Educational Measurement and Evaluation Studies

National Organization for Educational Testing

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

Identifying Differential Item Functioning in PIRLS 2016: A Rasch-Tree Based Approach

Elham Pirmoradian[1], Asghar Minaei[2]

1. Msc. Graduate, Islamic Azad University, Email:elham.pirmoradian@gmail.com.
2. Associate Professor, Educational and Psychological Measurement, Allameh Tabataba'i University; (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: A growing concern in unfairness of educational assessments is the possible presence of differential item functioning. Differential item functioning or bias will undermine the validity of the assessment. In this study, identification of differential item functioning in PIRLS 2016 test among Iranian students has been considered via the Rasch-Tree model.

Methods: The data used in this research are from PIRLS 2016 exam results that was also held in Iran. A total of 438∘ fourth grade Iranian students were involved consisting of 2143 girls and 2242 boys with the average age of 10.14. The analysis performed by the authors on this dataset was written in R programming language.

Results: The results indicate that out of 181 questions, based on raschtree model, one dichotomous question shows uniform differential item functioning. Moreover, among polytomous questions, block 16 shows differential item functioning.

Conclusion: It appears even though there is differential item functioning in the PIRLS 2016 test for Iranian students, the impact is negligible. Ideally, however, we would want to remove the questions with differential item functioning before conducting any analysis. Using test results requires care and discretion.

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

Cite this article: Pirmoradian, Elham; Minaei, Asghar (2023). Identifying Differential Item Functioning in PIRLS 2016: A Rasch-Tree Based Approach. *Educational Measurement and Evaluation Studies*, 13 (43): 28-44 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. دانش آموخته 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. چرا که کارکرد افتراقی یا سوگیری باعث تردید در روایی آزمون می‌شود. در مطالعهٔ حاضر شناسایی کارکرد افتراقی سوال‌های 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 دو مرحله‌ای تصادفی، تعداد ۴۳۸۵ نفر دانش‌آموز ایرانی شامل ۲۱۴۳ دانش‌آموز دختر (۴۸/۸۷ درصد) و ۲۲۴۲ دانش‌آموز پسر (۵۱/۱۳ در صد) 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 سوال در این مطالعه بر اساس جنسیت، بر اساس مدل درخت راش 1 سوال در میان سوالات دو ارزشی با کارکرد افتراقی یکنواخت شناسایی شد. به علاوه در بین سوالات چندارزشی، برای بلوک ۱۶ کارکرد افتراقی شناسایی شد.

نتیجه‌گیری: به نظر می‌رسد با وجود کارکرد افتراقی تاثیر آن قابل صرف‌نظر است. اگرچه بهتر است که سوال‌های با کارکرد افتراقی هنگام انجام تحلیل‌ها کنار گذاشته شود. عملکرد ضعیف دانش‌آموزان ایرانی می‌تواند متاثر 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 سوال، مدل راش، مدل درخت راش

استناد: پیرمرادیان، الهام. مینائی، اصغر (۱۴۰۲). کارکرد افتراقی در آزمون پرلز ۲۰۱۶ با استفاده از درخت راش. *مطالعات اندازه‌گیری و ارزشیابی آموزشی*،۱۳ (شماره۴۳)، ۲۸-۴۴ 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). با این حال، ممکن است که گروه‌های مختلف مردم به طور متفاوتی نسبت به یک آزمون واکنش نشان دهند، که اعتبار اندازه‌گیری‌ها را تحت‌الشعاع قرار می‌دهد. همچنین، اگر آزمون ها به نتایج متفاوتی برای گروه های متمایز از افراد منجر شوند، آزمون دچار سوگیری خواهد شد. کارکرد افتراقی سوال[3] به این معنی است که تغییرناپذیری سوال نقض می شود. به طور دقیق تر، کارکرد افتراقی سوال در صورتی وجود دارد که یک یا چند سوال برای یک گروه به طور قابل توجهی دشوارتر از گروه دیگر باشد. می توان بین کارکرد افتراقی یکنواخت و غیر یکنواخت تمایز قائل شد. کارکرد افتراقی یکنواخت به این معنی است که تفاوت بین گروه ها 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 سوال پاسخ[4] و روش‌های غیرپارامتریک طبقه‌بندی کرد. روش های سوال پاسخ که روش های پارامتریک نیز نامیده می شوند، روش هایی هستند که در آنها از مدل IRT برای تشخیص DIF استفاده می شود. ایده اصلی روش‌های سنتی تشخیص DIF در هر دو مدل IRT دو ارزشی[5] و چند ارزشی[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 مدل های IRT استفاده شد، آزمون نسبت درستنمایی[7] بود (اندرسن، ۱۹۷۳). یک رویکرد جایگزین که می تواند برای هر نوع مدل IRT استفاده شود، آزمون مربع کای لرد[8] است. در حالی که این آزمون به مقایسه دو گروه محدود می شود، بسط آن توسط کیم، کوهن و پارک، که با نام آزمون لرد تعمیم یافته[9] معروف است، می تواند برای بیش از یک گروه کانونی استفاده شود. رویکرد سوم روش راجو[10] است، که 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 بیش از یک متغیر کمکی القاکننده DIF در یک زمان دشوار است. (استلا بولمن و همکاران[11]، ۲۰۱۷)

امروزه سواد خواندن به واسطه پیامدهای تحصیلی، اجتماعی و اقتصادی مورد توجه خاص محققان قرار گرفته است. زیرا سواد خواندن نقش تعیین کننده‌ای در تحصیل بچه‌ها در مدرسه و همچنین در کارکرد بزرگسالان بعد از مدرسه ایفا می‌کند. مطالعه بین‌المللی پیشرفت سواد خواندن پرلز[12] 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 سال‌های۲۰۰۱، 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[13]، 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

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
11. Stella Bollmann, Moritz Berger, and Gerhard Tutz
12. Progress in International Reading Literacy Study (PIRLS)
13. Kabiri M., Karimi A., & Bakhshali-Zadeh S.

ضعیف دانش‌آموزان ایرانی در آزمون سواد خواندن مطالعه پرلز، یک علت حقیقی مربوط به ساختار نظام آموزشی است یا مسئله متن‌های به‌کار برده شده است (احمدی و همکاران[1]، ۱۳۹۶). انجام یک مطالعه با استفاده از رویکردهای جدید برای شناسایی کارکرد افتراقی روی داده‌های مطالعه بین‌المللی پرلز دارای اهمیت است. چرا که در نتیجه این‌گونه پژوهش‌ها، مشخص می‌شود، این ابزار کارکرد خود را به درستی انجام می‌دهد و برای همه کشورهای شرکت‌کننده نتایج کاربردی و قابل محاسبه‌ای به دست می‌دهد. در نتیجه تفاوت‌های مشاهده شده در نتایج، به ویژگی‌های آزمون و سوال‌های آن و آزمودنی‌ها ارتباطی ندارد و در صورت شناسایی کارکرد افتراقی، سوالات 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. تا کنون در زمینه شناسایی کارکرد افتراقی سوال‌های مطالعه پرلز ۲۰۱۶ بین دانش‌آموزان ایرانی مطالعه‌ای نشده است. لذا اهداف این تحقیق شناسایی کارکرد افتراقی سوال‌های پرلز۲۰۱۶ با استفاده از مدل درخت راش است. مساله‌ای که می‌خواهیم در این تحقیق به آن بپردازیم به شرح زیر است. 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[2] با هدف به دست آوردن درک عمیق‌تری از تاثیر 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 به وسیله این انجمن اداره می شود (مولیس،۲۰۱۷). نخستین مطالعه بین‌المللی پیشرفت سواد خواندن (پرلز) در سال ۲۰۰۱ انجام شد. بر اساس اهداف این مطالعه، پایه چهارم، زمانی مهم در رشد دانش‌آموزان است. در این پایه دانش‌آموزان چگونه خواندن و مطالعه کردن را یادگرفته‌اند 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 شرکت کننده از جمله ۵۰ کشور و ۱۱ نهاد به عنوان معیارسنجش 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 از جمله 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 ۴ دوره، پیشرفت سواد خواندن را ارزیابی کند. کشورهای شرکت کننده 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، جمع آوری می‌شود. علاوه بر آن در هر دوره دانشنامه‌ای به چاپ می‌رسد که کشورهای شرکت کننده 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 سوال است (مولیس، ۲۰۱۷). هدف پرلز ارائه بهترین اطلاعات مربوط به سیاست‌ها دربارهٔ چگونگی بهبود آموزش و یادگیری و کمک به دانش‌آموزان جوان است. مولیس و همکاران (۲۰۰۹) بیان می‌کنند، مطالعه پرلز به سه بعد از سواد خواندن دانش آموزان تمرکز دارد: ۱- اهداف خواندن، ۲- فرایندهای درک مطلب، ۳- رفتارها و نگرش‌های خواندن.

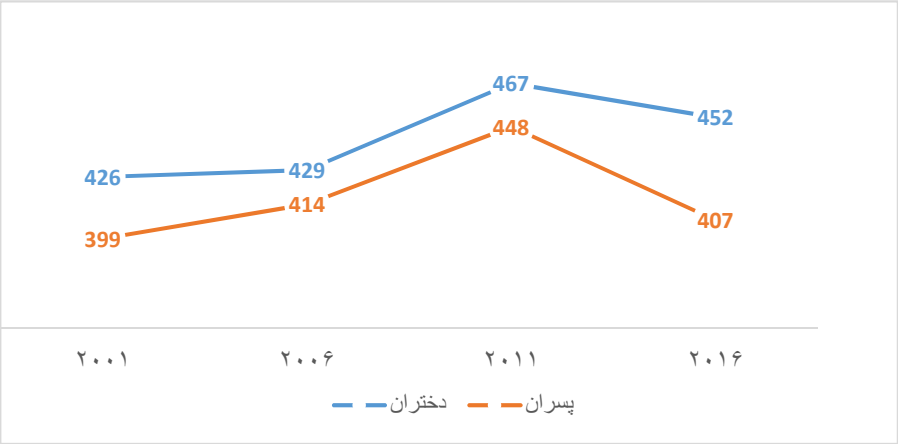
جدول ۱. اهداف، فرایندهای مورد ارزیابی و درصد تاکید بر آنها در مطالعه پرلز

اهداف خواندن[3]	درصد
تجربهٔ ادبی[4]	٪۵۰
کسب اطلاعات و استفاده از آن[5]	٪۵۰
فرایندهای درک مطلب[6]	درصد

1. Ahmadi Deh Qutbuddini M., Khodai E., Farzad V., Moghadam-Zadeh A., & Kabiri M
2. International Association for the Evaluation of Educational Achievement(IEA)
3. Purposes for Reading
4. Literary Experience
5. Acquire and Use Information
6. Process of Comprehension

تمرکز و بازیابی اطلاعاتی که به صراحت بیان شده[1]	٪۲۰
استنباط‌های مستقیم[2]	٪۳۰
تفسیر و تلفیق ایده‌ها و اطلاعات[3]	٪۳۰
بررسی و ارزیابی محتوا، زبان و عناصر متنی[4]	٪۲۰

از نظر آموزشی، توجه به تفاوت‌های جنسیتی در کشورهای توسعه یافته و به ویژه وضعیت دختران ممکن است قدیمی به نظر برسد. دختران با گسترش تحصیلات اجباری بسیار پیشرفت کرده‌اند (جراد ، ۲۰۰۷). در گزارش پیزا[5] 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 نمودار۱ روند تغییرات عملکرد خواندن دانش آموزان ایرانی به تفکیک دو جنس ارائه شده است. با اینکه در این شکل برتری دختران را در همهٔ این سال‌ها نشان می‌دهد، ولی لزوما روند عملکرد دو جنس به یک شکل نبوده است. به عنوان مثال 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 به دست آمده است (کبیری و همکاران، ۱۳۹۶).



نمودار 1. 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 مطالعهٔ تیمز[6]، اعلام می‌دارد که دانش‌آموزان پایه هشتم که در منزل به کتاب‌ها یا وسائل کمک‌آموزشی بیشتری دسترسی دارند یا پدر و مادرشان تحصیلات دانشگاهی دارند، مدارج بالاتری در ریاضیات و علوم به‌دست می‌آورند. در مجموع، 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 جزءکشورهای 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 نیز پایین

[1]. Focus on and Retrive Explicity Stated Information
[2]. Make Straightforward Inferences
[3]. Interpret and Integrate Ideas and Information
[4]. Evaluate and Critigue content and Textual Elements
[5]. Programme for International Student Assessment (PISA)
[6]. Trends in International Maths and Science Study (TIMSS)

است و در سال‌های اخیر نیز این میزان کاهش داشته است که البته این رویه کاهشی در بسیاری از کشورهای شرکت کننده در مطالعه پرلز نیز مشاهده 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[1] در سال ۱۹۶۰ مفهوم عینیت و تغییرناپذیری را در قالب یک مدل آماری مطرح ساخت و روشی برای آزمون آنها ابداع کرد. مدل راش در واقع خانواده‌ای از مدل‌ها است که از آن جمله می‌توان به مدل لجستیک ساده[2]، مدل مقیاس درجه‌بندی[3] و مدل امتیازدهی پاره‌ای[4] اشاره کرد(مینائی،۱۳۹۳). مدل‌های اندازه‌گیری راش به پژوهشگران امکان می‌دهد تا مشکلات زیربنایی اندازه‌گیری‌های مدل کلاسیک را حل کنند. لاینرس (۲۰۰۶) مدل‌های راش را در دو طبقه کلی دو ارزشی[5] و چند ارزشی[6] تقسیم‌بندی کرد(عسگری، ۱۹۸۵) . مدل راش، مورد خاصی از مدل لجستیکی ۱ پارامتری است که در آن تمیز روی ۱ تنظیم شده است. هدف مدل راش بدست آوردن سنجشی از متغیرهای پنهان و حداقل، که به صورت فاصله‌ای مقیاس‌بندی شده‌اند می‌باشد. مدل راش احتمال وقوع پاسخ خاص (به عنوان مثال پاسخ درست) را با دشواری سوال پیوند می‌دهد. مدل از این معادله استفاده می‌کند:

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

که در آن θ، نشان دهنده‌ی توانایی وb نشان دهنده دشواری سوال است و فرض می‌شود تمیز برابر ۱ است. تمیز روی یک فرض می‌شود چرا که مدل راش اینگونه فرض می‌کند که هر سوال میان افراد دارای توانایی بالا و پایین به شکل کاملی تمییز ایجاد می‌کند[7].

تغییرناپذیر[8] نبودن پارامترهای سوال و آزمون شوندگان، به عنوان مثال در بین دو گروه مردان و زنان، معمولا کارکرد افتراقی سوال[9] نامیده می شود و به این معنا است که افرادی از دو گروه مردان و زنان که از توانایی یکسان برخوردارند، در یک سوال خاص بطور متفاوتی عمل می کنند(هاگ کوئیست و آندریش ، ۲۰۰۴؛ به نقل از مینائی، ۱۳۹۳). روش افراز بازگشتی[10] مبتنی بر مدل یک رویکرد نیمه پارامتری است که از آزمون‌های آماری اقتصاد سنجی برای تغییر ساختاری استفاده می‌کند. هدف این است که تفاوت در پارامترهای یک مدل آماری بین گروه‌های افراد تعریف شده توسط (ترکیب) متغیرها را تشخیص دهد. افراز بازگشتی مبتنی بر مدل مربوط به روش طبقه‌بندی و درختان رگرسیون است که در آن فضای متغیر کمکی، به صورت بازگشتی تقسیم می شود تا گروه‌هایی از افراد با ارزش‌های مختلف یک متغیر پاسخ طبقه‌ای یا پیوسته را شناسایی کند (استروبل و همکاران[11]، ۲۰۱۳).

اخیرا رویکرد نیمه پارامتریک جدیدی از تست کارکرد افتراقی در مدل‌های راش بر اساس اِفراز بازگشتی توسط استروبل، کوپف، و زیلیس (۲۰۱۳) مطرح شده است. به بیانی ساده، رویکردهای نیمه پارامتریک، دارای مولفه‌های رویکردی پارامتریک و غیرپارامتریک می‌باشند. بخش غیرپارامتریک این است که افراز بازگشتی خطی فرض نشود. با استفاده از روش درخت راش، متغیرهای کمکی تعریف شده، یا آمیزه‌ای از متغیرهای کمکی را می‌توان به منظور شناسایی تفاوت‌ها در پارامترهای مدل استفاده نمود. رویکرد فعلی با استفاده از تکنیک‌های آماری نوین برای قوی‌تر بودن به نسبت درخت‌های طبقه بندی و رگرسیون، از پارمترهای مدل پارامتریک به جای یک متغیر پاسخ استفاده می‌کند. پاسخ‌ها به سوال با در نظر گرفتن تفاوت‌های گروهی مربوط به متغیرهای کمکی ثبت می شوند. مدل سپس به شکل یک تصویر یا درخت نشان داده می‌شود که هر ساقه‌ی انتهایی نشان دهنده‌ی براوردهای پارامتر می‌باشد. اگر بیش از یک ساقه وجود داشته باشد، فرضیه‌ی صفرمورد سنجش،

1. Georg William Rasch
2. simple logostic model
3. rating scale model
4. partial credit model
5. dichotomous
6. ploytomous
7. De Ayala, 2009
8. Invariance
9. differential item functioning(DIF)
10. recursive partitioning
11. Strobl, C., Kopf, J., & Zeileis, A.

که هیچ تفاوت گروهی وجود ندارد رد خواهد شد. یکی از مزیت‌های درخت‌های راش این است که نتایج به شکلی تصویری نمایش داده می‌شوند (استروبل و همکاران[1]،a۲۰۱۱). در این روش نیازی نیست متغیرها برای الگوریتم تخصیص داده شوند و این، در مقایسه با مدل‌های پارامتریک، در کار شناسایی گروه‌های متاثر از کارکرد افتراقی سهولت ایجاد می‌کند. به عبارتی دیگر، 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 طبقه بندی و رگر سیون ا ست دا شته و از آزمون‌ها برای تغییر ساختار در رویکردی نیمه پارامتریک ا ستفاده می‌کند. و متغیرهای کمکی فرد در گروه‌ها برای شناسایی تفاوت‌ها در پارامترهای مدل‌های آماری تحلیل می‌شوند.

در مدل راش فرض می‌شود که تخمین پارامتر سوال وابسته به گروه نمونه نیست. این فرض در صورتی که برای گروه‌های خاصی از افراد – بدون در نظر گرفتن توانایی واقعی آنها – پاسخ به سوالات خاصی آسان‌تر یا دشوارتر باشد ممکن است نقض شود؛ در این گونه مواقع است که ما از کارکرد افتراقی سخن به میان می‌آوریم. به منظور شناسایی کارکرد افتراقی سؤال با تابع درخت راش، پاسخ‌های سؤال و همه متغیرهای کمکی که باید از نظر DIF آزمایش شوند باید به عنوان ورودی در اختیار روش درخت راش قرار داده شوند (استروبل و همکاران[2]، b.۲۰۱۱).

مزیت اصلی رویکرد درخت راش این است که DIF بین گروهی از آزمودنی‌ها با بیش از یک متغیر کمکی قابل تشخیص است. به عنوان مثال، در مقایسه با سایر افراد سؤالات 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 به تنهایی در نظر گرفته شود (استروبل و همکاران، b.۲۰۱۱). 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) به تفصیل شرح داده شده است (استروبل و همکاران، b.۲۰۱۱).

علی میرزایی، مقدم زاده، مینائی، ایزانلو، و صالحی (۱۳۹۸) بیان کرده‌اند که ویژگی‌های روان سنجی مثل تغییرناپذیری اندازه‌گیری، نبود کارکرد افتراقی سؤال یا فقدان سوگیری باید برقرار با شد 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 حاکی از آن ا ست که اهمیت ا صلی شنا سایی منابع DIF در ساخت و تف سیر نتایج آزمون‌ها ا ست که برای مقایسهٔ بین گروه‌ها به کار برده می‌شود. سیاستگذاران باید هنگام تصمیم‌گیری دربارهٔ برنامه درسی، منابع یا آموزش بر مبنای مقایسه نمرات آزمون، مراقب روایی مقایسه‌پذیری نمرات باشند.

روکس[3] (۲۰۲۰) با استفاده از یافته‌های بین‌المللی پرلز ۲۰۱۶ مبنی بر ضعف شدید دانش‌آموزان جنوب آفریقا در درک مطلب نسبت به دیگر کشورهای شرکت‌کننده به برر سی 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 متخصصین، بیانگر معادل نبودن متون پرلز با زبان‌های مورد نظر است. با این حال، الگوی مشخصی در سوال‌های دارای کارکرد افتراقی وجود نداشته است و همچنین هیچگونه تبعیض یا بی‌علاقگی نسبت به زبان دیده نشد. با نگاهی عمیق به متون و سوال‌هایش مشخص شد که متن گل‌های روی پشت‌بام فقیرترین ترجمه به آفریقایی و ایسی زولو تلقی می‌شود. به طور کلی، همه متون برای زبان آموزان آفریقای جنوبی مناسب تلقی شدند. زیرا متون از واژگان غنی استفاده می‌کردند و دانش‌آموزان را با ایده‌ها و مفاهیم جدید آشنا می‌کردند.

الشهری[4] (۲۰۲۰) در تحقیقی به تجزیه و تحلیل کارکرد افتراقی سوال‌های پرلز ۲۰۱۶ بر اساس جنسیت در سه کشور که به زبان مشترک صحبت می‌کردند؛ پرداخت. در این مطالعه از رگر سیون دو جمله‌ای لج ستیک برای تشخیص کارکرد افتراقی یکنواخت و غیریکنواخت در سه کشور انتخاب شده به طور تصادفی استفاده شده است. یافته‌ها به وضوح نشان دهنده 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.

بررسی‌ها نشان می‌دهد که در خصوص کارکرد افتراقی سوال‌های آزمون بین المللی پرلز بر اساس جنسیت مطالعات متعددی همچون (جفرز ،۲۰۲۰)، (الشهری،۲۰۲۰)، (ساری،۲۰۲۰)، (احمدی و همکاران، ۲۰۱۷)، (جسکه و ازولا، ۲۰۱۰)، (کندی، ۲۰۰۸) صورت گرفته است. هر چند در

1. Strobl, Kopf, & Zeileis
2. Strobl, C., Kopf, J., & Zeileis, A.
3. Roux, k.
4. Alshehri, Yahya M

کشور ایران با توجه به رتبه‌ای که در این آزمون کسب شده است تا کنون مطالعه‌ای برای شناسایی کارکرد افتراقی سوال‌های سواد خواندن پرلز ۲۰۱۶ از طریق روش‌های مبتنی بر مدل درخت راش انجام نشده است .

سوال‌های پژوهش:

۱. آیا آزمون پرلز ۲۰۱۶ در دانش¬آموزان کشور ایران تک بعدی است؟
۲. آیا سوال‌های 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 سوال‌های آزمون پرلز ۲۰۱۶ بین دانش آموزان دختر و پسر ایرانی چگونه است؟

روش پژوهش

مطالعه پرلز از یک طرح نمونه‌گیری دو مرحله‌ای تصادفی استفاده می‌کند. در کشور شرکت کننده در مرحله اول از بین مدارس موجود در کشور با خصوصیات مختلف جنسیت، جغرافیایی (استانی)، نوع مدرسه (دولتی یا غیردولتی) و منطقه آن (شهری یا روستایی) تعدادی مدرسه به طور تصادفی انتخاب شده، سپس یک یا چند کلاس کامل از هر مدرسهٔ نمونه‌گیری شده به طور تصادفی به عنوان 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 به طور معمول 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 استاندارد شده 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 در صد) و ۲۲۴۲ دانش‌آموز پسر (معادل ۵۱/۱۳ در صد) با میانگین سنی ۱۰/۱ و انحراف استاندارد ۰/۴۸ هستند. آزمون پرلز ۲۰۱۶، دارای ۱۸۱ سوال است که در ۱۶ دفترچه ارائه شده‌اند. هر دفترچه حاوی دو متن و سوال‌های مربوط به آن است. جدول ۳ بیانگر اطلاعات کاملی از هر دفترچه می باشد. این دفترچه‌ها به صورت تصادفی بین دانش‌آموزان توزیع شدند. از بین ۱۸۱ سوال سواد خواندن در پرلز ۲۰۱۶، تعداد ۸۶ سوال دو ارزشی (چهارگزینه‌ای) و ۹۵ سوال چند ارزشی هستند. در پژوهش حاضر برای تحلیل داده‌ها سوال‌های چند ارزشی[1] و دو ارزشی[2] به صورت جداگانه بررسی شده اند. همچنین برای سوالات بدون پاسخ که دانش‌آموزان به دلیل ندانستن پاسخ صحیح آنها را خالی گذاشته‌اند یا سوال‌های جا مانده که به دلیل کمبود وقت قادر به پاسخگویی به سوال‌ها نبوده‌اند؛ نمره صفر داده شد.

جدول ۲. **توزیع دفترچه های آزمون پرلز۲۰۱۶، تعداد سوال، نوع متن و شکل سوال**

دفترچه	نام متن	تعداد سوال	نوع متن	سوال دو ارزشی	سوال چند ارزشی
۱	گل‌های روی پشت بام	۱۳	ادبی	۷	٦
	اسب‌های ایسلندی	۱۵	اطلاعاتی	٦	۹
۲	کهربای درخشان	۱٤	ادبی	۸	٦
	اسب‌های ایسلندی	۱۵	اطلاعاتی	٦	۹
۳	کهربای درخشان	۱٤	ادبی	۸	٦
	چگونه پرواز را یاد گرفتیم؟	۱۷	اطلاعاتی	۸	۹
٤	گلدان خالی	۲۰	ادبی	۱۱	۹
	کوسه‌ها	۱۲	اطلاعاتی	٦	٦
۵	گلدان خالی	۲۰	ادبی	۱۱	۹
	چگونه پرواز را یاد گرفتیم؟	۱۷	اطلاعاتی	۸	۹
٦	الیور و گریفین	۱۳	ادبی	٤	۹
	کوسه‌ها	۱۲	اطلاعاتی	٦	٦

[1]. polytomous
[2]. dichotomous

دفترچه	نام متن	تعداد سوال	نوع متن	سوال دو ارزشی	سوال چند ارزشی
۷	الیور و گریفین	۱۳	ادبی	٤	۹
	لئوناردو داوینچی	۱۲	اطلاعاتی	٦	٦
۸	پمبا شرپا	۱۷	ادبی	۸	۹
	لئوناردو داوینچی	۱۲	اطلاعاتی	٦	٦
۹	گل های روی پشت بام	۱۳	ادبی	۷	٦
	عسل کجاست؟	۱٦	اطلاعاتی	۷	۹
۱۰	پمبا شرپا	۱۷	ادبی	۸	۹
	عسل کجاست؟	۱٦	اطلاعاتی	۷	۹
۱۱	کهربای درخشان	۱٤	ادبی	۸	٦
	کوسه ها	۱۲	اطلاعاتی	٦	٦
۱۲	الیور و گریفین	۱۳	ادبی	٤	۹
	چگونه پرواز را یاد گرفتیم؟	۱۷	اطلاعاتی	۸	۹
۱۳	گل‌های روی پشت بام	۱۳	ادبی	۷	٦
	لئوناردو داوینچی	۱۲	اطلاعاتی	٦	٦
۱٤	پمبا شربا	۱۷	ادبی	۸	۹
	اسب‌های ایسلندی	۱٥	اطلاعاتی	٦	۹
۱٥	گلدان خالی	۲۰	ادبی	۱۱	۹
	عسل کجاست؟	۱٦	اطلاعاتی	۷	۹
۱٦	میسی و مرغ قرمز	۱٦	ادبی	۸	۸
	لاک‌پشت دریایی سبز	۱٦	اطلاعاتی	۷	۹

Error! Reference source not found.

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 **9/83** و **14/83** با میانگین سنی **10/14** و انحراف استاندارد ۰/۴۸ هستند. ۱۶ دفترچه برای آزمون پرلز ارائه شدند که هر دفترچه حاوی دو متن و سوال‌های 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.

سوال اول پژوهش:

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

بعدیت هر بلوک 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 تک بعدی است. به عنوان مثال مقدار ویژه[4]

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

بدست آمده توسط دستور DIMTESTS حاکی از غلبه یکی از ضرایب بر دیگر مقادیر ویژه است. به طوری که 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 بصورت گرافیکی نشان می دهد که در توافق با نتایج پکیج DIMTESTS است.

DIMTESTS results:

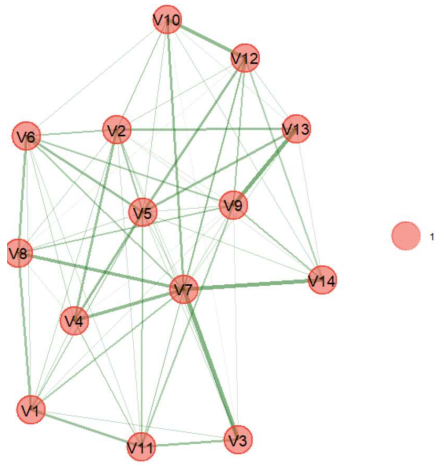
# of Factors:

EMPKC 1

HULL 1

RAWPAR 1

شکل ۱ خروجی دستور EAG در نرم‌افزار R برای یافتن بعدیت



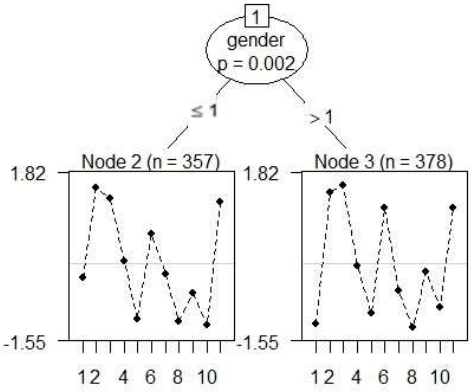
سوال دوم پژوهش:

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

برای شناسایی کارکرد افتراقی سوال‌های 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 سوال را–به ا ستثنای اولین سوال که همیشه محدود به صفر ا ست– ا ستخراج می‌کند. Itempar تمامی پارامترهای سوال را ا ستخراج کرده و به صورت پیش‌فرض مجموع آنان را به صفر محدود می‌کند.

در پژوهش حاضر از روش درخت راش در سوال‌های مطالعه پرلز ۲۰۱۶ کارکرد افتراقی تنها در بلوک ۱۶ تشخیص داده شد. در سوال‌های دو ارز شی بلوک ۱۶، سوال یک با کارکرد افتراقی یکنواخت بر ا ساس جنسیت شنا سایی شدند. بلوک ۱۶ دارای متن با هدف کسب اطلاعات و کاربرد آن است.

در ادامه نتایج شناسایی کارکرد افتراقی سوال در جدول ۳ با اعمال الگوریتم درخت راش بر روی بلوک ۱۶ پرسش‌نامه پرلز ۲۰۱۶ ارائه شده است. همانطور که در شکل ۲ ملاحظه می‌کنید این بلوک بر اساس جنسیت دو شاخه شده است که بیانگر وجود کارکرد افتراقی در این سوال ها است.

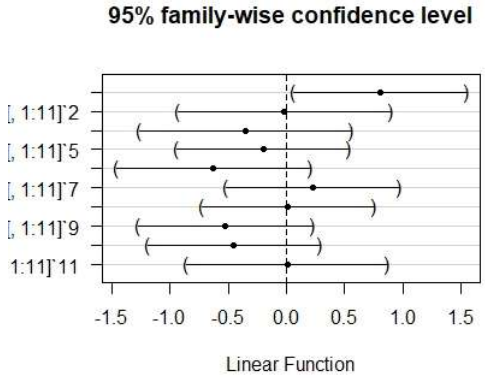


شکل ۲ کارکرد افتراقی بلوک ۱٦ درخت راش

هدف	بلوک	سوال	Itempar		P-value	Z value	Std.Error	Estimate	نوع DIF
			Node2	Node3					
اطلاعاتی	بلوک ۱٦	۱	-0.29	-1.20	۰/۰۲	۲/۹۷	۰/۲۷	۰/۸۰	یکنواخت

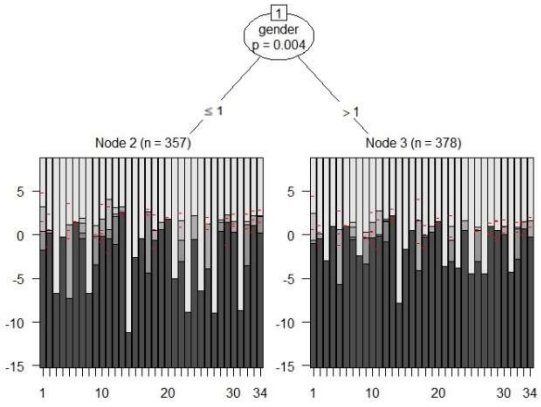
جدول ۳ پارامترهای سوال شناسایی شده با کارکرد افتراقی بر اساس جنسیت در مدل افراز بازگشتی درخت راش

مدل درخت راش در زمینه شناسایی کارکرد افتراقی سوال، صرفا به صورت کلی، با توجه به تمام متغیرها، اعلام می‌نماید آیا کارکرد افتراقی وجود دارد یا خیر. اما از تعیین شماره سوال‌ها ناتوان است. برای تشخیص اینکه در مدل راش دقیقا چه سوالهایی باعث ایجاد کارکرد افتراقی شده‌اند نیاز به استفاده از روشهای دیگر همچون پکیج psychotools، دستور anchortests هستیم. در این روش، تعدادی از سوال‌ها به عنوان لنگر (anchor) که عاری از کارکرد افتراقی هستند تشخیص داده شده و باقی سوالها با آنها مقایسه خواهند شد. این روش تنها برای سوال‌های دو ارزشی کاربرد دارد. نتیجه اعمال این روش بر بلوک ۱۶ آزمون پرلز در شکل ۳ آمده است. سوال ۴ به عنوان لنگر در نظر گرفته شده است. بر اساس شکل ۳، سوال یک از فاصله اطمینان بیرون است که بیانگر وجود کارکرد افتراقی یکنواخت در سوال یک می باشد.



شکل ۳ نمودار سطح اطمینان ۹۵٪ بلوک ۱٦

تحلیل سوالات چند ارزشی با استفاده از دستور pctree انجام شد:



شکل ٤ کارکرد افتراقی بلوک ١٦ درخت راش چند وجهی(Pctree)

نتیجه اعمال این روش بر سوال‌های چند ارزشی بلوک ۱۶ آزمون پرلز در شکل۴، بیانگرکارکرد افتراقی بر اساس جنسیت می باشد. در این مرحله الگوریتم راشpctree ، قادر به تشخیص دقیق سوال یا سوال‌های با کارکرد افتراقی نیست. در همین راستا پارامترهای سوال‌های چند ارزشی بلوک ۱۶برای هر شاخه بررسی شد. لگاریتم درستنمایی بیشینه[1] در شاخه ۲(دختران)، برابر 10020-، $(df = 92)$ و در شاخه ۳ (پسران) ، برابر 10020-، $(df = 92)$ می باشد. در ادامه پارامترهای سوال‌های چند ارزشی بلوک ۱۶ شاخه ۲ دختران را در جدول ۴قرار داده شده است. (Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1)

Partial credit model
Item category parameters: $۲

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	***
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	***

[1]. Log-likelihood

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	***
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	***
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	***

سوال چهارم پژوهش:

۴. وضعیت یا عملکرد نسبی روش درخت راش در شناسایی کارکرد افتراقی سوال‌های آزمون پرلز ۲۰۱۶ بین دانش¬آموزان دختر و پسر ایرانی چگونه است؟

مدل درخت راش قادر به شناسایی کارکرد افتراقی سوال با چندین متغیر به طور همزمان می‌باشد و حسن دیگر آن نتیجه ترسیمی‌اش است که تفسیر کار را بسیار ساده و کاربردی نشان می‌دهد. هر چند که قادر به شناسایی دقیق سوال دارای کارکرد افتراقی نمی‌باشد و نیاز به استفاده از روش دیگری برای تشخیص سوال می‌باشد.

بر اساس مدل درخت راش در بلوک 16 کارکرد افتراقی شناسایی شد. بر اساس تست لنگر کارکرد افتراقی یکنواخت در سوال 1 از سوال‌های دو ارزشی بر اساس جنسیت شناسایی شد.

بحث و نتیجه‌گیری

با توجه به اهمیت سواد خواندن به واسطه 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 به‌کار برده شده است. لذا کارکرد افتراقی سوال‌های ازمون پرلز ۲۰۱۶ در بین دانش‌آموزان دختر و پسر ایرانی پایه چهارم مورد بررسی قرار گرفت. در ادامه یافته‌های این پژوهش مورد بررسی قرار می‌گیرد.

۱. آیا آزمون پرلز ۲۰۱۶ در دانش¬آموزان کشور ایران تک بعدی است؟

پیش از آنکه بتوانیم مساله وجود کارکرد افتراقی را در آزمون بررسی کنیم، لازم است از تک بعدی بودن آزمون پرلز در دانش‌آموزان ایرانی مطمئن شویم. برای پاسخگویی به این سوال، بعدیت سوال‌های پرلز ۲۰۱۶ 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 تک بعدی است. نتایج تحلیل بعدیت سوال‌های پرلز ۲۰۱۶، بیانگر تک بعدی بودن این سوال‌ها هستند. این مسئله از آن جهت مهم است که مدل درخت راش تنها در صورت تک بعدی بودن سوال‌ها قادر به شناسایی کارکرد افتراقی است.

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

برای شناسایی کارکرد افتراقی سوال‌های آزمون پرلز ۲۰۱۶ از روش افراز بازگشتی درخت محور از طریق پکیج Psychotree در محیط R استفاده کردیم. با توجه به متغیرهای سن، جنسیت و مدرسه دانش آموزان، روش درخت راش قادر به تشخیص کارکرد افتراقی بر اساس جنسیت در بلوک 16 شد. در این بلوک برای سوال‌های دو ارزشی به وسیله دستور raschtree کارکرد افتراقی یکنواخت بر اساس جنسیت شناسایی شد. همچنین برای سوال‌های چند ارزشی به وسیله دستور pctree کارکرد افتراقی بر اساس جنسیت مشاهده شد.

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

در بررسی عملکرد روش درخت راش در شناسایی کارکرد افتراقی سوال‌های پرلز ۲۰۱۶ در بین دانش‌آموزان دختر و پسر ایرانی در می‌یابیم مدل راش قادر به تشخیص کارکرد افتراقی سوال با چندین متغیر است. به علاوه 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. مضافا سریعترین بررسی DIF و پارامترهای دشواری سوال تو سط مدل راش به سبب خاصیت ترسیمی‌اش صورت می‌گیرد. با این حال عدم توانایی تشخیص دقیق سوال دارای DIF در مدل راش و نیاز به روشی مجزا همچون Anchortest که تنها برای سوال‌های دو ارزشی کاربرد دارد از نکات منفی این روش است.

در تحلیل ذکر شده در بالا دیده شد که در بلوک ۱۶،کارکرد افتراقی در آزمون پرلز ۲۰۱۶ بین دانش‌آموزان ایرانی وجود دارد. با اینکه در سوال‌های دو ارزشی تنها یک سوال دارای کارکرد افتراقی یکنواخت شناسایی شد. اما در سوال‌های دو ارزشی نمی‌توانیم مشخص کنیم کدام سوال‌ها دارای کارکرد افتراقی هست. لذا برای قضاوت منصفانه بین گروه‌های مختلف آموزشی دانش‌آموزان ایرانی در پایه چهارم سوال‌های بلوک ۱۶ نیازمند بازبینی و یا حذف از آزمون دارند. همچنین در سوال‌هایی که اغلب از دشوارترین سوال‌های 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. مثلا ترجمه متون مناسب نبوده یا دانش‌آموزان هنوز به توانایی لازم برای این سطح از خواندن دست نیافته‌اند. این مساله شایان توجه است که ساختار آموزشی ما نیازمند فراهم کردن بستر مناسبی برای اصلاح و ارتقا سواد خواندن دانش‌آموزان ایرانی می‌باشد .

مطالعات محدودی در زمینه پرلز در ایران انجام شده است. از جمله محدودیت‌های این پژوهش نبود امکان الگوبرداری از کارهای داخلی بود، به خصوص در مورد پرلز ۲۰۱۶ ،که به بحث و ویژگی های روان سنجی آن پرداخته نشده بود. در نتیجه مجبور به یادگیری همه ملزومات تحقیق از صفر شده که سبب طولانی تر شدن مدت تحقیق شد. پیشنهاد می‌شود در مطالعه‌های بعدی، برای شناخت بهتر عوامل موثر در سواد خواندن از چهار پرسشنامه پیشینه کاوی پرلز ۲۰۱۶ شامل پرسشنامه دانش آموز، معلم، مدیرمدرسه و والدین استفاده شود.

---

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

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.*

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.*

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.*

Anderson, R. C., & Pearson, P. D. (1984). A Schema-Theoretic View of Basic Processes in Reading Comprehension. *Technical Report No. 306.*

Anderson, R. C., & Pearson, P. D. (1988). A schema-theoretic view of basic processes in reading comprehension. *Interactive approaches to second language reading, 37-55.*

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

Barro, R. J., & Lee, J. W. (2001). International data on educational attainment: updates and implications. *Oxford Economic papers, 53(3), 541-563.*

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning*. Psychological methods, 22(3), 507.*

Bollmann, S., Berger, M., & Tutz, G. (2017). Item-Focused Trees for the Detection of Differential Item Functioning in Partial Credit Models. *Educational and Psychological Measurement*, 1-24.

Coulombe, S., Tremblay, J. F., & Marchand, S. (2004). Literacy scores, human capital and growth across fourteen OECD countries (Vol. 89). *Geological Survey of Canada.*

Desjardins, C. D., & Bulut, O. (2018). Handbook of educational measurement and psychometrics using R. *CRC Press.*

Finch, W. H., & French, B. F. (2007). Detection of crossing differential item functioning: A comparison of four methods. *Educational and psychological Measurement, 67(4), 565-582.*

Finch, W. H., Hernández Finch, M. E., & French, B. F. (2016). Recursive partitioning to identify potential causes of differential item functioning in cross-national data. *International Journal of Testing, 16(1), 21-53.*

Fischer, G. H., & Molenaar, I. W. & (Eds.). (1995). Rasch Models: Foundations ،Recent Developments and Applications. *New York: Springer-Verlag.*

Frick, H., Strobl, C., & Zeileis, A. (2015). Rasch mixture models for DIF detection: A comparison of old and new score specifications*. Educational and Psychological Measurement, 75(2), 208-234.*

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.*

Gustafsson, J. E. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology, 33(2), 205-233.*

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory (Vol. 2). *Sage.*

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.*

Kamata, A., & Vaughn, B. K. (2004). An Introduction to Differential Item Functioning Analysis. *Learning Disabilities: A Contemporary Journal, 2(2), 49-69.*

Karami, H., & Salmani Nodoushan, M. A. (2011). Differential Item Functioning (DIF): Current Problems and Future Directions. *International Journal of Language Studies (IJLS), 5(3), 133-142.*

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.*

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*

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*

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. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.*

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. (۲۰۲۰). Decomposing Gender Differences in Reading: Evidence from Nordic PIRLS 2016 *(Master's thesis).*

Schulz-Heidorf, K., & Støle, H. (2018). Gender differences in Norwegian PIRLS 2016 and ePIRLS 2016 results at test mode, text and item format level. *Nordic Journal of Literacy Research, 4(1).*

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.*

Stubbe, T. C. (2011). How do different versions of a test instrument function in a single language? A DIF analysis of the PIRLS 2006 German assessments. *Educational Research and Evaluation, 17(6), 465-481.*

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models.

Westers, P., & Kelderman, H. (1992). Examining differential item functioning due to item difficulty and alternative attractiveness. *Psychometrika, 57(1), 107-118.*

Zeileis, A., Strobl, C., Wickelmaier, F., & Kopf, J. (2010). psychotree: Recursive partitioning based on psychometric models. *R package version 0.11-1, URL http://CRAN. R-project. org/package= psychotree. University of Innsbruck-Working Papers in Economics and Statistics Recent Papers.*

Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF). *Ottawa: National Defense Headquarters, 160.*