



## Diagnosis of science cognitive attributes of Iranian fourth grad students in TIMSS 2015

Masoud Kabiri<sup>1</sup>

1. Assistant Professor of Research Institute for Education, Email : maskabiri@yahoo.com.

---

### Article Info

### ABSTRACT

**Article Type:**  
Research Article

**Objective:** Nowadays, cognitive diagnostic assessment (CDA) is highly attended due to attain deeply information about examinees. In present study, CDA is used to provide more information about science responses of Iranian fourth graders who participated in TIMSS 2015.

**Methods:** 11 science attributes were identified and the Q matrix was created. Then, responses of 3823 students to 206 items was analyzed by G-DINA model.

**Received:**  
**Received in revised form:**

**Results:** The results showed that the model can produce suitable diagnostic evidences. Also, three attributes including: Recognize, Providing examples, and Describe had higher attribute probabilities than others and their mastery probabilities were more than 0.50. Whereas, only probability of Recognize was higher than 0.65 and this is considered as mastered attribute. In other hand, Predict and Scientific inquiry had lowest attribute probabilities.

**Accepted:**  
**Published online:**

**Conclusion:** In general, the results revealed that there are sizable deference between knowledge-based attributes and higher-order thinking attributes in students, that can associated to distance between intended and implemented curricular.

**Keywords:** Cognitive diagnostic assessment, Science education, G-DINA, TIMSS, Iran education.

---

**Cite this article:** Kabiri, Masoud (2023). Diagnosis of Science Cognitive Attributes of Iranian Fourth Grad Students in TIMSS 2015. *Educational Measurement and Evaluation Studies*, 11 (33): 14 pages.

DOI:

© The Author(s).

Publisher: National Organization of Educational Testing (NOET)





## تشخیص وضعیت دانش‌آموزان پایه چهارم ایران در خصیصه‌های شناختی آزمون علوم مطالعه تیمز ۲۰۱۵

مسعود کبیری<sup>۱</sup>

۱. استادیار پژوهشگاه مطالعات آموزش و پرورش، تهران، ایران. رایانامه: maskabiri@yahoo.com

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	<b>هدف:</b> امروزه سنجش شناختی تشخیصی برای کسب اطلاعات بیشتر در مورد یادگیری آزمودنی‌ها مورد توجه قرار گرفته است. در همین راستا، در مطالعه حاضر این شیوه به کار بسته شد تا اطلاعات بیشتری در مورد پاسخ‌های دانش‌آموزان پایه چهارم شرکت‌کننده در آزمون علوم مطالعه تیمز ۲۰۱۵ استخراج شود.
دریافت: اصلاح: پذیرش: انتشار:	<b>روش پژوهش:</b> ۱۱ خصیصه مهم در آموزش علوم شناسایی و با بررسی سؤالات ماتریس Q تشکیل شد. سپس، پاسخ ۳۸۲۳ دانش‌آموزی که ۲۰۶ سؤال آزمون علوم را داده بودند، با استفاده از مدل تشخیصی جی‌دینا تحلیل شد.
	<b>یافته‌ها:</b> نتایج نشان داد که این مدل قدرت تولید اطلاعات تشخیصی را دارد. در مقایسه احتمال پاسخ خصیصه‌ها مشخص شد که سه خصیصه بازشناسی، ارائه مثال، و توصیف احتمال بالاتری نسبت به سایر خصیصه‌ها نشان دادند و احتمال تسلط بر آن‌ها از ۰/۵۰ بیشتر بود. با این حال، تنها خصیصه بازشناسی از ملاک مقبول ۰/۶۵ بیشتر بود. از طرف دیگر، دو خصیصه پیش‌بینی و کاوشگری علمی کم‌ترین احتمال پاسخ را داشتند.
	<b>نتیجه‌گیری:</b> نتایج نشان داد که به‌طور کلی تفاوت زیادی بین احتمال پاسخ خصیصه‌های دانشی و خصیصه‌های دربرگیرنده تفکر سطح بالا در بین دانش‌آموزان ایران وجود دارد که می‌تواند به فاصله گرفتن برنامه اجرا شده علوم از برنامه مصوب آن مربوط باشد.
	<b>واژه‌های کلیدی:</b> سنجش شناختی تشخیصی، آموزش علوم، مدل جی‌دینا، مطالعه تیمز، آموزش و پرورش ایران.

استناد: کبیری، مسعود (۱۴۰۲). تشخیص وضعیت دانش‌آموزان پایه چهارم ایران در خصیصه‌های شناختی آزمون علوم مطالعه تیمز ۲۰۱۵. مطالعات اندازه‌گیری و

ارزشیابی آموزشی، ۱۱ (شماره ۳۳)، ۱۴ صفحه.

ناشر: سازمان سنجش آموزش کشور

DOI:

حق مؤلف © نویسندگان.



## مقدمه

سنجش شناختی تشخیصی یکی از آخرین توسعه‌های حوزه سنجش است که در دو دهه اخیر به طور فزاینده‌ای مورد توجه قرار گرفته تا اطلاعاتی در مورد قوت و ضعف‌های شناختی آزمودنی‌ها فراهم نماید. فرض اصلی این سنجش آن است که در پاسخ آزمودنی‌ها به سؤالات آزمون، اطلاعات بیشتری به جز برآورد تک‌بعدی نمره توانایی وجود دارد. بنابراین، با هدف نشان دادن عملکرد آزمودنی در یک سؤال بر اساس مهارت‌های لازم برای پاسخ‌گویی (روسوس، تمپلین و هنسون<sup>۱</sup>، ۲۰۰۷)، مجموعه‌ای از زیرتکالیف غیرقابل مشاهده، مهارت‌ها یا خصیصه‌های مکنون لازم برای پاسخ‌گویی به سؤالات آزمون کشف می‌شود (دی کارلو<sup>۲</sup>، ۲۰۱۱). در نتیجه، اطلاعات بیشتری در مورد آزمودنی‌ها فراهم می‌شود که مورد استفاده طیف وسیعی از ذی‌نفعان قرار می‌گیرد. این اطلاعات در رابطه با دانش و مهارت‌های جزئی شده دانش‌آموزان بر اساس پاسخ‌های آنان برآوردهای درونی<sup>۳</sup> (چن، گورین، تامپسون و تاتسوآکا<sup>۴</sup>، ۲۰۰۸، الف) نیز نامیده می‌شود که در تقابل با برآوردهای بیرونی<sup>۵</sup> است که در نمره کل تولیدشده در سنجش‌های سنتی حاصل می‌شود.

رایج‌ترین کارکرد سنجش شناختی تشخیصی ترسیم نیمرخ‌های مهارتی است؛ به طوری که فضای توانایی مکنونی متشکل از مجموعه حالت‌های دانش، مهارت‌ها یا خصیصه‌ها را فرض کرده و عملکرد دانش‌آموز را در فضای چندبعدی قرار داده و احتمال رسیدن به تسلط در هر مهارت را به صورت نیمرخ‌های مهارتی نشان می‌دهد (تاتسوآکا، کورتر و تاتسوآکا<sup>۶</sup>، ۲۰۰۴؛ کیم<sup>۷</sup>، ۲۰۱۱؛ ون داویر<sup>۸</sup>، ۲۰۰۹). به عبارت دیگر، از طریق تجزیه نمره کل فرد به مجموعه‌ای از نمرات مهارت‌ها یا خصیصه‌ها، نمرات چندگانه‌ای تولید کرده و اطلاعات بیشتری برای بازخوردهای تشخیصی و طرح‌ریزی آموزش‌های جبرانی فراهم می‌کند. به طور مشخص، در این کارکرد سه موضوع میزان اکتساب خصیصه‌ها در جامعه، توزیع کلاس‌های مکنون در جامعه، و اکتساب خصیصه‌ها توسط افراد بررسی می‌شود (جورج و روبیتس<sup>۹</sup>، ۲۰۱۵).

برای اجرای سنجش شناختی تشخیصی، مدل‌های شناختی تشخیصی لازم هستند. مدل‌های شناختی تشخیصی گونه‌ای از مدل‌های روان‌سنجی هستند که برآورد معتبری را برای سازه‌های چندبعدی به منظور فراهم کردن نگاه چندمتغیری به نقاط قوت و ضعف دانش‌آموزان مهیا می‌کنند (راپ، تمپلین و هنسون<sup>۱۰</sup>، ۲۰۱۰). این کار از طریق تبدیل نظریات شناختی به مدل‌های احتمالاتی (سینهارا و الموند<sup>۱۱</sup>، ۲۰۰۷) به منظور اضافه کردن اطلاعات تشخیصی (چن و همکاران، ۲۰۰۸، ب) صورت می‌پذیرد. در ادبیات سنجش شناختی تشخیصی، مدل‌های متعددی معرفی شده‌اند که بسته به شکل پاسخ (دوبخشی در مقابل چندبخشی)، مقیاس‌پردازی خصیصه (تسلط یا عدم تسلط در مقابل حالت‌های چندبخشی و پیوسته) و قانون تخلیص<sup>۱۲</sup> یا ارتباط بین خصیصه‌ها از لحاظ پاسخ‌گویی (به شکل جبرانی بودن یا غیرجبرانی بودن) تسلط بر خصیصه‌های لازم برای پاسخ‌گویی به سؤال) تقسیم‌بندی می‌شوند (راپ، تمپلین و هنسون، ۲۰۱۰). با وجود تعدد مدل‌های شناختی تشخیصی، سه مدل عمومی مطرح شده‌اند که می‌توان بقیه مدل‌ها را حالت‌های ویژه‌ای از آن‌ها در نظر گرفت. این سه مدل شامل مدل تشخیصی کلی<sup>۱۳</sup> (ون داویر، ۲۰۰۸)، مدل شناختی تشخیصی لوگ‌خطی<sup>۱۴</sup> (هنسون، تمپلین و ویلسه<sup>۱۵</sup>، ۲۰۰۹)، و جی‌دینا<sup>۱۶</sup> (دلاتوره<sup>۱۷</sup>، ۲۰۱۱) هستند.

1. Roussos, Templin, & Henson

2. DeCarlo

3. internal estimate

4. Chen, Gorin, Thompson, & Tatsuoka

5. external estimate

6. Tatsuoka, Corter, & Tatsuoka

7. Kim

8. von Davier

9. George & Robitzsch

10. Rupp, Templin, & Henson

11. Sinharay & Almond

12. condensation rule

13. Cognitive diagnostic model (GDM)

14. log-linear cognitive diagnosis model (LCDM)

15. Henson, Templin, & Willse

16. generalized deterministic inputs, noisy and gate (G-DINA)

17. de La Torre

مدل جی‌دینا که چارچوب متحد‌کننده مدل‌های شناختی تشخیصی است، برای درک بهتر ماهیت منحصر به فرد مدل‌های تشخیصی و ارتباط آن‌ها با هم‌دیگر پیشنهاد شده است (دل‌توره و مینشن، ۲۰۱۹). منطق این مدل بر اساس مفهوم تحلیل واریانس تنظیم شده است. برای درک این مدل به معادله اصلی آن توجه کنید:

$$P(\alpha_{ij}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'} \alpha_{lk} \alpha_{lk'} \dots + \delta_{j12\dots k_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}$$

در این معادله  $\delta_{j0}$  عرض از مبدأ سؤال  $j$ ،  $\delta_{jk}$  اثر اصلی  $\alpha_k$ ،  $\delta_{jkk'}$  اثر تعاملی  $\alpha_{lk}$  و  $\alpha_{lk'}$  و  $\delta_{j12\dots k_j^*}$  اثر تعاملی  $\alpha_{1k} \dots \alpha_{K_j^*k}$  محسوب می‌شوند. این پارامترها می‌توانند بدین گونه تفسیر شوند:  $\delta_{j0}$  احتمال خطاپایه را نشان می‌دهد (احتمال پاسخ درست زمانی که هیچ‌یک از خصیصه‌های لازم وجود نداشته باشد)؛  $\delta_k$  احتمال پاسخ درست به عنوان نتیجه‌ای از تسلط به یک خصیصه (مثل  $\alpha_k$ )؛  $\delta_{kk'}$  اثر تعاملی سطح اول (احتمال پاسخ درست به خاطر تسلط بر هر دو  $\alpha_k$  و  $\alpha_{k'}$  که بیشتر از اثر افزایشی تسلط بر هر دو خصیصه است)؛ و  $\delta_{12\dots k_j^*}$  تغییر در احتمال پاسخ درست به دلیل تسلط بر همه خصیصه‌های لازم را نشان می‌دهد که از اثر افزایشی اثرات اصلی و تعاملی سطح پایین‌تر بیشتر است (دل‌توره، ۲۰۱۱). به دلیل اینکه مدل جی‌دینا همه اثرات را دربر گرفته است به عنوان مدل اشباع‌شده شناخته می‌شود. در مقایسه، مدلی چون مدل یک‌پارچه با پارامترپردازی کاهش‌یافته<sup>۲</sup> تنها عرض از مبدأ و اثر اصلی را داشته و فاقد اثرات تعاملی است (یاماگوشی و اوکادا، ۲۰۱۸).

با این حال، هنگامی که خصیصه‌ها و پاسخ‌ها دوبخشی باشند، مدل جی‌دینا که از تابع ارتباطی همانی<sup>۳</sup> استفاده می‌کند، با مدل شناختی تشخیصی لوگ‌خطی و مدل تشخیصی کلی (با تابع ارتباطی لوجیت<sup>۴</sup>) معادل می‌شوند (دل‌توره و مینشن، ۲۰۱۹). همین رابطه در مورد ارتباط مدل جی‌دینا و مدل‌های زیرمجموعه آن نیز وجود دارد؛ به طوری که با صفر شدن یا پارامترپردازی مجدد برخی از مؤلفه‌های معادله فوق یا انتخاب تابع ارتباطی مناسب، مدل‌های دیگری به دست می‌آید. برای مثال، مدل دینا<sup>۵</sup> تنها دارای مؤلفه‌های عرض از مبدأ و بالاترین اثر تعاملی و مدل دینو<sup>۶</sup> دارای محدودیت‌هایی در اثرات اصلی و تعاملی است (یاماگوشی و اوکادا، ۲۰۱۸؛ دل‌توره و مینشن، ۲۰۱۹).

فارغ از بحث درباره مدل‌های شناختی تشخیصی، مدل‌سازی بر اساس داده‌های از پیش موجود که به بازبرازش<sup>۸</sup> معروف است، یکی از علایق اصلی پژوهشگران سنجش شناختی تشخیصی را تشکیل می‌دهد. دو مطالعه‌ای که بیش از همه مطالعات در داخل کشور برای بازبرازش استفاده شده‌اند، مطالعه بین‌المللی روندها در ریاضی و علوم (تیمز)<sup>۹</sup> و مطالعه بین‌المللی پیشرفت سواد خواندن (پرلز)<sup>۱۰</sup> هستند. هر دو مطالعات از جمله قدیمی‌ترین و گسترده‌ترین مطالعات تطبیقی سنجش پیشرفت تحصیلی در دنیا هستند که از سال ۱۹۹۵ در بیش از ۵۰ کشور دنیا به همراه ایران با تأکید بر سنجش دروس ریاضی، علوم و خواندن به صورت دوره‌ای اجرا می‌شود و به دلیل معرف بودن مشارکت‌کنندگان، گستردگی سنجش موضوعات، در اختیار بودن داده‌ها و وجود آرشیو داده‌های گذشته موردقبال بسیاری از پژوهشگران قرار گرفته است.

با وجود محبوبیت استفاده از بازبرازش در داده‌های مطالعات بین‌المللی، به دلیل نیاز به معرفی قابلیت‌های سنجش شناختی تشخیصی، امکان پرداختن به همه جزئیات مدل‌سازی شناختی تشخیصی وجود نداشت ولی مدل‌سازی‌های بعدی هم کم‌تر به همه ویژگی‌های داده‌های مورد استفاده توجه می‌کردند. به عنوان مثال، با اینکه سنجش مطالعات تیمز و پرلز از دفترچه‌های متعددی تشکیل شده است ولی برخی از پژوهشگرانی که از داده‌های مطالعات تیمز یا پرلز استفاده کرده بودند، خود را به یک یا چند دفترچه محدود نموده بودند (ایم و پارک<sup>۱۱</sup>، ۲۰۱۰؛ بیرنباوم، تاتسواکا و ژین<sup>۱۲</sup>، ۲۰۰۵؛ تاتسواکا، کورتر و تاتسواکا، ۲۰۰۴؛ چن و همکاران، ۲۰۱۰، ۲۰۰۸، الف، ب؛ لی، پارک و تایلان، ۲۰۱۱). هم‌چنین،

1. Minchen  
 2. reduced reparametrized unified model (R-RUM)  
 3. identity link function  
 4. logit link function  
 5. deterministic-input noisy and gate (DINA)  
 6. deterministic-input noisy or gate (DINO)  
 7. Yamaguchi, Okada  
 8. retrofit  
 9. Trends in International Mathematics and Science Study (TIMSS)  
 10. Progress in International Reading Literacy Study (PIRLS)  
 11. Im Park  
 12. Birenbaum, Tatsuoka, & Xin

استفاده از مدل‌هایی که تنها قابلیت تحلیل داده‌های دوبخشی دارند، باعث شده است، برخی از پژوهشگرانی که از داده‌های مطالعه تیمز استفاده کرده‌اند، یا سوالات چندبخشی را کنار بگذارند و یا آن‌ها را به سوالات دوبخشی تبدیل کنند (ایم و پارک، ۲۰۱۰؛ بیرنهام، تاتسواکا و ژین، ۲۰۰۵؛ تاتسواکا، کورتر و تاتسواکا، ۲۰۰۴؛ چن و همکاران، ۲۰۱۰، ۲۰۰۸ الف و ب؛ چپو و سئو، ۲۰۰۹؛ دوگان<sup>۲</sup> و تاتسواکا، ۲۰۰۸؛ مینایی و همکاران، ۱۳۹۳؛ لی، پارک و تایلان<sup>۳</sup>، ۲۰۱۱). این راهبرد باعث کاهش دقت مدل‌های تولیدشده می‌شود. افزون بر این، در پژوهش‌های اشاره شده در قبل که داده‌های مطالعه تیمز را به کار برده بودند، از وزن‌های نمونه‌گیری که یکی از استلزامات هر گونه مدل‌سازی بر اساس این داده‌ها است، استفاده نشده بود. مجموعه این شرایط پژوهشگران را ترغیب به در نظر گرفتن همه شرایط و ویژگی‌های داده‌های مورد استفاده در بازبرازش سنجش شناختی تشخیصی می‌نماید. در کنار این موارد استفاده همراه با دقت نه چندان زیاد از مدل‌های شناختی تشخیصی نیز قابل ذکر است. به عنوان مثال، استفاده از آماره‌های نسبی برازش مدل ملاک اطلاعاتی آکاپک<sup>۴</sup> و ملاک اطلاعات بیزی<sup>۵</sup> در حالت‌هایی گزارش شده است که هیچ مدل رقیب دیگری وجود ندارد (سعدتی و همکاران، ۱۳۹۹؛ شه‌میرزادی و همکاران، ۱۳۹۹). در نهایت، می‌توان به معرفی مدل به عنوان مدل نهایی در شرایطی اشاره کرد که ضرایب لغزش یا حدس به اندازه‌ای زیادی هستند که در نتیجه ضریب تمیز سؤال به شدت پایین می‌آید: برای مثال ضریب تمیز ۰/۰۲۵ و ۰/۰۶۳ برای سوالات ۲ و ۱۶ در مطالعه سعدتی و همکاران (۱۳۹۹)؛ ۰/۱۰۳ و ۰/۰۹۶ برای سوالات ۱۶ و ۱۷ در مطالعه رحیمی، یونسی، مکرمی (۱۳۹۷).

پژوهشگران داخلی در سالیان اخیر به‌طور رایج‌تری از سنجش شناختی تشخیصی استفاده کرده‌اند. گستره عموم این پژوهش‌ها بیشتر در زمینه کنکورهای سراسری یا آزمون‌های درک مطلب زبان انگلیسی بوده است (جاویدان‌مهر و عنانی‌سراب، ۱۳۹۶؛ رحیمی و همکاران، ۱۳۹۹؛ مقدم و همکاران، ۱۳۹۵). البته در حوزه آموزش عمومی، در دوره اول متوسطه و با استفاده از داده‌های مطالعه بین‌المللی تیمز نیز پژوهش‌های دیگری صورت پذیرفته است (کبیری و همکاران، ۲۰۱۷؛ مینایی و همکاران، ۱۳۹۳). دلیل رواج بازبرازش داده‌های گسترده موجود احتمالاً امکان مدل‌سازی بهتر با تکیه بر داده‌های برگرفته از نمونه‌ها و سوالات گسترده باشد. با این حال، پژوهش‌های دیگری هم به طراحی آزمون‌ها بر اساس سنجش شناختی تشخیصی پرداخته‌اند (افضلی و همکاران، ۱۳۹۳؛ تقییان و همکاران، ۱۳۹۷؛ رنجبران و علوی، ۱۳۹۵؛ محسن‌پور، ۱۳۹۸؛ محسن‌پور و همکاران، ۱۳۹۴). با این همه، این مطالعات کاربرد سنجش شناختی تشخیصی را در دوره ابتدایی مورد بررسی قرار نداده‌اند، حال آنکه پژوهش‌های تشخیصی در این دوره به دلیل امکان کاربست سریع‌تر از اهمیت بیشتری برخوردار است. در این مطالعه در نظر است تا با استفاده از داده‌های مطالعه تیمز، نیمرخ خصیصه‌های دانش‌آموزان پایه چهارم ایران در درس علوم مورد بررسی قرار گیرد و میزان تسلط دانش‌آموزان ایرانی بر هر یک از خصیصه‌ها تعیین شود.

## روش پژوهش

### مشارکت‌کنندگان

نمونه مورد بررسی در این مطالعه برگرفته از نمونه پایه چهارم مطالعه تیمز ۲۰۱۵ است که با استفاده از روش نمونه‌گیری خوشه‌ای طبقه‌ای دومرحله‌ای انتخاب شده است. طبقات مورد استفاده برای طبقه‌بندی مدارس شامل نوع مدرسه (دولتی و غیردولتی)، جنسیت مدرسه (دخترانه، پسرانه، و مختلط)، استان (شش استان بزرگ شهر تهران، شهرستان‌های تهران، خراسان رضوی، خوزستان، اصفهان، و فارس، به همراه ۳ دسته دیگر از استان‌ها حاصل شده از گروه‌بندی) بودند (لاروشه، یونکاس و فوی<sup>۶</sup>، ۲۰۱۶). تعداد نمونه ۳۸۲۳ دانش‌آموز بود که از ۲۴۸ مدرسه انتخاب شده بود. از این تعداد ۱۸۶۳ نفر دختر (۴۸/۷ درصد) و ۱۹۶۰ نفر پسر (۵۱/۳ درصد) بودند. میانگین سنی دانش‌آموزان در زمان گردآوری داده‌ها ۱۰/۷ سال بود.

### ابزار پژوهش

1. Chiu & Seo
2. Dogan
3. Lee, Park, & Taylan
4. Akaike's Information Criterion (AIC)
5. Bayesian Information Criterion (BIC)
6. LaRoche, Joncas, & Foy

در این پژوهش از سؤالات علوم مطالعه تیمز ۲۰۱۵ در پایه چهارم استفاده شده است. در این درس ۱۹۶ سؤال وجود داشتند که ۱۱۳ سؤال آن چند گزینه‌ای و ۸۳ سؤال تشریحی هستند. از مجموع سؤالات تشریحی ۱۰ سوال ۲ نمره‌ای (چندبخشی) بودند که در مدل جی‌دینا به طور طبقات جداگانه در نظر گرفته شده و در نتیجه ۲۰۶ سؤال خروجی داده شدند. اعتبار و روایی سؤالات توسط مرکز بین‌المللی مطالعات محاسبه و گزارش شده است.

در کنار سؤالات، پرسشنامه‌های پیشینه‌کاوی برای گردآوری اطلاعات زمینه‌ای از دانش‌آموزان، والدین، مدیر و معلم کلاس نیز به کار رفت. در هر یک از این پرسشنامه‌ها اطلاعات متعددی در زمینه عوامل مؤثر بر یادگیری گردآوری شد. برای مثال، در پرسشنامه معلم درباره ویژگی‌های معلم (سن، سابقه، رشته و مدرک تحصیلی)، جو مدرسه، تأکید مدرسه بر موفقیت علمی، تعامل معلم با همکاران، رضایت‌شغلی معلم، ویژگی‌ها و شرایط دانش‌آموزان مدرسه، فعالیت‌های تدریس، محدودیت‌های تدریس به دلیل نیازهای دانش‌آموزان، زمان تدریس، خودکارآمدی تدریس، میزان درخواست از دانش‌آموزان حیت تدریس، استفاده از رایانه در جریان تدریس، موضوعات تدریس شده به کلاس، چگونگی تخصیص تکلیف به دانش‌آموزان، سنجش از دانش‌آموزان، مشارکت در دوره‌های ضمن خدمت و میزان آمادگی برای تدریس موضوعات درسی به کلاس پرسیده شد. هر یک از این موضوعات که با مجموعه‌ای از گویه‌ها سنجیده شد با لحاظ کردن مراحل مقیاس‌سازی به نمراتی تبدیل می‌شدند تا برای تبیین یادگیری مورد استفاده قرار گیرند.

### مدل شناختی تشخیصی

مدلی برای برای تحلیل داده‌های مطالعه تیمز مناسب است که هر دو نوع سؤال با پاسخ‌های دویخی و چندبخشی را بتواند تحلیل کند، وزن‌های نمونه‌گیری را در تحلیل‌ها دخیل کند، نمونه‌گیری ماتریسی موجود در ساختار داده‌ها را پوشش دهد، و در مورد درس علوم ارتباطات جبرانی خصیصه‌ها را عملیاتی کند. چنین قابلیت‌هایی تنها در سه مدل شناختی تشخیصی لوگ‌خطی، مدل تشخیصی کلی، و مدل جی‌دینا فراهم است. در مدل شناختی تشخیصی لوگ‌خطی تحلیل پاسخ‌های چندبخشی با افزار سؤالات چندبخشی به دو سؤال ممکن می‌شد، ولی به دلیل نیاز به دستورنویسی بسیار طولانی در نرم‌افزار MPLUS و تعداد زیاد سؤالات عملاً ناکارآمد بود. به ویژه آنکه برای اصلاح چندباره ماتریس Q احتمال بروز خطا در دستورات و در نتیجه عدم دسترسی به خروجی نامناسب به شدت بالا می‌رفت. مدل تشخیصی کلی با نرم‌افزار mdlm گزینه مطلوب دیگری بود که مشکلات مدل قبلی را نداشت. اشکالی کار با این مدل آن بود که در مدل‌های بالای ۱۰ خصیصه متوقف می‌شد. هر چند سازنده نرم‌افزار عقیده داشت که چنین مشکلی در نرم‌افزار وجود ندارد (ام. ون‌داویر، ارتباط شخصی، ۱۸ دسامبر، ۲۰۱۸) ولی نتایج بررسی اسکریپت، ویکینز و هین<sup>۱</sup> (۲۰۱۶) حاکی از آن بود که نرم‌افزار در شرایط بیش از ۱۰ خصیصه در هم‌گرایی مشکل پیدا می‌کند. در نتیجه مدل جی‌دینا به عنوان مدل تحلیلی این مطالعه در نظر گرفته شد و داده‌ها با نرم‌افزار R و بسته CDM تحلیل شدند. استفاده از بسته CDM به دلیل پرهیز از سرعت کُندتر و بی‌نظمی‌های اندک در برآوردهای خطای استاندارد است که در بسته GDINA گزارش شده است (راپ و ون‌ریجن، ۲۰۱۸).

### تدوین ماتریس Q

برای تولید ماتریس Q، ابتدای خصیصه‌های مورد انتظار علوم از دانش‌آموزان پایه چهارم مشخص شدند. این کار بر اساس چهارچوب سنجش تیمز ۲۰۱۵ (مولیس و همکاران<sup>۳</sup>، ۲۰۱۳) و برنامه درسی علوم تجربی در دوره ابتدایی انجام شد. علاوه بر این، دسته‌بندی خصیصه‌های علوم به انواع شایستگی‌های مورد انتظار از دانش‌آموزان بر اساس کار کبیری، طباطبایی و بازرگان (۱۳۹۶) صورت پذیرفته شده است. خصیصه‌های مورد انتظار به همراه توضیحات آن‌ها در زیر ارائه شده‌اند:

شایستگی‌های دانشی و کاربردی علوم تجربی:

● بازشناسی: تعیین بیانیه‌های دقیق در مورد دقایق علمی، روابط، و مفاهیم و همچنین تعیین کردن ویژگی‌ها یا خصوصیات ارگانیزم‌های علمی، مواد و فرایندها و تعیین کردن استفاده مناسب از تجهیزات و رویه‌های علمی و بازشناسی و استفاده از واژگان، نمادها، حروف اختصاری، واحدها و مقیاس‌های علمی؛

1. Skaggs, Wilkins, & Hein

2. van Rijn

3. Mullis, et al.

- توصیف: توصیف یا تعیین کردن توصیف‌گرهای خصوصیات، ساختارها و عملکرد ارگانیزم‌ها، مواد فیزیکی و فرایندها و پدیده‌ها؛
- استفاده از مدل‌ها: استفاده از نمودار یا سایر مدل‌ها برای نشان دادن دانش مفاهیم علوم، نشان دادن رابطه فرایند چرخشی یا سیستم، یا پیدا کردن راه‌حل برای مسائل علوم؛
- ارائه مثال: فراهم کردن یا تعیین مثال‌های ارگانیزم‌ها، مواد فیزیکی و فرایندها که ویژگی‌های مشخصی را دارند و توضیح جملاتی از حقایق یا مفاهیم با مثال‌های مناسب؛
- مقایسه و مقابله: تعیین یا توصیف مشابهت‌ها یا تفاوت‌های بین گروه‌هایی از ارگانیزم‌ها، مواد یا فرایندها و همچنین تمایزگذاری، طبقه‌بندی یا مرتب کردن تک تک موضوعات، مواد، ارگانیزم‌ها و فرایندها بر اساس ویژگی‌ها و خصوصیات معین؛
- شایستگی‌های مفاهیم سطح بالا در علوم تجربی:
- مرتبط کردن: ارتباط دادن دانش یک مفهوم علمی زیربنایی به خصوصیت یا رفتار مشاهده شده یا تفسیر شده، یا استفاده از موضوعات، ارگانیزم‌ها یا مواد؛

- تبیین پدیده‌ها: فراهم کردن یا تعیین کردن تبیینی برای مشاهده یا پدیده‌های طبیعی با استفاده از مفهوم یا اصل علوم؛
- تحلیل: تعیین کردن عناصر مسئله علمی و استفاده از اطلاعات، مفاهیم، روابط و الگوهای داده‌ای برای پاسخ‌گویی به سؤالات و حل مسائل؛

● استدلال علمی: استفاده از شواهد و ادراک علمی برای حمایت از منطقی بودن درباره تبیین‌ها، راه‌حل‌های مسائل و نتیجه‌گیری‌ها؛

● پیش‌بینی: استفاده از شواهد و ادراکات یا روابط مفهومی برای پیش‌بینی کردن درباره اثر تغییرات در شرایط بیولوژیکی یا فیزیکی؛

شایستگی‌های کاوشگری علمی در علوم تجربی:

- کاوشگری: کاوشگری علمی به فعالیت‌های دانش‌آموزان در توسعه دانش و درک از ایده‌های علمی اشاره دارد (شورای ملی پژوهش آمریکا، ۱۹۹۶) و به دنبال درک دنیای طبیعی از طریق تدوین و آزمودن نظریات و به‌کارگیری آن در مسائل عملی است. به‌خصوص زمانی که دانش قبلی در موضوعی محدود باشد، کسب اطلاعات از طریق کاوشگری بسیار سودمند است. به همین دلیل در بیشتر برنامه‌های درسی همانند برنامه درسی انگلیس (راتکلیف<sup>۱</sup>، ۱۹۹۸)، چارچوب برنامه درسی آمریکا (شورای ملی پژوهش آمریکا<sup>۲</sup>، ۱۹۹۶) و ایران (دفتر برنامه‌ریزی و تألیف کتب درسی، ۱۳۸۶) به این شایستگی‌ها توجه ویژه‌ای شده است. فعالیت‌های دانش‌آموزان در توسعه دانش و درک از ایده‌های علمی به‌خصوص در مواقعی که دانش پیشینه‌ای در مورد موضوع محدود باشد، شامل زیرمهارت‌های فرضیه‌سازی و تعیین سؤالات، طراحی تحقیق، گردآوری اطلاعات، ارزشیابی شواهد، و تحلیل داده‌ها و نتیجه‌گیری.

برای تشکیل ماتریس Q دو نفر متخصص آموزش و برنامه‌ریزی، تک تک سؤالات علوم تیمز ۲۰۱۵ را بررسی و خصیصه‌های مورد نیاز برای پاسخ‌گویی آن‌ها را مشخص کردند و سپس پاسخ‌های خود را با هم‌دیگر مقایسه نمودند. توافق حاصل به عنوان ورودی ماتریس Q در نظر گرفته شد. با این حال، برای هماهنگ کردن ماتریس Q با پاسخ‌های دانش‌آموز، هر یک از سؤالات به‌طور جداگانه پس از اجرای مدل مورد بررسی قرار گرفت و حالت‌های دیگر مربوط به دخالت دادن یا کم کردن از خصیصه‌های احتمالی با اجرای چند باره مدل بررسی شد تا در این زمینه اطمینان حاصل شود. ملاک‌های مورد توجه برای این بخش برآزش سؤال و احتمال پاسخ متناظر به هر یک از خصیصه‌ها بودند.

### یافته‌ها

آخرین تغییرات ماتریس Q و تغییر مداوم درایه‌های آن برای به‌دست آوردن بهترین ترکیب مربوط به هر سؤال مشخص کرد که ۱۵۳ سؤال تنها در یک خصیصه و ۵۳ سؤال دیگر در دو خصیصه بار داشتند. برای اینکه دقت و تجانس طبقه‌بندی هر یک از خصیصه‌ها مشخص شوند، این دو آماره برای همه خصیصه‌ها با استفاده از روش بیشینه‌سازی پسین<sup>۳</sup> محاسبه شدند. نتیجه در جدول ۱ ارائه شده است. علاوه‌براین، همبستگی تتراکوریک بین خصیصه‌ها نیز محاسبه شد که نتیجه در جدول ۲ ارائه شده است.

1. Ratcliffe

2. National Research Council (NRC)

3. MAP

جدول ۱. دقت و تجانس طبقه‌بندی خصیصه‌های مورد بررسی در مدل

خصیصه‌ها	تعداد سؤالات	دقت طبقه‌بندی	تجانس طبقه‌بندی
استفاده از مدل‌ها	10	۰/۸۴۱	۰/۷۹۶
بازشناسی	46	۰/۸۶۹	۰/۸۳۹
توصیف	52	۰/۸۸۳	۰/۸۳۷
ارائه مثال	25	۰/۸۶۲	۰/۸۳۰
مقایسه	14	۰/۷۷۶	۰/۷۴۴
مرتبط کردن	7	۰/۸۰۱	۰/۷۷۲
تیبین پدیده‌ها	28	۰/۸۷۰	۰/۸۱۰
تحلیل	27	۰/۸۱۴	۰/۷۵۱
استدلال علمی	21	۰/۸۵۷	۰/۸۰۵
پیش‌بینی	13	۰/۸۲۴	۰/۷۵۳
کاوشگری علمی	16	۰/۷۸۸	۰/۷۴۵

جدول ۲. همبستگی بین خصیصه‌ها

خصیصه‌ها	مثال‌ها	بازشناسی	توصیف	ارائه مثال	مقایسه	مرتبط کردن	تیبین پدیده‌ها	تحلیل	استدلال	پیش‌بینی
استفاده از مدل‌ها	۱									
بازشناسی	۰/۷۳۷	۱								
توصیف	۰/۹۶۱	۰/۸۵۸	۱							
ارائه مثال	۰/۹۰۹	۰/۸۹۱	۰/۹۰۹	۱						
مقایسه	۰/۸۰۴	۰/۶۴۶	۰/۶۰۲	۰/۶۴۷	۱					
مرتبط کردن	۰/۹۳۷	۰/۵۲۲	۰/۸	۰/۷۵۹	۰/۹۰۳	۱				
تیبین پدیده‌ها	۰/۸۸۶	۰/۷۳۷	۰/۹۹۷	۰/۹۳۱	۰/۷۹	۰/۸۳۸	۱			
تحلیل	۰/۴۵۲	۰/۶۱	۰/۶۱۵	۰/۶۸۷	۰/۵۳۲	۰/۷۳۹	۰/۸۰۱	۱		
استدلال علمی	۰/۸۱۹	۰/۶۲۳	۰/۹۱۱	۰/۹۳۶	۰/۸۳۷	۰/۷۸۳	۰/۹۹۸	۰/۷۶۵	۱	
پیش‌بینی	۰/۴۳۳	۰/۴۵۴	۰/۷۴۹	۰/۹۷۳	۰/۳۸۸	۰/۵۴	۰/۶۶۸	۰/۷۶۲	۰/۷۱۷	۱
کاوشگری علمی	۰/۷۸۷	۰/۴	۰/۵۶۸	۰/۲۱۳	۰/۵۴۳	۰/۵۴۲	۰/۵۵۱	۰/۷۲۷	۰/۵۲۹	۰/۶

نتایج جدول ۱ نشان داده که دقت و تجانس خصیصه‌ها در حد قابل قبولی است. همچنین، بیشترین همبستگی در این جدول بین خصیصه‌های استدلال علمی و تیبین پدیده‌ها به‌طور علمی به میزان ۰/۹۹۸ است و کم‌ترین همبستگی مربوط به ارائه مثال و کاوشگری علمی به میان ۰/۲۱۳ است. سایر مقادیر کم همبستگی‌ها بیشتر بین خصیصه‌های دانشی و خصیصه‌های تفکر سطح بالا است و این نکته نشان می‌دهد که تفکیک بین خصیصه‌ها به خوبی صورت پذیرفته است. همچنین، میانگین همبستگی‌های بین خصیصه‌ها برابر با ۰/۷۰۹ بود.

برای بررسی برازش مدل از آماره‌های برازش مدل مطلق استفاده شد. شاخص‌های MADcor برابر با ۰/۰۵۸۵، شاخص SRMSR برابر با ۰/۰۷۶، و شاخص MAD Q3 برابر با ۰/۰۵۶۶ حاصل شد که با توجه نزدیکی به صفر و در نظر گرفتن تعداد سؤالات و خصیصه‌ها مناسب تلقی می‌شوند. در مورد شاخص‌های برازش سؤال نیز میانگین برازش سؤالات برابر با ۰/۰۸۶۱ با انحراف استاندارد ۰/۰۳۵۴ به دست آمد.



ارائه تک تک شاخص‌های برازش به علت محدودیت فضا ممکن نیست. کم‌ترین شاخص برازش سؤال ۰/۰۳۹ و بیشترین آن ۰/۱۹ بود. سؤالاتی که برازش آن‌ها برازش بالایی را نشان ندارد، یا تغییراتی در خصیصه‌های آنان داده شد یا از مجموع سؤالات کنار گذاشته شدند. یک از مواردی که در بررسی سؤالات مورد بررسی قرار گرفت، میزان احتمال هر یک از ترکیبات خصیصه‌های درگیر در سؤالات بود. این احتمالات در فضای مورد بررسی در مدل جی‌دینا نیز قابل بحث است که شامل عرض از مبدأ (A0 و A00)، اثرات اصلی (A1، A01 و A10) و اثر تعاملی (A11) بود. ارائه همگی نتایج احتمالات مربوط به سؤالات در اینجا میسر نیست ولی میانگین احتمالات عرض از مبدأها ۰/۲۹۵ و انحراف استاندارد ۰/۲۰۳، میانگین احتمالات اثرات اصلی ۰/۵۳۵ با انحراف استاندارد ۰/۳۱۲، و میانگین احتمالات اثرات تعاملی ۰/۶۲۱ با انحراف استاندارد ۰/۲۴۸ بود. برای آشنایی بیشتر آماره‌های ۱۰ سؤال منتخب در جدول ۳ ارائه شده است.

جدول ۳. احتمال پاسخ و برازش سؤال ۱۰ سؤال منتخب

سؤال	خصیصه‌ها	A0/A00	A1/A10	A01	A11	برازش سؤال
S01.01	بازشناسی	۰/۳۳۸	۰/۸۷۱			۰/۰۵۸
S01.05	بازشناسی	۰/۰۶۲	۰/۱۳۳			۰/۰۶۲
S01.10A	استدلال	۰/۰۰۱	۰/۱۳۸			۰/۰۵۶
S01.10B	استدلال	۰/۰۰۰	۰/۱۳۰			۰/۰۴۱
S02.01	مثال - مقایسه	۰/۳۸۲	۰/۷۰۵	۰/۱۳۹	۰/۸۰۶	۰/۰۸۴
S02.09	مرتبط کردن - استدلال	۰/۱۴۳	۰/۰۰۰	۰/۳۰۳	۰/۷۱۰	۰/۰۹۶
S03.06	توصیف - پیش‌بینی	۰/۰۷۵	۰/۴۶۱	۰/۶۲۱	۰/۶۸۴	۰/۰۶۹
S03.11	بازشناسی - استدلال	۰/۳۰۳	۰/۶۲۳	۰/۶۵۴	۰/۷۸۷	۰/۰۹۹
S06.01	تبیین	۰/۲۱۲	۰/۴۱۲			۰/۰۶
S06.12	بازشناسی - استدلال	۰/۲۰۱	۰/۴۹۷	۰/۶۱۶	۰/۶۶۰	۰/۰۸۲

با توجه به قابل قبول بودن مدل تشخیصی، میزان تسلط دانش‌آموزان ایران بر خصیصه‌های شناختی علوم در جدول ۴ ارائه شده است.

جدول ۴. احتمال تسلط دانش‌آموزان به خصیصه‌های مورد بررسی

خصیصه‌ها	احتمال تسلط
استفاده از مدل‌ها	۰/۴۹۰
بازشناسی	۰/۶۸۶
توصیف	۰/۵۲۲
ارائه مثال	۰/۶۱۷
مقایسه	۰/۴۸۴
مرتبط کردن	۰/۴۱۹
تبیین پدیده‌ها	۰/۳۹۷
تحلیل	۰/۳۶۶
استدلال علمی	۰/۴۲۷
پیش‌بینی	۰/۱۸۵
کاوشگری علمی	۰/۲۶۶

نتایج این جدول که نتایج پرسش اصلی این مطالعه را ارائه می‌دهد، نشان می‌دهد که بیشترین تسلط دانش‌آموزان در خصیصه بازشناسی، ارائه مثال، و توصیف است. این سه خصیصه، احتمال پاسخ بالاتر از ۰/۵۰ را کسب کرده‌اند. با این حال، بسیاری از پژوهشگرانی که در زمینه مدل‌های تشخیصی مطالعه کرده‌اند، مقادیر بالاتری از احتمال را در مقایسه با مقدار ۰/۵۰ برای قضاوت درباره تسلط بر خصیصه معرفی کرده‌اند. بر اساس توافق بین نظرات پژوهشگران، حد ۰/۶۵ به عنوان ملاک قضاوت درباره تسلط بر خصیصه مورد قبول قرار گرفته است. در نتیجه، دانش‌آموزان ایرانی تنها بر خصیصه بازشناسی تسلط پیدا کرده‌اند. از طرف دیگر، کم‌ترین احتمال تسلط دانش‌آموزان در خصیصه‌های پیش‌بینی، کاوشگری، و تحلیل دیده می‌شود. هر سه خصیصه در زمینه نیاز به تفکر سطح بالا مشترک هستند و بنابراین می‌توان ضعف دانش‌آموزان در بروز این نوع تفکر

را به عنوان علت کاهش احتمال خصیصه‌های این چینی در نظر گرفت. میزان وقوع ۱۰ مورد از شایع‌ترین کلاس‌های مکنون در جدول ۵ مطرح شده است.

جدول ۵. درصد وقوع ۱۰ کلاس مکنون شایع‌تر

کلاس مکنون	درصد وقوع کلاس مکنون
0000000000	۲/۲۲۹
1111111111	۰/۹۱۹
1111111100	۰/۶۵۷
1111111101	۰/۶۴۰
11111110100	۰/۶۳۰
1111000000	۰/۵۱۸
0101000000	۰/۴۴۶
0100100000	۰/۴۱۳
01110011110	۰/۳۰۵
01010001000	۰/۲۷۳

ترتیب خصیصه‌ها: کاوشگری علمی، پیش‌بینی، استدلال علمی، تحلیل، تبیین پدیده‌ها، مرتبط کردن، مقایسه، ارائه مثال، توصیف، بازشناسی، استفاده از مدل‌ها.

در این جدول به‌جز دو کلاس مکنونی که دانش‌آموزان در همه خصیصه‌ها مسلط یا غیرمسلط هستند، کلاسی که دانش‌آموزان در همه خصیصه‌ها به‌جز کاوشگری علمی و پیش‌بینی مسلط نیستند بیش از بقیه کلاس‌ها است (در ۰/۶۵۷ درصد از نمونه). در دو کلاس شایع بعدی هم یا تنها خصیصه پیش‌بینی به حد تسلط نرسیده یا به غیر از پیش‌بینی، خصیصه‌های کاوشگری علمی و تحلیل هم به حد تسلط نرسیده و بقیه به حد تسلط رسیده است. این مورد نیز شاهد دیگری برای ادعای قبلی است که دانش‌آموزان ایرانی در به تسلط رسیدن خصیصه‌هایی با دشواری روبه‌رو هستند که تفکر سطح بالایی را می‌طلبند.

### بحث و نتیجه‌گیری

این پژوهش برای شناسایی میزان تسلط دانش‌آموزان پایه چهارم بر خصیصه‌های شناختی علوم تجربی طرح‌ریزی شده است. انتخاب پایه چهارم و درس علوم به دلیل کمبود مطالعاتی است که در این زمینه در جامعه دانش‌آموزان ایرانی صورت پذیرفته است. همچنین، به دلیل نیاز به داده‌های مرتبط با جامعه دانش‌آموزان ایران، از نمونه مطالعه تیمز استفاده شده است که معرف جامعه دانش‌آموزان ایرانی است. برای اینکه اطلاعات کامل‌تری در مورد خصیصه‌های شناختی جدا از اطلاعات تولیدشده در گزارش اصلی مطالعه تیمز فراهم گردد، از سنجش شناختی تشخیصی استفاده شد و به دلیل لزوم پوشش انواع سؤالات دوبخشی و چندبخشی، تعداد نسبتاً زیاد خصیصه‌ها، لزوم توجه به ویژگی‌های داده‌های مطالعه تیمز هم‌چون وزن‌های نمونه‌گیری و دسترسی به نرم‌افزار قابل اعتماد از مدل شناختی تشخیصی جی‌دینا استفاده شد.

نتایج نشان داد که مدل، قدرت تولید اطلاعات تشخیصی را دارد. در مقایسه احتمال پاسخ خصیصه‌ها مشخص شد که سه خصیصه بازشناسی، ارائه مثال، و توصیف احتمال بالاتری نسبت به سایر خصیصه‌ها نشان داده و از ۰/۵۰ بیشتر هستند. با این حال، تنها خصیصه بازشناسی از ملاک مقبول ۰/۶۵ بیشتر بود و بنابراین، می‌توان استنباط کرد که دانش‌آموزان پایه چهارم ایران تنها در بازشناسی به تسلط رسیده‌اند. از طرف دیگر، دو خصیصه پیش‌بینی و کاوشگری علمی کم‌ترین احتمال پاسخ را داشتند.

تفاوت بین میزان تسلط دانش‌آموزان بر خصیصه‌های مبتنی بر دانستن و خصیصه‌های مبتنی بر تفکر و تأمل که می‌توان آن را خصیصه‌های تفکر سطح بالا دانست یکی از یافته‌های اصلی این مطالعه است. پیش‌تر نتایج مشابهی در مورد دانش‌آموزان پایه هشتم ایرانی در تیمز ۲۰۱۱ حاصل شده بود که در آن دانش‌آموزان ایرانی در خصیصه‌های استفاده از علوم و استفاده از مدل‌ها احتمال تسلط بیشتری نسبت به خصیصه‌های تفکر سطح بالا شامل کاوشگری علمی، تبیین پدیده‌ها، و استدلال نشان داده بودند (کبیری و همکاران، ۲۰۱۷). به همین دلیل شاید ابتدایی‌ترین استدلال ربط دادن احتمال تسلط پایین دانش‌آموزان در خصیصه‌های سطح بالا به دشواری ذاتی این‌گونه خصیصه‌ها باشد. مشابه چنین نتیجه‌ای

در ریاضی توسط مینایی و همکاران (۱۳۹۳) گزارش شده است به طوری که کم‌ترین احتمال تسلط در دانش‌آموزان پایه هشتم ایرانی در کاربردهای محاسباتی، استدلال منطقی، و ارزیابی و بررسی صحت گزینه‌ها گزارش شده که از سطح شناختی بالاتری نسبت به سایر خصیصه‌ها برخوردارند. از طرف دیگر، نتایج پژوهش چن و همکاران (۲۰۰۸) الف و ب) نشان داده که دانش‌آموزان چین تاییه در خصیصه‌های دربرگیرنده سطح بالای تفکر ریاضی به همان میزان خوب بودند که در خصیصه‌های دانشی به خوبی عمل کرده بودند؛ به طوری که کم‌ترین میزان احتمال در تشخیص الگوها (با احتمال ۰/۶۵) گزارش شد که احتمال کم‌تری نسبت به خصیصه‌های سطح بالا هم‌چون استدلال منطقی به‌شمار می‌آید. نظیر چنین یافته‌ای در پژوهش دوگان و تاتسواکا (۲۰۰۸) دیده شد که دانش‌آموزان پایه هشتم هر دو کشور ترکیه و آمریکا در خصیصه شناسایی و گسترش الگوها و روابط کم‌ترین میزان احتمال را نشان دادند، در حالی که انتظار می‌رفت در خصیصه‌ای هم‌چون حل مسائل جدید و ناآشنا که منطقی‌دشوارتر به نظر می‌رسند و مستلزم عملیات ذهنی پیچیده‌تر هستند، احتمال کم‌تری حاصل آید.

بنا به توضیحات فوق می‌توان نتیجه گرفت که دانش‌آموزان ایرانی در خصیصه‌های مستلزم تفکر سطح بالا نسبت به سایر خصیصه‌ها از ضعف برخوردارند که این ضعف قابل تعمیم به الگوها و نیمرخ‌های خصیصه‌ای کشورهای دیگری که مورد بررسی قرار گرفته‌اند، نیست. از آنجا که برنامه درسی علوم به بیشتر خصیصه‌های دانشی توجه نشان داده است (کبیری، قاضی طباطبایی و بازرگان، ۱۳۹۶)، بنابراین، احتمالاً نوع آموزش دانش‌آموزان در علوم تجربی به سمت دانش محوری سوق دارد تا پرورش تفکر دانش‌آموزان. به عبارت دیگر، معلمان تمایل دارند که مفاهیم علوم را به سمت ارائه مفاهیم دانشی پیش ببرند و انتظارات خود را از دانش‌آموزان به دانستن برخی از مفاهیم و اصطلاحات محدود نمایند. پاسخ‌هایی که از پرسشنامه‌های معلمان همین دانش‌آموزان به‌دست آمده است تا اندازه‌ای گویای اتفاقی است که در کلاس‌های درس اتفاق می‌افتد؛ در حالی که معلمان ۷۹/۵ درصد از دانش‌آموزان گوش دادن به توضیحات معلم در مورد محتوای جدید علوم، ۵۷/۴ درصد مطالعه کتاب درسی یا سایر منابع آموزشی، و ۴۲/۱ درصد حفظ اصول و مفاهیم پایه و اساسی را در «تمام درس‌ها یا تقریباً در همه درس‌ها» گزارش داده بودند، از طرف دیگر معلمان ۳۴/۱ درصد از دانش‌آموزان تفسیر داده‌های حاصل از یک آزمایش یا تحقیق را در تمام درس‌ها از دانش‌آموزان درخواست کرده و معلمان ۳۶ درصد از دانش‌آموزان راحتی و اطمینان «خیلی زیادی» در رشد و توسعه مهارت‌های سطوح بالاتر تفکر در دانش‌آموزان را گزارش داده بودند (کم‌ترین میزان در بین سایر مهارت‌های معلمی). این نکته نشان می‌دهد که تأکید بیشتری از طرف معلمان بر انتقال دانش نسبت به پرورش مهارت‌های دربرگیرنده تفکر وجود دارد؛ چنانچه معلمان تنها ۹ کشور از ۴۷ کشور شرکت‌کننده در پایه چهارم مطالعه تیمز میزان بالاتری از گوش دادن به توضیحات معلم در مورد محتوای جدید علوم و معلمان ۱۰ کشور میزان بالاتر درخواست از دانش‌آموزان برای حفظ اصول و مفاهیم پایه و اساسی را گزارش داده بودند. علاوه‌براین، خود معلمان نیز برای تغییر تدریس به سمت پرورش مهارت‌های تفکر سطح بالا با دشواری‌هایی روبه‌رو هستند؛ به طوری که کم‌ترین میزان اعتماد به مهارت‌های تدریس در بین سایر موارد در همین موضوع وجود داشت.

توجه به کاوشگری علمی به عنوان یکی از وجوه ممیزه برنامه درسی جدید علوم تجربی نسبت به برنامه‌های پیشین علوم است تا جایی که در برنامه جدید کاوشگری علمی به‌طور مفصل توضیح داده شده است. با وجود این، به نظر می‌رسد این حجم از تأکیدات توسط معلمان و مدارس چندان جدی گرفته نشده است. پاسخ‌های معلمان پایه چهارم نشان می‌دهد که معلمان ۵۹/۲ درصد از دانش‌آموزان در همه دروس علوم از شاگردان خود می‌خواهند به آن‌ها هنگام نشان دادن یک آزمایش یا تحقیق نگاه کنند، در مقابل معلمان ۴۳/۲ درصد از دانش‌آموزان انجام یک آزمایش یا تحقیق را از دانش‌آموزان خود در همه دروس علوم گزارش داده‌اند. همین نکته نشان می‌دهد که توجه به کنشگری دانش‌آموز که امری مهم در جریان کاوشگری علمی به‌شمار می‌رود، عملاً توسط معلمان یا نادیده گرفته شده و یا کم‌تر مورد توجه است، به شکلی که کاوشگری به امری تحت کنترل و برنامه‌ریزی معلم تنزل یافته است که قطعاً با خواست برنامه‌ریزان درسی در تناقض است. به عنوان شاهد دیگر به مقایسه مجهز بودن مدارس ابتدایی به آزمایشگاه علوم در دو سال ۲۰۱۱ و ۲۰۱۵ توجه کنید: مدیران ۴۸ درصد از دانش‌آموزان پایه چهارم ایران در سال ۲۰۱۱ بیان کرده بودند که مدرسه آنان آزمایشگاه دارد ولی همین شاخص در سال ۲۰۱۵ به ۳۷ درصد کاهش می‌یابد. این در حالی است که در این چهار سال مدارس جدیدتر ساخته شده‌اند که فرض می‌شود برخورداری از آزمایشگاه‌ها به عنوان یکی از استانداردهای آن مورد توجه بوده است. جدا از امکانات سخت‌افزاری در بین این دو سال اهتمام به کاوشگری نیز افول داشته است؛ در سال ۲۰۱۱ معلمان ۸۰ درصد از دانش‌آموزان بیان کرده بودند که در بیش از نیمی از دروس علوم بر کاوشگری علمی تأکید داشته‌اند، در حالی که این میزان در سال ۲۰۱۵ به ۷۲ درصد کاهش پیدا کرده است. همه این اتفاقات در حالی است که تغییرات کتاب‌های درسی طی این مدت بر اساس برنامه‌ای تدوین شده است که کاوشگری علمی را به‌طور ویژه ارج نهاده است.

با توجه به مطالبی که گذشت به نظر می‌رسد تفاوت بین برنامه مصوب و برنامه اجرا شده، حداقل در آموزش علوم جدی و قابل اعتنا باشد. لذا توصیه می‌شود توجهات مسئولان برنامه‌های درسی وزارت آموزش و پرورش به سمت کاهش این فاصله از طریق نزدیک کردن شرایط اجرای برنامه‌ها به سمت موارد مطلوب برنامه‌های مصوب اقدام کنند. جدی‌تر کردن نظارت بر چگونگی اجرای برنامه‌ها می‌تواند از جمله این اقدامات باشد. علاوه بر این، اجرای دوره‌های ضمن خدمت موثر به شکلی که معلمان را هم از لحاظ اجرایی و هم از لحاظ نگرشی آماده کنشگری مؤثر در شرایط کلاس درس نماید نیز از جمله اقدامات خواهد بود که با این تمهیدات ضعف‌های احتمالی برگزاری و تأثیرگذاری دوره‌های آتی رفع گردد.

## References

- Afzali. F., Delavar. A., Falsafinezhad. M.R., Farrokhi. N.A., & Boriali. A. (2014). Using CDMs in determining the essence of gender differences of mathematic performance in first grade high school, *Psychological Achievements*, 21(4), 89-104. [in Persian]
- Birenbaum. M., Tatsuoka. C., & Xin. T. (2005). Large-scale diagnostic assessment: comparison of eighth graders' mathematics performance in the United States, Singapore and Israel. *Assessment in Education*, 12(2), 167-181 .
- Chen. Y. H., Gorin. J. S., Thompson. M. S., & Tatsuoka. K. K. (2008). Cross-cultural validity of the TIMSS-1999 mathematics test: Verification of a cognitive model. *International Journal of Testing*, 8(3), 251-271.
- Chen. Y. H., Gorin. J. S., Thompson. M. S., & Tatsuoka. K. K. (2008). An alternative examination of Chinese Taipei mathematics achievement: Application of the rule-space method to TIMSS 1999 data. In M. v. Davier & D. Hastedt (Eds.), *Issues and Methodologies in Large-Scale Assessments* (Vol. 1, pp. 23-49). Hamburg/ Princeton: IEA-ETS Research Institute.
- Chen. Y. H., Ferron. J. M., Thompson. M. S., Gorin. J. S., & Tatsuoka. K. K. (2010). Group comparisons of mathematics performance from a cognitive diagnostic perspective. *Educational Research and Evaluation*, 16(4), 325-343 .
- Chiu. C. Y & ..Seo. M. (2009). Cluster analysis for cognitive diagnosis: An application to the 2001 PIRLS reading assessment. In M. v. Davier & D. Hastedt (Eds.), *Issues and Methodologies in Large-Scale Assessments* (Vol. 2, pp. 137-159). Hamburg: IEA-ETS Research Institute .
- Curriculum Center in Iran. (2007). Guide of science curriculum in primary and middle high- school, Tehran: Organization of Educational Research and Planning. [in Persian]
- DeCarlo. L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, 35(1), 8-26 .
- de La Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2), 179-199.
- de la Torre. J., & Minchen. N. D. (2019). The G-DINA Model Framework. In M. von Davier & Y. Lee (Eds.), *Handbook of diagnostic classification models: Models and model extensions, applications, software packages* (pp. 155-170). Switzerland, Cham: Springer.
- Dogan. E., & Tatsuoka. 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.
- George. A. C., & Robitzsch. A. (2015). Cognitive diagnosis models in R: A didactic. *The quantitative methods for psychology*, 11(3), 189-205.
- Henson. R. A., Templin. J. L., & Willse. J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74(2), 191-201.
- Im, S., & Park. H. J. (2010). A comparison of US and Korean students' mathematics skills using a cognitive diagnostic testing method: Linkage to instruction. *Educational Research and Evaluation*, 16(3), 287-301.
- Kabiri. M., Ghazi-Tabatabaei. M., Bazargan. A., Shokoohi-Yekta. M., & Kharrazi. K. (2017). Diagnosing competency mastery in science: An application of GDM to TIMSS 2011 data. *Applied Measurement in Education*, 30(1), 27-38 .
- Kabiri. M., Ghazi Tabatabae. M., & Bazargan. A. (2017). Specification of key competencies for eight graders in science: Conformity with of science curriculum requirements in Iran, *Journal of Curriculum Studies*, 44, 109-144. [in Persian]
- Kim. Y. H. (2011). Diagnosing EAP writing ability using the reduced reparameterized unified model. *Language Testing*, 28(4), 509-541 .

- Javidanmehr, Z. & Anani Sarab, M.R. (2018). An investigation of the prevalence and difficulty of reading comprehension's sub-skills by the G-DINA model, *Critical Language and Literary Studies*, 19, 99-117. [in Persian]
- LaRoche, S., Joncas, M., & Fov, P. (2016). Sample design in TIMSS 2015. In M. O. Martin, I. V. S. Mullis, & M. Hooper (Eds.). *Methods and Procedures in TIMSS* (Vol. 2016, pp. 3.1-3.37). Boston: TIMSS & PIRLS International Study Center.
- Lee, Y. S., Park, Y. S., & Tavlan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and the US national sample using the TIMSS 2007. *International Journal of Testing*, 11(2), 144-177.
- Minaei, A., Delavar, A., Falsafinezhad, M., Kiamanesh, A., & Mohaier Y. (2014). Cognitive diagnostic modeling of Iranian eight grade student to mathematics items of TIMSS 2007 using reduced noncompensatory reparameterized unified model and comparison between girls and boys, *Quarterly of Educational Measurement*, 16, 130-170. [in Persian]
- Mohsenpour, M., Goova, Z., Shokuhi Yekta, M., Kiamanesh, A., & Bazargan, A. (2015). A diagnostic test for math literacy cognitive competencies, *Quarterly Journal of Educational Innovations*, 53, 7-33. [in Persian]
- Mohsenpour M. (2019). Assessing polytomous cognitive attributes of mathematics literacy of 9th grade students: Supplying PGDINA model, *Educational Measurement and Evaluation Studies*, 26, 109-134. [in Persian]
- Moghadam, A., Falsafinehad, M., Farokhi, N., & Estaii, M. (2016). Diagnostic analysis of general English reading comprehension's items of PhD entrance exam using non-compensatory fusion model, *Quarterly of Educational Measurement*, 22, 41-68. [in Persian]
- Mullis, I. V. S., & Martin, M. O. (2013). *TIMSS 2015 assessment frameworks*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- NRC. (1996). *National science education standards*. Washington, DC: National Academies Press.
- Rahimi, R., Younesi, J., & mokarami, M. (2018). Application of a cognitive diagnostic assessment to analysis English comprehension's items of master's degree, *Quarterly of Educational Measurement*, 32, 17-40. [in Persian]
- Rahemei, R., Delavar, A., Younesi, J., & Naserei, Z. (2020). Choosing an appropriate cognitive diagnostic model for reading comprehension tests: A case study of the graduate entrance exam of the English language, *Educational Measurement and Evaluation Studies*, 29, 205-227. [in Persian]
- Ranibaran, F., & Alavi, S. M. (2017). Cognitive diagnostic assessment of a reading comprehension test battery for formative diagnostic feedback, *Journal of Foreign Language Research*, 6(3), 321-342. [in Persian]
- Ratcliffe, M. (1998). The purpose of science education. In M. Ratcliffe (Ed.), *ASE guide to secondary science education*. London: Association for Science Education.
- 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.
- Rupp, A. A., & van Riin, P. W. (2018). GDINA and CDM packages in R. *Measurement: Interdisciplinary Research and Perspectives*, 16(1), 71-77.
- Saadati, S., Moghadamzadeh, A., Minaei, A., & Geraminour, M. (2020). Differential item functioning in the framework of cognitive diagnostic assessment: Questions related to the differential and integral calculus of the Iranian national university entrance examination 2018, *Biquarterly Journal of Cognitive Strategies in Learning*, 15, 19-35. [in Persian]
- Shahmirzadi, N., Sivvari, M., Marashi H., & Geraminour, M. (2020). Test fairness analysis in reading comprehension PhD nationwide admission test items under CDA, *Journal of Foreign Language Research*, 10(1), 152-165. [in Persian]
- Sinharav, S., & Almond, R. G. (2007). Assessing fit of cognitive diagnostic models: A case study. *Educational and Psychological Measurement*, 67(2), 239-257 .
- Skaggs, G., Wilkins, J. L. M., & Hein, S. F. (2016): Grain Size and Parameter Recovery with TIMSS and the General Diagnostic Model, *International Journal of Testing*, 16(4), 310-330.
- Taghivan, H., Khodavee, F., Bazargan, A., Moghaddamzadeh, A., & Kabiri, M. (2018). Developing a reading test for sixth grade students using cognitive diagnostic assessment model, *Journal of Teaching Persian to Speakers from Other Languages*, 7(1), 3-30. [in Persian]

- Tatsuoka. K. K., Corter. J. E., & 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.
- von Davier. M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61(2), 287-307 .
- von Davier. M. (2009). Using the general diagnostic model to measure learning and change in a longitudinal large-scale assessment (RR-09-28). Retrieved from Princeton, NJ :
- Yamaguchi. K., & Okada. K. (2018) Comparison among cognitive diagnostic models for the TIMSS 2007 fourth grade mathematics assessment. *PLoS ONE*, 13(2): e0188691.