی، کد ۱ یعنی فرد به سؤال پاسخ صحیح داده، کد ۰ یعنی فرد به سؤال پاسخ نادرست داده و کد ۹ یعنی فرد به سؤال پاسخ نداده است.

احتمال تعلق پنج آزمودنی را برای چهار صلاحیت شناختی مورد نیاز در آزمون نشان می‌دهد.

جدول (۵) احتمالات تسلط بر صلاحیت

کد افراد	الگوی پاسخ	نیمرخ صلاحیت	احتمال	صلاحیت ۱	صلاحیت ۲	صلاحیت ۳	صلاحیت ٤
٦	١١١١١٠١١١١١١١١١١١١١	٣٣٣٣	٠/٨٤	١/٠٠	١/٠٠	١/٠٠	١/٠٠
٨٠	٠٠٠٠٠٠٠١٠٠٠١٠١١٠١٠٠١	٠٠٣٠	٠/٦٥	٠/٠٢	١/٠٠	٠/٠٢	٠/٠٦
٣١٣	١١٠١٠٠٠١١١١١١٠٠٠٠١١١	١٣٣٣	٠/٦٠	١/٠٠	١/٠٠	١/٠٠	٠/٧٩
٥٦١	٠٠٩١٠٠٠٠٠١٠١٠٠١٩١١١١	١٠١٠	٠/٢٧	٠/٠٧	١/٠٠	٠/٠٧	١/٠٠
١٦١	٠٠٩٩٠٠٠٠٠١٠١٠٠٠٠٠٠١١	٠٠٠٠	٠/٦٨	٠/٠٦	٠/١٤	٠/٤٢	٠/١٤

بر اساس جدول (۵)، به‌عنوان نمونه آزمودنی شماره ٥٦١ با الگوی پاسخگویی ٠٠٩١٠٠٠٠٠١٠١٠٠١٩١١١١ (این فرد از بین ۲۰ سؤال، فقط به سؤال‌های ۹،٦،٤،۳،۲،۱، ۱۱ و ۱۷ پاسخ صحیح داده است) با احتمال ۲۷٪ به نیمرخ صلاحیتی ۱۰۱۰ تعلق دارد. یعنی از میان چهار صلاحیت، به صلاحیت‌های دوم و چهارم تسلط دارد و به‌طور دقیق‌تر، احتمال تسلط بر هر یک از صلاحیت‌های اول تا چهارم برای این فرد به ترتیب برابر ۰/۰۷، ۱/۰۰، ۰/۰۷ و ۱/۰۰ است. بدین ترتیب می‌توان بر اساس الگوی پاسخگویی هر فرد، تعیین کرد هر فرد دارای چه نیمرخ صلاحیتی است.

بحث و نتیجه‌گیری

به‌طور کلی، سنجش تشخیصی شناختی دارای چندین نقطه قوت اساسی نسبت به مدل‌های مرسوم اندازه‌گیری است. نخست اینکه، نتایج این سنجش علاوه بر ارائه اطلاعات در مورد صلاحیت‌هایی که افراد در آن به تسلط نرسیده‌اند، می‌تواند برای ارتقای استنباط‌های تکوینی در مورد یادگیری دانش‌آموزان یعنی صلاحیت‌هایی که باید در دور بعدی آموزش، تدریس شود نیز استفاده شود. همچنین یک ویژگی مناسب این سنجش، ارائه خرده نمره‌های تشخیصی در مقایسه با ارائه کلان نمره کل در رویکردهای مرسوم است. یعنی به جای اختصاص یک نمره به هر یک از آزمودنی‌ها، سنجش شناختی تشخیصی نیمرخی از نمره‌های تشخیصی که اطلاعاتی در مورد میزان تسلط صلاحیت‌ها که به‌طور خاص و مجزا تعریف شده‌اند، ارائه

می‌دهد. این رویکرد فرصتی را در اختیار آزمودنی‌ها قرار می‌دهد تا با دریافت اطلاعات مفصل و دقیق در مورد نقاط قوت و ضعف حل مسئله شناختی خود، در یک حیطه محتوایی خاص، یادگیری خود را بهبود بخشند. نقطه قوت دیگر سنجش شناختی تشخیصی، ارائه اطلاعاتی غنی‌تر و جزئی‌تر در مورد دانش و صلاحیت‌های اندازه‌گیری شده توسط یک آزمون است. درنتیجه امکان ارائه بازخورد تشخیصی وجود دارد؛ زیرا درک پیچیده‌تر صلاحیت‌های حل مسئله و روش مستقیم‌تری برای ارزیابی دانش راهبردی آزمودنی وجود دارد (الوس[1]، ۲۰۱۲).

هدف از مطالعه حاضر، به کارگیری مدل *پی‌جی‌دینا* برای سنجش سواد ریاضی در میان دانش‌آموزان پایه دهم (ابتدای سال تحصیلی) بود. نتایج این مطالعه نشان داد که بیشترین درصد تعلق آزمودنی‌ها به نیمرخ‌های صلاحیتی مربوط به دو نیمرخ صلاحیت (۰۰۰۰) و (۳۳۳۳) به ترتیب ۸٪ و ۵٪ بود. در میان همبستگی‌های پلی‌کوریک بین صلاحیت‌های شناختی، صلاحیت‌های *اول و سوم* بیشترین میزان همبستگی (۰/۵۱۵) و دو صلاحیت *دوم* و سوم کمترین میزان همبستگی (۰/۱۰٦) را داشتند. همچنین نتایج نشان داد صلاحیت استدلال به ترتیب از ریاضی‌وار کردن و بازنمایی، *ارتباطات و استفاده از زبان و عملیات نمادین و طراحی راهبرد ساده‌تر است.*

در میان سؤال‌های یک‌صلاحیتی ۱ ،۱۲ و ۱۳ آزمون تفاوت بین عدم تسلط بر صلاحیت مورد نیاز (عرض از مبدأ) و تسلط بر صلاحیت مورد نیاز به ترتیب برابر ۰/۰۵، ۰/٦٤ و ۰/۵۸ بود. در سؤال‌های دوصلاحیتی، تفاوت اثر عرض از مبدأ و اثر تعاملی برای سؤال‌های ۲، ٦، ۹، ۱۱ و ۱٦ به ترتیب ۰/۵۷، ۰/۳۹، ۰/۳۹، ۰/۲ و ۰/۷٤ بود. در سؤال‌های سه‌صلاحیتی ۳، ۷، ۱۷ و ۱۹ این تفاوت به ترتیب برابر ۰/٦۱، ۰/٦۹، ۰/٦٦ و ۰/۵٤ بود. در بین سؤال‌های چهارصلاحیتی چهارم، پنجم، هشتم، دهم، چهاردهم، پانزدهم، هیجدهم و بیستم تفاوت بین تفاوت اثر تعاملی با عرض از مبدأ به ترتیب ۰/۷۹، ۰/۸۹، ۰/۹۳، ۰/۸، ۰/۵۷، ۰/۵۳، ۰/٦٦ و ۰/۷۱ بود. به‌طور کلی سؤال‌های ۱ و ۱۱ کمترین اطلاعات تشخیصی را فراهم می‌کردند و هردو پارامتر حدس بسیار بالایی (بیش از ۰/۷) داشتند. از نظر چن و دی‌لتوره (۲۰۱۳) این نتیجه

---

1. Alves

شاید حاصل بد تعریف شدن[1] صلاحیت‌های مورد نیاز برای سؤال باشد. همچنین تفاوت مذکور در سؤال‌های ٦ و ۹ کمتر از ۰/۵ بود که البته هر دو سؤال دارای پارامتر حدس متوسطی (بین ۰/۵ تا ۰/٦) بودند. از نظر چن و دی‌لتوره (۲۰۱۳) این نتیجه احتمالاً در اثر تعریف کردن نارسا[2] سطوح صلاحیت‌های شناختی مورد نیاز برای سؤال است. نکته قابل توجه اینکه همه این چهار سؤال، چندگزینه‌ای بودند که این خود به وجود عامل حدس، منجر شده است.

در تحلیل پارامترهای برآورد شده سؤال‌ها و در سؤال‌های دوصلاحیتی، یک ویژگی مشترک بین سؤال‌ها، بالا بودن اثرات اصلی صلاحیت‌های شناختی مورد نیاز برای پاسخگویی به سؤال‌ها در مقایسه با اثر تعاملی بود. در سؤال‌های دوم و ششم، صلاحیت *ارتباطات و استفاده از زبان و عملیات نمادین* تعیین‌کننده‌تر، در سؤال نهم صلاحیت استدلال پررنگ‌تر (هرچند سطح بازنمایی نسبتاً بالا بود) و در سؤال یازدهم، صلاحیت راهبرد پررنگ‌تر و در سؤال شانزدهم، تقریباً اثرات یکسانی وجود داشت. نکته مهم در این مجموعه سؤال‌ها، به سؤال ششم مربوط می‌شد جایی که اثر اصلی صلاحیت *اول* یعنی احتمال پاسخگویی درست با تسلط بر این صلاحیت، از اثر تعاملی دو صلاحیت *اول* و دوم بیشتر بود. شاید با توجه به ساختار نیمرخ صلاحیتی این سؤال و این نکته که سطح صلاحیت *اول* بالاتر از سطح صلاحیت دوم بود این نتیجه تا حدودی قابل توجیه باشد. در سؤال نهم، اثر اصلی هر یک از دو صلاحیت دوم و سوم بالا بود و تفاوت بین دو اثر اصلی تفاوت اندکی داشت (۰/۱۰، به نفع صلاحیت استدلال). در سؤال یازدهم، میزان اثر اصلی هر یک از دو صلاحیت *ارتباطات و استفاده از زبان و عملیات نمادین* و طراحی راهبرد بالا و نزدیک به اثر تعاملی بود و تفاوت بین دو اثر اصلی تفاوت اندکی داشت (۰/۱۰، به نفع صلاحیت طراحی راهبرد). یعنی تسلط بر صلاحیت راهبرد تعیین‌کننده پاسخگویی درست به سؤال بود. در سؤال شانزدهم، اثر اصلی هر یک از دو صلاحیت یک و سه نسبتاً بالا بود و تفاوت بین دو اثر اصلی تفاوت اندکی داشت (۰/۰۳، به نفع صلاحیت اول).

در بین سؤال‌های سه صلاحیتی، سؤال **سوم**، اثرات اصلی صلاحیت‌ها اغلب بیش از اثرات تعاملی دوبه‌دو بود و اثر اصلی صلاحیت *ارتباطات و استفاده از زبان و*

---

[1]. Misspecification
[2]. Underspecified

عملیات نمادین نیز حتی بیش از اثر تعاملی سه صلاحیت دیگر بود. این نتیجه به نقش کلیدی صلاحیت *ارتباطات و استفاده از زبان و عملیات نمادین* در این سؤال اشاره داشت. در سؤال هفتم، اثر تعاملی صلاحیت‌های یک و چهار بالا و مساوی اثر تعاملی سه صلاحیت بود. یعنی فرد با داشتن دو صلاحیت *ارتباطات و استفاده از زبان و عملیات نمادین* و طراحی راهبرد احتمال پاسخگویی درست به سؤال را داشت و این یعنی نقض ویژگی غیر جبرانی بودن. در سؤال **هفدهم**، اگرچه اثرات اصلی صلاحیت‌ها پایین یا متوسط بود اما اثر تعاملی *ارتباطات و استفاده از زبان و عملیات نمادین* و طراحی راهبرد حتی بالاتر از اثر تعاملی سه صلاحیت بود یعنی نقض ویژگی غیر جبرانی بودن. در سؤال **نوزدهم**، اثر اصلی صلاحیت یک بالاتر از اثر تعاملی هر سه صلاحیت بود.

در میان سؤال‌های چهارصلاحیتی این پیچیدگی بیشتر شده و در سؤال چهارم، اثر اصلی صلاحیت یک و چهار بیشتر از اثر تعاملی چهار صلاحیت بود. در سؤال پنجم، این نتیجه برای اثر اصلی صلاحیت یک به دست آمد. در سؤال چهاردهم، اثر تعاملی یک و چهار بیشترین احتمال پاسخگویی درست را در برداشت. در سؤال پانزدهم، اثر تعاملی دوصلاحیتی بیش از اثر تعاملی کل صلاحیت‌ها بود. در سؤال هجده، اثر تعاملی سه صلاحیت دوم، سوم و چهارم بیش از اثر تعاملی کلی بود. در سؤال بیست، اثر تعاملی سه صلاحیت اول، دوم، سوم بیشتر اثر تعاملی کلی بود.

به‌طور کلی، برخلاف ادعای روسوس، تمپلین و هنسون[1]، (۲۰۰۷) که اصولاً در سنجش صلاحیت‌های ریاضی، کاربرد مدل‌های شناختی با رویکرد غیر جبرانی، متداول‌تر است؛ نتایج به دست آمده نشان داد که احتمال موفقیت در هر یک از سؤال‌های آزمون سواد ریاضی می‌تواند کاملاً وابسته به ساختار صلاحیت‌های شناختی مورد نیاز برای سؤال باشد. یعنی این امکان وجود دارد که در برخی سؤال‌ها ساختار صلاحیت‌ها جبرانی و در برخی دیگر غیر جبرانی باشند. این موضوع عمدتاً زمانی بیشتر بروز پیدا می‌کند که تعداد صلاحیت‌های بیشتر و سطوح صلاحیت‌های شناختی بالاتری برای سؤال‌ها مورد نیاز است. از سوی دیگر، هر اندازه پیچیدگی آزمون و تنوع صلاحیت‌های شناختی مورد نیاز برای هر سؤال بیشتر باشد، توصیف سؤال‌ها و استنباط درباره نتایج به دست آمده دشوارتر است. بنابراین به نظر می‌رسد کاربرد این

---

[1]. Roussos, Templin & Henson

نوع سنجش می‌تواند در زمینه سنجش‌های کلاسی و شناسایی نقاط قوت و ضعف دانش‌آموزان کاربردی‌تر و قابل تفسیرتر باشد.

در پایان، با توجه به این دیدگاه که قضاوت درباره سؤال‌ها در سطح اطلاعات به دست آمده در سطح سؤال، کار دشواری است (روجاس، دی‌لتوره و اولیا[1]، ۲۰۱۲، به نقل از راوند، ۲۰۱۵)، پیشنهاد می‌شود که برای تحلیل دقیق‌تر سؤال‌ها، بر اساس ساختار روابط میان سؤال‌ها از مدل‌های شناختی تشخیصی متناسب با آن سؤال استفاده شود.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

[1]. Rojas, De la Torre & Olea

منابع

محسن‌پور، مریم؛ گویا، زهرا؛ شکوهی‌یکتا، محسن؛ کیامنش، علیرضا و بازرگان، عباس(۱۳۹۳). طراحی و ساخت آزمونی برای صلاحیت‌های شناختی سواد ریاضی دانش‌آموزان ایرانی بر مبنای مطالعات پیزا. *دوفصلنامه نظریه و عمل در برنامه درسی*، ۲ (٤)، ۵ – ۳٤.

Chen, J. & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. ***Applied Psychological Measurement***, 37, 419-437.

De la Torre, J. (2011). The generalized DINA model framework. ***Psychometrika***, 76, 179–199.doi:10.1007/s11336-011-9207-7

Di Bello, L. V.; Roussos, L. A. & Stout, W. (2007). Review of Cognitively Diagnostic Assessment and a Summary of Psychometric Models. In C. V. Rao & S. Sinharay (Eds.), ***Handbook of statistics*** (Vol. 26, Psychometrics, pp. 979-1027). Amsterdam, the Netherlands: Elsevier.

Haberman, S. J.; von Davier, M. & Lee, Y. (2008). ***Comparison of multidimensional item response models: multivariate normal ability distributions versus multivariate polytomous distributions*** (ETS Research Rep. No. RR-08-45). Princeton, NJ: ETS.

Jang, E. (2008). A Framework for Cognitive Diagnostic Assessment. In C. A. Chapelle, Y.-R. Chung, & J. Xu (Eds.), ***towards adaptive CALL: Natural language processing for diagnostic language assessment*** (pp. 117-131). Ames, IA: Iowa State University.

Karelitz, T. M. (2004). ***Ordered category attribute coding framework for cognitive assessments*** (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Kunina-Habenicht, Olga; Rupp, André A. & Wilhelm, Oliver (2012). The Impact of Model Misspecification on Parameter Estimation and Item-Fit Assessment in Log-Linear Diagnostic Classification Models. ***JEM, 49*** (1), 59 – 81.

Leighton, J. P. Gierl, M. J. & Hunka, S. M. (2004). The Attribute Hierarchy Method for Cognitive Assessment: A Variation on Tatsuoka's Rule-Space Approach. ***Journal of Educational Measurement***, ***41*** (3), 205-237.

Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models (with discussion). *Measurement: Interdisciplinary Research and Perspectives, 11*, 71-137.

Niss, M. & Højgaard, T. (Ed.). (2011). *Competencies and Mathematical Learning, Ideas and inspiration for the development of mathematics teaching and learning in Denmark*, Roskilde: Roskilde University.

Niss, M. & Jensen, T. H. (eds) (2002). *Kompetencer og matematiklæring –Ideer og inspiration til udvikling af matematik undervisning i Danmark, number 18 in Uddannelsesstyrelsens temahæfteserie*. The Ministry of Education, Copenhagen, Denmark.

OECD (2009*). Learning Mathematics for Life: A Perspective from PISA*. Paris: OECD Publications.

OECD (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy.* Paris: OECD Publications.

Ravand, H. (2015). Application of a Cognitive Diagnostic Model to a High-Stakes Reading Comprehension Test. *Journal of Psychoeducational Assessment*, *34* (8) 782–799.

Robitzsch, A.; Kiefer, T. George, A. & Uenlue, A. (2014). *Package CDM*, Date/Publication 2014-04-11 12:27:06 UTC

Rojas, G.; de la Torre, J. & Olea, J. (2012). *Choosing between general and specific cognitive diagnosis models when the sample size is small*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Vancouver, British Columbia, Canada.

Roussos, L. A.; Templin, J. L. & Henson, R. A. (2007). Skills Diagnosis Using IRT-Based Latent Class Models. *Journal of Educational Measurement,* 44 (4), 293–311.

Rupp, A. A.; Templin, J. & Henson, R.A. (2010). *Diagnostic Measurement, Theory, Methods, and Applications*. New York: The Guilford Press.

Stacey, K. (2012). *The International Assessment of Mathematical literacy: PISA 2012 Framework and Items.* 12th International Congress on Mathematical Education. 8 July – 15 July, COEX, Seoul, Korea.

Turner, R.; Blum, W. & Niss, M. (2015). Using Competencies to Explain Mathematical Item Demand: A Work in Progress. In K. Stacey & R. Turner (Eds.), *Assessing Mathematical Literacy*, the

PISA Experience (Vol. 26, Psychometrics, pp. 85-116. Amsterdam, the Netherlands: Springer.

Von Davier, M. (2005). *mdltm: Software for the General Diagnostic Model and for Estimating Mixtures of Multidimensional Discrete Latent Traits Models*. [Computer software]. Princeton, NJ: ETS.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

پیوست (۱)

ارتباطات:				
تعریف متغیر	**سطح ۰**	**سطح ۱**	**سطح ۲**	**سطح ۳**
مشخص کردن اجزا مسئله	درک ارتباط بین اجزای مسئله	استخراج اطلاعات حاصل شده از ارتباطات	استفاده مکرر از چرخه یادگیری در سطوح قبل برای ارائه یک توصیف یا توضیح منسجم	ارائه یک توصیف یا توضیح شفاف، منسجم، کامل و باصرفه برای یک راه حل، فرایند یا استدلال
ریاضی‌وار کردن:				
تعریف متغیر	**سطح ۰**	**سطح ۱**	**سطح ۲**	**سطح ۳**
استفاده از مدل‌های مفروض یا ساخته‌شده	عدم نیاز به استفاده از مدل	استنباط مستقیم از یک مدل مفروض	جرح‌وتعدیل مدل	تبیین مدل
طراحی راهبرد برای حل مسئله:				
تعریف متغیر	**سطح ۰**	**سطح ۱**	**سطح ۲**	**سطح ۳**
حل مسئله در دنیای ریاضی	انجام دادن عملیاتی بدون استفاده از راهبرد	انتخاب راهبرد مناسب	ساختن راهبرد مورد نیاز	ساختن راهبرد تفصیلی برای یک راه حل جامع
استدلال:				
تعریف متغیر	**سطح ۰**	**سطح ۱**	**سطح ۲**	**سطح ۳**
فرایندهای فکری که اجازه استنباط یا بررسی استدلال‌های ارائه شده را بدهد	استنباط کردن مستقیم از دستورالعمل‌های داده شده	تلاش برای ترکیب اطلاعات جهت استنباط کردن	تجزیه‌وتحلیل اطلاعات برای استدلال چندمرحله‌ای	ارزشیابی از زنجیره‌های استدلالی برای توجیه استنباط‌ها
بازنمایی:				
تعریف متغیر	**سطح ۰**	**سطح ۱**	**سطح ۲**	**سطح ۳**
استفاده و حرکت بین بازنمایی‌های مختلف	استفاده مستقیم از یک بازنمایی	انتخاب و تفسیر یک بازنمایی استاندارد	استفاده از دو یا چند بازنمایی و حرکت بین آن‌ها	فهمیدن و استفاده از یک بازنمایی غیر استاندارد که نیازمند رمزگشایی‌ها و تفسیرهای اساسی است

استفاده از زبان و عملیات نمادین، رسمی و فنی:				
تعریف متغیر	**سطح ۰**	**سطح۱**	**سطح۲**	**سطح۳**
درک کردن، **دست‌کاری کردن** و **استفاده** عبارات نمادین در درون یک زمینه ریاضیاتی (شامل عبارات حسابی و عملیاتی) که **قراردادها و قواعد** ریاضی بر آنها حاکم است، **درک کردن** و **به‌کاربردن سازه‌های** رسمی مبتنی بر تعاریف، قواعد و **نظام‌های رسمی**	قواعد ریاضیاتی یا عبارات نمادین نیازی نیست که فعال شوند فراتر از محاسبات حسابی، عملیات با اعداد کوچک و ؟	استفاده مستقیم از یک رابطه تابعی ساده، خواه صریح یا غیر صریح (به‌عنوان مثال روابط خطی آشنا)؛ استفاده از نمادهای ریاضیاتی رسمی (به‌عنوان مثال، به وسیله جایگزین کردن مستقیم یا محاسبات حسابی پایدار شامل کسرها و اعداد اعشاری) یا فعال کردن و استفاده مستقیم از یک تعریف ریاضیاتی رسمی، قرارداد یا مفهوم نمادین.	استفاده صریح و دست‌کاری نمادها (به‌عنوان مثال، دوباره مرتب کردن جبری یک فرمول)؛ فعال‌سازی و استفاده از قواعد ریاضیاتی، تعاریف، قراردادها، رویه‌ها یا فرمول‌ها با استفاده ترکیبی روابط چندگانه یا مفاهیم نمادین.	کاربرد چندمرحله‌ای رویه‌های ریاضیاتی رسمی، کار کردن به‌گونه‌ای انعطاف‌پذیر با روابط جبری تابعی؛ استفاده از تکینک‌های ریاضیاتی و دانش به‌منظور ایجاد نتایج.

.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی