Educational Measurement and Evaluation Studies

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

National Organization for Educational Testing

Identifying Differential Item Functioning in PIRLS Test 2016: Based on RaschTree Method

Elham Pirmoradian[1], Asghar Minaei[2]

1. Msc. Graduate of Psychometrics, Islamic Azad University, Tehran, Iran. Email: elham.pirmoradian@gmail.com
2. Associate Professor of Measurement and Evaluation, Allameh Tabataba'i University Tehran, Iran; (corresponding author), Email:asghar.minaei@yahoo.com

Article Info

Article Type:
Research Article

Received:
2023.03.29

Received in revised form:
2023.06.29

Accepted:
2023.08.29

Published online:
2023.09.24

ABSTRACT

Objective: The growing concern about unfairness of educational assessments shows the possibility of differential functioning. The differential function or bias causes doubt in the validity of the test. In this study, identification of differential item functioning in PIRLS 2016 test among Iranian students has been investigated through the Rasch-Tree model.

Methods: The data used in this research are the results of the 2016 Pearls international study in Iran, which with a two-stage random sampling of 4385 Iranian students, including 2143 female students (48/87 percent) and 2242 male students (51/13 percent) of the fourth grade with an average age of 10/14, were selected for the study and participated in the test. The data of this study were analyzed and reviewed in the R environment (based on psychotree library.)

Results: The results indicated that out of 181 questions, based on Raschtree, one question was identified as a uniform differential function among two-valued questions. Moreover, among polytomous questions for block 16 differential functions were identified.

Conclusion: It seems that despite its differential function, its effect can be ignored. However, it is better to leave out questions with differential function when conducting analyses. Using test results requires delicacy and special consideration.

Keywords: *Differential Item Functioning, PIRLS 2016, Rasch-tree model.*

Cite this article: Pirmoradian, Elham; Minaei, Asghar (2023). Identifying Differential Item Functioning in PIRLS Test 2016: Based on RaschTree Method. *Educational Measurement and Evaluation Studies*, 13 (43): 39-62 pages. DOI: 10.22034/EMES.2023.555049.2389

© 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

شناسایی کارکرد افتراقی در آزمون پرلز ۲۰۱۶ با استفاده از درخت راش

الهام پیرمرادیان[۱]، اصغر مینائی[۲]

۱. دانش‌آموخته کارشناسی ارشد رشته روان‌سنجی، دانشگاه آزاد اسلامی، تهران، ایران. رایانامه: elham.pirmoradian@gmail.com

۲. دانشیار گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبایی، تهران، ایران؛ (نویسنده مسئول)، رایانامه: asghar.minaei@yahoo.com

اطلاعات مقاله

نوع مقاله:
مقاله پژوهشی

دریافت: ۱۴۰۲/۰۱/۰۹
اصلاح: ۱۴۰۲/۰۴/۰۸
پذیرش: ۱۴۰۲/۰۶/۰۷
انتشار: ۱۴۰۲/۰۷/۰۲

چکیده

هدف: دغدغهٔ رو به افزایشی که در ناعادلانه بودن آزمون‌ها شکل گرفته، امکان وجود کارکرد افتراقی است. چراکه کارکرد افتراقی یا سوگیری باعث تردید در روایی آزمون می‌شود. در مطالعهٔ حاضر، شناسایی کارکرد افتراقی سؤال‌های سواد خواندن پرلز ۲۰۱۶ بین دانش‌آموزان ایرانی از طریق مدل درخت راش بررسی شده است.

روش پژوهش: داده‌های استفاده شده در این پژوهش، نتایج مطالعه بین‌المللی پرلز ۲۰۱۶ در ایران است که با نمونه‌گیری دومرحله‌ای تصادفی، تعداد ۴۳۸۵ دانش‌آموز ایرانی شامل ۲۱۴۳ دانش‌آموز دختر (۴۸/۸۷ درصد) و ۲۲۴۲ دانش‌آموز پسر (۵۱/۱۳ درصد) پایه چهارم با میانگین سنی ۱۰/۱۴ برای شرکت در مطالعه انتخاب شدند و در آزمون شرکت کردند. داده‌های این مطالعه در محیط R (تحت کتابخانه psychotree) تحلیل و بررسی شد.

یافته‌ها: یافته‌ها نشان می‌دهند که از ۱۸۱ سؤال در این مطالعه بر اساس جنسیت، بر اساس مدل درخت راش 1 سؤال در میان سؤال‌های دو ارزشی با کارکرد افتراقی یکنواخت شناسایی شد. به‌علاوه در بین سؤال‌های چندارزشی، برای بلوک ۱۶ کارکرد افتراقی شناسایی شد.

نتیجه‌گیری: به نظر می‌رسد با وجود کارکرد افتراقی تأثیر آن قابل صرف‌نظر است. اگرچه بهتر است که سؤال‌های با کارکرد افتراقی هنگام انجام تحلیل‌ها کنار گذاشته شود. عملکرد ضعیف دانش‌آموزان ایرانی می‌تواند متأثر از این باشد که ترجمه متون مناسب نبوده یا دانش‌آموزان هنوز به توانایی لازم برای این سطح از خواندن دست نیافته‌اند. استفاده از نتایج آزمون پرلز نیازمند ظرافت و ملاحظات خاص است.

واژه‌های کلیدی: پرلز ۲۰۱۶، کارکرد افتراقی سؤال، مدل راش، مدل درخت راش

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

استناد: پیرمرادیان، الهام. مینائی، اصغر (۱۴۰۲). کارکرد افتراقی در آزمون پرلز ۲۰۱۶ با استفاده از درخت راش. مطالعات اندازه گیری و ارزشیابی آموزشی، ۱۳ (شماره۴۳)، ۳۹-۶۲ صفحه. DOI: 10.22034/EMES.2023.555049.2389
ناشر: سازمان سنجش آموزش کشور
حق‌مؤلف © نویسندگان.

مقدمه

در آزمون‌های روان‌سنجی، عموماً فرض می‌شود که ویژگی‌های اندازه‌گیری در بین افراد پایدار است. این ویژگی با نام تغییرناپذیری[1] اندازه‌گیری شناخته می‌شود (میلسپ[2]، ۲۰۱۲). بااین‌حال، ممکن است که گروه‌های مختلف مردم به‌طور متفاوتی نسبت به یک آزمون واکنش نشان دهند که اعتبار اندازه‌گیری‌ها را تحت‌الشعاع قرار می‌دهد. همچنین، اگر آزمون‌ها به نتایج متفاوتی برای گروه‌های متمایز از افراد منجر شوند، آزمون دچار سوگیری خواهد شد. کارکرد افتراقی سؤال[3] به این معنی است که تغییرناپذیری سؤال نقض می‌شود. به‌طور دقیق‌تر، کارکرد افتراقی سؤال در صورتی وجود دارد که یک یا چند سؤال برای یک گروه به‌طور قابل توجهی دشوارتر از گروه دیگر باشد. می‌توان بین کارکرد افتراقی یکنواخت و غیر یکنواخت تمایز قائل شد. کارکرد افتراقی یکنواخت به این معنی است که تفاوت بین گروه‌ها در سطوح پیوسته پنهان فرد ثابت است. ولی اگر تفاوت‌ها به توانایی یا ویژگی فرد بستگی داشته باشد آنگاه کارکرد افتراقی غیر یکنواخت وجود دارد. روش‌های تشخیص DIF را می‌توان به روش‌های نظریه سؤال پاسخ[4] و روش‌های غیر پارامتریک طبقه‌بندی کرد. روش‌های سؤال پاسخ که روش‌های پارامتریک نیز نامیده می‌شوند، روش‌هایی هستند که در آنها از مدل IRT برای تشخیص DIF استفاده می‌شود. ایده اصلی روش‌های سنتی تشخیص DIF در هر دو مدل IRT دو ارزشی[5] و چندارزشی[6]، به این شکل خواهد بود که ابتدا دو گروه از افراد انتخاب شده و سپس تعیین می‌کنیم آیا برآورد پارامتر آیتم بین این گروه‌ها متفاوت است یا خیر. اولین روشی که برای تشخیص DIF در مدل‌های IRT استفاده شد، آزمون نسبت درستنمایی[7] بود (اندرسن، ۱۹۷۳). یک رویکرد جایگزین که می‌تواند برای هر نوع مدل IRT استفاده شود، آزمون مربع کای لرد[8] است. درحالی‌که این آزمون به مقایسه دو گروه محدود می‌شود، بسط آن توسط کیم، کوهن و پارک، که با نام آزمون لرد تعمیم یافته[9] معروف است، می‌تواند برای بیش از یک گروه کانونی استفاده شود. رویکرد سوم روش راجو[10] است و بر این ایده استوار است که تفاوت بین شکل منحنی‌های پاسخ آیتم (IRCs) بین دو گروه نشان‌دهنده DIF است. این روش‌های کلاسیک وجه مشترکی دارند که محدود به چند زیرگروه هستند که باید توسط کاربر از قبل مشخص شوند. علاوه بر این، در نظر گرفتن

1. Invariance
2. Millsap
3. Invariance
4. Item Response Theory (IRT)
5. dichotomous
6. polytomous
7. Likelihood Ratio Test (LRT)
8. Lord’s chi square test
9. generalized Lord test
10. Raju method

بیش از یک متغیر کمکی القاکننده DIF در یک زمان دشوار است. (بولمن و همکاران[1]، ۲۰۱۷) امروزه سواد خواندن به‌واسطه پیامدهای تحصیلی، اجتماعی و اقتصادی مورد توجه خاص محققان قرار گرفته است. زیرا سواد خواندن نقش تعیین‌کننده‌ای در تحصیل بچه‌ها در مدرسه و همچنین در کارکرد بزرگسالان بعد از مدرسه ایفا می‌کند. مطالعه بین‌المللی پیشرفت سواد خواندن پرلز[2] مطالعه‌ای است که سنجش جامعی از مهارت‌های خواندن دانش‌آموزان پایه چهارم ابتدایی به دست می‌دهد. حضور کشور ایران در هر چهار دورهٔ مطالعه پرلز در سال‌های ۲۰۰۱، ۲۰۰۶، ۲۰۱۱ و ۲۰۱۶ موجب شده است، مطالعهٔ روند پیشرفت تحصیلیِ خواندن قابل بررسی باشد. متوسط عملکرد دانش‌آموزان ایرانی به ترتیب ۴۱۴، ۴۲۱، ۴۵۷ و ۴۲۸ بوده است که از متوسط عملکرد ۵۰۰ پایین‌تر است. نتایج بررسی روند پیشرفت تحصیلی خواندن نشان می‌دهد که با وجود افزایش عملکرد خواندن از ۲۰۰۱ تا ۲۰۱۱، عملکرد خواندن دانش‌آموزان ایرانی از سال ۲۰۱۱ به ۲۰۱۶ کاهش چشمگیر ۲۹ نمره‌ای داشته است. از ۴۱ کشوری که در دوره‌های ۲۰۱۱ و ۲۰۱۶ مطالعه پرلز شرکت کرده‌اند؛ ۱۰ کشور ازجمله ایران، عملکرد کاهشی داشته‌اند. علاوه بر آن، بیشترین میزان کاهش عملکرد پیشرفت تحصیلی خواندن در ایران ثبت شده است (کبیری و همکاران، ۱۳۹۶). همچنین عملکرد سواد خواندن دختران ایرانی ۴۶ نمره بهتر از پسران شده است که این تفاوت باعث شده است، ایران به همراه عمان، سومین کشور از لحاظ عملکرد دو جنس به شمار برود. مسئله‌ای که پاسخ مشخصی برای آن وجود ندارد، این است که عملکرد ضعیف دانش‌آموزان ایرانی در آزمون سواد خواندن مطالعه پرلز، یک علت حقیقی مربوط به ساختار نظام آموزشی است یا مسئله متن‌های به‌کاربرده شده است (احمدی و همکاران، ۱۳۹۶). انجام یک مطالعه با استفاده از رویکردهای جدید برای شناسایی کارکرد افتراقی روی داده‌های مطالعه بین‌المللی پرلز دارای اهمیت است؛ چراکه در نتیجه این‌گونه پژوهش‌ها مشخص می‌شود که این ابزار کارکرد خود را به‌درستی انجام می‌دهد و برای همه کشورهای شرکت‌کننده نتایج کاربردی و قابل محاسبه‌ای به دست می‌دهد. درنتیجه تفاوت‌های مشاهده‌شده در نتایج به ویژگی‌های آزمون و سؤال‌های آن و آزمودنی‌ها ارتباطی ندارد و در صورت شناسایی کارکرد افتراقی، سؤال‌های با کارکرد افتراقی مشخص شده و می‌توانیم آگاهانه بدون سوگیری جنسیتی، سنی، موقعیت جغرافیایی یا اجتماعی، ... تصمیمات آموزشی و سیاست‌های لازم در جهت پیشرفت سواد خواندن دانش‌آموزان اتخاذ کنیم. تاکنون در زمینه شناسایی کارکرد افتراقی سؤال‌های مطالعه پرلز ۲۰۱۶ بین دانش‌آموزان ایرانی مطالعه‌ای نشده است. بنابراین، هدف‌های این پژوهش شناسایی کارکرد افتراقی سؤال‌های پرلز ۲۰۱۶ با استفاده از مدل درخت راش است. مسئله‌ای که می‌خواهیم

1. Bollmann et al
2. Progress in International Reading Literacy Study (PIRLS)

در این تحقیق به آن بپردازیم به این شرح است که آیا آزمون پرلز در میان دانش‌آموزان ایرانی نسبت به متغیرهایی همچون جنسیت و سن کارکرد افتراقی نشان می‌دهد یا خیر؟ در بین کدام زیرگروه‌ها از دانش‌آموزان، آزمون پرلز به دلیل کارکرد افتراقی ناکافی است؟

مبانی نظری و پیشینه پژوهش

انجمن بین‌المللی ارزشیابی پیشرفت تحصیلی[1] با هدف به دست آوردن درک عمیق‌تری از تأثیر سیاست‌گذاری نظام‌های آموزشی مختلف از دهه ۱۹۶۰، بیش از ۵۰ سال است که به ارزیابی سواد خواندن و عوامل مرتبط با آن پرداخته است. مطالعات تیمز و پرلز به‌وسیله این انجمن اداره می‌شود (مولیس و همکاران[2]، ۲۰۱۷). نخستین مطالعه بین‌المللی پیشرفت سواد خواندن (پرلز) در سال ۲۰۰۱ انجام گرفت. بر اساس هدف‌های این مطالعه، پایه چهارم، زمانی مهم در رشد دانش‌آموزان است. در این پایه دانش‌آموزان چگونه خواندن و مطالعه کردن را یاد گرفته‌اند و از خواندن برای یادگیری استفاده می‌کنند. بر همین اساس آزمون پرلز از سال ۲۰۰۱ هر ۵ سال یک بار برای دانش‌آموزان پایه چهارم با هدف سنجش پیشرفت سواد خواندن اجرا می‌شود. ۶۱ شرکت‌کننده ازجمله ۵۰ کشور و ۱۱ نهاد به‌عنوان معیار سنجش در پرلز ۲۰۱۶ شرکت داشتند. برای کشورهایی که از سال ۲۰۰۱ در ارزیابی قبلی شرکت کرده‌اند، نتایج پرلز ۲۰۱۶ فرصتی برای ارزیابی پیشرفت در دستیابی به خواندن در چهار مقطع زمانی ۲۰۰۱، ۲۰۰۶، ۲۰۱۱ و ۲۰۱۶ فراهم می‌کند. ایران ازجمله کشورهایی است که از سال ۲۰۰۱ در این مطالعه شرکت کرده است و این امکان را دارد تا با بررسی روند عملکرد دانش‌آموزان در این چهار دوره، پیشرفت سواد خواندن را ارزیابی کند. کشورهای شرکت‌کننده در این مطالعه اطلاعات مرتبط با بافت و زمینه یادگیری برای خواندن را در اختیار انجمن بین‌المللی ارزشیابی پیشرفت تحصیلی قرار می‌دهند. در مطالعه پرلز همواره اطلاعات مفیدی از طریق پرسشنامه‌های مدرسه، معلمان، دانش‌آموزان و والدین گردآوری می‌شود. علاوه بر آن، در هر دوره دانشنامه‌ای به چاپ می‌رسد که کشورهای شرکت‌کننده برنامه درسی خواندن و تدریس آن را در کشور خود توضیح می‌دهند. بدین سبب داده‌های پرلز منبعی غنی و مفید مرتبط با سیاست‌ها در اختیار سیاست‌گذاران قرار می‌دهد (کبیری و همکاران، ۱۳۹۶). پرلز بر اساس ۱۲ متن (۶ ادبیاتی و ۶ اطلاعاتی) و تقریباً ۱۸۰ سؤال است (مولیس و همکاران، ۲۰۱۷). هدف پرلز ارائه بهترین اطلاعات مربوط به سیاست‌ها دربارهٔ چگونگی بهبود آموزش و یادگیری و کمک به دانش‌آموزان جوان است. مولیس و همکاران (۲۰۰۶) بیان می‌کنند، مطالعه پرلز به سه بعد از سواد خواندن دانش‌آموزان تمرکز دارد: ۱- هدف‌های خواندن؛ ۲- فرایندهای درک مطلب؛ ۳- رفتارها و نگرش‌های خواندن.

1. International Association for the Evaluation of Educational Achievement(IEA)
2. Mullis

جدول (۱) **هدف‌ها، فرایندهای مورد ارزیابی و درصد تأکید بر آنها در مطالعه پرلز**

هدف‌های خواندن[1]	درصد
تجربهٔ ادبی[2]	٪۵۰
کسب اطلاعات و استفاده از آن[3]	٪۵۰
فرایندهای درک مطلب[4]	درصد
تمرکز و بازیابی اطلاعاتی که به‌صراحت بیان شده[5]	٪۲۰
استنباط‌های مستقیم[6]	٪۳۰
تفسیر و تلفیق ایده‌ها و اطلاعات[7]	٪۳۰
بررسی و ارزیابی محتوا، زبان و عناصر متنی[8]	٪۲۰

از نظر آموزشی، توجه به تفاوت‌های جنسیتی در کشورهای توسعه‌یافته و به‌ویژه وضعیت دختران ممکن است قدیمی به نظر برسد. دختران با گسترش تحصیلات اجباری بسیار پیشرفت کرده‌اند (جراد[9]، ۲۰۰۷). در گزارش پیزا[10] (۲۰۰۶) بیان شده است که خواندن حوزه‌ای است که شکاف جنسیتی در آن زیاد است و دختران نمرات بالاتری نسبت به پسران کسب می‌کنند (احمدی و همکاران، ۱۳۹۶). همچنین بر اساس نتایج مطالعه پرلز ۲۰۱۶، میانگین عملکرد دختران ۱۹ نمره بهتر از پسران بوده است. در نمودار (۱) روند تغییرات عملکرد خواندن دانش آموزان ایرانی به تفکیک دو جنس ارائه شده است. با اینکه این شکل، برتری دختران را در همهٔ این سال‌ها نشان می‌دهد، ولی لزوماً روند عملکرد دو جنس به یک شکل نبوده است. به‌عنوان‌مثال در فاصله سال‌های ۲۰۱۱ تا ۲۰۱۶ پسران کاهش عملکرد بیشتری نسبت به دختران داشته‌اند یا در فاصله سال‌های ۲۰۰۱ تا ۲۰۰۶ رشد عملکرد پسران بیش از دختران بوده است. کمترین فاصله عملکردی بین دختران و پسران در سال ۲۰۰۶ و در حدود ۱۵ نمره و بیشترین فاصله عملکردی در سال ۲۰۱۶ با ۴۶ نمره به‌دست‌آمده است (کبیری و همکاران، ۱۳۹۶).

1. Purposes for Reading
2. Literary Experience
3. Acquire and Use Information
4. Process of Comprehension
5. Focus on and Retrive Explicity Stated Information
6. Make Straightforward Inferences
7. Interpret and Integrate Ideas and Information
8. Evaluate and Critigue content and Textual Elements
9. Giraud
10. Programme for International Student Assessment (PISA)

نمودار (۱) روند تغییرات پیشرفت تحصیلی خواندن دختران و پسران ایرانی

IEA در سال ۱۹۹۱ دربارهٔ سواد خواندن در ۳۲ کشور دنیا، ارتباط مشخصی بین تعداد کتاب‌های دانش‌آموزان در منزل و میزان رشد خواندنشان نشان داد. همچنین بر اساس مطالعهٔ تیمز[1]، دانش‌آموزان پایه هشتم که در منزل به کتاب‌ها یا وسایل کمک‌آموزشی بیشتری دسترسی دارند یا پدر و مادرشان تحصیلات دانشگاهی دارند، مدارج بالاتری در ریاضیات و علوم به دست می‌آورند. درمجموع، کشورهایی که گزارش داده بودند در منزل به کتاب‌های بیشتری دسترسی دارند؛ مدارج بالاتری در آزمون خواندن داشتند (مولیس و همکاران، ۲۰۰۶). در پرلز ۲۰۱۶، بررسی‌های مربوط به پشتیبانی از یادگیری خواندن از خانه و اینکه ایران جزء کشورهای انتهای فهرست منابع یادگیری در منزل قرار دارد، بیانگر این است که شرایط یادگیری در منزل برای دانش‌آموزان ایرانی چندان مهیا نیست. علاوه بر این، میزان علاقه والدین ایرانی به خواندن پایین است و در سال‌های اخیر نیز این میزان کاهش داشته است که البته این رویه کاهشی در بسیاری از کشورهای شرکت‌کننده در مطالعه پرلز نیز دیده شده است. با وجود این بین میزان علاقه والدین به مطالعه و عملکرد فرزندان در خواندن رابطه مثبتی وجود دارد (کبیری و همکاران، ۱۳۹۶).

مرور کلی مدل درخت راش

ریاضیدان دانمارکی به نام جورج راش[2] در سال ۱۹۶۰ مفهوم عینیت و تغییرناپذیری را در قالب یک مدل آماری مطرح ساخت و روشی برای آزمون آنها ابداع کرد. مدل راش درواقع خانواده‌ای از مدل‌ها

1. Trends in International Maths and Science Study (TIMSS)
2. Georg William Rasch

اســت که از آن جمله می‌توان به مدل لجســتیک ســاده[1]، مدل مقیاس درجه‌بندي[2] و مدل امتیازدهی پاره‌ای[3] اشــاره کــرد (مینائی، ۱۳۹۳). مدل‌هــای اندازه‌گیري راش به پژوهشــگران امکان مي‌دهد تا مشــکلات زیربنایي اندازه‌گیري‌هاي مدل کلاسیك را حل کنند. لیناکر[4] (۲۰۰۶) مدل‌هاي راش را در دو طبقه کلي دو ارزشــي[5] و چندارزشي[6] تقسیم‌بندي کرد (نقل در عسگری، ۱۳۸۶). مدل راش، مورد خاصی از مدل لجســتیکی یک پارامتری اســت که در آن تمیز روی ۱ تنظیم شــده اســت. هدف مدل راش به دســت آوردن سنجشی از متغیرهای پنهان و حداقل است که به‌صورت فاصله‌ای مقیاس‌بندی شــده‌اند. مدل راش احتمال وقوع پاســخ خاص (به‌عنوان‌مثال پاسخ درست) را با دشواری سؤال پیوند می‌دهد. مدل از این معادله استفاده می‌کند:

$$P(correct) = \frac{e^{(\theta - b)}}{1 + e^{(\theta - b)}}$$

که در آن θ، نشــان‌دهنده توانایی و b نشــان‌دهنده دشواری سؤال اســت و فرض می‌شود تمیز برابر ۱ اســت. تمیز روی یک فرض می‌شــود چراکه مدل راش این‌گونه فرض می‌کند که هر ســؤال میان افراد دارای توانایی بالا و پایین به شکل کاملی تمیز ایجاد می‌کند[7].

تغییرناپذیر[8] نبودن پارامترهای ســؤال و آزمون‌شــوندگان، به‌عنوان‌مثال در بین دو گروه مردان و زنان، معمولاً کارکرد افتراقی ســؤال[9] نامیده می‌شــود و به این معنا اســت که افرادی از دو گروه مردان و زنان که از توانایی یکســان برخوردارند، در یک ســؤال خاص به‌طور متفاوتی عمل می‌کنند (هاگ‌کوئیست و آندریــش[10]، ۲۰۰۴؛ بــه نقل از مینائی، ۱۳۹۳). روش افراز بازگشــتی[11] مبتنی بــر مدل یک رویکرد نیمه‌پارامتری اســت که از آزمون‌های آماری اقتصادســنجی برای تغییر ســاختاری اســتفاده می‌کند. هدف این اســت که تفــاوت در پارامترهای یــک مدل آماری بین گروه‌های افراد تعریف‌شــده توســط (ترکیب) متغیرها را تشخیص دهد. افراز بازگشتی مبتنی بر مدل مربوط به روش طبقه‌بندی و درختان رگرسیون است که در آن فضای متغیر کمکی، به‌صورت بازگشتی تقسیم می‌شود تا گروه‌هایی از افراد با ارزش‌های مختلف یک متغیر پاسخ طبقه‌ای یا پیوسته را شناسایی کند (استروبل و همکاران[12]، ۲۰۱۳).

1. simple logostic model
2. rating scale model
3. partial credit model
4. Linacre
5. dichotomous
6. ploytomous
7. De Ayala, 2009
8. Invariance
9. differential item functioning(DIF)
10. Hagquist & Andrich
11. recursive partitioning
12. Strobl et al

اخیراً رویکرد نیمه پارامتریک جدیدی از تست کارکرد افتراقی در مدل‌های راش بر اساس اِفراز بازگشتی توسط استروبل و همکاران (۲۰۱۳) مطرح شده است. به بیانی ساده، رویکردهای نیمه‌پارامتریک، دارای مؤلفه‌های رویکردی پارامتریک و غیرپارامتریک هستند. بخش غیرپارامتریک این است که افراز بازگشتی خطی فرض نشود. با استفاده از روش درخت راش، متغیرهای کمکی تعریف‌شده، یا آمیزه‌ای از متغیرهای کمکی را می‌توان به‌منظور شناسایی تفاوت‌ها در پارامترهای مدل استفاده کرد. رویکرد فعلی با استفاده از تکنیک‌های آماری نوین برای قوی‌تر بودن به نسبت درخت‌های طبقه‌بندی و رگرسیون، از پارامترهای مدل پارامتریک به جای یک متغیر پاسخ استفاده می‌کند. پاسخ‌ها به سؤال با در نظر گرفتن تفاوت‌های گروهی مربوط به متغیرهای کمکی ثبت می‌شوند. مدل سپس به شکل یک تصویر یا درخت نشان داده می‌شود که هر ساقه انتهایی نشان‌دهنده برآوردهای پارامتر است. اگر بیش از یک ساقه وجود داشته باشد، فرضیه صفر مورد سنجش که هیچ تفاوت گروهی وجود ندارد، رد خواهد شد. یکی از مزیت‌های درخت‌های راش این است که نتایج به شکل تصویری نمایش داده می‌شوند (استروبل و همکاران، ۲۰۱۱a). در این روش نیازی نیست متغیرها برای الگوریتم تخصیص داده شوند و این در مقایسه با مدل‌های پارامتریک، در کار شناسایی گروه‌های متأثر از کارکرد افتراقی سهولت ایجاد می‌کند. به‌عبارتی‌دیگر، سایر مدل‌ها مستلزم این هستند که محقق بداند کدام متغیرها کارکرد افتراقی را قبل از برازش مدل برای داده نشان می‌دهند. این مدل، ریشه در افراز بازگشتی مبتنی بر مدل که مرتبط با درخت‌های طبقه‌بندی و رگرسیون است، داشته و از آزمون‌ها برای تغییر ساختار در رویکردی نیمه پارامتریک استفاده می‌کند و متغیرهای کمکی فرد در گروه‌ها برای شناسایی تفاوت‌ها در پارامترهای مدل‌های آماری تحلیل می‌شوند.

در مدل راش فرض می‌شود که تخمین پارامتر سؤال وابسته به گروه نمونه نیست. این فرض درصورتی‌که برای گروه‌های خاصی از افراد - بدون در نظر گرفتن توانایی واقعی آنها - پاسخ به سؤال‌های خاصی آسان‌تر یا دشوارتر باشد ممکن است نقض شود؛ در این‌گونه مواقع است که ما از کارکرد افتراقی سخن به میان می‌آوریم. به‌منظور شناسایی کارکرد افتراقی سؤال با تابع درخت راش، پاسخ‌های سؤال و همه متغیرهای کمکی که باید از نظر DIF آزمایش شوند باید به‌عنوان ورودی در اختیار روش درخت راش قرار داده شوند (استروبل و همکاران، ۲۰۱۱ b.).

مزیت اصلی رویکرد درخت راش این است که DIF بین گروهی از آزمودنی‌ها با بیش از یک متغیر کمکی قابل تشخیص است. به‌عنوان‌مثال، در مقایسه با سایر افراد سؤال‌های خاصی برای مردان بالای ۴۰ سال ممکن است راحت‌تر باشد. در این حالت DIF در ارتباط با تعامل متغیرهای جنسیت و سن است به جای آنکه با هرکدام به‌تنهایی در نظر گرفته شود (همان). علاوه بر این، در این رویکرد نیاز

به برش‌های از پیش تعیین شده در متغیرهای پیوسته نیست. روش درخت راش مقداری که مربوط به قوی‌ترین تغییر پارامتر است را جستجو می‌کند و نمونه را در آن ارزش تقسیم می‌کند. تکنیک‌های آماری خاصی برای انجام این کار لازم است که در استروبل و همکاران (۲۰۱۰) به‌تفصیل شرح داده شده است (استروبل و همکاران، ۲۰۱۱ b).

علی میرزایی و همکاران (۱۳۹۸) بیان کرده‌اند که ویژگی‌های روان‌سنجی مثل تغییرناپذیری اندازه‌گیری، نبود کارکرد افتراقی سؤال یا فقدان سوگیری باید برقرار باشد تا نمرات یک آزمون برای گروه‌های مختلف آزمودنی مقایسه‌پذیر شوند. تحلیل منابع کارکرد افتراقی سؤال به محققان اجازه می‌دهد فرضیه‌هایی در ارتباط با منابع اصلی و بالقوه سوگیری و واریانس سازه مزاحم را فرمول‌بندی کنند. یافته‌ها نشان‌دهندهٔ آن است که اهمیت اصلی شناسایی منابع DIF در ساخت و تفسیر نتایج آزمون‌ها است که برای مقایسهٔ بین گروه‌ها به‌کاربرده می‌شود. سیاست‌گذاران باید هنگام تصمیم‌گیری دربارهٔ برنامه درسی، منابع یا آموزش بر مبنای مقایسه نمرات آزمون، مراقب روایی مقایسه‌پذیری نمرات باشند.

روکس[1] (۲۰۲۰) با استفاده از یافته‌های بین‌المللی پرلز ۲۰۱۶ مبنی بر ضعف شدید دانش‌آموزان جنوب آفریقا در درک مطلب نسبت به دیگر کشورهای شرکت‌کننده به بررسی معادل بودن متون پرلز ۲۰۱۶ با سه زبان (انگلیسی، آفریقایی و ایسی زولو) در جنوب آفریقا پرداخت. نتایج بررسی از طریق مدل راش و استفاده از نظر متخصصان، بیانگر معادل نبودن متون پرلز با زبان‌های مورد نظر است. بااین‌حال، الگوی مشخصی در سؤال‌های دارای کارکرد افتراقی وجود نداشته است و همچنین هیچ‌گونه تبعیض یا بی‌علاقگی نسبت به زبان دیده نشد. با نگاهی عمیق به متون و سؤال‌هایش مشخص شد که متن گل‌های روی پشت‌بام فقیرترین ترجمه به آفریقایی و ایسی زولو تلقی می‌شود. به‌طورکلی، همه متون برای زبان‌آموزان آفریقای جنوبی مناسب تلقی شدند. زیرا متون از واژگان غنی استفاده می‌کردند و دانش‌آموزان را با ایده‌ها و مفاهیم جدید آشنا می‌کردند.

الشهری[2] (۲۰۲۰) در تحقیقی به تجزیه‌وتحلیل کارکرد افتراقی سؤال‌های پرلز ۲۰۱۶ بر اساس جنسیت در سه کشور که به زبان مشترک صحبت می‌کردند؛ پرداخت. در این مطالعه از رگرسیون دوجمله‌ای لجستیک برای تشخیص کارکرد افتراقی یکنواخت و غیریکنواخت در سه کشور انتخاب شده به‌طور تصادفی استفاده شده است. یافته‌ها به‌وضوح نشان‌دهنده وجود DIF یکنواخت و غیریکنواخت بودند. بزرگی اثر از کوچک تا بزرگ متغیر است. الشهری این مطالعه را در جهت هشدار به روان‌شناسان، متخصصان و سایر ذی‌نفعان در مورد پیش‌بینی DIF مبتنی بر جنسیت انجام داده است.

1. Roux
2. Alshehri

بررسی‌ها نشان می‌دهد که در خصوص کارکرد افتراقی سؤال‌های آزمون بین‌المللی پرلز بر اساس جنسیت مطالعات متعددی همچون (جفرز[1]، ۲۰۲۰)، (الشهری، ۲۰۲۰)، (ساری[2]، ۲۰۲۰)، (احمدی و همکاران، ۲۰۱۷)، (جسکه و ازولا[3]، ۲۰۱۰)، (کندی[4]، ۲۰۰۸) صورت گرفته است. هرچند در کشور ایران با توجه به رتبه‌ای که در این آزمون کسب شده است تاکنون مطالعه‌ای برای شناسایی کارکرد افتراقی سؤال‌های سواد خواندن پرلز ۲۰۱۶ از طریق روش‌های مبتنی بر مدل درخت راش انجام نگرفته است.

سؤال‌های پژوهش:

۱. آیا آزمون پرلز ۲۰۱۶ در دانش‌آموزان کشور ایران تک‌بعدی است؟

۲. آیا سؤال‌های پرلز ۲۰۱۶ در بین دانش‌آموزان دختر و پسر کشور ایران بر اساس مدل درخت راش دارای کارکرد افتراقی است؟

۳. وضعیت یا عملکرد نسبی روش‌های افراز بازگشتی درخت راش در شناسایی کارکرد افتراقی سؤال‌های آزمون پرلز ۲۰۱۶ بین دانش‌آموزان دختر و پسر ایرانی چگونه است؟

روش پژوهش

مطالعه پرلز از یک طرح نمونه‌گیری دومرحله‌ای تصادفی استفاده می‌کند. در کشور شرکت‌کننده در مرحله اول از بین مدرسه‌های موجود در کشور با خصوصیات مختلف جنسیت، جغرافیایی (استانی)، نوع مدرسه (دولتی یا غیردولتی) و منطقه آن (شهری یا روستایی) تعدادی مدرسه به‌طور تصادفی انتخاب شده، سپس یک یا چند کلاس کامل از هر مدرسهٔ نمونه‌گیری‌شده به‌طور تصادفی به‌عنوان مرحله دوم نمونه‌گیری انتخاب می‌شود. این روش نمونه‌گیری دارای دو مزیت مهم است: اول اینکه پرلز توجه خاصی به برنامه درسی دانش‌آموزان و تجارب آموزشی دارد و این به‌طورمعمول بر اساس کلاس درس سازمان‌دهی می‌شود؛ مزیت دیگر، این روش نمونه‌گیری برخلاف نمونه‌گیری فردی دانش‌آموز، موجب می‌شود که در برنامه روزانه مدرسه اختلالی ایجاد نشود. با توجه به این موارد، بر اساس روش استانداردشده توسط IEA، ۲۷۱ مدرسه با حجم ۴۳۸۵ نفر دانش‌آموز پایه چهارم برای شرکت در مطالعه پرلز ۲۰۱۶ انتخاب شدند که از این تعداد ۲۱۴۳ نفر دانش‌آموز دختر (معادل ۴۸/۸۷ درصد) و ۲۲۴۲ دانش‌آموز پسر (معادل ۵۱/۱۳ درصد) با میانگین سنی ۱۰/۱ و انحراف استاندارد ۰/۴۸ هستند. آزمون پرلز ۲۰۱۶، دارای ۱۸۱ سؤال است که در ۱۶ دفترچه ارائه شده‌اند. هر دفترچه حاوی دو متن و سؤال‌های مربوط به آن است. جدول (۳) بیانگر اطلاعات کاملی از هر دفترچه است. این

1. Jeffers
2. Sari
3. Geske & Ozola
4. Kennedy

دفترچه‌ها به‌صورت تصادفی بین دانش‌آموزان توزیع شـدند. از بین ۱۸۱ سـؤال سواد خواندن در پرلز ۲۰۱۶، تعداد ۸۶ سـؤال دو ارزشی (چهارگزینه‌ای) و ۹۵ سؤال چندارزشی هستند. در پژوهش حاضر برای تحلیل داده‌ها سـؤال‌های چندارزشی[1] و دو ارزشی[2] به‌صورت جداگانه بررسی شده‌اند. همچنین برای سـؤال‌های بدون پاسخ که دانش‌آموزان به دلیل ندانستن پاسخ صحیح آنها را خالی گذاشته‌اند یا سـؤال‌های جامانده که به دلیل کمبود وقت قادر به پاسخگویی به سـؤال‌ها نبوده‌اند؛ نمره صفر داده شد.

جدول (۲) **توزیع دفترچه‌های آزمون پرلز ۲۰۱۶، تعداد سؤال، نوع متن و شکل سؤال**

دفترچه	نام متن	تعداد سؤال	نوع متن	سؤال دو ارزشی	سؤال چندارزشی
۱	گل‌های روی پشت بام	۱۳	ادبی	۷	۶
	اسب‌های ایسلندی	۱۵	اطلاعاتی	۶	۹
۲	کهربای درخشان	۱۴	ادبی	۸	۶
	اسب‌های ایسلندی	۱۵	اطلاعاتی	۶	۹
۳	کهربای درخشان	۱۴	ادبی	۸	۶
	چگونه پرواز را یاد گرفتیم؟	۱۷	اطلاعاتی	۸	۹
۴	گلدان خالی	۲۰	ادبی	۱۱	۹
	کوسه‌ها	۱۲	اطلاعاتی	۶	۶
۵	گلدان خالی	۲۰	ادبی	۱۱	۹
	چگونه پرواز را یاد گرفتیم؟	۱۷	اطلاعاتی	۸	۹
۶	الیور و گریفین	۱۳	ادبی	۴	۹
	کوسه‌ها	۱۲	اطلاعاتی	۶	۶
۷	الیور و گریفین	۱۳	ادبی	۴	۹
	لئوناردو داوینچی	۱۲	اطلاعاتی	۶	۶
۸	پمبا شرپا	۱۷	ادبی	۸	۹
	لئوناردو داوینچی	۱۲	اطلاعاتی	۶	۶

1. polytomous
2. dichotomous

دفترچه	نام متن	تعداد سؤال	نوع متن	سؤال دو ارزشی	سؤال چندارزشی
۹	گل‌های روی پشت بام	۱۳	ادبی	۷	۶
	عسل کجاست؟	۱۶	اطلاعاتی	۷	۹
۱۰	پمبا شرپا	۱۷	ادبی	۸	۹
	عسل کجاست؟	۱۶	اطلاعاتی	۷	۹
۱۱	کهربای درخشان	۱۴	ادبی	۸	۶
	کوسه‌ها	۱۲	اطلاعاتی	۶	۶
۱۲	الیور و گریفین	۱۳	ادبی	۴	۹
	چگونه پرواز را یاد گرفتیم؟	۱۷	اطلاعاتی	۸	۹
۱۳	گل‌های روی پشت بام	۱۳	ادبی	۷	۶
	لئوناردو داوینچی	۱۲	اطلاعاتی	۶	۶
۱۴	پمبا شربا	۱۷	ادبی	۸	۹
	اسب‌های ایسلندی	۱۵	اطلاعاتی	۶	۹
۱۵	گلدان خالی	۲۰	ادبی	۱۱	۹
	عسل کجاست؟	۱۶	اطلاعاتی	۷	۹
۱۶	میسی و مرغ قرمز	۱۶	ادبی	۸	۸
	لاک‌پشت دریایی سبز	۱۶	اطلاعاتی	۷	۹

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

یافته‌ها

برای شناسایی کارکرد افتراقی در آزمون پرلز ۲۰۱۶، داده‌های پژوهش از طریق نرم‌افزار R تحلیل شد. در این پژوهش ۲۷۱ مدرسه با حجم ۴۳۸۵ نفر دانش‌آموز پایه چهارم برای شرکت در مطالعه پرلز ۲۰۱۶ انتخاب شدند که از این تعداد ۲۱۴۳ نفر دانش‌آموز دختر (معادل ۴۸/۸۷ درصد) و ۲۲۴۲ دانش‌آموز پسر (معادل ۵۱/۱۳ درصد) که کمترین و بیشترین سن به ترتیب ۹/۸۳ و ۱۴/۸۳ با میانگین سنی ۱۰/۱۴ و انحراف استاندارد ۰/۴۸ هستند. ۱۶ دفترچه برای آزمون پرلز ارائه شد که هر دفترچه حاوی دو متن و سؤال‌های مربوط به آن بود. نتایج مطلوب همسانی درونی هر بلوک که از طریق بسته ltm در محیط R به دست آمد. بر اساس این نتایج، بلوک ۱ با ۰/۹۱ و بلوک ۹ با ۰/۶۸، به ترتیب بیشترین و کمترین آلفای کرونباخ را در بین بلوک‌ها دارا بودند.

پرسش اول پژوهش: آیا آزمون پرلز ۲۰۱۶ در دانش‌آموزان کشور ایران تک‌بعدی است؟
بعدیت هر بلوک با دستور EAG و DIMTESTS بر اساس معیار تجربی کایزر[1]، روش هال[2]، تحلیل موازی[3] از طریق پکیج EFA.dimensions مورد بررسی قرار گرفت. بر اساس نتایج تحلیل بعدیت، آزمون پرلز ۲۰۱۶ تک‌بعدی است. به‌عنوان‌مثال، مقدار ویژه[4] به‌دست‌آمده توسط دستور DIMTESTS حاکی از غلبه یکی از ضرایب بر دیگر مقادیر ویژه است. به‌طوری‌که بزرگ‌ترین مقدار ویژه برابر ۶/۸۹۷ بوده و بقیه مقادیر ویژه کمتر از ۱ بودند که بیانگر تک‌بعدی بودن داده است. شکل (۱) خروجی دستور EAG را به‌صورت گرافیکی نشان می‌دهد که در توافق با نتایج پکیج DIMTESTS است.

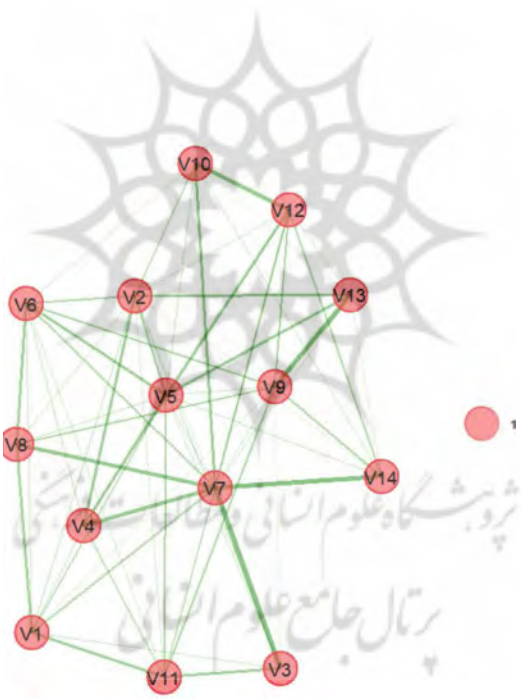
DIMTESTS results:

# of Factors:

EMPKC 1

HULL 1

RAWPAR 1



شکل (۱) خروجی دستور EAG در نرم‌افزار R برای یافتن بعدیت

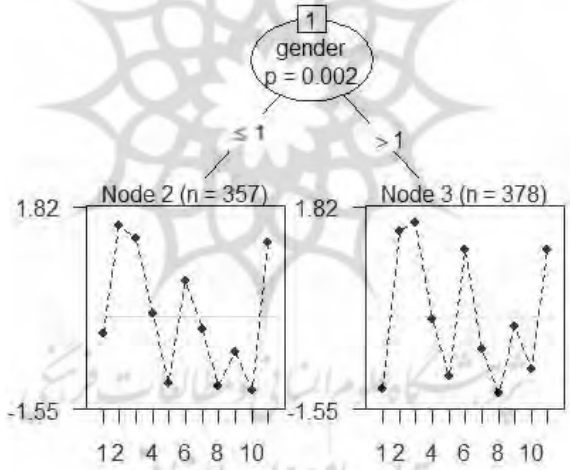
پرسش دوم پژوهش: آیا سؤال‌های پرلز ۲۰۱۶ در بین دانش‌آموزان دختر و پسر کشور ایران بر اساس مدل درخت راش دارای کارکرد افتراقی است؟

1. The Empirical Kaiser Criterion
2. Hull method
3. Parallel Analysis
4. Eigenvalue

برای شناسایی کارکرد افتراقی سؤال‌های آزمون پرلز ۲۰۱۶ از روش افراز بازگشتی درخت محور از طریق پکیج Psychotree در محیط R استفاده کردیم. برای درخت راش، تنظیمات پیش‌فرض اجرا شد. به این معنی که ساقه‌های درخت coef تمامی پارامترهای سؤال را-به استثنای اولین سؤال که همیشه محدود به صفر است- استخراج می‌کند. Itempar تمامی پارامترهای سؤال را استخراج کرده و به‌صورت پیش‌فرض مجموع آنان را به صفر محدود می‌کند.

در پژوهش حاضر از روش درخت راش در سؤال‌های مطالعه پرلز ۲۰۱۶ کارکرد افتراقی تنها در بلوک ۱۶ تشخیص داده شد. در سؤال‌های دو ارزشی بلوک ۱۶، سؤال یک با کارکرد افتراقی یکنواخت بر اساس جنسیت شناسایی شدند. بلوک ۱۶ دارای متن با هدف کسب اطلاعات و کاربرد آن است.

در ادامه نتایج شناسایی کارکرد افتراقی سؤال در جدول (۳) با اعمال الگوریتم درخت راش بر روی بلوک ۱۶ پرسشنامه پرلز ۲۰۱۶ ارائه شده است. همان‌طور که در شکل (۲) دیده می‌شود این بلوک بر اساس جنسیت دو شاخه شده است که بیانگر وجود کارکرد افتراقی در این سؤال‌ها است.



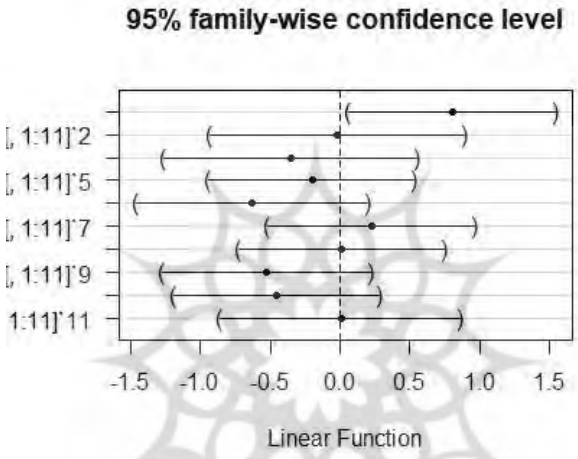
شکل (۲) کارکرد افتراقی بلوک ۱۶ درخت راش

جدول (۳) پارامترهای سؤال شناسایی‌شده با کارکرد افتراقی بر اساس جنسیت در مدل افراز بازگشتی درخت راش

هدف	بلوک	سؤال	Itempar		P-value	Z value	Std.Error	Estimate	نوع DIF
			Node3	Node2					
اطلاعاتی	بلوک ۱۶	۱	-1.20	-0.29	۰/۰۲	۲/۹۷	۰/۲۷	۰/۸۰	یکنواخت

مدل درخت راش در زمینه شناسایی کارکرد افتراقی سؤال، صرفاً به‌صورت کلی، با توجه به تمام متغیرها اعلام می‌نماید آیا کارکرد افتراقی وجود دارد یا خیر اما از تعیین شماره سؤال‌ها ناتوان است.

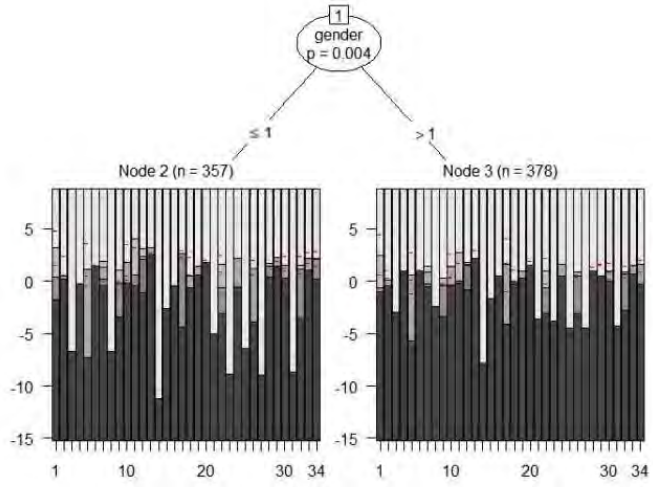
برای تشـخیص اینکه در مدل راش دقیقاً چه سـؤال‌هایی باعث ایجاد کارکرد افتراقی شـده‌اند نیاز به اسـتفاده از روش‌های دیگر همچون پکیج psychotools، دسـتور anchortests هستیم. در این روش، تعدادی از سـؤال‌ها به‌عنوان لنگر (anchor) که عاری از کارکرد افتراقی هسـتند تشخیص داده شده و باقی سـؤال‌ها با آنها مقایسه خواهند شد. این روش تنها برای سؤال‌های دو ارزشی کاربرد دارد. نتیجه اعمال این روش بر بلوک ۱۶ آزمون پرلز در شـکل (۳) آمده اسـت. سؤال ۴ به‌عنوان لنگر در نظر گرفته شـده است. بر اساس شـکل (۳)، سـؤال یک از فاصله اطمینان بیرون اسـت که بیانگر وجود کارکرد افتراقی یکنواخت در سؤال یک است.



شکل (۳) نمودار سطح اطمینان ۹۵٪ بلوک ۱۶

تحلیل سؤال‌های چندارزشی با استفاده از دستور pctree انجام گرفت:

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



شکل (٤) کارکرد افتراقی بلوک ١٦ درخت راش چندوجهی (□□□□□□)

نتیجه اعمال این روش بر سؤال‌های چندارزشی بلوک ۱۶ آزمون پرلز در شکل (۴) بیانگر کارکرد افتراقی بر اساس جنسیت است. در این
مرحله، الگوریتم راش pctree قادر به تشخیص دقیق سؤال یا سؤال‌های با کارکرد افتراقی نیست. در همین راستا پارامترهای سؤال‌های چندارزشی بلوک ۱۶ برای هر شاخه بررسی شد. لگاریتم درستنمایی بیشینه[1] در شاخه ۲ (دختران)، برابر (df = 92) ،10020- و در شاخه ۳ (پسران)، برابر (df = 92) ،10020- است. در ادامه پارامترهای سؤال‌های چندارزشی بلوک ۱۶ شاخه ۲ دختران در جدول (۴) قرار داده شده است. (Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1)

Partial credit model

Item category parameters:2$

Item#	Estimate Std.	Error	z value	Pr(>|z|)	
item1-C2	2.1153	0.4352	4.861	1.17E-06	***
item1-C3	8.6689	0.9044	9.586	< 2e-16	***
item1-C4	11.946	1.2836	9.307	< 2e-16	***
item2-C1	1.924	0.4994	3.853	0.000117	***
item2-C2	4.2438	0.8453	5.02	5.16E-07	***

1. Log-likelihood

Item#	Estimate Std.	Error	z value	Pr(>|z|)	
item2-C3	8.2982	1.2436	6.673	2.51E-11	***
item2-C4	8.6053	1.6099	5.345	9.04E-08	***
item3-C1	3.6164	0.4128	8.761	< 2e-16	***
item4-C1	5.0985	0.4322	11.798	< 2e-16	***
item5-C1	-0.422	0.5793	-0.728	0.466372	
item5-C2	4.8509	0.9447	5.135	2.82E-07	***
item5-C3	6.0832	1.2671	4.801	1.58E-06	***
item5-C4	8.2648	1.643	5.03	4.90E-07	***
item6-C1	3.2212	0.4174	7.717	1.19E-14	***
item6-C2	6.2866	0.8052	7.808	5.83E-15	***
item7-C1	1.2954	0.5104	2.538	0.011151	*
item7-C2	3.1907	0.8499	3.754	0.000174	***
item7-C3	6.2888	1.2295	5.115	3.14E-07	***
item7-C4	9.8787	1.619	6.102	1.05E-09	***
item8-C1	4.9091	0.4277	11.478	< 2e-16	***
item9-C1	3.2094	0.558	5.752	8.82E-09	***
item9-C2	4.9663	0.8521	5.828	5.60E-09	***
item9-C3	6.2969	1.2193	5.164	2.41E-07	***
item9-C4	9.0695	1.609	5.637	1.73E-08	***
item10-C1	2.2365	0.4996	4.476	7.59E-06	***
item10-C2	3.0991	0.8326	3.722	0.000198	***
item10-C3	8.3068	1.2331	6.736	1.62E-11	***
item10-C4	10.1828	1.6106	6.322	2.57E-10	***
item11-C1	1.2418	0.4801	2.587	0.009689	**
item11-C2	6.1628	0.8956	6.881	5.95E-12	***
item11-C3	5.8605	1.2182	4.811	1.50E-06	***
item11-C4	11.584	1.6256	7.126	1.03E-12	***
item12-C1	0.6112	0.4548	1.344	0.179019	
item12-C2	4.4416	0.8288	5.359	8.36E-08	***
item12-C3	8.4981	1.2226	6.951	3.63E-12	***
item12-C4	13.344	1.6542	8.067	7.22E-16	***
item13-C1	4.4178	0.4214	10.483	< 2e-16	***
item13-C2	8.4632	0.823	10.283	< 2e-16	***

Item#	Estimate Std.	Error	z value	Pr(>|z|)	
item13-C3	13.4012	1.2947	10.351	< 2e-16	***
item14-C1	3.8687	0.4142	9.341	< 2e-16	***
item15-C1	3.016	0.4114	7.332	2.27E-13	***
item16-C1	4.2964	0.418	10.279	< 2e-16	***
item17-C1	1.6312	0.4226	3.86	0.000113	***
item17-C2	5.9702	0.8154	7.321	2.45E-13	***
item17-C3	10.6447	1.2331	8.632	< 2e-16	***
item17-C4	14.6788	1.6563	8.862	< 2e-16	***
item18-C1	1.0454	0.5715	1.829	0.067395	.
item18-C2	3.3692	0.8917	3.778	0.000158	***
item18-C3	7.3307	1.276	5.745	9.19E-09	***
item18-C4	7.4797	1.6334	4.579	4.66E-06	***
item19-C1	2.3352	0.4204	5.555	2.78E-08	***
item19-C2	5.5045	0.8072	6.819	9.17E-12	***
item20-C1	3.6819	0.4172	8.825	< 2e-16	***
item20-C2	6.9828	0.8068	8.655	< 2e-16	***
item21-C1	3.6678	0.413	8.88	< 2e-16	***
item22-C1	2.294	0.4931	4.652	3.29E-06	***
item22-C2	3.4104	0.8303	4.107	4.00E-05	***
item22-C3	7.5423	1.2211	6.177	6.55E-10	***
item22-C4	10.1221	1.6086	6.292	3.13E-10	***
item23-C1	2.9811	0.4113	7.247	4.26E-13	***
item24-C1	4.1758	0.4187	9.973	< 2e-16	***
item24-C2	8.1043	0.8168	9.922	< 2e-16	***
item25-C1	3.3933	0.412	8.237	< 2e-16	***
item26-C1	1.1898	0.4613	2.579	0.009898	**
item26-C2	4.8946	0.8362	5.853	4.82E-09	***
item26-C3	7.2255	1.2154	5.945	2.77E-09	***
item26-C4	10.1985	1.6064	6.349	2.17E-10	***
item27-C1	2.923	0.4114	7.106	1.20E-12	***
item28-C1	5.0251	0.4372	11.494	< 2e-16	***
item28-C2	8.1265	0.8161	9.958	< 2e-16	***
item28-C3	11.5829	1.2115	9.561	< 2e-16	***

Item#	Estimate Std.	Error	z value	Pr(>|z|)	
item29-C1	3.63	0.501	7.245	4.32E-13	***
item29-C2	6.2841	0.8456	7.432	1.07E-13	***
item29-C3	6.8338	1.2058	5.668	1.45E-08	***
item29-C4	10.7939	1.6018	6.739	1.60E-11	***
item30-C1	2.0535	0.4262	4.818	1.45E-06	***
item30-C2	6.174	0.8196	7.533	4.97E-14	***
item30-C3	9.1572	1.2081	7.58	3.46E-14	***
item30-C4	11.8055	1.5989	7.384	1.54E-13	***
item31-C1	4.8354	0.4262	11.346	< 2e-16	***
item32-C1	3.0523	0.4255	7.174	7.29E-13	***
item32-C2	5.9542	0.8087	7.363	1.80E-13	***
item32-C3	10.0576	1.2104	8.309	< 2e-16	***
item32-C4	12.4254	1.5977	7.777	7.43E-15	***
item33-C1	3.4063	0.4321	7.884	3.18E-15	***
item33-C2	5.4521	0.8059	6.765	1.33E-11	***
item33-C3	9.8782	1.2081	8.176	2.92E-16	***
item33-C4	13.2614	1.6041	8.267	< 2e-16	***
item34-C1	1.9652	0.4222	4.654	3.25E-06	***
item34-C2	5.7389	0.8117	7.07	1.54E-12	***
item34-C3	10.3051	1.2212	8.439	< 2e-16	***
item34-C4	13.4535	1.6115	8.348	< 2e-16	***
item34-C4	13.4535	1.6115	8.348	< 2e-16	***

پرسش سوم پژوهش: وضعیت یا عملکرد نسبی روش درخت راش در شناسایی کارکرد افتراقی سؤال‌های آزمون پرلز ۲۰۱۶ بین دانش‌آموزان دختر و پسر ایرانی چگونه است؟

مدل درخت راش قادر به شناسایی کارکرد افتراقی سؤال با چندین متغیر به‌طور هم‌زمان است و حسن دیگر آن نتیجه ترسیمی‌اش است که تفسیر کار را بسیار ساده و کاربردی نشان می‌دهد. هرچند که قادر به شناسایی دقیق سؤال دارای کارکرد افتراقی نیست و به استفاده از روش دیگری برای تشخیص سؤال نیاز است.

بر اساس مدل درخت راش در بلوک ۱۶ کارکرد افتراقی شناسایی شد. بر اساس تست لنگر کارکرد افتراقی یکنواخت در سؤال ۱ از سؤال‌های دو ارزشی بر اساس جنسیت شناسایی شد.

بحث و نتیجه‌گیری

با توجه به اهمیت سواد خواندن به‌واسطه پیامدهای تحصیلی، اجتماعی و اقتصادی آن در پژوهش حاضر تلاش شد تا مشخص شود عملکرد ضعیف دانش‌آموزان ایرانی در آزمون سواد خواندن مطالعه پرلز، یک علت حقیقی مربوط به ساختار نظام آموزشی است یا مسئله متن‌های به‌کاربرده شده است. بنابراین کارکرد افتراقی سؤال‌های آزمون پرلز ۲۰۱۶ در بین دانش‌آموزان دختر و پسر ایرانی پایه چهارم بررسی شد. در ادامه یافته‌های این پژوهش مورد بررسی قرار می‌گیرد.

۱. آیا آزمون پرلز ۲۰۱۶ در دانش‌آموزان کشور ایران تک‌بعدی است؟

پیش از آنکه مسئله وجود کارکرد افتراقی را در آزمون بررسی کنیم، لازم است از تک‌بعدی بودن آزمون پرلز در دانش‌آموزان ایرانی مطمئن شویم. برای پاسخگویی به این پرسش، بعدیت سؤال‌های پرلز ۲۰۱۶ با دستورEAG و DIMTESTS بر اساس معیار تجربی کایزر[1]، روش هال[2]، تحلیل موازی[3] از طریق پکیج EFA.dimensions مورد بررسی قرار گرفت. بر اساس نتایج تحلیل بعدیت، آزمون پرلز ۲۰۱۶ تک‌بعدی است. نتایج تحلیل بعدیت سؤال‌های پرلز ۲۰۱۶، بیانگر تک‌بعدی بودن این سؤال‌ها هستند. این مسئله ازآن‌جهت مهم است که مدل درخت راش تنها در صورت تک‌بعدی بودن سؤال‌ها قادر به شناسایی کارکرد افتراقی است.

۲. آیا سؤال‌های پرلز ۲۰۱۶ در بین دانش‌آموزان دختر و پسر کشور ایران بر اساس مدل درخت راش دارای کارکرد افتراقی است؟

برای شناسایی کارکرد افتراقی سؤال‌های آزمون پرلز ۲۰۱۶ از روش افراز بازگشتی درخت‌محور از طریق پکیج Psychotree در محیط R استفاده کردیم. با توجه به متغیرهای سن، جنسیت و مدرسه دانش‌آموزان، روش درخت راش قادر به تشخیص کارکرد افتراقی بر اساس جنسیت در بلوک ۱۶ شد. در این بلوک برای سؤال‌های دو ارزشی به‌وسیله دستور raschtree کارکرد افتراقی یکنواخت بر اساس جنسیت شناسایی شد. همچنین برای سؤال‌های چندارزشی به‌وسیله دستور pctree کارکرد افتراقی بر اساس جنسیت مشاهده شد.

۳. وضعیت یا عملکرد نسبی روش درخت راش در شناسایی کارکرد افتراقی سؤال‌های آزمون پرلز ۲۰۱۶ بین دانش‌آموزان دختر و پسر ایرانی چگونه است؟

در بررسی عملکرد روش درخت راش در شناسایی کارکرد افتراقی سؤال‌های پرلز ۲۰۱۶ در بین دانش‌آموزان دختر و پسر ایرانی درمی‌یابیم مدل راش قادر به تشخیص کارکرد افتراقی سؤال با چندین

1. The Empirical Kaiser Criterion
2. Hull method
3. Parallel Analysis

متغیر است. به‌علاوه مدل راش نیاز به تعیین نقطه برش ندارد. مدل راش این امکان را دارد که احتمال پاسخ صحیح در بین آزمون‌دهندگان با توانایی یکسان را مورد ملاحظه قرار دهد. مضافاً سریع‌ترین بررسی DIF و پارامترهای دشواری سؤال توسط مدل راش به سبب خاصیت ترسیمی‌اش صورت می‌گیرد. بااین‌حال، ناتوانایی در تشخیص دقیق سؤال دارای DIF در مدل راش و نیاز به روشی مجزا همچون Anchortest که تنها برای سؤال‌های دو ارزشی کاربرد دارد از نکات منفی این روش است.

در تحلیل ذکرشده در بالا دیده شد که در بلوک ۱۶ کارکرد افتراقی در آزمون پرلز ۲۰۱۶ بین دانش‌آموزان ایرانی وجود دارد. با اینکه در سؤال‌های دو ارزشی تنها یک سؤال دارای کارکرد افتراقی یکنواخت شناسایی شد. اما در سؤال‌های دو ارزشی نمی‌توانیم مشخص کنیم کدام سؤال‌ها دارای کارکرد افتراقی است. بنابراین برای قضاوت منصفانه بین گروه‌های مختلف آموزشی دانش‌آموزان ایرانی در پایه چهارم سؤال‌های بلوک ۱۶ نیازمند بازبینی و یا حذف از آزمون دارند. همچنین در سؤال‌هایی که اغلب از دشوارترین سؤال‌های متن بودند هر دو گروه دختر و پسر به نسبت میانگین بین‌المللی عملکرد ضعیف‌تری داشته‌اند. این عملکرد ضعیف می‌تواند دلایل دیگری غیر از جنسیت داشته باشد. مثلاً ترجمه متون مناسب نبوده یا دانش‌آموزان هنوز به توانایی لازم برای این سطح از خواندن دست نیافته‌اند. این مسئله شایان توجه است که ساختار آموزشی ما نیازمند فراهم کردن بستر مناسبی برای اصلاح و ارتقای سواد خواندن دانش‌آموزان ایرانی است.

مطالعات محدودی در زمینه پرلز در ایران انجام گرفته است. ازجمله محدودیت‌های این پژوهش نبود امکان الگوبرداری از کارهای داخلی بود، به‌خصوص در مورد پرلز ۲۰۱۶ که به بحث و ویژگی‌های روان‌سنجی آن پرداخته نشده بود. درنتیجه مجبور به یادگیری همه ملزومات تحقیق از صفر شده که سبب طولانی‌تر شدن مدت تحقیق شد. پیشنهاد می‌شود در مطالعه‌های بعدی برای شناخت بهتر عوامل مؤثر در سواد خواندن از چهار پرسشنامه پیشینه‌کاوی پرلز ۲۰۱۶ شامل پرسشنامه دانش‌آموز، معلم، مدیر مدرسه و والدین استفاده شود.

تقدیر و تشکر

بدین‌وسیله از آقای دکتر سید جواد کاظمی‌تبار، استادیار دانشکده برق و کامپیوتر دانشگاه صنعتی نوشیروانی بابل جهت همکاری در این تحقیق تشکر و قدردانی می‌شود.

References

Ahmadi Deh Qutbuddini M., Khodai E., Farzad V., Moghadam-Zadeh A., & Kabiri M. (2017). Applying Bi-factor Multidimensional Item-response Theory Model for Dimensionality and Differential items Functioning Analysis on Testlet-Based Tests. *Quarterly of*

Educational Measurement, 7(28), 121-153. [Persian]

Alimirzaie, M., Moghadam-Zadeh, A., Minaei, A., Ezanloo, B., & Salehi, K. (2019). Sources of Question Differential Function and its Application in Education. *Journal of Research in Teaching, 7*(1), 133-153. [Persian]

Alshehri, Y. M. (2020). *Empirical Evaluation for Gender-Based DIF in an International Assessment: The Case of PIRLS*. Doctoral dissertation, University of Kansas.

American Educational Research Association. (2018). *Standards for educational and psychological testing*. American Educational Research Association.

Asgari, A. (2007). New Approaches to Psychometrics Part Three: Models of Item Response Theory, Rasch Models. *Journal of Developmental Psychology: Iranian Psychologist, 4*(13), 80-83. [Persian]

Bollmann, S., Berger, M., & Tutz, G. (2017). Item-Focused Trees for the Detection of Differential Item Functioning in Partial Credit Models. *Educational & Psychological Measurement*, *78*(5),781-804.

Doi: 10.1177/0013164417722179. Epub 2017 Sep 25.

Geske, A., & Ozola, A. (2010). Differential item functioning in the aspect of gender differences in reading literacy. *In The 4th IEA International Research Conference.*

Giraud, F. (2007). *Christian Baudelot, Roger Establet, Quoi de neuf chez les filles?* Entre stéréotypes et libertés. Lectures.

Hagquist, C., & Andrich, D. (2004). Is the sense of coherence instrument applicable on adolescents? A latent trait analysis using Rasch-modelling. *Personality and Individual Differences*, 36, 955- 968.

Jeffers, H. (2020). *Using rasch tree to detect uniform differential item functioning and item difficulty parameters in the progress in international reading literacy data (PIRLS).* Ball State University.

Kabiri, M., Karimi A., & Bakhshali-Zadeh, S. (2017). *PIRLS 2016 findings*, Madreseh Publications, Tehran, Iran. [Persian]

Kennedy, A. M. (2008). Examining gender and fourth graders' reading habits and attitudes in PIRLS 2001 and 2006. *In 3rd IEA International Research Conference, Taipei, Chinese Taipei.*

Linacre, J. M. (2006). *A user's guide to Winsteps: Rasch model computer programs*. Chicago: Winsteps.

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. Routledge.

Minaie, A. (2015). Aapplication of Rasch measurement model to evaluate measurement properties of the Test of Visual- Motor Skills-Revised. *Educational Measurement, 5*(18), 77-114. [Persian]

Minaei, A., & Ghafari, Z. (2015) Differential Items functioning of Mathematics Test of TIMSS 2011 in Grade 8 between Girls and Boys using Item-Response Theory Approach (IRT). *Educational Measurement, 6*(21), 21-40. [Persian]

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 International Results in Reading. Retrieved from Boston College,* TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/pirls2016/international-results/

Mullis, I. V., Kennedy, A. M., Martin, M. O., & Sainsbury, M. (2006). *Assessment Framework and Specifications. PIRLS 2006.* International Association for the Evaluation of Educational Achievement.

Roux, K. (2020). *Examining the equivalence of the PIRLS 2016 released texts in South Africa across three languages.* Doctoral dissertation, University of Pretoria.

Sari, I. M. (2020). *Decomposing Gender Differences in Reading: Evidence from Nordic PIRLS 2016* Master's thesis.

Strobl, C., Kopf, J., & Zeileis, A. (2011a). *A new method for detecting differential item functioning in the Rasch model.* Innsbruck: University of Innsbruck Working Papers in Economic and Statistic.

Strobl, C., Kopf, J., & Zeileis, A. (2011b*). Using the raschtree function for detecting differential item functioning in the Rasch model.* Innsbruck: University of Innsbruck Working Papers in Economic and Statistic.

Strobl, C., Kopf, J., & Zeileis, A. (2015). Rasch trees: A new method for detecting differential item functioning in the Rasch model. *Psychometrika, 80*(2), 289-316.