Educational Measurement and Evaluation Studies

National Organization for Educational Testing

Print ISSN: 2476-2865 Online ISSN: 2783-0942

Basic and Content Competencies in Mathematics for the Humanities Group: Predicting the Difficulty of Items in the National Exam

Enayatollah Zamanpour[1] , Abdolkarim Shadmehr[2], Reza Mohamadi[3] Shoeayb Qasemi[4]

1. Assistant Professor, Department of Assessment and Measurement, Allameh Tabataba'i University, Tehran, Iran, (Corresponding Author), Email: zamanpour@atu.ac.ir
2. Assistant Professor, National Organization of Educational Testing, Email: shadmehrabdolkarim@yahoo.com
3. Associate Professor, Faculty of Psychology and Educational Sciences, University of Tehran, Tehran, Iran. Email: remohammadis@ut.ac.ir
4. National Organization of Educational Testing, Tehran, Iran. Email: q.shoeayb@gmail.com

Article Info

Article Type:
Research Article

Received
Received in revised form
Accepted
Published online

ABSTRACT

Objective: The present study aims to identify the basic and content competencies effective in answering the mathematics items of the national university entrance exam for the Humanities group based on a cognitive approach, and to predict the level of difficulty of the items according to these competencies.

Methods: The study was conducted using a mixed method. The first population consisted of first-year university students, from which 3 individuals were selected. Another population included subject matter experts and item writers, with 8 individuals purposefully selected. The final population comprised the participants in two sessions of the 2023 national exam, with 71,730 randomly selected for the first session and 79,221 for the second session.

Results: Through student think-aloud protocols, literature review, and analysis of humanities group mathematics textbooks by subject matter experts and item writers, 13 basic competencies and 6 content competencies were identified. The competencies of applying algebraic rules and solving multi-step problems were the most frequently used in the first test, while understanding question complexity was the most frequently used in the second test. Basic competencies accounted for 56%, content competencies for 31%, and the combination of both for 64% of the variance in question difficulties.

Conclusion: Basic and content competencies outperform conventional methods in predicting question difficulty prior to exam administration. Their use in constructing the national exam can help minimize differences in question difficulty between the two sessions and contribute to the fairness of the exam.

Keywords: National Exam, Mathematics, Item Difficulty, Basic Competencies, Content Competencies

Cite this article: Zamanpour, Enayatollah; Shadmehr, Abdolkarim; Mohamadi, Reza; Qasemi, Shoeayb (2025). Basic and Content Competencies in Mathematics for the Humanities Group: Predicting the Difficulty of Items in the National Exam. *Educational Measurement and Evaluation Studies*, 15 (49):111-123pages.

DOI: 10.22034/emes.2025.2058071.2648

© The Author(s).

Publisher: National Organization of Educational Testing (NOET)

مطالعات اندازه‌گیری و ارزشیابی آموزشی

سازمان سنجش آموزش کشور

شاپا چاپی: ۲۸۶۵-۲۴۷۶ شاپا الکترونیکی: ۰۹۴۲-۲۷۸۳

شایستگی‌های پایه و محتوایی ریاضی گروه انسانی: پیش‌بینی دشواری سوال‌های آزمون سراسری

عنایت‌اله زمانپور[۱]، عبدالکریم شادمهر[۲] ، رضا محمدی[۳] و شعیب قاسمی[۴]

۱. استادیار سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران، ؛ (نویسنده مسئول)، رایانامه: zamanpour@atu.ac.ir

۲. استادیار، سازمان سنجش آموزش کشور، تهران، ایران. رایانامه: shadmehrabdolkarim@yahoo.com

۳. دانشیار روش‌ها و برنامه‌های درسی و آموزشی، دانشگاه تهران، تهران، ایران، رایانامه: remohammadis@ut.ac.ir

۴. سازمان سنجش آموزش کشور، تهران، ایران. Q.shoeayb@gmail.com

اطلاعات مقاله

نوع مقاله:
مقاله پژوهشی

دریافت:

اصلاح:

پذیرش:

انتشار :

چکیده

هدف: پژوهش حاضر با هدف شناسایی شایستگی‌های پایه و محتوایی مؤثر در پاسخگویی به سؤال‌های درس ریاضی کنکور سراسری گروه انسانی بر اساس رویکرد شناختی و تبیین میزان دشواری سؤال‌ها بر اساس این شایستگی‌ها است.

روش پژوهش: پژوهش به روش ترکیبی کیفی و کمی اجرا شد و چند جامعه داشت. جامعه نخست، دانشجویان سال اول دانشگاه بودند که ۶ نفر انتخاب شدند. جامعه دیگر شامل متخصصان موضوعی و طراحان سؤال بود که ۸ نفر به‌صورت هدفمند و با رویکرد تنوع حداکثری انتخاب شدند. آخرین جامعه آماری، پاسخنامه شرکت‌کنندگان در دو نوبت آزمون سراسری ۱۴۰۲ بود. برای نوبت اول ۷۱۷۳۰ و نوبت دوم ۷۹۲۲۱ به صورت تصادفی انتخاب گردید.

یافته‌ها: با بررسی شیوه تفکر بلند دانشجویان، پیشینه پژوهش و کتاب‌های درسی ریاضی گروه انسانی توسط متخصصان موضوعی و طراحان، ۱۳ شایستگی پایه و ۶ شایستگی محتوایی شناسایی شدند. شایستگی‌های کاربرد قواعد جبری و مسائل چندمرحله‌ای پرکاربردترین شایستگی‌ها در نوبت اول بودند و در نوبت دوم درک پیچیدگی سوال پرکاربردترین شایستگی بود. شایستگی‌های پایه ۵۶٪ ، شایستگی‌های محتوایی ۳۱٪ و مجموع هر دو شایستگی ۶۴٪ درصد واریانس دشواری سوالات را تبیین می‌کردند.

نتیجه‌گیری: شایستگی‌های پایه و محتوایی در مقایسه با روش‌های مرسوم در پیش‌بینی دشواری سوالات پیش از اجرا، عملکرد بهتری دارند و استفاده از آنها در ساخت آزمون سراسری می‌تواند به کمینه کردن تفاوت دشواری سوال‌های دو نوبت و عادلانه بودن آزمون کمک کند.

واژه‌های کلیدی: آزمون سراسری، ریاضی، دشواری سوال، شایستگی‌های پایه، شایستگی‌های محتوایی

استناد: زمانپور، عنایت‌اله؛ شادمهر، عبدالکریم؛ محمدی، رضا؛ قاسمی، شعیب (۱۴۰۴). شایستگی‌های پایه و محتوایی ریاضی گروه انسانی: پیش‌بینی دشواری سوال‌های آزمون سراسری . *مطالعات اندازه‌گیری و ارزشیابی آموزشی*، ۱۵ (شماره ۴۹)، ۱۱۱-۱۲۳ صفحه. DOI: 10.22034/emes.2025.2058071.2648

ناشر: سازمان سنجش آموزش کشور

حق‌مؤلف © نویسندگان.

مقدمه

پیشنهاد روشی برای تخمین دقیق دشواری نسبی آزمون‌ها قبل از اینکه آزمودنی آنها را پاسخ دهد، مدت‌ها است به‌عنوان یک هدف ارزشمند دست‌نیافتنی در ساخت آزمون مطرح بوده است. موضوع تشابه سؤال‌ها به‌لحاظ ویژگی‌های روان‌سنجی و به‌خصوص ضریب دشواری سؤال‌ها و آزمون، دغدغه‌ای جهانی است و متخصصان روش‌هایی را به‌منظور همتا کردن سؤال‌ها و در نهایت همتاسازی آزمون‌ها پیشنهاد داده‌اند که در اکثر آنها، باید سؤال‌ها قبل از اجرای نهایی به‌صورت آزمایشی، در مقیاسی کوچک اجرا شود که این امر مشکلات امنیتی و هزینه‌ای را به همراه دارد (بنتون[1]، ۲۰۲۰).

بررسی دشواری نسبی سؤال‌ها و آزمون‌ها بدون اجرای واقعی بر روی داوطلبان، در سازمان‌ها و در زمان‌های مختلف انجام شده است، به‌طور مثال اتالی[2] و همکاران (۲۰۱۴) در سازمان خدمات سنجش آموز شی (ETS) در آمریکا و در هلند، تو سط ون‌اونا[3] و همکاران (۲۰۱۹) در مؤ سسه سیتو[4] این کار انجام شده؛ اجرای آزمای شی و اعتباریابی سؤال‌های یک آزمون، در یک حیطه محدود نیز چندان مطلوب نی ست، چرا که بارِ آن بر دوش مدارس و دانش‌آموزان قرار می‌گیرد و نکته مهم‌تر آنکه، امنیت آزمون تحت‌تأثیر قرار می‌گیرد. اینکه بتوان آزمون‌هایی سـاخت که قبل از اجرای نهایی، راجع به دشواری سؤال‌های آن، به‌طور نسبتاً دقیق اظهارنظر کرد، علاوه بر اینکه منجر به افزایش امنیت آزمون می‌شود، هزینه‌های اجرای آزمای شی را نیز به شدت پایین می‌آورد. البته این نکته را باید مدّنظر دا شت که پیش‌بینی‌پذیری د شواری سؤال‌ها، به من صفانه بودن آزمون، به‌ویژه اگر از انواع آزمون‌های سرنو شت ساز[5] با شد (همانند کنکور سرا سری) نیز کمک شایانی به مؤ سسات مجری آزمون می‌کند. روش های گوناگونی به منظور برآورد د شواری پی شنهاد شده ا ست که مهمترین آنها به مبتنی بر قـضاوت متخصـصان بوده ا ست. روش قـضاوتی، ا ستفاده از طبقه‌بندی بلوم و قـضاوت مقایسه‌ای از جمله روش‌هایی هستند که برای برآورد د شواری ا ستفاده شده‌اند؛ با این وجود تلاش‌های تجربی تخمین دشواری سؤال‌ها به‌ندرت بیش از ۲۵ درصد واریانس را تبیین کرده است (المصری[6] و همکاران، ۲۰۱۷).

یکی از روش‌های پیش‌بینی ضرایب د شواری ا ستفاده از مهارت و توانایی‌های پایه و شناختی زیربنایی در سوال‌ها ا ست که می توان آنها را به صورت کلی شایستگی[7] نامید. ا ستفاده از شایستگی‌ها برای پیش‌بینی د شواری سوالات جزو روندهای پژوهشی جدید و امیدبخش در این حوزه است که پژوهش‌های اندکی بر روی آن انجام شده است (ترنر[8] و همکاران، ۲۰۱۳). ریاضی از حوزه‌های محتوایی است که به دلیل ماهیت خود، با مدل‌های تشخیصی شناختی[9] هم‌خوانی دارد و شناسایی شایستگی‌های پایه برای آن مناسب‌تر است، تحقیقات اولیه در شناسایی شایستگی‌های پایه نیز روی ریا ضیات انجام شده ا ست (تاتـسوکا[10]، ۲۰۰۹). از سوی دیگر درک ریا ضی در آمادگی افراد جوان برای زندگی اهمیت زیادی دارد (OCED[11]، ۲۰۱۳)، همچنین عملکرد مناسب در درس ریاضی نقش به‌سزایی در موفقیت در آزمون سراسری دارد.

تاکنون در ایران پژوهشی که به بررسی کارآمدی شایستگی‌ها در پیش‌بینی دشواری سوالات بپردازد اجرا نشده است. گذشته از کمبود مطالعات و نیاز به تحقیقات بی شتر جهت شناخت زوایای مختلف ا ستفاده از شای ستگی‌های پایه در تخمین د شواری سوال‌ها، یکی از مواردی که می‌تواند محرک نیاز به پژوهش در این زمینه با شد اجرای کنکور سراسری در خارج و داخل کشور است که نیاز به نسخه‌های همتا دارد. از این رو، با توجه به اهمیت پیش‌بینی ضرایب د شواری سؤال‌ها قبل از اجرای آزمون و ناکارامدی روش‌های مر سوم، پژوهش حا ضر به دنبال پا سخ به این م سئله پژوهشی است که مهارت‌ها و توانایی‌های حل مسائل و نیز سطوح محتوای سؤال‌ها یا به عبارت دیگر شایستگی‌های پایه و محتوایی کدام‌اند و تا چه اندازه می تواند ضرایب دشواری سؤالی آزمون ریاضی را تبیین کنند.

مبانی نظری و پیشینه پژوهش

برآورد دشواری سؤال‌ها به روش قضاوتی یکی از پرکاربردترین روش‌های برای تخمین دشواری دشواری سوال‌ها بدون اجرای ازمایشی است که مبتنی بر روش آنگوف[12] است. روش آنگوف مشهورترین روش‌های تنظیم استاندارد یا همان استانداردگزینی است. این روش، یک روش استانداردگزینی آزمون–محور است که ارزیابان به‌جای ارزیابی شرکت‌کنندگان، در مورد سؤال‌های آزمون و احتمال پاسخ به آنها نظر می‌دهند. در

1. Benton
2. Attali
3. Van Onna
4. Cito
5. high-stakes test
6. ElMasri.
7. competency
8. Turner
9. cognitive diagnostic models
10. Tatsuoka
11. Organization for Economic Cooperation and Development
12. Angoff

این روش، متخصصان موضوعی با ملاحظه حداقل کفایت[1] یا حداقل صلاحیت[2] یا مرزی[3] برای آزمودنی‌ها تصمیم می‌گیرند که احتمال پاسخ به تک‌تک سؤال‌ها چقدر است؛ این احتمال بین صفر تا یک برای هر سؤال ارزیابی می‌شود و میانگین کل نظرات متخصصان به‌عنوان نمره برش آزمون لحاظ می‌شود. (پیتونیاک و سیژک[4]، ۲۰۱۶). در پژوهش‌های اولیه‌ای که درباره برآورد دشواری سوال‌ها به روش قضاوتی به‌عمل آمده است، لورج و کروگلف[5] (۱۹۵۲، ۱۹۵۳) به این نتیجه رسیده‌اند که داوران ضمن بررسی یک سلسله از داده‌های تجربی، در ارزشیابی دشواری نسبی سؤال‌ها به یک توافق نسبی دست یافته‌اند. در ارزشیابی دشواری سؤال، هر داور دارای خطای ثابت مخصوص به خود بوده است و چنین می‌نماید که برای کنترل دشواری مطلق سؤال و درنتیجه میانگین نمره‌های آزمون، نمی‌توان به این داوری‌ها تکیه کرد (ثرندایک[6]، ۱۹۹۶). عمدتاً داوران تمایل به کم‌برآورد کردن دشواری سؤال‌های جدید دارند (رزیگالا[7]، ۲۰۲۴). بدون هیچ‌گونه آموزش در زمینه طراحی سؤال به طراحان، قضاوت‌های انجام‌شده در مورد سطح دشواری سؤال‌ها دچار مشکل است، که به بیش‌برآورد دشواری سؤال‌های ساده و زیربرآورد دشورای سؤال‌های دشوار می‌انجامد (ون د وترینگنگ و ون در ریجت[8]، ۲۰۰۶).

از روش‌های دیگری که از یک‌سو آزمون–محور و از سوی دیگر هنجار–مرجع باشند، می‌توان به پیش‌بینی ضرایب دشواری بر اساس طبقه‌بندی بلوم[9] اشاره داشت. روابط بین سطوح طبقه‌بندی بلوم، توانایی داوران آزمون برای پیش‌بینی دشواری سؤال‌های آزمون و ویژگی‌های روان‌سنجی واقعی سؤال‌های آزمون، به‌اندازه کافی موردمطالعه قرار نگرفته‌اند. این مقایسه نیز در مطالعات اندکی بررسی شده است که یکی از آنها پژوهشی بوده است که بر روی داده‌های حاصل از آزمون پیشرفت تحصیلی ده دانشکده پزشکی در برزیل انجام شده است؛ در این شیوه ویژگی‌های روان سنجی سؤال‌ها بر اساس سطح و زمینه‌های دانش آنها در طبقه‌بندی بلوم مورد ارزیابی قرار گرفت. اگرچه نتایج پژوهش، به‌منظور برآورد ضریب تمیز صرفاً در طبقات بالای بلوم، شاخص‌های بالاتری را گزارش کردند و راجع به دشواری سؤال‌های سطح بالا هم تخمین نسبتاً قابل‌قبولی ارائه شد، لیکن در مورد دیگر طبقات بلوم، نمی‌توان چنین انتظاری را داشت (هاماموتو فیلهو[10] و همکاران، ۲۰۲۰). این پژوهش همچنان نشان داد که نمی‌توان تمایزی بین دسته‌بندی سؤال‌ها بر اساس طبقه‌بندی و دشواری واقعی محاسبه شده قائل شد. به‌عبارت‌دیگر طبقه‌بندی بلوم نمی‌تواند دشواری سؤال‌ها را پیش‌بینی کند و پیش‌بینی داورها دقیق‌تر بوده است (المصری و همکاران، ۲۰۱۷).

قضاوت مقایسه‌ای [11] روشی است که بنتون (۲۰۲۰) به‌منظور تعیین دشواری‌های سؤال‌ها درنظر گرفته است که نسبت به روش انگوف به‌منظور تعیین دشواری سؤال‌ها مناسب‌تر عمل می‌کند (آتالی و همکاران، ۲۰۱۴). در این روش، دشواری سؤال‌ها به ترتیب درجه سختی سؤال به‌صورت بصری تعیین می‌شود؛ برای این منظور، دشواری تخمینی سؤال‌ها به‌صورت زوجی یا چهارتایی به داوران ارائه می‌شود و آنها بر اساس میزان سختی سؤال آنها را مرتب می‌کنند و براین‌اساس، همه سؤال‌ها رتبه‌بندی می‌شود و در نهایت، میزان دشواری سؤال‌ها برای همه سؤال‌ها بر اساس مقیاس رتبه‌بندی تعیین می‌شود. این اقدام بر اساس مدل بردلی–تری[12] یا مدل‌های مشابه انجام می‌شود و سپس دشواری‌های تعیین‌شده بر اساس قضاوت مقایسه‌ای با مقداری واقعی دشواری بر اساس مدل راش در نظریه سؤال–پاسخ یا روش‌های آماری ساده‌تر مقایسه شده است (بنتون، ۲۰۲۰). بنتون (۲۰۲۰) در پژوهش خود از داده‌های مربوط به اداره صلاحیت‌ها و مقررات آزمون‌ها[13] استفاده کرده است وی تحلیل خود را بر روی شش دسته سؤال‌های آزمون ریاضی انجام داده و به‌طورکلی، این نتیجه استنتاج شد که دشواری‌های قضاوت‌شده، پیش‌بینی‌کننده خوبی برای دشواری واقعی سؤال‌ها نیستند. ناکارآمدی روش‌های بیان شده، پژوهشگران را به استفاده از شایستگی‌های پایه برای پیش‌بینی سوق داده است.

در آزمون پیسا[14]، شش شایستگی پایه لحاظ شده است (ترنر و همکاران، ۲۰۱۳) که عبارت‌اند از: استدلال و برهان، رابطه، مدل‌سازی، بازنمایی، حل مسائل و استفاده از زبان و عملیات نمادین، رسمی و فنی. همان‌طور که ملاحظه می‌شود این شایستگی‌ها قرار نیست که به طور کامل از یکدیگر مستقل باشند، بلکه تا حدی با هم همپوشانی دارند و عمدتاً باید در فرایند حل مسائل ریاضی، به صورت مشترک فعال و مورداستفاده واقع شوند. هریک از شایستگی‌های شش‌گانه، خود دارای ۴ سطح بوده که در نخستین سطح، کمترین نیاز به آن شایستگی برای حل مسئله درنظر گرفته شده و سطح ۴ نیز دارای بالاترین میزان به‌کارگیری شایستگی مدّنظر بوده است. نتایج پژوهش ترنر و همکاران (۲۰۰۹)، شواهد کافی و قوی از تأثیر

1. minimally qualified
2. minimally competent
3. borderline
4. Pitoniak & Cizek
5. Lorge and Kurglov
6. Thorndike
7. Rezigala
8. van de Watering & van der Rijt
9. Bloom's taxonomy
10. Hamamoto Filho
11. comparative judgement
12. Bradley-terry model
13. the Office of Qualifications and Examinations Regulation
14. PISA

شایستگی‌های ریاضی، در تبیین تغییرپذیری میزان دشواری سؤال‌های ریاضی پیسا ارائه داشته است. علاوه بر این، در پژوهش مشابهی که ترنر و آدامز[1] (۲۰۱۲) بر روی داده‌های جدیدتر آزمون ریاضی پیسا و مشابه با همان تقسیم‌بندی شایستگی‌های پایه ریاضی در شش مؤلفه و هر کدام در ۴ سطح داشته انجام داده‌اند، نشان داده شد که می‌توان بین ۶۰ تا ۷۰ درصد واریانس دشواری سؤال را بر اساس شش شایستگی تبیین کرد.

روش پژوهش

پژوهش حاضر بر مبنای رویکرد روش‌های آمیخته[2] انجام شده است. مطالعه به روش‌های آمیخته، مستلزم گردآوری و تحلیل داده‌های کمّی و کیفی در یک مطالعه است (کرسول و همکاران[3]، ۲۰۰۳). در گام نخست پژوهش از راهبرد کیفی تحلیل مضمون به‌منظور شناسایی شایستگی‌های پایه استفاده گردید. در مرحله کیفی پژوهش جهت اعتباریابی از روش «همسوسازی[4]» استفاده شد. در بخش کمی، نخست برای تحلیل آماری داده‌ها از ضریب دشواری کلاسیک سوال بهره برده شده و در نهایت به منظور پیش‌بینی ضرایب دشواری بر اساس شایستگی های پایه و محتوایی از رگرسیون خطی چندگانه استفاده شد.

باتوجه‌به اینکه پژوهش حاضر چندمرحله‌ای است به ترتیب اجرا می‌توان چند جامعه برای آن متصور شد. جامعه اول سوال‌های ریاضی گروه انسانی در سال‌های مختلف است. کلیه سؤال‌های درس ریاضی گروه آزمایشی انسانی کنکور سراسری سال ۱۴۰۲ در هر دو نوبت مورد بررسی قرار گرفته و ویژگی‌های روانسنجی آنها محاسبه شده است. انتخاب سوال‌های سال ۱۴۰۲ به این دلیل بود که این سوال‌ها با آخرین تغییرات کتاب‌های درسی منطبق بود. درس ریاضی در هر نوبت دارای ۲۰ سوال است که در مجموع ۴۰ سوال مورد بررسی قرار گرفت. ضریب آلفای کرونباخ برای نوبت اول و دوم به‌ترتیب ۰/۵۶ و ۰/۶۲ بود. جامعه دوم داوطلبانی است که در سال ۱۴۰۲ در گروه آزمایشی انسانی در آزمون سراسری کارشناسی شرکت کرده‌اند. در این راستا ۶ نفر دانشجوی سال اول که در گروه علوم انسانی آزمون ۱۴۰۲ شرکت کرده و به سؤال‌های درس ریاضی پاسخ داده بودند (نمره درس ریاضی بالای ۵۰ درصد) به صورت هدفمند با رویکرد همگن انتخاب شدند. علاوه بر این از جامعه متخصصان موضوعی نیز به‌صورت هدفمند و با رویکرد نمونه‌های با تنوع حداکثری (سوری[5]، ۲۰۱۱) ۵ دبیر با تجربه تدریس درس ریاضی کنکور و هم نمونه‌ای از ۳ مدرس دانشگاهی در تخصص‌های ریاضی، آمار و آزمون‌سازی انتخاب شده است. در نهایت، جامعه آماری بخش کمی پژوهش، پاسخنامه شرکت‌کنندگان گروه آزمایشی انسانی است که در دو نوبت آزمون سراسری سال ۱۴۰۲ شرکت کرده‌اند. اطلاعات این نمونه به‌صورت گزارش ضرایب دشواری توسط سازمان سنجش در اختیار محققان قرار گرفت. در محاسبه ضرایب دشواری تعداد ۷۱۷۳۰ نفر در نوبت اول و ۷۹۲۲۱ نفر در نوبت دوم به‌عنوان نمونه به‌صورت تصادفی انتخاب و تحلیل‌ها بر روی آن‌ها انجام شده است تا برآوردی از دشواری سؤال‌ها بر اساس نظریه کلاسیک آزمون محاسبه و ارتباط آن با شایستگی‌های پایه و سرفصل محتوایی در ماتریس Q بررسی شود.

یافته‌ها

به‌منظور تدوین شایستگی‌های دخیل در حل مسائل ریاضی، نخست مطالعات جهانی این حوزه مورد مداقه قرار گرفت. در جدول ۱ فهرست شایستگی‌های ریاضی برخی از مطالعات گزارش شده است. براساس تحلیل صورت گرفته، شایستگی‌های شناسایی شده در مطالعات پیشین در دو گروه شایستگی‌های پایه و محتوایی قرار گرفتند. شایستگی‌های محتوایی مورد اندازه‌گیری در آزمون‌های مختلف، تفاوت قابل ملاحظه‌ای دارند، این امر تا حدود زیادی به پایه تحصیلی داوطلبان آزمون بستگی دارد. برای مثال در آزمون تیمز[6] (کورتر[7] و همکاران، ۲۰۰۶) مفاهیم پایه و عملیات در اعداد کامل و صحیح جزو شایستگی‌های محتوایی است و در آزمون داوطلبان معلمی (چاندیا[8] و همکاران، ۲۰۲۳) استفاده از حساب دیفرانسیل برای تجزیه و تحلیل توابع یکی از شایستگی‌ها در حوزه حساب است. برخی از شایستگی‌ها مانند استدلال و تفکر منطقی با عناوین متفاوت در بیشتر مطالعات مشاهده می‌شوند، این نوع شایستگی‌ها تحت عناوینی مانند توانایی پردازش شناختی، یا متغیرهای فرایندی قرار می‌گیرند و در این پژوهش شایستگی‌های پایه نام‌گذاری شدند.

جدول ۱. شایستگی های پایه و محتوایی حل مسائل ریاضی در پیشینه پژوهش

1. Turner & Adams
2. mixed methods approach
3. Creswell, Plano Clark, Gutman, & Hanson
4. triangulation
5. Suri
6. TIMSS
7. Corter
8. Chandia

مطالعه	آزمون	شایستگی‌ها
ترنر (۲۰۱۲، پیسا ۲۰۱۳)		**شش شایستگی (هرکدام در ٤ سطح):** ارتباط؛ ریاضی‌سازی؛ بازنمایی؛ استدلال و اثبات؛ ابداع راهبردها؛ استفاده از زبان نمادین، رسمی و فنی؛
کورتز و همکاران (۲۰۰٦)	تیمز پایه هشتم	**دانش محتوای ریاضی:** مفاهیم پایه و عملیات در اعداد کامل و صحیح؛ مفاهیم پایه و عملیات در کسر، اعشار و نسبت‌ها؛ مفاهیم پایه و عملیات در جبر مقدماتی؛ مفاهیم پایه و عملیات در هندسه؛ تکنیک‌های داده، احتمال و آمار مقدماتی؛ اندازه‌گیری و برآورد؛ **توانایی‌های پردازش شناختی:** تبدیل عبارات (تنه سؤال) از طریق فرمول‌بندی معادلات و عبارات جبری؛ کاربردهای محاسباتی دانش در حساب، جبر و هندسه؛ کاربردهای قضاوتی دانش حساب، جبر و هندسه؛ کاربرد قواعد پیچیده در جبر؛ استدلال منطقی؛ کاوش مسئله–فضا؛ ایجاد تصویرسازی یا تبدیل شکل‌ها و نمودارها؛ تشخیص و ارزیابی صحت ریاضیاتی؛ مدیریت داده‌های پیچیده، روندها، اهداف فرعی یا شرایط؛ خوانش منطقی و کمی؛ **مهارت‌های حل مسئله آزمون‌دادن:** تبدیل واحد؛ کاربرد خواص و روابط اعداد؛ استفاده از شکل، نمودار و جدول؛تقریب/برآورد؛ ارزیابی/صحت‌سنجی/ چک‌کردن حل‌ها؛ تشخیص الگوهای سریالی؛ استدلال نسبتی؛ سوالات بازپاسخ؛خواند/درک‌کردن مسئله‌های با متن پیچیده یا بافت جهان–واقعی؛
چاندیا و همکاران (۲۰۲۳)	آزمون داوطلبان معلمی	**خصیصه‌های محتوایی در پنج حوزه:** **اعداد** (۱. شناسایی ویژگی ها و معانی عناصر هر سیستم اعداد، ۲. درک ویژگی‌ها و ساختارهای سیستم‌های اعداد، و ۳. برقراری روابط سطح متوسط بین موضوعات عددی و جبری)؛ **جبر** (۱. استفاده از عبارات جبری و اجرای عملیات جبر ابتدایی، ۲. حل معادلات و نامعادله های خطی و درجه دوم)، ۳. درک مفهوم تابع، خواص آن و رابطه آن با معادلات)؛ **هندسه** (۱. استفاده از ویژگی‌های هندسی، ۲. انجام تبدیل‌های هندسی، ۳. درک هندسه اقلیدسی در صفحه دکارتی)؛ **آمار و احتمال** (۱. خواندن، سازماندهی و نمایش داده‌ها، ۲. استنتاج در مورد جمعیت‌ها بر اساس نمونه‌ها، ۳. انتخاب روش‌های آماری مناسب برای تحلیل داده‌ها، ٤. درک و کاربرد مفاهیم اولیه احتمالی)؛ **حساب** (۱. تفسیر و استفاده از محدودیت‌ها و توابع پیوسته، ۲. استفاده از حساب دیفرانسیل برای تجزیه و تحلیل توابع، ۳. استفاده از سری برای مدل سازی مسائل، ٤. استفاده از حساب انتگرال برای تعیین مساحت و حجم) ؛
تاتسوکا و بودو (۲۰۰۰)	بخش کمّی آزمون GRE	**متغیرهای محتوایی:** اعداد حسابی و عملیات اصلی؛ جبر مقدماتی؛ جبر متوسط؛ هندسه و هندسه تحلیلی؛ **متغیرهای فرایندی:** تبدیل مسائل متنی/کلامی؛ کاربرد دانش؛ کاربرد قواعد و الگوریتم‌ها؛ انتخاب قواعد و قضایا؛ استدلال و تفکر منطقی؛ تفکر تحلیلی و ساختاربندی مسئله؛ مهارت‌های آزمون‌دهی؛ درجه پیچیدگی؛ **متغیرهای بافتاری:** فرمت مقایسه‌ای؛ کار با شکل‌ها، جدول‌ها و نمودارها؛
سعادتی و همکاران (۲۰۲۰)	مباحث مشتق و انتگرال در آزمون سراسری گروه ریاضی	**۱۲ مهارت:** تعاریف و دانش مفاهیم و قواعد؛ مدل‌سازی تنه سؤال؛ کاربرد اعمال جبری مقدماتی سریع؛ کاربرد تکنیک‌های متداول؛ کاربرد اعمال جبری پیشرفته؛ کاربرد قضایا؛ خواندن اشکال و نمودارها؛ تجزیه مفاهیم؛ تحلیل مفاهیم؛ ترکیب مفاهیم؛ تفکر خلاق؛ تجسم فضایی؛

هم‌راستا با مطالعه پیشینه، سؤال‌های درس ریاضی آزمون گروه علوم انسانی در هر دو نوبت در اختیار ۶ دانشجوی سال اول قرار داده شد و از هر شش نفر آن‌ها خواسته شد آزمون‌های دو نوبت را حل کنند. این ۶ نفر به شیوه تفکر با صدای بلند سؤال‌ها را حل کردند. با تحلیل مضمون صوت مربوط به حل سؤال‌ها، شایستگی‌های شناختی شناسایی شد. مهمترین شایستگی‌های استخراج شده در این بخش شامل مهارت حل مسئله، فهم دقیق سؤال، به یادآوردن فرمول‌ها و قضایا، درک شکل و نمودارها، توانایی استدلال، مهارت‌های ریاضی و توانایی حل سؤال‌ها بر اساس گزینه‌ها. اکثر این شایستگی‌ها در پیشینه نیز با نام‌های مختلف نیز وجود داشت. ترکیبی از شایستگی‌های شناسایی شده حاصل از شیوه تفکر با صدای بلند به همراه شایستگی های حاصل از پیشینه مبنای کار قرار گرفت.

جدول ۲. شایستگی های پایه ریاضی آزمون سراسری کارشناسی گروه آزمایشی علوم انسانی

کد	نام شایستگی	تعریف
C1	مدل‌سازی جبری	دانش‌آموز بتواند یک مسئله متنی را به یک عبارت یا معادله جبری تبدیل کند که در آن تشخیص عبارت متغیر و ثابت (ها) لزوماً آشکار نیست. مدل‌سازی جبری گاهی سرراست است اما بیشتر اوقات نیازمند شناسایی متغیرها و روابطی است که به‌طور ضمنی در سؤال نهفته است.
C2	مدل‌سازی تجسمی	دانش‌آموز بتواند برای یک مسئله متنی یک شکل یا الگو ترسیم کند و با آن مسئله را مدل‌سازی کند. به‌طور مثال برای حل سؤال‌ها بتواند نمودارهای توابع پایه یا نمودار ون ترسیم کند.
C3	درک شکل، نمودار و جدول	دانش‌آموز بتواند شکل، نمودار و یا جدول ارائه‌شده در سؤال را درک کند و اطلاعات آن برای حل سؤال استفاده کند. مواردی که نیازمند بازتولید و دست‌کاری شکل، نمودار و یا جدول باشد در این شایستگی کدگذاری نمی‌شود.
C4	کاربردهای محاسباتی	دانش‌آموز بتواند عملیات حساب پایه شامل مهارت‌هایی مانند تقسیم دو کسر، به توان رساندن، گرفتن ریشه، محاسبه نسبت‌ها، تفضیل نسبت در صورت و مخرج و عملیات آمار پایه شامل محاسبه نما، میانه، میانگین، دامنه تغییرات، واریانس و انحراف معیار و عملیات هندسه پایه شامل محاسبه محیط، مساحت، حجم و غیره را انجام دهد.
C5	کاربرد قواعد جبری	دانش‌آموز بتواند با استفاده از قواعد جبری، با عبارت‌های جبری کار کند به‌طور مثال اعداد و یا متغیرها را در عبارت و معادله‌های جبری جایگزین کند.مقدار عددی یک تابع را محاسبه کند. همچنین بتواند از عبارت و یا معادله‌های جبری فاکتور بگیرد و انواع معادلات را حل کند.
C6	کاربرد قضایای ریاضیاتی	دانش‌آموز بتواند قاعده و یا قضیه مناسب برای حل سؤال‌های که استراتژی و فرایند حل آن‌ها روشن است را از یک دانش پایه وسیع جبر، هندسه و آمار تشخیص دهد و آن را به کار برد. به‌طور مثال از قواعد دمورگان، اصل جمع، اصل ضرب، متمم پیشامد در آمار. چنانچه حل سؤال مستلزم انتخاب استراتژی‌های مکنون یا ضمنی باشد در این شایستگی کدگذاری نمی‌شود. سؤال‌هایی که فقط مستلزم قواعد ارائه‌شده در C5 هستند در اینجا کدگذاری نمی‌شوند.
C7	مسائل چندمرحله‌ای	دانش‌آموز بتواند کار با مسائلی که شامل دو یا چند گام /هدف فرعی بوده و محاسباتی که نیازمند مدیریت نتایج میانی هستند را انجام دهد. این گام‌ها می‌تواند شامل هدف‌های فرعی ضمنی یا آشکار باشد و لزوماً مکنون نیستند. به‌عبارت‌دیگر دانش‌آموز باید بتواند: هدف‌های فرعی مسئله را به‌صورت گام‌به‌گام برای رسیدن به پاسخ نهایی پایه‌ریزی کند. به‌طور مثال در حل مسئله ریشه‌های معادله درجه‌دو ابتدا مجموع و حاصل‌ضرب ریشه‌ها را به دست آورد سپس به کمک اتحادها یک عبارت میانی را محاسبه کند و درنهایت حاصل عبارت داده‌شده در تنه سؤال را محاسبه کند.
C8	مهارت‌های آزمون‌دهی	دانش‌آموز بتواند از گزینه‌ها رو به عقب حرکت کند تا به یک جواب برسد این‌چنین دانش‌آموزی ممکن است سؤال را به یکی از این شیوه‌ها حل کند: حرکت رو به عقب از گزینه‌ها، تولید یک یا چند مثال و استنباط گزینه صحیح از آن‌ها؛ تحلیل گزینه‌ها، جایگذاری اعداد برای انتخاب گزینه صحیح؛ حذف یک یا چند گزینه و حدس از بین گزینه‌های باقی‌مانده، به دست آوردن پاسخ سؤال به روش حدس و آزمایش.
C9	استدلال و تفکر منطقی	دانش‌آموز بتواند سؤال‌ها را با استفاده از استدلال حل کند. مورد اصلی این خصیصه زمانی است که در حل مسئله، باید شرایط مختلفی در نظر گرفته شود، برای مثال بررسی شرایط $x>0$، $x<0$، $x=0$. دانش‌آموز بتواند شرایط لازم و کافی را درک کند و برای برقراری و اجرای فعالیت‌های حل مسئله مناسب از آن‌ها استفاده کند. سایر مهارت‌های الزامی شامل پیدا کردن یک مثال نقض؛ چک کردن اینکه حل در حالت کلی صدق می‌کند در مقایسه باحالتی که فقط در موارد خاص صادق است (یعنی چک کردن شرایط لازم و کافی)؛ فهمیدن شرایط برای حل کردن دستگاه‌ها و مسائل. این شایستگی در مواردی که شرایط به‌صورت آشکار در گزینه‌های سؤال‌های چندگزینه‌ای یا در تنه سؤال داده‌شده باشد (مانند مسائل مربوط به منطق و استدلال ریاضی)، الزاماً کدگذاری نمی‌شود.
C10	تفکر تحلیلی و ساختاربندی مسئله	دانش‌آموز بتواند سؤال‌ها را به مؤلفه‌ها یا عناصر تشکیل‌دهنده آن‌که آشکار نیستند یا مکنون و یا ضمنی هستند تجزیه کند؛ مؤلفه‌ها را به‌گونه‌ای سازمان دهد که سؤال قابل‌حل شود. مثلاً اینکه دانش‌آموز بداند وقتی تابع بر محور Xها مماس بوده به معنای داشتن ریشه مضاعف است و یا اینکه وقتی بیان می‌شود نمودار دو تابع فقط در یک نقطه یکدیگر را قطع می‌کنند بداند این عبارت معادل این است که حاصل تساوی دو معادله، ریشه مضاعف دارد.
C11	درک پیچیدگی سؤال	دانش‌آموز بتواند پیچیدگی سؤال ناشی از جملات پیچیده، طولانی، جملات سلبی و یا حتی استفاده از اصطلاحاتی مانند «حداقل»، «حداکثر»، «معادل»، «باید»، «می‌تواند (ممکن است)» که درک سؤال را سخت‌تر می‌کند را در حل سؤال در نظر بگیرد. برای مثال در بحث احتمال سؤال‌هایی که حداقل تعداد حالت‌های مطلوب خواستار هستند، در این شایستگی کدگذاری می‌شود.

کد	نام شایستگی	تعریف
C12	تبدیل و بازنمایی شکل و نمودار و جدول	دانش‌آموز بتواند نمودارها، اشکال هندسی و جدول‌های ارائه‌شده را برای حل مسئله دست‌کاری کند. مثلاً به کمک انتقال یک نمودار پایه نمودار جدید تولید کند یا با رسم خط در یک شکل شرایط استفاده از قضیه تالس را فراهم کند و یا معادله‌ها را به‌صورت نموداری حل کند.
C13	تفکر خلاق	دانش‌آموز بتواند از روش‌های مبتکرانه و فارغ از قواعد کلاس و کتاب درسی برای حل مسائل استفاده کند. به‌طور مثال استفاده از مثال عددی در تنه سؤال به‌منظور ارزیابی گزینه‌ها، استفاده از روش‌های مرسوم جبری در حل مسائل هندسی و یا برعکس.

در ادامه به‌منظور نهایی کردن شایستگی‌های پایه و محتوایی، شایستگی‌های به‌دست‌آمده از مطالعات پیشین به همراه شایستگی‌های حاصل از تفکر با صدای بلند دانشجویان و کتاب‌های درسی ریاضی گروه انسانی در مقطع متوسطه دوم در اختیار پنل متخصصان موضوعی قرار گرفت. در قدم اول شایستگی‌های پایه موردنیاز برای پاسخگویی به سؤال‌های درس ریاضی کنکور سراسری تعیین شدند. فرایند به این صورت بود که ابتدا یک فهرست اولیه از شایستگی‌های پایه که از مطالعات پیشین و تحلیل مضمون به‌دست‌آمده بود در اختیار اعضای پنل قرار گرفت، اعضای پنل به بحث در مورد هرکدام از شایستگی‌ها می‌پرداختند، برخی از شایستگی‌ها که با سطح آزمون سراسری انطباق نداشتند حذف یا بازنگری شدند، همچنین شایستگی‌های جدید بر اساس الزامات کنکور سراسری تعریف شد. پس از برگزاری ۴ جلسه، فهرست شایستگی‌های پایه نهایی و به تأیید متخصصان رسید که شامل ۱۳ شایستگی پایه بود. جدول ۲ این شایستگی‌ها را به همراه تعریف و برخی مصادیق آن نشان می‌دهد.

جدول ۳. شایستگی های محتوایی ریاضی آزمون سراسری کارشناسی گروه آزمایشی علوم انسانی

حیطه	محتوای پایه	شایستگی‌های محتوایی سطح ۱	شایستگی‌های محتوایی سطح ۲	شایستگی‌های محتوایی سطح ۳
جبر (A)	عبارت‌های جبری (A1)	A11 حل معادلات جبری درجه ۱	A12 حل دستگاه معادلات خطی؛ حل معادله درجه ۲ به روش دلتا، حل معادله با روش مربع کامل	A13 معادله و مسائل توصیفی (کاربرد معادله درجه ۲)؛ تجزیه عبارت جبری؛ معادله‌های شامل عبارت‌های گویا؛ محاسبه ریشه nام و توان گویا
	تابع (A2)	A21 مفهوم تابع و بازنمایی در قالب جدول، نمودار و ضابطهٔ جبری	A22 حل مسائلی که شامل برخی انواع توابع خطی (تابع ثابت، تابع چندضابطه‌ای، تابع همانی، تابع پلکانی، تابع علامت، تابع جزء صحیح، تابع قدر مطلق) باشد	A23 حل مسائلی که شامل برخی انواع توابع غیرخطی (تابع نمایی)
	مجموعه و دنباله‌ها (A3)	A31 مفهوم مجموعه و اعمال روی آن‌ها (پیدا کردن اجتماع و اشتراک دو مجموعه مشخص)	A32 حل مسائل مربوط به مجموعه‌ها، حل مسائل مربوط به الگوها و دنباله‌ها	
	منطق و استدلال ریاضی (A4)	A41 جدول ارزش و ترکیب گزاره‌ها	A42 استدلال ریاضی	
آمار و احتمال (S)	آمار (S1)	S11 محاسبه شاخص‌های گرایش به مرکز محاسبه خط فقر	S12 محاسبه شاخص‌های پراکندگی محاسبه تورم	S13 نمودارهای یک متغیّره و چند متغیّره حل مسائل چرخه آمار درون و برون‌یابی، سری زمانی
	شمارش و احتمال (S2)	S21 اصول شمارش (اصل جمع و ضرب)	S22 جایگشت، ترکیب، احتمال	

برای شناسایی شایستگی‌های محتوایی همانند شایستگی‌های پایه از نظرات متخصصان استفاده شده است. به دلیل تفاوت زیاد محتوای سوال‌های کنکور سراسری با محتوی ریاضی در مطالعات پیشین، مطالعات پیشین نقش کم رنگی در شناسایی شایستگی‌های محتوایی داشتند. اعضای پنل پس از برگزاری ۳ جلسه در مورد شایستگی‌های محتوایی به اتفاق‌نظر رسیدند. شایستگی‌های محتوایی (جدول ۳) در دو حوزه «جبر» و «آمار و احتمال» قرار گرفتند. حوزه جبر دارای چهار محتوای پایه است که در درون هر محتوای پایه سطح‌بندی از ساده به دشوار لحاظ شده است. برای مثال محتوای پایه «منطق و استدلال ریاضی» دارای دو سطح است، سطح اول «جدول ارزش و ترکیب گزاره‌ها» و سطح دوم «استدلال ریاضی» را شامل می‌شود.

جدول ٤. بخشی از ماتریس Q شایستگی‌های پایه ریاضی نوبت اول

شماره سؤال	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
۱	۰	۰	۱	۰	۰	۱	۰	۱	۰	۰	۰	۰	۰
۲	۱	۱	۰	۰	۱	۰	۱	۱	۰	۰	۱	۰	۰
۳	۰	۰	۰	۰	۰	۱	۰	۰	۱	۰	۰	۰	۰
٤	۰	۰	۰	۰	۰	۱	۱	۰	۰	۱	۱	۰	۰

پس از شناسایی شایستگی‌های پایه و محتوایی، ماتریس Q برای درس ریاضی دو نوبت آزمون سراسری سال ۱۴۰۲ تهیه شد. در ماتریس Q سوال‌ها در ردیف و شایستگی‌ها در ستون‌های ماتریس قرار می‌گیرند. در هر سوال برای درایه‌های ماتریس برای شایستگی‌هایی که در فرایند پاسخ‌دهی به آن سوال درگیر هستند عدد ۱ و سایر درایه‌ها عدد صفر وارد می‌شود. ابتدا هرکدام از متخصصان به صورت مستقل برای هر دو نوبت آزمون ماتریس Q اولیه تدوین کردند. ماتریس‌های Q اولیه در یک جلسه توسط متخصصان مورد بحث و بازبینی قرار گرفت و ماتریس‌های Q نهایی آماده شد. بخشی از ماتریس Q برای شایستگی‌های پایه نوبت اول در جدول ۴ ارائه شده است.

میانگین دشواری سوال‌ها در نوبت اول ۰/۰۸ و در نوبت دوم ۰/۰۷ بود. انحراف معیار ضرایب دشواری در نوبت اول (۰/۰۵) بیشتر از نوبت دوم (۰/۰۲) است، این امر تاحدودی ناشی از این است که کمترین ضریب دشواری در هر دو نوبت نزدیک به هم است (۰/۳۰ در نوبت اول و ۰/۰۴ نوبت دوم)، امّا بیشترین ضریب دشواری در نوبت اول ۰/۲۴ و در نوبت دوم ۰/۱۴ است. در ادامه برای به‌دست‌آوردن قدرت پیش‌بینی شایستگی‌های پایه و محتوایی ابتدا یک رگرسیون خطی چندگانه به طور مجزا اجرا شد. هرکدام از شایستگی‌های پایه به‌صورت یک متغیر مجازی صفر و ۱ وارد معادله رگرسیون شدند. کدگذاری شایستگی‌های محتوایی به‌صورت یک نمره وزنی (سطح اول × ۱ + سطح دوم × ۲ + سطح سوم × ۳ = نمره وزنی) برای هر محتوای پایه محاسبه شد. بنابراین ۱۳ متغیر پیش‌بین برای شایستگی‌های پایه و ۶ متغیر پیش‌بین برای شایستگی‌های محتوایی در معادله‌های رگرسیون وجود داشت. در هر تحلیل رگرسیون، ورود متغیرهای پیش‌بین به صورت هم‌زمان[1] بود.

جدول ۵. رگرسیون دشواری سوال‌های درس ریاضی روی شایستگی‌های پایه و محتوایی

مدل	F	درجه‌های آزادی	سطح معناداری	R	R^2	$Adj.R^2$
فقط شایستگی‌های پایه	۲/۵۵	۱۳، ۲٦	۰/۰۲۰	۰/۷٤	۰/٥٦	۰/۳٤
فقط شایستگی‌های محتوایی (وزنی)	۲/٤۷	٦، ۳۳	۰/۰٤٤	۰/٥٦	۰/۳۱	۰/۱۸
شایستگی‌های پایه و محتوایی (وزنی)	۱/۸٦	۱۹، ۲۰	۰/۰۸۸	۰/۸۰	۰/٦٤	۰/۳۰

جدول ٦. خلاصه ضرایب رگرسیون سؤال‌های گروه انسانی روی شایستگی‌های پایه و محتوایی

	شایستگی پایه				شایستگی محتوایی				شایستگی پایه و محتوایی			
	b	Beta	P	part	b	Beta	P	part	b	Beta	P	part
عرض از مبدأ	-۲/۱۲	–	۰/۰۰۱	–	-۲/۱۲	–	۰/۰۰۱	–	-۲/۱۲	–	۰/۰۰۱	–
C1	-۰/۳۰	-۰/۲٦	۰/۰۹۸	-۰/۲۲	–	–	–	–	-۰/۳۳	-۰/۲۹	۰/۱۲٥	-۰/۲۱
C2	۰/۱۱	۰/۰۹	۰/٦۰۳	۰/۰۷	–	–	–	–	۰/۱٦	۰/۱۳	۰/۴۹۷	۰/۰۹
C3	۰/۷۳	۰/۳۸	۰/۰۴۸	۰/۲۷	–	–	–	–	۰/٦۱	۰/۳۲	۰/۲۰۳	۰/۱۸
C4	-۰/۱۷	-۰/۱٦	۰/۳۴۰	-۰/۱۳	–	–	–	–	-۰/۱٦	-۰/۱٥	۰/۴٦۲	-۰/۱۰
C5	-۰/۳۳	-۰/۳۳	۰/۰۴۷	-۰/۲۷	–	–	–	–	-۰/۲۹	-۰/۲۹	۰/۱۰۰	-۰/۲۳
C6	-۰/۰٥	-۰/۰٥	۰/۷۹۷	-۰/۰۳	–	–	–	–	-۰/۰۸	-۰/۰۸	۰/۷۰٥	-۰/۰٥
C7	-۰/۲۱	-۰/۲۰	۰/۲۰۴	-۰/۱۷	–	–	–	–	-۰/۰۹	-۰/۰۸	۰/۶۹٦	-۰/۰٥

1. enter

C8	-۰/۱۴	-۰/۱۱	۰/۴۸۵	-۰/۰۹	–	–	–	–	-۰/۱۶	-۰/۱۳	۰/۴۵۷	-۰/۱۰
C9	-۰/۳۹	-۰/۳۶	۰/۰۳۰	-۰/۳۰	–	–	–	–	-۰/۴	-۰/۴۵	۰/۰۳۲	-۰/۳۱
C10	-۰/۲۴	-۰/۲۳	۰/۱۵۰	-۰/۱۹	–	–	–	–	-۰/۱۸	-۰/۱۶	۰/۳۴۸	-۰/۱۳
C11	۰/۲۲	۰/۲۱	۰/۱۶۵	۰/۱۹	–	–	–	–	۰/۱۴	۰/۱۳	۰/۴۳۲	۰/۱۱
C12	۰/۲۴	۰/۱۰	۰/۵۷۳	۰/۰۷	–	–	–	–	۰/۱۲	۰/۰۵	۰/۸۳۱	۰/۰۳
C13	-۰/۱۲	-۰/۰۸	۰/۶۷	-۰/۰۶	–	–	–	–	۰/۰۱	۰/۰۱	۰/۹۸۱	۰/۰۰
A1	–	–	–	–	-۰/۱۳	-۰/۴۲	۰/۰۱۰	-۰/۳۹	-۰/۰۳	-۰/۰۹	۰/۶۷۶	-۰/۰۶
A2	–	–	–	–	-۰/۱۱	-۰/۲۶	۰/۰۹۵	-۰/۲۵	-۰/۰۹	-۰/۲۳	۰/۱۶۹	-۰/۱۹
A3	–	–	–	–	-۰/۲۱	-۰/۲۶	۰/۱۰۳	-۰/۲٤	-۰/۱۵	-۰/۱۸	۰/۳۸۱	-۰/۱۲
A4	–	–	–	–	۰/۱۲	۰/۱۷	۰/۲۵۷	۰/۱۷	۰/۰۷	۰/۱۱	۰/۶۳۰	۰/۰۷
S1	–	–	–	–	۰/۰۶	۰/۱۳	۰/۴۱۲	۰/۱۲	۰/۰۵	۰/۱۱	۰/۶۶۳	۰/۰۶
S2	–	–	–	–	-۰/۲۱	-۰/۲۰	۰/۲۰۶	-۰/۱۹	۰/۰۱	۰/۰۱	۰/۹۹۴	۰/۰۰

برای اطمینان از نرمال بودن توزیع ضرایب دشواری از آزمون شاپیرو–ویلک استفاده شد. نتایج نشان داد ضرایب دشواری دارای توزیع نرمال نیستند (W= ۰/۸۵، P<۰/۰۰۱)، لذا ضرایب دشواری با لگاریتم طبیعی تبدیل شدند که ضرایب تبدیل‌شده دارای توزیع نرمال بودند (W= ۰/۹۷، P=۰/۳۵۰). از این رو در تحلیل‌های رگرسیون از ضرایب تبدیل شده بهره گرفته شد. نتایج تحلیل‌های رگرسیون در جدول‌های ۵ و ۶ ارائه شده است. در هر تحلیل رگرسیون مفروضه‌های همگنی پراکندگی و نرمال بودن توزیع باقیمانده‌ها، استقلال مشاهدات و همخطی چندگانه بررسی و از برقرار بودن آنها اطمینان حاصل شد. به‌دلیل پایین بودن اندازه نمونه، در تفسیر ضرایب رگرسیون معناداری آماری آنها مدنظر قرار نگرفت.

همبستگی چندگانه بین شایستگی‌های پایه و ضریب دشواری سؤال‌ها برابر با ۰/۷۵ است که حاکی از میزان همبستگی تقریباً بالایی است و شایستگی‌های پایه به‌تنهایی توانایی برآورد ۵۶ درصد واریانس ضرایب دشواری سؤال‌ها را به عهده دارند. ضرایب رگرسیون استاندارد فارغ از علامت خود نشان از میزان اهمیت و تأثیر آن شایستگی بر پیش‌بینی ضرایب دشواری است. ضرایب مثبت این‌گونه تفسیر می‌شود که حضور آن شایستگی در یک سؤال باعث آسان‌تر شدن آن می‌شود و ضریب منفی باعث سخت‌تر شدن سؤال می‌شود. شایستگی «درک شکل، نمودار و جدول» با مقدار ۰/۳۸ بیشترین تأثیر را بر دشواری داشته است، سؤال‌های نیازمند این شایستگی ساده‌تر از سوال‌هایی هستند که نیازمند این شایستگی نیستند. شایستگی «استدلال و تفکر منطقی» بیشترین تأثیر منفی ۰/۳۶- را بر دشواری سوال‌ها داشته است و سوال‌های نیازمند این شایستگی برای داوطلبین دشوارتر بوده است.

همبستگی چندگانه بین شایستگی‌های محتوایی و ضریب دشواری سؤال‌ها برابر با ۰/۵۶ است که حاکی از میزان همبستگی تقریباً بالایی است و شایستگی‌های محتوایی به‌تنهایی برآورد ۳۱ درصد واریانس ضرایب دشواری سؤال‌ها را به عهده دارند. محتوای «منطق و استدلال ریاضی» با ۰/۱۷ بیشترین تأثیر مثبت را بر دشواری داشته است و سؤالات این مبحث ساده‌تر هستند. محتوای «عبارت‌های جبری» بیشترین تأثیر منفی ۰/۴۲- را بر دشواری سوال‌ها داشته است که نشان می‌دهد سوالات نیازمند این محتوا به‌طور معناداری دشوارتر از سوالاتی هستند که فاقد این محتوا هستند.

همبستگی چندگانه بین شایستگی‌های پایه و محتوایی با ضریب دشواری سؤال‌ها برابر با ۰/۸۰ است شایستگی‌های پایه و محتوایی در مجموع ۶۴ درصد واریانس ضرایب دشواری سؤال‌ها را تبیین می‌کنند. شایستگی «استدلال و تفکر منطقی» با ۰/۴۵- بیشترین تأثیر منفی و شایستگی «درک شکل، نمودار و جدول» با ۰/۳۲ بیشترین تأثیر مثبت را بر دشواری سوال‌ها داشته‌اند. میزان مجذور همبستگی تعدیل‌شده برای مدل رگرسیون که فقط شامل شایستگی‌های پایه بود ($adj.R^2$= ۰/۳۴) بود بیشتر از مدلی است که هردو شایستگی‌های پایه و محتوایی را شامل می‌شود ($adj.R^2$= ۰/۳۰).

بحث

اجرای آزمون خارج از کشور همزمان با داخل، آزمون‌های مجدد که در شرایط خاص اجرا می شود و حتی همین برگزاری همه ساله کنکو سراسری و نیز اجرای دو نوبت در چند سال اخیر بر اهمیت یکسان‌سازی دشواری سؤال‌های آزمون می‌افزاید. این پژوهش با هدف پیش‌بینی دشواری سؤال‌های درس ریاضی گروه علوم انسانی کنکور سراسری بر اساس شایستگی‌های پایه و محتوایی مبتنی‌بر الگوی تشخیصی شناختی اجرا شد. شایستگی‌های پایه به‌دست‌آمده در این پژوهش با شایستگی‌های تعیین‌شده در آزمون تیمز (کورتز و همکاران، ۲۰۰۶؛ تاتسوکا و همکاران، ۲۰۰۴؛ مینائی و همکاران، ۲۰۱۴)، آزمون داوطلبان معلمی ریاضی (چاندیا و همکاران، ۲۰۲۳)، آزمون پیسا (ترنر و همکاران، ۲۰۱۳) و درنهایت بخش حساب و دیفرانسیل آزمون درس ریاضی گروه آزمایشی علوم ریاضی (سعادتی و همکاران، ۲۰۲۰) همخوان بوده است. لیکن می‌توان ادعا کرد که این

مجموعه ۱۳ تایی از شایستگی‌های پایه برای همه انواع سؤال‌های ریاضی در گروه‌های سه‌گانه قابل استفاده است. در تعیین شایستگی‌های محتوایی تمامی سؤال‌های نوبت اول و دوم آزمون‌های سراسری کارشناسی گروه انسانی توسط دانشجویان سال نخست مشخص و به تأیید متخصصان رسید.

در آزمون گروه علوم انسانی نوبت نخست، دو شایستگی «کاربرد قواعد جبری» و«مسائل چندمرحله‌ای» با ۱۳ تکرار، پرکاربردترین شایستگی‌ها بودند. شایستگی «درک پیچیدگی سؤال» در نوبت اول ۱۱ بار مورد استفاده قرار گرفته بود. در مقابل شایستگی‌های «تبدیل و بازنمایی شکل، نمودار و جدول» و «تفکر خلاق» تنها ۱ بار تکرار شده بودند. در نوبت دوم شایستگی «درک پیچیدگی سؤال» با ۱۴ بار تکرار بیشترین و بعد از آن شایستگی «مسائل چندمرحله‌ای» با ۱۳ بار تکرار در رتبه بعدی قرار داشت. همچنین مشابه با نوبت نخست شایستگی «درک شکل، نمودار و جدول» تنها یک‌بار تکرار شده بود. اگرچه توزیع شایستگی‌ها در دو نوبت آزمون دارای اختلاف اندکی است، لیکن می‌توان این‌گونه استنباط کرد که در شایستگی‌های پرکاربرد، فراوانی کاربرد شایستگی‌ها نزدیک به هم است. توازن شایستگی‌های محتوایی نیز در دو نوبت آزمون رعایت شده و فراوانی کاربرد آنها در دو نوبت نزدیک به هم است. این امر بیانگر آن است که الزام سازه یکسان برای همترازسازی دو فرم درس ریاضی گروه علوم انسانی از نظر توزیع شایستگی‌های پایه و محتوایی رعایت شده است (شادمهر و همکاران، ۲۰۲۴). لازم به توضیح است که شایستگی درک پیچیدگی سؤال در هر دو نوبت پر تکرار بوده و بدین معنی است که داوطلبان برای پاسخ به اکثر سؤال‌ها بایستی بتوانند جملات طولانی و یا احتمالاً سلبی را به درستی درک کند.

میزان تبیین ضرایب دشواری بر اساس ترکیب شایستگی‌های پایه و محتوای وزنی ۶۶ درصد است. این مقدار واریانس تبیین شده برای ضرایب دشواری قابل‌مقایسه با روش‌هایی مانند قضاوت مقایسه‌ای نیست و مقدار قابل‌توجهی است. به‌عبارت‌دیگر روش‌های دیگر حداکثر تا ۲۵ درصد واریانس ضرایب دشواری را تبیین می‌کردند (هاماموتو فیلهو و همکاران، ۲۰۲۰؛ المصری و همکاران، ۲۰۱۷؛ بنتون، ۲۰۲۰؛ آتالی و همکاران، ۲۰۱۴). این در حالی است که میزان تبیین بیش از ۶۰ درصد پژوهش حاضر مشابه نتایج پژوهش‌هایی است که از شایستگی‌های ریاضی در تبیین دشواری سؤال‌ها استفاده شده است (ترنر و همکاران، ۲۰۰۹؛ ترنر و آدامز، ۲۰۱۲). در پژوهش‌های ترنر و همکارانش که بر روی سؤال‌های ریاضی پیسا انجام شده است ۶ مهارت پایه که هرکدام در ۴ سطح تقسیم‌بندی شده است می‌تواند بین ۶۰ تا ۷۰ درصد واریانس دشواری سؤال‌ها تبیین کند. نکته قابل‌تأمل این است که در دو گروه آزمایشی انسانی سهم تبیین‌کنندگی شایستگی‌های پایه بیشتر از شایستگی‌های محتوایی بوده است و به‌عبارت‌دیگر در این گروه آزمایشی فرایندهای ذهنی و شناختی که برای حل سؤال‌ها استفاده می‌شود قدرت تبیین دشواری سؤال‌ها را بیشتر از محتوای آن‌ها دارند.

در گروه علوم انسانی سؤال‌هایی که بر اساس محتوای مربوط به عبارت‌های جبری است که دربرگیرنده حل معادلات جبری درجه ۱ و ۲، حل دستگاه معادلات خطی، کاربرد معادله درجه ۲، تجزیه عبارت جبری و محاسبه ریشه nام و توان گویا است، منجر به سخت‌تر شدن سؤال شده و سؤال‌های مطرح‌شده از محتوای منطق و استدلال ریاضی که شامل جدول ارزش و ترکیب گزاره‌ها است، باعث ساده‌تر شدن آن سؤال می‌شود. به‌طورکلی می‌توان اذعان داشت که هر چه سؤال‌ها سخت و دشوارتر می‌شوند، فرایندهای شناختی سطح بالاتر مثل استدلال، تفکر تحلیلی و توانایی درک سؤال بیشترین نقش را در حل آن‌ها ایفا می‌کنند.

مهم‌ترین محدودیت این پژوهش نسبت تعداد سؤال‌ها به تعداد شایستگی‌ها بوده است که به دلیل محدودیت‌های اجرایی امکان تدوین ماتریس Q برای سوال‌های سال‌های دیگر فراهم نبود. محدودیت دیگر مرتبط با گروه آزمایشی بود. پژوهشگران در این پژوهش بر روی سوال‌های گروه علوم انسانی متمرکز بودند. با توجه به نتایج پیشنهاد می‌شود طراحی آزمون‌های همتا ازنظر دشواری بر اساس ضرایب رگرسیونی پیشنهادشده در سؤال درس ریاضی گروه علوم انسانی انجام شود. این امر به یکسان شدن آزمون‌ها از نظر سطح دشواری و همچنین محتوا و شایستگی‌های موردنیاز کمک خواهد کرد. پژوهش‌های آتی می‌توانند روی گروه‌های دیگر آزمایشی می‌تواند قدرت پیش‌بین شایستگی‌های پایه در گروه‌های مختلف را نشان دهد. باتوجه‌به اینکه نقض قواعد و اصول طراحی سؤال‌های چندگزینه‌ای می‌تواند با دشواری و قدرت تمیز سؤال مرتبط باشد، همبلتون و جیرکا[1] (۲۰۱۴)، انجام مطالعاتی که ترکیب شایستگی‌های پایه، محتوایی و قواعد طراحی سوال را در پیش‌بینی دشواری سوال بررسی کند می‌تواند در بهبود پیش‌بینی مؤثر باشد.

نتیجه‌گیری

یافته‌های این پژوهش چشم‌اندازی نوین برای درک دشواری سؤال‌های ریاضی در آزمون سراسری گشوده است. نتایج نشان داد که دشواری هر سؤال صرفاً به محتوای ریاضی آن وابسته نیست، بلکه نوع شایستگی مورد نیاز، یعنی شیوه تفکر و پردازش ذهنی داوطلب، نقش تعیین‌کننده‌تری در تعیین دشواری سؤال دارد. تحلیل‌ها بیانگر آن بود که سؤال‌هایی که مستلزم ترکیب چند شایستگی پایه، مانند استدلال، مدل‌سازی و تحلیل نمادین هستند، بیش از سایر سؤال‌ها، داوطلبان را به چالش می‌کشند.

مدل پیش‌بینی ارائه‌شده در این پژوهش توانست با دقت مناسبی میزان دشواری سؤال‌ها را تخمین بزند؛ موضوعی که می‌تواند در فرآیند طراحی، بازبینی و هم‌ترازی آزمون‌ها بسیار کاربردی باشد. از سوی دیگر، نتایج پژوهش نشان داد که توجه به شایستگی‌ها موجب طراحی علمی‌تر سؤال‌ها

1. Hambleton & jirka

و سنجش عادلانه‌تر می‌شود؛ همچنین از نظر کاربردی نیز می‌توان از چارچوب شایستگی‌ها به عنوان ابزاری برای طراحی بانک سؤال‌های استاندارد و پیش‌بینی دشواری سؤال در مراحل اولیه استفاده کرد. علاوه بر این، رویکرد شایستگی محور می‌تواند در فرایند تدریس نیز رهگشا باشد.

در مجموع، نتایج به روشنی نشان دادند که تحلیل شایستگی‌ها می‌تواند جایگزینی علمی و عملی برای روش‌های قضاوتی سنتی در برآورد دشواری باشد. این رویکرد، گامی مهم در جهت علمی‌تر شدن ارزیابی‌های ملی و حرکت به سوی نظام سنجشی است که به جای حفظ، تفکر را می‌سنجد.

تقدیر و تشکر

بدینوسیله از سازمان سنجش آموزش کشور به خاطر حمایت مالی و کلیه دوستان که در درانجام این پژوهش همکاری داشته‌اند، سپاسگزاری می‌شود.

References

Attali, Y., Saldivia, L., Jackson, C., Schuppan, F., & Wanamaker, W. (2014). Estimating item difficulty with comparative judgments. ETS Research Report Series, 2014(2) , 1-8.

Benton, T. (2020). How Useful Is Comparative Judgement of Item Difficulty for Standard Maintaining?. Research Matters, 29, 27-35.

Chandía, E., Sanhueza, T., Mansilla, A., Morales, H., Huencho, A., & Cerda, G. (2023). Nonparametric cognitive diagnosis of profiles of mathematical knowledge of teacher education candidates. Current Psychology, 42(36), 32498-32511.

Corter, J. E., Tatsuoka, K., Guerrero, A., Dean, M., & Dogan, E. (2006). Revised coding manual for identifying item involvement of content, context, and process subskills for the TIMSS and TIMSS-R 8th grade mathematics tests. Technical Report MES-06-01, Department of Human Development, Teachers College, Columbia University.

Creswell, J. W., Clark, V. L. P., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed. Handbook of mixed methods in social & behavioral research, 209, 209-240.

ElMasri, Y. H., Ferrara, S., Foltz, P. W., & Baird, J. A. (2017). Predicting item difficulty of science national curriculum tests: the case of key stage ۲ assessments. The Curriculum Journal, 28(۱) , 59-82.

Hamamoto Filho, P. T., Silva, E., Ribeiro, Z. M. T., Hafner, M. D. L. M. B., Cecilio-Fernandes, D., & Bicudo, A. M. (2020). Relationships between Bloom's taxonomy, judges' estimation of item difficulty and psychometric properties of items from a progress test: a prospective observational study. Sao Paulo Medical Journal, 138, 33-39.

Hambleton, R. K., & Jirka, S. J. (2014). Anchor-based methods for judgmentally estimating item statistics (Yadegarzadeh, Gh., Sharifi veganeh, N., Khodaii, E, Trans). In Handbook of test development (pp. 413-434). Routledge. (Original work published 2006(

Minaei, A., Delavar, A., Falsafinezhad, M. R., Kiamanesh, A. R., & Mohajer, Y. (2014). Cgnitive 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, 5(16), 138-170.

OECD. (2013). PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy. OECD Publishing.

Pitoniak, M. J., & Cizek, G. J. (2016). Standard setting. In C. S. Wells & M. Faulkner-Bond (Eds.) , Educational measurement: From foundations to future (pp. 38–61). The Guilford Press.

Rezigalla, A. A. (2024). AI in medical education: uses of AI in construction type A MCQs. BMC medical education, 24(1), 247.

Saadati, S., Moghadamzadeh, A., Minaei, A., & Geramipour, 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, 8(15), 19-35.

Shadmehr, A., Zamanpour, E., & Qasemi, S. (2024). The equating requirements of scores in alternative forms of high-stakes tests: the case study of the national entrance exam of the foreign language applicants. *Quarterly of Educational Measurement*, *15*(57), 33-53.

Suri, H. (2011). Purposeful sampling in qualitative research synthesis. *Qualitative research journal*, *11*(2), 63-75.

Tatsuoka, K. K. (2009). Cognitive assessment: An introduction to the rule space method. Routledge.

Tatsuoka, K. K., & Boodoo, G. M. (2000). Subgroup differences on the GRE quantitative test based on the underlying cognitive processes and knowledge. In Handbook of research design in mathematics and science education (pp. 821-857). Routledge.

Thorndike, R. L. (1996). Apllied psychometrics (Hooman, H,A, Trans). Houghton Mifflin School.(Original work published 1982(

Turner, R., & Adams, R. J. (2012). Some drivers of test item difficulty in mathematics: an analysis of the competency rubric.

Turner, R., Dossey, J., Blum, W., & Niss, M. (2013). Using mathematical competencies to predict item difficulty in PISA: A MEG study. In Research on PISA: Research outcomes of the PISA Research Conference 2009 (pp. 23-37). Springer Netherlands.

Van de Watering, G., & van der Rijt, J. (2006). Teachers' and students' perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty levels of assessment items. Educational Research Review, 1(2), 133-147.

Van Onna, m., Lampe, t., & Crompvoets, E. (2019). Equating by pairwise comparison. Presentation at the 20th annual AEA-Europe conference, Lisbon, Portugal