



## Differential Item Functioning of Minnesota Multiphasic Personality Inventory-2 (Depression) Items in among Recruitment Candidates: A Comparison Between the Rasch Model and Mantel-Haenszel

Mahdi Rahmani Malek Abad<sup>1</sup> , Mohammad Reza Falsafinejad<sup>2\*</sup> , Asghar Minaei<sup>3</sup> , Noor Ali Farrokhi<sup>4</sup> 

1. Department of Educational Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabataba'i University, Tehran, Iran. E-mail: [m.rahmani\\_psy@yahoo.com](mailto:m.rahmani_psy@yahoo.com)
2. Corresponding Author, Department of Educational Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabataba'i University, Tehran, Iran. E-mail: [falsafinejad@atu.ac.ir](mailto:falsafinejad@atu.ac.ir)
3. Department of Educational Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabataba'i University, Tehran, Iran. E-mail: [as\\_minaei@atu.ac.ir](mailto:as_minaei@atu.ac.ir)
4. Department of Educational Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabataba'i University, Tehran, Iran. E-mail: [farrokhi@atu.ac.ir](mailto:farrokhi@atu.ac.ir)

### ARTICLE INFO

**Article type:**  
Research Article

**Article History:**

Received: 08 Aug 2023  
Revised: 27 Oct 2023  
Accepted: 19 Nov 2023  
Online Published: 01 Aug 2024

**Keywords:**

Depression,  
Differential Item Functioning,  
Mantel-Haenszel Method,  
Rasch Model,  
Recruitment candidates.

### ABSTRACT

The purpose of the present study is to identify the questions of the depression component of the Minnesota Multifaceted Personality Questionnaire-2, which are biased toward a certain group among job candidates based on gender and age. This research is survey research, and the statistical population was all the Recruitment candidates of Isfahan city in 2022 and spring 2023. The sample size was 5997 people (57.31% men and 42.69% women) who were selected by one-step cluster method. The tool used was MMPI-2 questionnaire with 370 questions. The fit of the Items with the model was confirmed using outfit MnSq and infit MnSq, and DIF according to gender and age was checked and compared with Mantel-Haenszel and Rasch model by Winsteps version 3.64. After confirming the fit of the data with the model, Differential Item Functioning was performed. The results showed that more than half of the items (64.9%) were biased against women and young people. Mantel-Haenszel also provided a similar finding to Rasch's model, except for the difference in two items. Using depression component questions at once can cause non-bias. However, selective questions or the use of Harris-Lingoes subscales will cause severe DIF to be heavily biased towards women and young people. Also, the Rasch model has a relative advantage over Mantel-Haenszel in detecting DIF.

**Cite this article:** Rahmani Malek Abad, M., Falsafinejad, M. R., Minaei, A., & Farrokhi, N. A. (202..). Differential Item Functioning of Minnesota Multiphasic Personality Inventory-2 (Depression) Items in among Recruitment Candidates: A Comparison Between the Rasch Model and Mantel-Haenszel. *Journal of Applied Psychological Research*, (In press). doi: 10.22059/japr.2024.363263.644703



**Publisher:** University of Tehran Press

© The Author(s).

DOI: <https://doi.org/10.22059/japr.2024.363263.644703>

## بررسی کنش‌افتراقی عملکرد داوطلبین استخدامی در پرسشنامه شخصیت چندوجهی مینه سوتا-۲ (مولفه افسردگی): مقایسه مدل راش و منتل‌هنزل<sup>۱</sup>

مهدی رحمانی ملک‌آباد<sup>۱</sup>، محمدرضا فلسفی‌نژاد<sup>۲\*</sup>، اصغر مینائی<sup>۳</sup>، نورعلی فرخی<sup>۴</sup>

۱. دانشجوی دکتری سنجش و اندازه‌گیری، گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: [m.rahmani\\_psy@yahoo.com](mailto:m.rahmani_psy@yahoo.com)
۲. نویسنده مسئول، دانشیار، گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: [falsafinejad@atu.ac.ir](mailto:falsafinejad@atu.ac.ir)
۳. دانشیار، گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: [as\\_minaei@atu.ac.ir](mailto:as_minaei@atu.ac.ir)
۴. دانشیار، گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: [farrokhi@atu.ac.ir](mailto:farrokhi@atu.ac.ir)

### چکیده

### اطلاعات مقاله

هدف پژوهش حاضر، شناسایی سوالات مولفه افسردگی پرسشنامه شخصیت چندوجهی مینه سوتا-۲ است که بر حسب جنسیت و سن در بین داوطلبان استخدامی به سمت گروه خاصی تورش دارند. روش این پژوهش از نوع پیمایشی و جامعه آماری کلیه داوطلبان استخدامی شهر اصفهان در سال ۱۴۰۱ و بهار ۱۴۰۲ بود. حجم نمونه ۵۹۹۷ نفر (۵۷/۳۱ درصد مرد و ۴۲/۶۹ درصد زن) بود که به روش خوشه‌ای یک‌مرحله‌ای انتخاب شد. ابزار مورد استفاده پرسشنامه MMPI-2 با ۳۷۰ سوال بود. برازش سوالات با مدل با استفاده از شاخص‌های outfit MnSq و infit MnSq تأیید شد و کنش‌افتراقی سوالات بر حسب جنسیت و سن با دو روش منتل‌هنزل و مدل‌راش توسط نرم افزار Winsteps نسخه ۳.۶۴ بررسی و مقایسه شد. پس از تأیید برازش داده‌ها با مدل، تحلیل کنش‌افتراقی انجام شد. نتایج نشان داد که بیش از نیمی سوالات (۶۴/۹ درصد) در عملکرد خود دارای تورش به ضرر زنان و افراد جوان بوده‌است. روش منتل‌هنزل نیز به جز اختلاف در دو سوال، یافته مشابهی را نسبت به مدل‌راش ارائه داد. بکارگیری سوالات انتخابی و یا استفاده از خرده‌مقیاس‌های هریس-لینگوز موجب تورش شدید سوالات به جهت زنان و افراد جوان گردید اما استفاده از سوالات مولفه افسردگی به صورت یکجا می‌تواند موجب عدم تورش گردد. همچنین مدل‌راش برتری نسبی به روش منتل‌هنزل در تشخیص DIF داشته است.

### نوع مقاله:

پژوهشی

### تاریخ‌های مقاله:

تاریخ دریافت: ۱۴۰۲/۰۵/۱۷

تاریخ بازنگری: ۱۴۰۲/۰۸/۰۵

تاریخ پذیرش: ۱۴۰۲/۰۸/۲۸

تاریخ انتشار آنلاین: ۱۴۰۲/۰۵/۱۱

### کلیدواژه‌ها:

افسردگی، داوطلبان استخدامی، روش منتل‌هنزل، کنش‌افتراقی، مدل‌راش.

**استناد:** رحمانی ملک‌آباد، م.، فلسفی‌نژاد، م.، مینائی، ا.، و فرخی، ن. ع. (۱۴۰۰). بررسی کنش‌افتراقی عملکرد داوطلبین استخدامی در پرسشنامه شخصیت چندوجهی مینه سوتا-۲ (مولفه افسردگی): مقایسه مدل راش و منتل‌هنزل. فصلنامه پژوهش‌های کاربردی روانشناختی، (پذیرش شده/ درنوبت انتشار). Doi: 10.22059/japr.2024.363263.644703

ناشر: انتشارات دانشگاه تهران

DOI: <https://doi.org/10.22059/japr.2024.363263.644703>

© نویسندگان



## ۱. مقدمه

گزینش و استخدام نیروی کارآمد، مؤثر و سالم جزء یکی از اصلی ترین نیازهای یک سازمان و مؤثر بر عملکرد و بازده کاری<sup>۱</sup> یک فرد و سازمان است. بدین جهت همواره ابتدا سلامت و توانمندی های داوطلبان مشاغل مورد نیاز مورد سنجش قرار می گیرد و پس از آن افراد به خدمت گرفته می شوند. بدین جهت به منظور سنجش سلامت جسمانی و روانی داوطلبان و به کارگیری آنان در جایگاه متناسب با توانمندی هایشان، معایناتی با عنوان معاینات بدو استخدام صورت می پذیرد (سلطانی شال و همکاران، ۱۳۹۹). براساس دستورالعمل سنجش و ارزیابی سلامت روان کارکنان معاونت حقوقی ریاست جمهوری در سال ۱۳۹۸ نتایج معاینات بدو استخدام به عنوان سندی قانونی جهت ثبت وضعیت سلامت فرد در زمان استخدام جهت تشخیص بیماری های مؤثر در محیط کار و رد یا تایید افراد مورد استفاده قرار گرفته است. در این سند، سلامت افراد به معنای نداشتن اختلال روانی و جسمی در نظر گرفته شده است. یکی از مهمترین شاخص های سلامت روان داوطلبان بدو استخدام، تشخیص نداشتن افسردگی<sup>۲</sup> در بین داوطلبان استخدامی است که توسط پرسشنامه چندوجهی مینه سوتا<sup>۳</sup> و پیرایش دوم (MMPI-2) ارزیابی می شود. پرسشنامه شخصیت چندوجهی مینه سوتا-۲ از مهمترین و معتبرترین پرسشنامه ها در زمینه بررسی شخصیت و برخی اختلالات روانی است (چری<sup>۴</sup>، ۲۰۲۲). از طرفی افسردگی یکی از شایع ترین اختلال های روانی و در حال حاضر چهارمین بیماری شایع در جهان محسوب می شود و یکی از مسائل اساسی بهداشت همگانی است که همانند حالات مزمن دیگر، حالات عود و بازگشت دارد (اون ولی و همکاران<sup>۵</sup>، ۲۰۱۵) و باعث کاهش عملکرد اجتماعی و شغلی می شود بنحوی که زمینه ساز اختلالات دیگر شغلی همچون استرس و اضطراب شغلی نیز می گردد و از آنجائی که افسردگی می تواند پایه و یکی از ویژگی های هر اختلال خلقی و روانی باشد (کاپلان و سادوک<sup>۶</sup>، ۲۰۲۲)، و بررسی مولفه های دیگر پرسشنامه در این مقال نمی گنجد، این پژوهش تنها به بررسی مولفه افسردگی پرداخته است.

در کنار داشتن ابزاری استاندارد، یکی دیگر از مهمترین مسائل در مرحله سنجش داوطلبان استخدامی، ارزیابی صحیح و دقیق، داشتن انصاف در گزینش و عدم سوءگیری سوالات در اعلام نتایج است. گاهی برخی سوالات در تعیین سطح و مشخص کردن صفتی در افراد یک گروه نسبت به گروهی دیگر با همان سطح از صفت دچار سوءگیری می گردد و نتایج را به صورت جهت دار گزارش می نماید که این امر باعث تبعیض در استخدام و حتی گزینش اشتباه افراد می شود. این امر موجب می گردد تا افرادی که سالم هستند در مرحله ارزیابی رد شوند و افرادی که دارای افسردگی و یا بیمار هستند مورد تأیید قرار گرفته و استخدام شوند که در نتیجه این امر سازمان ها دچار مشکل و آسیب می شوند و حتی گاهی نسبت به آزمون های روانشناختی بدبین می گردند. بدین جهت برای مشخص نمودن سوءگیری سوالات پرسشنامه چندوجهی مینه سوتا باید کنش افتراقی سوالات را بررسی نمود تا مشخص شود چه سوالاتی در تشخیص صفت دارای سوءگیری است و به چه سمت جهت دارد و کدام سوال بدون تورش به ارزیابی صفات روانشناختی داوطلبان استخدامی می پردازد.

یکی از مسائلی که موجب ایجاد تبعیض در انتخاب می شود، تبعیض های جنسیتی است، بنحوی که بکر و همکاران<sup>۷</sup> (۲۰۲۳) تفاوت های جنسیتی در نحوه پاسخگویی افراد به سوالات آزمون های روانشناختی را یکی از مسائل مهم تاثیرگذار در ارزیابی ها عنوان نموده اند. این تاثیرگذاری تا حدی است که پزشکان و محققان پس از انتشار نسخه اول پرسشنامه چندوجهی مینه سوتا شروع به زیر سوال بردن صحت این پرسشنامه کردند. منتقدان اشاره کردند که نتایج نشان دهنده سوگیری احتمالی پرسشنامه است، در

1. work efficiency
2. Depression
3. Minnesota Multiphasic Personality Inventory
4. Cherry
5. Avenevoli, Swendsen, Burstein, Merikangas
6. Kaplan & Sadock
7. Baker et al.

حالی که دیگران احساس کردند که خود پرسشنامه حاوی سوالات جنسیتی و نژادپرستانه است که می‌تواند منجر به عدم انصاف در نمره‌گذاری و اعلام ناصحیح نتایج شود (تالریکو و همکاران<sup>۱</sup>، ۲۰۲۳). تبعیض جنسیتی یکی از گسترده‌ترین شکل تبعیض‌ها است، زیرا زنان در هر سنی، اگر رقیبشان مرد باشد، شانس بسیار کمتری برای دریافت شغل دارند، حتی اگر مردان دارای صلاحیت کمتر یا بی‌تجربه باشند. حال اگر در کنار تبعیض‌های موجود در جامعه، سوالات نیز دارای تبعیض و سوءگیری باشند، اطمینان کردن به نتایج بسیار سخت و دشوار می‌شود و هم افراد و سازمان‌ها متضرر می‌گردند. بنابراین تا می‌شود باید سوالات دارای تورش را شناسایی کرد و راهکارهایی را در جهت رفع سوءگیری سوالات بکار برد.

در کنار جنسیت، تعصبات تبعیض آمیزی نیز در مورد سن نیز دیده شده، که در هر دو مورد به افراد در جهت حذف از انتخاب منصفانه فشار آورده می‌شود (بوربای و خامزینا<sup>۲</sup>، ۲۰۱۹). سن افراد یکی از عواملی است که می‌تواند موجب تغییر در نگرش و یا نوع پاسخدهی افراد شود، بنوعی که افراد در سنین بالا ممکن است نگرش و جهان بینی را به دست آورده باشند که به موجب آن دارای رفتارها و پاسخ‌هایی باشد که نزدیک به علائم افسردگی باشد اما در واقع افسرده نیستند و یا افراد در سنین کم نشانه‌هایی از شیدائی را نشان می‌دهند و سوالاتی در این زمینه را بیشتر پاسخ مثبت می‌دهند ولی در واقع شیدا نیستند و بنوعی سن افراد در نحوه پاسخدهی آنان می‌تواند موثر باشد و این فراوانی در افراد کمتر از ۲۵ سال و بالای ۲۵ سال نشان داده شده است (کاپلان و سادوک<sup>۳</sup>، ۲۰۲۲). از آنجائی که ویژگی جنسیت و سن در مولفه‌های روان‌شناختی و به تبع آن عملکرد شغلی موثر است و تبعیض جنسیتی و سنی می‌تواند به صورت‌های مختلف در آموزش، اشتغال، شرایط اجتماعی و توانمندسازی افراد بروز یابد (زاهد و همکاران<sup>۴</sup>، ۱۳۹۹) و از طرفی این تفاوت می‌تواند به عنوان یک عامل موثر بر نحوه پاسخدهی باشد و تا کنون تاثیر کنش افتراقی سوالات در انصاف و دقت در گزینش برای داوطلبان استخدامی مورد بررسی قرار نگرفته است، نیاز به بررسی کنش افتراقی سوالات بر اساس جنسیت و سن است، چرا که یکی از موانع و تهدیدات اصلی علیه اعتبار<sup>۳</sup> آزمون وجود واریانس برگرفته از عوامل بی‌ربط و سازه‌های نامرتبط<sup>۴</sup> است، بدین معنا که عملکرد فرد در آزمون، نباید تحت تأثیر عواملی قرار گیرد که متفاوت از سازه مورد نظر آزمون است. در غیر این صورت، آزمون، داری کنش‌افتراقی<sup>۵</sup> (DIF) خواهد بود (کرمی<sup>۶</sup>، ۲۰۱۳).

یکی از راه‌ها برای پی بردن به اینکه آزمونی غیر منصفانه است کشف پرسش‌های دارای کنش‌افتراقی است. دورانز و هالند<sup>۷</sup> (۱۹۹۳) به نقل از وایبرگ<sup>۸</sup> (۲۰۰۷) کنش‌افتراقی را چنین تعریف می‌نمایند: "منظور از کنش‌افتراقی (DIF)، تفاوت‌های موجود در عملکرد هر پرسش از گروه‌ها با توجه به توانایی یا ویژگی‌ای که پرسش‌ها از افراد گروه بطور مفهومی می‌خواهد. کنش‌افتراقی یک تفاوت ناخواسته بین گروه‌های امتحان دهنده است که قرار است براساس تفاوت در توانایی‌های شان در هر پرسش و هر امتحان قیاس شوند". کمیلی و گنکدون<sup>۹</sup> (۱۹۹۹) کنش‌افتراقی را در قالب مفهوم بی طرفی آزمون<sup>۱۰</sup> توضیح داده است. روش‌های آماری آشکارسازی کنش‌افتراقی (DIF) را می‌توان دو دسته کلی تقسیم نمود. دسته اول براساس نظریه کلاسیک آزمون<sup>۱۱</sup> (CTT) است. نمرات مشاهده شده در این روش به عنوان متغیر هم‌تا برای پیش‌بینی نمرات واقعی در نظر رفته خواهد شد (لرد و نایوک<sup>۱۲</sup>، ۱۹۶۸).

1. Talerico et al.
2. Buribayev and Khamzina
3. Reliability
4. factors irrelevant-construct
5. Differential Item Functioning
6. Karami
7. Dorans & Holland
8. Wiberg
9. Camilli & Congdon
10. Test fairness
11. Classic Test Theory
12. Lord & Novick

از جمله روش‌هایی که در این دسته قرار می‌گیرند عبارت از روش منتل هنزل<sup>۱</sup> (MH)، روش منتل هنزل تعمیم داده‌شده<sup>۲</sup> (GMH) (سو و وانگ<sup>۳</sup>، ۲۰۰۵)، روش اختلاف میانگین استاندارد شده<sup>۴</sup> و روش رگرسیون لوجستیک است (زومبو<sup>۵</sup>، ۲۰۰۳).

دسته دوم شامل روش‌های مبتنی بر نظریه سوال- پاسخ<sup>۶</sup> است، زیرا نمره متغیر مکنون به عنوان متغیر هم‌تا برای برآورد نمرات متغیر مکنون مورد استفاده قرار می‌گیرد. روش‌های این دسته شامل روش‌های مجذور کای-دو لرد<sup>۷</sup>، راجو و روش‌های آزمون نسبت درست‌نمایی<sup>۸</sup> است (امبرتسون و ریس<sup>۹</sup>، ۲۰۰۰ و کرمی، گرامی پور و مینائی، ۱۴۰۰). روش‌های بررسی کنش افتراقی مبتنی بر نمرات مشاهده‌شده و متغیرهای مکنون می‌توانند در چهارچوب مدل‌های مختلط خطی و غیرخطی تعمیم داده‌شده ترکیب شود و مدل‌های سلسله‌مراتبی<sup>۱۰</sup> نظریه سوال- پاسخ و رگرسیون لوجستیک<sup>۱۱</sup> را به وجود بیاورد. در این مدل‌ها اندازه اثر کنش افتراقی (DIF) و جایگزینی داده‌های گمشده امکان‌پذیر است (اسکروندال و رابه-هیسکت و بوکا راتن<sup>۱۲</sup>، ۲۰۰۴). از آنجائی که روش منتل هنزل از نظر کاربرد و اعتبار بالاتر و پرکاربردتر از روش‌های دیگر کلاسیک بود و همچنین مدل راش با توجه به ساختار دو ارزشی مدل‌های پارامتری دارای برازش مناسب سوالات و اعتبار بالای شاخص‌ها می‌باشد، از این جهت در این پژوهش به مقایسه دو روش مدل راش و منتل هنزل پرداخته شده است. گوری و مارین<sup>۱۳</sup> (۲۰۱۵) در پژوهشی به بررسی کنش افتراقی سوالات اسکیزوفرنی، اضطراب و پرخاشگری پرسشنامه MMPI-2 در بین دانشجویان تازه وارد دانشگاه با مدل راش پرداخت و نشان داد برخی تفاوت‌ها در نحوه برازش و تشخیص سوالات وجود داشت. مک نالتی و همکاران<sup>۱۴</sup> (۲۰۱۵) در مطالعه‌ای به منظور بررسی شاخص‌های روانی مقیاس MMPI-2 درز یک نمونه مشخص شده از بین اتباع خارجی زندانی شده در ایالت‌های امریکا با استفاده از نظریه کلاسیک انجام شد. نتایج نشان داد که زنان بیشتر از مردان و قفقازی‌ها بیشتر از آمریکایی‌های آفریقایی تبار محتواهای جعلی و غیر محتوا تولید کرده اند. همچنین در پژوهشی دیگر چیلدز، داهلسترون، کیمپ و پانتتر<sup>۱۵</sup> (۲۰۰۰) به بررسی ارزیابی شخصیت با استفاده از نظریه سوال پاسخ جهت بررسی مقیاس افسردگی در آزمون MMPI-2 پرداختند. در این تحقیق از مدل نظریه سوال- پاسخ دو پارامتری استفاده شد. نتایج این پژوهش نشان داد که مقیاس افسردگی، دارای سوالات بسیار شدید تبعیض‌آمیز است. با توجه به اهمیت رعایت انصاف در گزینش نیروی انسانی سالم و مناسب در آزمون‌های بدو ورود و انتخاب بدون سوءگیری افراد، سوال اول این پژوهش این است که آیا سوالات مولفه افسردگی پرسشنامه MMPI-2 با مدل راش برازش دارد؟ سوال دوم این است که کدام سوالات مولفه افسردگی با مدل راش و روش منتل هنزل دارای کنش افتراقی تشخیص داده می‌شود؟ و سوال سوم اینکه آیا تفاوتی در تشخیص سوالات دارای سوءگیری بین دو روش منتل هنزل و مدل راش وجود دارد یا خیر؟

1. Mantel- Haenszel
2. Generalized Mantel- Haenszel
3. Su & Wang
4. Standardized Mean Difference
5. Zumbo
6. Item Response Theory
7. Lord's chi-square
8. likelihood ratio test
9. Embretson & Reis
10. Hierarchical models
11. logistic regression
12. Skronidal & Rabe-Hesketh & Boca Raton
13. Gori and Marin
14. McNulty et al.
15. Childs, Dahlstrom, Kemp & Panter

## ۲. روش

## ۲-۱. جامعه آماری، نمونه و روش نمونه‌گیری

پژوهش از نوع کاربردی و همچنین از طرفی توسعه ابزار و کمی است، همچنین در قسمت تحقیقات توصیفی و از نوع پیمایشی قرار می‌گیرد، چرا که در تحقیقات توصیفی و نوع پیمایشی محقق به دنبال چگونگی بودن موضوع است و می‌خواهد بداند پدیده، متغیر، شیء یا مطلب چگونه است و از طرفی به بررسی توزیع ویژگی‌های یک جامعه آماری می‌پردازد، و نوعی به مطالعه متغیرهای مختلف و بررسی آن در یک گروه پرداخته می‌شود چرا که این نوع تحقیقات وضع موجود را بررسی کرده و به توصیف منظم و نظام دار وضعیت فعلی آن پرداخته و ویژگی‌ها و صفات آن را مطالعه و ارتباط بین متغیرها را بررسی نموده است.

جامعه آماری این پژوهش کلیه داوطلبان استخدامی متقاضی شغل معلمی (یا دانشجو معلمی) شهر اصفهان به تعداد ۱۰ هزار نفر بود که در آزمون‌های بدو ورود سال ۱۴۰۱ و بهار ۱۴۰۲ به مراکز طب کار در شهر اصفهان در مراجعه نمودند. روش نمونه‌گیری این پژوهش به روش خوشه‌ای یک مرحله‌ای بود. این روش زمانی کاربرد دارد که فهرست کامل افراد جامعه در دسترس نباشد و انتخاب افراد به صورت خوشه‌ای فراهم است و انتخاب فردی امکان پذیر نباشد. حجم نمونه با توجه به شیوه تجزیه و تحلیل داده‌ها و استفاده از مدل راش حجم نمونه هرچه بزرگ‌تر باشد و بیشتر از ۴۰۰۰ نفر باشد، نتایج قابل اعتمادتر است (شاین و انیل، ۲۰۱۷) که با توجه به نمونه‌گیری خوشه‌ای تعداد ۶۲۱۳ نفر در ابتدا از بین سه مرکز طب کار انتخاب شدند که پس از حذف تعداد افرادی که ناقص به سوالات پاسخ داده‌اند (طبق دستور العمل آزمون افرادی که به تعداد ۳۵ سوال و بیشتر پاسخ نداده‌اند، پاسخ نامه آن‌ها قابل استفاده نیست) تعداد ۵۹۹۷ نفر در تحلیل مورد استفاده قرار گرفت. معیار ورود به پژوهش عبارت است از اینکه داوطلبین استخدام در آموزش و پرورش باشد، مراحل غربالگری طب کار را تا به مرحله ارزیابی روانشناختی سپری کرده باشد و به پرسشنامه با دقت و صحت پاسخ داده باشد، و معیارهای خروجی از پژوهش عبارت است از اینکه پرسشنامه بیش از ۳۵ سوال بدون پاسخ باشد، دارای پرونده درمانی فعال باشد و اینکه شاخص‌های روانی آزمون نا معتبر باشد.

## ۲-۲. شیوه اجراء

هدف این پژوهش بررسی و مقایسه کنش افتراقی سوالات مولفه افسردگی پرسشنامه چندوجهی شخصیتی مینه‌سوتا با استفاده از روش منتل‌هنزل و مدل راش برای داوطلبان استخدامی بود. بدین منظور ابتدا فهرست مراکز طب کار شهر اصفهان استخراج و از بین ۱۷ مرکز ۹ مرکز اعلام آمادگی برای همکاری کردند که از این بین ۳ مرکز به تصادف و با توجه به حجم مراجعین استخدامی انتخاب و پس از صحبت با مسئول مرکز مقرر شد ضمن حفظ محرمانه بودن اسامی داوطلبان شرکت داده‌شده در طرح پژوهش و محرمانه ماندن نتایج به دست آمده از آزمون‌های هر فرد از مراجعه کنندگان برای معاینات استخدامی، آزمون MMPI-2 در محل مرکز گرفته شود و نتایج آزمون‌ها به صورت هفتگی در اختیار دستیار پژوهشگر قرار گیرد. دستیار پژوهشگر نیز پاسخ‌نامه‌های افراد را وارد نرم افزار کرده و پس از تجمیع در اختیار پژوهشگر قرار داده است.

## ۲-۳. ابزار پژوهش

۲-۳-۱. پرسشنامه چندوجهی شخصیتی مینه‌سوتا-۲ (MMPI-2): این پرسشنامه توسط هاتاوی و مک‌کینلی<sup>۲</sup> (۱۹۴۳)

ساخته شده و در سال ۱۹۷۰ و در نهایت سال ۲۰۰۸ توسط دانشگاه مینه‌سوتا مورد آخرین بازنگری قرار گرفت و باز انتشار شد. ابزار استفاده شده در این پژوهش نسخه MMPI-2 است که بر اساس ویرایش سال ۲۰۰۸ توسط هاتاوی و مک‌کینلی و با سرپرستش دانشگاه مینه‌سوتا دارای ۳۷۰ ماده است (مارنات و رایت<sup>۳</sup>، ۲۰۱۶). این مقیاس دارای ۳ شاخص روائی دروغ<sup>۴</sup> (L) غیر معمولی<sup>۵</sup> (F) و انکار<sup>۶</sup> (K) و ۱۰ شاخص بالینی هیپوکندریا<sup>۷</sup> (HS)، افسردگی<sup>۸</sup> (D)، هسیتری<sup>۹</sup> (Hy)، انحراف روانی-اجتماعی<sup>۱۰</sup> (Pd)، زنانگی/مردانگی<sup>۱۱</sup> (MF)، پارانوئیا<sup>۱۲</sup> (Pa)، ضعف روانی<sup>۱۳</sup> (Pt)، اسکیزوفرنی<sup>۱۴</sup> (Sc)، هیپومانیا<sup>۱۵</sup> (Ma) و درونگرایی اجتماعی<sup>۱۶</sup> (Si) است. در این پژوهش با توجه به حجم سوالات و عدم امکان بررسی تمام مولفه‌ها تنها به مولفه افسردگی از پرسشنامه MMPI-2 با ۵۷ سوال پرداخته شد. پاسخ به این ماده‌ها به صورت دو گزینه‌ای بلی/درست (۱) و خیر/غلط (۰) است. از بین ۵۷ سوال ۲۰ سوال به صورت مستقیم نمره گذاری شد و ۳۷ سوال به صورت معکوس نمره گذاری (بلی=۰ و خیر=۱) شد. روائی این مقیاس در مطالعات فراوانی با استفاده از تحلیل عاملی متمایل و چرخش ایلیمین و همچنین تشخیص افتراقی روانشناسان بالینی بررسی و مورد تأیید قرار گرفته است و نشان داده شد که مقیاس افسردگی با افرادی که دارای نشانگان بالینی افسردگی دارد همخوان است و دارای روائی قابل قبولی است (مارنات و رایت، ۲۰۱۶). در مطالعه دیسکوئیتس و همکاران<sup>۱۷</sup> (۲۰۱۶) اعتبار مقیاس‌های آزمون اصلی با روش دو نیمه کردن<sup>۱۸</sup> از ۰/۵۱ به بالا بوده است که بعضی از این ضرایب برابر با ۰/۹۵ بدست آمد، ضرایب همبستگی مقیاس‌ها از راه باز آزمایی نیز بین ۰/۵۰ تا ۰/۹۰ بود. ولیانپور، مدرس غروی و مهرا<sup>۱۹</sup> (۲۰۲۰) در مطالعه‌ای که بر روی جامعه سالم و جامعه بالینی ایرانی انجام دادند، اعتبار مقیاس افسردگی پرسشنامه MMPI-2 را با روش آلفای کرونباخ<sup>۲۰</sup> برابر با ۰/۸۰ و ضریب اعتبار مقیاس از راه باز آزمایی برابر ۰/۸۲ گزارش کردند. اعتبار این مقیاس در این پژوهش با روش آلفای کرونباخ برابر با ۰/۷۸ و با روش دو نیمه کردن برابر با ۰/۷۳ به دست آمد، که این سطح از اعتبار در مقایسه با منابع اولیه کمی پائین‌تر بوده که علت آن را می‌توان پائین بودن واریانس صفت در بین اعضای جامعه به علت همگن بودن افراد به جهت گزینش شدن در فرایندهای ابتدائی آزمون‌های استخدامی فرهنگیان دانست، ولی با این حال این ضرایب اعتبار در سطح مطلوبی قرار دارد.

1. Minnesota Multiphasic Personality Inventory-2 (MMPI-2)
2. Hathaway & Mckinely
3. Marnat & Wright
4. Lie (L)
5. Infrequency (F)
6. Defensiveness (K)
7. Hypochondriasis (Hs)
8. Depression (D)
9. Hysteria (Hy)
10. Psychopatic Deviate (Pd)
11. Masculinity / Femininity (MF)
12. Paranoia (Pa)
13. Psychasthenia (Pt)
14. Schizophrenia (Sc)
15. Hypomania (Ma)
16. Social Introversion (Si)
17. Deskovitz et al.
18. split-half reliability
19. Valianpour, Modarres Gharavi & Mahram
20. cronbach alpha reliability

## ۲-۳. روش تجزیه و تحلیل

برای محاسبه آماره‌های توصیفی میانگین و انحراف استاندارد و ضریب آلفای کرونباخ و اعتبار مقیاس از طریق نظریه کلاسیک از نرم افزار SPSS نسخه ۲۷ استفاده شد و برای بررسی برازش سوال از آماره  $\text{outfit Mnsq}$  و  $\text{infit Mnsq}$  و با استفاده نرم افزار Winsteps نسخه ۳.۶۴ بهره گرفته شد، همچنین ضریب مطلوبیت سوالات نیز با شاخص Measure بیان شد. در بررسی تناسب شاخص‌ها ابتدا به شاخص  $\text{Mnsq}^1$  مراجعه می شود و در صورت عدم تائید به شاخص  $\text{ZSTD}^2$  باید رجوع کرد. مقادیر  $\text{MNSQ}$  از ۰ تا مثبت بی نهایت قرار دارد و مقدار مطلوب برابر ۱ است. هر چه این مقدار به ۰ نزدیک شود نشان از بیش برازش الگوهای پاسخ با مدل است که به معنای وجود وابستگی در میان پاسخ‌ها یا سوالات است و مقادیر بالاتر از ۱ بیانگر انحراف از تک بعدی بودن است. در کل مقادیر بیشتر از ۲ غیر مفید و حاکی از اعتبار و برازش پائین است. مقادیر بین ۱.۵ تا ۲ با اینکه بالاست اما گمراه کننده نیست و قابل اغماض است. مقادیر بین ۰.۵ تا ۱.۵ برای برازش مقیاس اندازه‌گیری مناسب و میزان اعتبار مقیاس را افزایش می‌دهد و در نهایت مقادیر کمتر از ۰.۵ برای برازش مقیاس اندازه‌گیری بهره پائینی دارد و میزان اعتبار را به صورت غیر واقعی بالا می‌برد. مقادیر  $\text{ZSTD}$  نیز بهتر است بین -۲ تا +۲ گزارش شود، البته با توجه به اینکه این آماره به حجم نمونه بسیار حساس است (لینکر<sup>۳</sup>، ۲۰۱۲)، و حجم نمونه این پژوهش بالاست، در این پژوهش جهت بررسی برازش سوال با مدل از آماره  $\text{Mnsq}$  که به دو شکل  $\text{outfit}$  و  $\text{infit}$  آمده است، استفاده شد.

## ۳. یافته‌ها

### ۳-۱. توصیف جمعیت شناختی

از بین اعضای نمونه پژوهش ۵۷/۳۱ درصد مرد و ۴۲/۶۹ درصد را زنان تشکیل داده اند. همچنین میانگین سنی اعضای نمونه برابر با ۲۸/۹۷ سال و با انحراف استاندارد ۴/۱۵ سال بود. تحصیلات تمام افراد بالای دیپلم و از این بین ۳۷/۱۷ درصد دیپلم، ۴۲/۸۷ درصد کارشناسی، ۱۵/۸۱ درصد دارای تحصیلات کارشناسی ارشد و یا دکتری و در نهایت ۴/۱۵ درصد دارای تحصیلات حوزوی بودند.

### ۳-۲. بررسی برازش مدل با داده‌ها

با توجه به اینکه هدف این پژوهش بررسی کنش افتراقی سوالات پرسشنامه  $\text{MMPI-2}$  در بین داوطلبین استخدامی با مدل راش در کنار روش متل‌هنزل بود، ابتدا به تبیین تناسب مدل و داده‌ها و مفروضات نظریه راش پرداخته شد و سپس شاخص‌های کمی سنجش سوالات به دست آمده ارائه شد. شاخص‌های کمی در جدول ۱ آمده است.

1. mean-square t standardized fit  
2. standardized as a z-score  
3. Lincare



جدول ۱. پیش‌فرض‌های مدل راش و بررسی تناسب مدل با داده‌ها

ماره سوال	Measure	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	شماره سوال	Measure	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
۴۵	۲,۵۵	۰,۹۹	-۰,۲	۰,۹۳	-۰,۹	۲۰	۰,۶۸	۱,۰۰	۰,۲	۰,۹۹	-۰,۲
۷۵	۲,۴۴	۰,۹۴	-۱,۰	۰,۶۸	-۵,۱	۹	۰,۶۱	۰,۹۳	-۳,۸	۰,۹۰	-۴,۴
۳۹	۲,۳۸	۰,۹۵	-۰,۹	۰,۷۹	-۳,۳	۱۴۶	۰,۳۳	۱,۰۰	-۰,۲	۱,۰۱	۰,۵
۱۰	۲,۱۹	۰,۹۶	-۰,۷	۰,۹۱	-۱,۴	۱۸	۰,۳۲	۱,۰۴	۲,۵	۱,۰۵	۲,۷
۲	۲,۱۴	۰,۹۸	-۰,۴	۰,۸۷	-۲,۳	۳۸	۰,۱۳	۰,۹۸	-۱,۶	۱,۰۰	-۰,۱
۱۷۵	۲,۱۲	۰,۹۴	-۱,۱	۰,۷۶	-۴,۴	۲۴۵	۰,۰۲	۱,۰۰	۰,۰	۱,۰۱	۰,۴
۱۵	۲,۰۳	۰,۹۶	-۰,۸	۰,۸۸	-۲,۳	۱۴۳	-۰,۲۰	۱,۰۰	۰,۳	۱,۰۱	۰,۶
۹۲	۲,۰۳	۱,۰۰	۰,۰	۱,۰۴	۰,۸	۵	-۰,۶۲	۱,۰۱	۲,۰	۱,۰۲	۲,۶
۱۰۹	۲,۰۰	۰,۹۳	-۱,۶	۰,۷۶	-۴,۸	۲۶۰	-۰,۷۵	۱,۰۶	۹,۴	۱,۰۹	۹,۹
۳۱	۱,۹۹	۰,۹۵	-۱,۰	۰,۹۰	-۱,۹	۱۴۸	-۰,۹۹	۰,۹۸	-۳,۶	۰,۹۷	-۳,۴
۱۳۰	۱,۹۸	۰,۹۴	-۱,۴	۰,۸۵	-۲,۸	۱۱۸	-۱,۲۱	۱,۰۰	-۰,۱	۱,۰۳	۳,۱
۱۶۵	۱,۸۹	۰,۹۵	-۱,۱	۰,۸۵	-۳,۰	۲۶۷	-۱,۳۵	۱,۰۲	۲,۵	۱,۰۷	۶,۵
۱۴۱	۱,۸۹	۰,۹۳	-۱,۶	۰,۷۷	-۴,۷	۷۶	-۱,۶۷	۱,۰۸	۷,۱	۱,۲۳	۹,۹
۷۳	۱,۷۷	۰,۹۱	-۲,۲	۰,۷۶	-۵,۴	۳۳	-۱,۷۹	۱,۰۳	۲,۰	۱,۱۴	۸,۱
۱۴۷	۱,۷۲	۰,۹۶	-۱,۱	۰,۸۹	-۲,۴	۲۲۶	-۱,۹۱	۱,۰۱	۰,۸	۱,۱۰	۵,۶
۹۵	۱,۶۶	۰,۹۰	-۲,۸	۰,۷۱	-۷,۱	۱۱۷	-۲,۲۸	۱,۰۷	۳,۸	۱,۲۳	۸,۹
۳۳۰	۱,۶۲	۰,۹۹	-۰,۲	۰,۹۸	-۰,۵	۲۹	-۲,۳۸	۱,۰۵	۲,۶	۱,۳۴	۹,۹
۱۸۸	۱,۶۰	۰,۹۶	-۱,۲	۰,۸۵	-۳,۶	۵۵	-۲,۶۲	۱,۰۳	۱,۳	۱,۳۱	۹,۳
۲۱۵	۱,۵۵	۰,۹۱	-۲,۵	۰,۸۰	-۵,۰	۲۱۲	-۲,۷۳	۱,۰۲	۰,۶	۱,۲۱	۵,۹
۴۶	۱,۵۲	۰,۹۸	-۰,۷	۰,۹۴	-۱,۵	۱۸۱	-۲,۸۴	۱,۰۶	۲,۰	۱,۳۴	۸,۷
۱۴۰	۱,۵۱	۰,۹۴	-۱,۷	۰,۸۵	-۳,۸	۱۷۸	-۳,۰۲	۱,۰۵	۱,۵	۱,۵۴	۹,۹
۵۶	۱,۴۶	۰,۹۰	-۳,۱	۰,۷۷	-۶,۳	۲۲۱	-۳,۱۲	۱,۰۳	۰,۸	۱,۴۸	۹,۹
۱۷۰	۱,۴۰	۰,۹۹	-۰,۳	۰,۹۹	-۰,۳	۲۴۸	-۳,۱۷	۱,۰۲	۰,۷	۱,۳۴	۷,۱
۲۲۳	۱,۳۰	۰,۹۸	-۰,۶	۰,۹۴	-۱,۷	۱۸۹	-۳,۱۷	۱,۰۰	-۰,۱	۱,۲۲	۴,۶
۴۳	۱,۱۷	۱,۰۰	۰,۰	۰,۹۵	-۱,۴	۲۳۸	-۳,۵۲	۱,۰۱	۰,۳	۱,۳۷	۶,۳
۲۳۳	۱,۰۱	۰,۹۵	-۲,۱	۰,۹۰	-۳,۳	۳۷	-۳,۶۰	۱,۰۳	۰,۷	۱,۷۳	۹,۹
۴۹	۰,۸۷	۰,۹۶	-۱,۶	۰,۹۳	-۲,۶	۱۳۴	-۳,۶۶	۱,۰۱	۰,۳	۱,۵۶	۸,۶
۱۴۲	۰,۸۰	۱,۰۴	۱,۶	۱,۶	۲,۳	۶۸	-۴,۷۴	۱,۰۰	۰,۰	۱,۲۵	۲,۴
۱۲۷	۰,۶۹	۰,۹۷	-۱,۵	۰,۹۷	-۱,۲	--	--	--	--	--	--

در جدول ۱ ستون اول از سمت راست به شماره سوال در پرسشنامه MMPI-2 تعلق دارد و اینکه ترتیب هر سوال در پرسشنامه برای شرکت کنندگان به چه صورت بوده است. در ستون دوم (Measure) به میزان و درجه سختی هر پرسش در واحد لوجیت<sup>۱</sup> اشاره دارد که در پرسشنامه‌های شخصیتی به عنوان ضریب دشواری یا مطلوبیت معرفی می‌گردد (لینکر، ۲۰۱۲). در این مولفه سخت ترین سوال متعلق به سوال ۴۵، ۷۵ و ۳۹ بود و سوالی که بیشتر افراد به آن پاسخ مثبت داده‌اند به ترتیب سوالات ۶۸، ۱۳۴، ۳۷ و ۲۳۸ بوده است. ستون سوم و چهارم مربوط به شاخص‌های تناسب Infit و ستون پنجم و ششم متعلق به شاخص‌های تناسب Outfit است. با توجه به حجم بالای نمونه در این پژوهش جهت بررسی برازش سوال با مدل از آماره MnSq استفاده

<sup>۱</sup>. Logit

شد. با توجه به نتایج گزارش شده در جدول شماره ۱ می‌توان بیان داشت که از بین ۵۷ سوال مقیاس افسردگی شاخص‌های برآزش ۵۳ سوال در سطح مناسب (بین ۰.۵ تا ۱.۵) قرار داشته است و تنها ۴ سوال (۳۷، ۱۳۴، ۱۴۲ و ۱۷۸) دارای مقدار  $Outfit\ MnSq$  بالاتر از ۱.۵ بوده است و به علت اینکه این مقادیر از ۲ کمتر است، گمراه کننده نبوده و قابل اغماض است. در مجموع می‌توان بیان داشت شاخص‌های برآزش تمامی سوالات در دامنه مطلوبی قرار دارد و سوالات برآزش خوبی با مدل راش دارد.

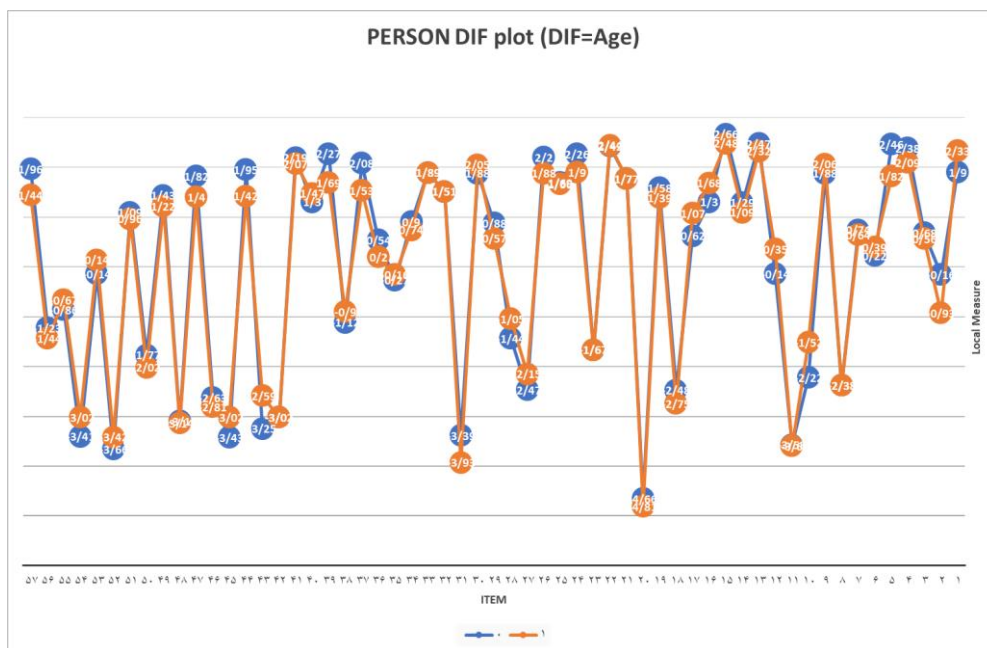
### ۳-۳. بررسی کنش افتراقی سوالات

پس از تأیید برآزش داده‌ها با مدل، در ادامه برای تشخیص و مقایسه سوالات دارای کنش افتراقی بر حسب جنسیت و سن از دو شیوه متل‌هنزل و مدل راش استفاده شد و نتایج آن در جدول شماره ۲ آورده شد. در ستون اول این جدول شماره سوالات مولفه افسردگی پرسشنامه MMPI-2 آورده شد. سپس نتایج به دست آمده از رویکرد متل‌هنزل ارائه شد. در این روش مقادیر کنش افتراقی با  $Size$  و سطح معناداری آن با  $Prob.$  آورده شد. پس از آن نتایج به دست آمده از کنش افتراقی مدل راش گزارش شد که با شاخص‌های  $DIF\ Contrast$  و  $DIF\ Size$  مشخص شد و معناداری کنش افتراقی با  $Prob.$  نشان داده شد. برای بررسی وجود کنش افتراقی باید سطح  $Prob.$  کمتر از ۰.۰۵ گزارش شود. اگر مقادیر  $Size$  کمتر از ۰.۴۳ باشد نشان از عدم کنش افتراقی (A) است، اگر این شاخص بین ۰.۴۳ تا ۰.۶۴ باشد دارای کنش افتراقی متوسط (B) است و اگر بالاتر از ۰.۶۴ باشد دارای کنش افتراقی بزرگ (C) است (لینکر، ۲۰۱۲). کنش افتراقی بزرگ گواهی بر ناعادلانه بودن سوال است، این بدین معناست که اگر پاسخ به سؤال‌های آزمون با کنترل توانایی آزمودنی‌ها تابعی از ویژگی‌های گروهی که آزمودنی‌ها به آن تعلق دارند، مانند جنسیت و سن باشد، آن سؤال‌ها نسبت به آن افراد دارای کنش افتراقی و سوءگیری است.

جدول ۲. مقایسه توانش مدل راش و متل‌هنزل در تشخیص کنش افتراقی پرسش‌ها بر حسب جنسیت و سن

شماره سوال	متل‌هنزل		مدل راش			نتیجه	جهت تورش
	Prob.	Size	Welch Prob.	DIF Size	DIF Contrast		
۲	۰/۰۰۰۰	۰/۷۹	۰/۰۰۰۰	۰/۴۵	۰/۷۶	C	U
۹	۰/۰۰۰۲	۰/۳۸	۰/۰۰۰۱	۰/۳۰	۰/۴۹	B	F
۱۰	۰/۰۰۰۰	۰/۶۴	۰/۰۰۰۰	۰/۴۰	۰/۶۲	C	F / U
۱۵	۰/۰۰۰۰	۰/۷۵	۰/۰۰۰۰	۰/۴۷	۰/۷۷	C	F
۲۹	۰/۰۰۰۰	۰/۵۲	۰/۰۰۰۰	۰/۳۳	۰/۵۳	B	F
۳۱	۰/۰۰۰۰	-۰/۶۶	۰/۰۰۰۰	-۰/۴۳	-۰/۷۰	C	L
۳۷	۰/۰۰۰۰	-۰/۵۴	۰/۰۰۰۰	-۰/۲۸	-۰/۴۹	B	L
۴۵	۰/۰۰۰۰	-۰/۴۴	۰/۰۰۰۰	-۰/۲۱	-۰/۳۸	B	L
۴۶	۰/۰۰۰۰	-۰/۵۴	۰/۰۰۰۰	-۰/۲۵	-۰/۴۵	B	L
۴۹	۰/۰۰۰۰	-۰/۴۹	۰/۰۰۰۰	-۰/۳۴	-۰/۵۵	B	M
۵۵	۰/۰۰۰۰	۰/۷۸	۰/۰۰۰۰	۰/۴۴	۰/۷۰	C	F
۵۶	۰/۰۰۰۰	-۱/۳۹	۰/۰۰۰۰	-۱/۳۳	-۱/۷۷	C	M
۷۵	۰/۰۰۰۰	-۰/۵۲	۰/۰۰۰۰	-۰/۳۵	-۰/۵۹	B	M
۷۶	۰/۰۰۰۰	۰/۷۳	۰/۰۰۰۰	۰/۵۴	۰/۸۳	C	F
۹۲	۰/۰۰۰۱	۰/۴۴	۰/۰۰۰۲	۰/۲۳	۰/۳۸	B	F
۹۵	۰/۰۰۰۰	۰/۵۷	۰/۰۰۰۰	۰/۳۶	۰/۵۸	B	F
۱۰۹	۰/۰۰۰۰	-۰/۵۰	۰/۰۰۰۰	-۰/۳۳	-۰/۵۴	B	M





شکل ۲. نمودار کنش افتراقی سوالات افسردگی بر حسب ویژگی سن (= ۱۸ تا ۲۵ سال و = ۲۵ تا ۶۰ سال)

با توجه به نتایج ارائه شده در جدول شماره ۲ می توان بیان داشت که از بین ۵۷ سوال مولفه افسردگی پرسشنامه MMPI-2 که بر داوطلبان استخدامی اجراء شده است، تعداد ۲۸ سوال دارای کنش افتراقی است و مابقی سوالات کنش افتراقی بین زنان و مردان نداشته است. از بین سوالات دارای کنش افتراقی تعداد ۱۸ مورد (سوالات ۹، ۲۹، ۴۹، ۷۵، ۹۲، ۹۵، ۱۰۹، ۱۳۰، ۱۴۱، ۱۶۵، ۱۷۵، ۱۷۸، ۱۸۹، ۱۷۸، ۲۲۱، ۲۳۳، ۲۳۸، ۲۴۵ و ۲۴۸) دارای کنش افتراقی متوسط و ۱۰ مورد (سوالات ۱۰، ۱۵، ۵۵، ۵۶، ۷۶، ۱۴۳، ۱۴۶، ۱۸۸، ۲۱۵ و ۲۲۳) کنش افتراقی بزرگی را بر حسب ویژگی جنسیت گزارش داده‌اند. از طرفی با توجه به پیشینه پژوهش‌های انجام شده در حوزه افسردگی بیان شده که فراوانی افسردگی در افراد کمتر از ۲۵ سال و بالای ۲۵ سال متمایز است (کاپلان و سادوک<sup>۱</sup>، ۲۰۲۲)، از این جهت داوطلبان استخدامی از نظر سنی به طبقات ۱۸ تا ۲۵ سال (L) و گروه بالای ۲۵ سال (U) تقسیم شدند. در بررسی کنش افتراقی بر حسب سن نیز از بین سوالات مولفه افسردگی ۹ مورد (سوالات ۳۷، ۴۵، ۴۶، ۱۳۰، ۱۴۸، ۱۷۸، ۱۸۱، ۲۱۲ و ۲۶۷) دارای کنش افتراقی متوسط و ۵ مورد (سوالات ۲، ۱۰، ۳۱، ۱۴۶، ۲۷۸) دارای کنش افتراقی بزرگ بود. در مقایسه بین دو رویکرد متل‌هنزل و مدل راش در بررسی کنش افتراقی نیز می توان بیان داشت که تمام سوالاتی که مدل راش دارای کنش افتراقی تشخیص داده شدند نیز در مدل متل‌هنزل نیز دارای کنش افتراقی تشخیص داده شدند به جز سوال ۱۶۵ و ۲۳۳ که در روش راش دارای کنش افتراقی بود اما در روش متل‌هنزل شناسائی نشد. با توجه به حجم سوالات و تعداد نسبتاً یکسان در شناسائی سوالات دارای کنش افتراقی توسط دو روش می توان بیان داشت که دو روش متل‌هنزل و مدل راش عملکرد یکسانی را در شناسائی سوالات دارای کنش افتراقی داشتند. همچنین این نتیجه نشانه دیگری از برآزش مطلوب داده‌ها با مدل راش دارد (لینکر، ۲۰۱۲). در شکل شماره ۱ و ۲ نتایج کنش افتراقی سوالات افسردگی بر حسب ویژگی جنسیت و سن به صورت نمودار آمده است.

<sup>۱</sup>. Kaplan & Sadock

در ستون آخر جدول جهت تورش سوالات داری کنش‌افتراقی نشان داده شده است. اگر سوال تورش دار، مردان بیشتر جواب مثبت داده باشند، آن سوال با M نشان داده شده است و نشان از این است که این سوال برای مردان دارای کنش‌افتراقی بیشتری است و اگر در سوال تورش دار، زنان جواب مثبت بیشتری داده بودند، آن سوال با F نشان داده شده است و نشان از این است که زنان به این سوال از مولفه افسردگی پاسخ مثبت بیشتری داده‌اند. با توجه به ستون آخر جدول شماره ۲ نشان داده شد که ۱۳ سوال دارای تورش به جهت زنان و ۱۵ سوال دارای تورش به جهت مردان داشته است و تقریباً تعداد سوالات دارای کنش‌افتراقی به یک اندازه به جهت دو جنس تورش داشته‌اند. همچنین از این بین ۸ سوال در جهت اعمال نمره بالا به گروه بالای سنی و ۵ سوال به ضرر گروه پائین سنی تشخیص داده شد.

تاثیر وجود کنش‌افتراقی بر انصاف در آزمون را می‌توان از منظر تفاوت‌های عملکردی شرکت‌کنندگان در پرسشنامه مشخص نمود. وجود پرسش‌هایی با ویژگی کنش‌افتراقی بدان مفهوم است که سازه‌های غیرمرتبطی به پاسخ افراد تاثیر می‌گذارد و مالک توانایی شرکت‌کنندگان در آن آزمون نیست. در مقیاس افسردگی پرسشنامه MMPI-2، به دلیل وجود تعداد پرسش‌ها با ویژگی کنش‌افتراقی می‌توان بیان کرد که ۴۹/۱ درصد سوالات امکان تاثیرپذیری از عامل نامرتبط جنسیت و ۲۲/۸ درصد سوالات متاثر از سن شرکت‌کنندگان بود، که در برخی از موارد به نفع آنان و در برخی اوقات به ضرر آنان بوده است. پر واضح است که به صورت کلی بخش بسیار اعظمی از سوالات پرسشنامه حاضر دارای تبعیضی برای شرکت‌کنندگان بوده است و انصاف در آزمون را تا حد زیادی نقض کرده است و تنها در ۲۰ سوال (۶۴/۹ درصد) از نظر تبعیض جنسیتی انصاف در تشخیص برای مولفه افسردگی رعایت شده بود. لیکن باید در نظر داشت که مفهوم انصاف در آزمون مفهومی چند بعدی است و وجود یا فقدان آن متاثر از عوامل مختلفی است که از زمان طراحی آزمون باید آنان را کنترل کرد و سرانجام در زمان تفسیر نتایج، اثرگذاری آن عوامل را در نظر گرفت.

#### ۴. بحث و نتیجه‌گیری

هدف از این پژوهش بررسی و تبیین قدرت هر یک از شیوه‌های سنجش و تشخیص کنش‌افتراقی سوالات مولفه افسردگی پرسشنامه چندوجهی مینه‌سوتا-۲ بر اساس مدل راش و رویکرد منتل‌هنزل بود که بیشتر معرفی شد. هدف دیگر ارائه شده در این پژوهش، اثرگذاری این شیوه‌ها بر روایی و صحت تحلیل‌های آماری انجام شده بر نحوه پاسخگوئی داوطلبان استخدامی با توجه به خصیصه جنسیت و سن بود. بر اساس نتایج گزارش شده در جدول ۱ می‌توان بیان داشت که تقریباً تمامی سوالات مولفه افسردگی با مدل راش یک پارامتری دارای برازش مناسبی است و تنها ۴ سوال از مدل مورد برازش کمی انحراف داشته است که این انحراف‌ها قابل اغماض بوده و در کل داده‌ها با مدل راش برازش داشته است. بنابراین نتایج به دست آمده از کنش‌های افتراقی قابل اتکا و استفاده است.

در جدول ۲ نشان داده شد که از بین ۵۷ سوال مقیاس افسردگی آزمون MMPI-2 اجراء شده بر داوطلبان استخدامی ۲۸ سوال دارای کنش‌افتراقی بین زنان و مردان بود که از این بین ۱۳ سوال به ضرر مردها بوده است و در این سوالات مردان نمرات بالاتری را گرفتند و ۱۵ سوال به ضرر گروه زنان بوده است و این سوالات به نفع مردان دارای تورش است. با بررسی سوالات دارای تورش مشخص شد که سوالاتی بیشتر به ضرر مردان دارای کنش‌افتراقی بود که مربوط به سنجش ملال ذهنی (۹، ۱۰، ۱۵، ۸۲ و ...) و سوالاتی به ضرر زنان دارای تورش بوده است که بیشتر به سنجش افسردگی ذهنی و آشفتگی فکری (۴۹، ۵۶، ۷۵، ۹۵، ۱۳۰ و ...) پرداخته است. اگر به طور دقیق بیان شود، تعداد سوالات دارای کنش‌افتراقی، تا حدی بیشتر به ضرر زنان است و دلیل آن این

است که این سوالات به سنجش حالات هیجانی و عاطفی افسردگی پرداخته است که به طور طبیعی زنان در این سوالات اقبال بیشتری را از خود نشان می‌دهند. با اینحال اگر با کمی اغماض به سوالات دارای کنش افتراقی نگاه کرد، می‌توان بیان داشت که سوالات به یک اندازه به ضرر مردان و زنان دارای تورش است. بنابراین اگر کل سوالات مقیاس افسردگی با هم اجراء شود تا حدی جلوی تورش را در کل آزمون می‌گیرد، اما اگر سوالات به صورت گزینشی اجراء شود و یا اینکه از خرده مقیاس‌های هریس-لینگوز<sup>۱</sup> استفاده شود امکان تورش شدید در نتایج بیشتر است و نتایج تا حد زیادی ناعادلانه و غیر قابل اتکاء خواهد بود و این دلیل تفاوت در یافته‌های به دست آمده از [گوری و مارین \(۲۰۱۵\)](#)، [مک نالتی و همکاران \(۲۰۱۵\)](#) و [چیلدز و همکاران \(۲۰۰۰\)](#) با یافته‌های به دست آمده از این پژوهش بوده است، چرا که این پژوهش‌ها از خرده مقیاس‌های هریس-لینگوز استفاده نموده‌اند. البته در تبیین این نکته که چرا در مقیاس افسردگی سوالات زیادی دارای کنش افتراقی تشخیص داده شده‌است را می‌توان چنین بیان داشت که نمونه مورد هدف این پژوهش، داوطلبین استخدامی بودند و این گروه نمونه با توجه به اینکه در حیطة غیر بالینی و جامعه نرمال قرار گرفته‌اند، شاخص‌های خصیصه‌ای بیشتر از شاخص‌های بالینی نمایان می‌گردد و این یکی از دلایلی است که می‌توان برای بالا بودن تعداد سوالات دارای کنش افتراقی در این مقیاس بیان داشت. البته سوالاتی نیز که دارای کنش افتراقی نبودند نیز می‌توان به‌عنوان سوالات منصفانه و سوالاتی معرفی کرد که گروه‌های سنی و یا جنسیتی تأثیری در نتایج آنان ندارد و این سوالات به دور از هر گونه سوءگیری ناشی از گروه‌بندی‌های جنسیتی و سنی نتایج را گزارش می‌نمایند.

همچنین در مورد سوالات دارای کنش افتراقی بر اساس سن می‌توان بیان داشت که از مجموع ۵۷ سوال مقیاس افسردگی تنها تعداد ۱۴ سوال دارای کنش افتراقی تشخیص داده شد، که از این بین ۹ سوال دارای کنش افتراقی به نفع افراد با سن بالاتر از ۲۵ سال و ۵ سوال به نفع افراد بین ۱۸ تا ۲۵ سال بوده است که در مجموع می‌توان بیان داشت که سوالات بیشتری به ضرر گروه جوان دارای سوءگیری است. از بین این سوالات، مواردی که به بررسی بی‌ثباتی خلقی در مقیاس افسردگی بیشتر می‌پردازند (برای مثال سوالات ۲، ۱۰، ۱۳۰، ۱۴۶ و ...) به ضرر گروه جوان دارای تورش بوده است و سوالاتی که مربوط به عملکرد جسمانی و تمرکز بوده است (برای مثال سوالات ۳، ۳۷، ۴۵ و ...) به ضرر گروه مسن‌تر دارای کنش افتراقی بود. البته بیشتر سوالات (۷۵ درصد) مقیاس افسردگی بدور از کنش افتراقی بودند و نشان داده شد که این سوالات به طور منصفانه به بررسی مولفه افسردگی در بین سنین مختلف پرداخته‌اند. در بحث مقایسه شیوه‌های سنجش کنش افتراقی مطرح شد در تشخیص پرسش‌هایی با سطح کنش افتراقی متوسط و یا بالا (۳۷ سوال از ۵۷ سوال) مدل مثل‌هنزل کارآمدی مناسبی از خود ارائه داد و نتایج نشان داد که مدل مثل‌هنزل نسبتاً به اندازه مدل راش در تشخیص کنش افتراقی توانمند است. البته این نکته را باید بیان داشت که روش مثل‌هنزل برای سوالات با تمام سطوح توانائی نسبتاً در یک حد توانمند بوده است که این نتیجه با یافته [کرمی و خودی \(۱۳۹۹\)](#) مبنی بر اینکه این روش برای سوالات بسیار ساده و یا بسیار سخت کارآمدتر از سوالاتی با میزان سختی متوسط است، ناهمسو بوده است که علت این امر می‌تواند این باشد که آزمون مورد استفاده در پژوهش [کرمی و خودی \(۱۳۹۹\)](#) از نوع آزمون‌های شناختی و مربوط به درس زبان بوده که محتوای آن کاملاً متفاوت از آزمون‌های غیرشناختی است.

یافته‌های این پژوهش برای روان‌سنج و متخصصان طراحی آزمون و تحلیل‌گران نتایج آزمون، رویکردهای جدید سنجش کنش افتراقی را توصیه می‌نماید، چرا که مدل‌های راش برتری نسبی خوبی نسبت به نظریه کلاسیک اندازه‌گیری دارد ([ویلسون و هاسکنس، ۲۰۰۵](#)). به نظر [لی، ژانگ و بین<sup>۳</sup> \(۲۰۱۰\)](#) دقت همبستگی‌های بین خرده مقیاس‌ها و اعتبار آزمون با استفاده از مدل

<sup>۱</sup>. Harris - Lingo

<sup>۲</sup>. Wilson & Hoskens

<sup>۳</sup>. Lee, Zhang & Yin

راش بهبود می‌یابد زیرا این مدل خطای اندازه‌گیری را در برآورد به حساب می‌آورد. مدل راش با فراهم آوردن یک روش مناسب برای تبدیل غیرخطی داده‌های خام رتبه‌ای به اندازه‌های فاصله‌ای بر یکی از نگرانی‌ها و دغدغه‌های رایج در ساخت ابزارهای اندازه‌گیری روانی و تربیتی، یعنی دستیابی به عینیت خاص یا تغییر ناپذیری مقیاس‌ها غلبه کرده‌است (مینائی، ۱۳۹۳؛ تنانت و کانگان، ۲۰۰۷). البته در تعمیم‌پذیری نتایج به دست آمده از مدل راش باید دقت لازم را در نظر داشت چرا که مولفه افسردگی پرسشنامه چند وجهی شخصیتی مینه‌سوتا به کار رفته در این پژوهش دارای سازه تک بعدی<sup>۲</sup> است و مفروضات لازم نظریه سوال- پاسخ را رعایت کرده است. اما اگر سازه‌ای مفروضه عدم وابستگی مکانی<sup>۳</sup> در آن رعایت نشده باشد، ممکن است در سنجش با شیوه‌های ذکر شده نتایج متفاوتی را بدست آورد. همچنین باید در نظر گرفت اگر در مدل‌های نظریه جدید اندازه‌گیری فراتر از مدل راش رفته و مدل‌های دو یا سه پارامتری را استفاده کنیم، محتمل است که نتایج تغییر یافته و حساسیت نظریه سوال- پاسخ افزایش یابد.

همچنین با توجه به نتایج حاصل از این پژوهش به روانسنج و روان‌شناسانی که با مراکز طب کار جهت بررسی سلامت روان داوطلبان استخدامی همکاری دارند و به ویژه مجریان آزمون MMPI-2 توصیه می‌شود که نگاهی به نتایج به دست آمده از کنش‌های افتراقی سوالات بر اساس سن و جنسیت کرده و در تفسیر و نمره‌گذار به این ویژگی‌های دموگرافیک داوطلبان توجه بیشتری نمایند. در نهایت به پژوهشگران پیشنهاد می‌شود تا کنش‌افتراقی مولفه‌های دیگر پرسشنامه چندوجهی مینه‌سوتا مانند پرخاشگری، اضطراب، پارانوئیا و اسکیزوفرنی را در بین داوطلبان استخدامی بررسی نموده و همچنین از روش‌های دیگر بررسی کنش‌های افتراقی نیز در کنار مدل راش بهره ببرند. همچنین به پژوهشگران توصیه می‌گردد که به بررسی نقاط برش مولفه افسردگی و دگر مولفه‌های این پرسشنامه به تفکیک جنسیت و سن پرداخته تا نهایت رعایت انصاف در آزمون صورت پذیرد، و پیشنهاد می‌شود تا از مسیرهای دیگر کنش‌افتراقی سوالات افسردگی مورد بررسی قرار گیرد و نتایج آن با نتایج پژوهش حاضر مقایسه شود.

## ۵. ملاحظات اخلاقی

در این پژوهش سعی شد که نام و مشخصات داوطلبان شرکت داده‌شده در طرح پژوهشی و نتایج به دست آمده از آزمون‌ها به صورت محرمانه باقی بماند. همچنین نتایج به دست آمده از آزمون‌های شرکت کنندگان صرفاً در این پژوهش مورد استفاده قرار گرفت و در فرایند شغلی و استخدامی له یا علیه افراد ترتیب اثر داده نشد. داده‌های به دست آمده با اطلاع و هماهنگی با مراکز طب کار استفاده شد.

## ۶. تقدیر و تشکر و حمایت مالی

این مقاله بر گرفته از رساله دکتری نویسنده اول در رشته سنجش و اندازه‌گیری دانشگاه علامه طباطبائی با عنوان کاربرد مدل راش و مقایسه آن با نظریه کلاسیک اندازه‌گیری در افزایش دقت و روایی پرسشنامه شخصیت چندوجهی مینه‌سوتا (MMPI) در فرایند استخدام‌های شغلی بوده است و برای نگارش این مقاله حمایت مالی و معنوی از سازمان و یا شرکتی دریافت نشده است.

## ۷. تعارض منافع

1. Tennant & Conaghan

2. unidimensiona

3. local Independence

این مقاله تعارض منافی با مجله و یا سازمانی دیگر ندارد.

## منابع

- امبرتسون، س.، و رایس، اس. پی. (۱۳۸۸). نظریه‌های جدید روانسنجی برای روانشناسان. ترجمه حسن پاشا شریفی، ولی‌اله فرزاد، مجتبی حبیبی عسگرآباد، و بلال ایزانلو. تهران: رشد.
- دستورالعمل سنجش و ارزیابی سلامت روانی کارکنان وظیفه در مراکز آموزشی نیروهای مسلح. (۱۳۹۶). ۷۰۳/۳۲۱۲/۱/۲۲۱۵. <https://dotic.ir/news/2979>
- زاهد، ز.، نصراللهی، ط.، و مهینی‌زاده، م. (۱۳۹۹). از تبعیض جنسیتی به سوی برابری و رشد اقتصادی. زن در توسعه و سیاست، ۱۸(۴)، ۶۴۳-۶۶۲. <http://dx.doi.org/10.22059/jwdp.2021.311205.1007897>
- سلطانی شال، ر.، سعادت بین جواهری، ف.، و زبردست، ع. (۱۳۹۹). بررسی سطح بهزیستی و ویژگی‌های روانسنجی مقیاس بهزیستی پرستاران در محیط کار. طب کار، ۱۲(۱)، ۵۵-۶۸. <http://dx.doi.org/10.18502/tkj.v12i1.3659>
- کریمی، ح.، ر.، گرامی پور، م.، و مینائی، ا. (۱۴۰۰). نرخ آشکارسازی کنش افتراقی سؤال با استفاده از مدل درخت‌های راش یک مطالعه شبیه‌سازی شده و داده‌های واقعی آزمون‌های سرنوشت‌ساز ناچا. اندازه‌گیری تربیتی، ۱۱(۴۴)، ۱-۳۰. <http://dx.doi.org/10.22054/jem.2021.61694.2190>
- کریمی، ح.، و خودی، ع. (۱۳۹۹). بررسی کنش افتراقی پرسش‌ها و عملکرد در آزمون: مقایسه رگرسیون لجستیک، مدل رش و منتل-هنزل. پژوهش‌های زبان‌شناختی در زبان‌های خارجی، ۱۰(۴)، ۸۴۲-۸۵۳. <http://dx.doi.org/10.22059/jflr.2021.315079.783>
- مینائی، ا. (۱۳۹۳). کاربرد مدل اندازه‌گیری راش در ارزشیابی ویژگی‌های اندازه‌گیری آزمون مهارت‌های دیداری-حرکتی (R-TVMS). فصلنامه اندازه‌گیری تربیتی، ۱۸(۱)، ۷۷-۱۱۴. <http://dx.doi.org/10.22054/jem.2015.790>
- همیلتون، ر.ک.، سوامیناتان، ا.ج.، و راجرز، ج. (۱۳۸۹). میانی نظریه پرسش پاسخ. ترجمه محمد رضا فلسفی نژاد. تهران: انتشارات دانشگاه علامه طباطبائی.

## References




- Avenevoli, S., Swendsen, J. P., Burstein M., & Merikangas, K.R. (2015). Major depression in the national comorbidity survey adolescent supplement: prevalence, correlates and treatment. *Journal American Academic Children Adolescent Psychiatry*, 54(1), 37-44. <http://dx.doi.org/10.1016/j.jaac.2014.10.010>
- Baker. C. A, Baum. L. J & Francis. J. C. (2023). Assessment of test bias on the MMPI-2-RF higher order and restructured clinical scales as a function of gender and race. *Professional Psychology Research and Practice* 54 (4), 314-325. <http://dx.doi.org/10.1037/pro0000517>
- Bethune M. M. (2011). Predictors of Performance in a Professional Counselor Masters Program, Doctoral Thesis of Counseling Psychology Program, College of Education and Behavioral Sciences, University of Northern Colorado. <https://digscholarship.unco.edu/dissertations/75>
- Buribayev, Y. A., & Khamzina Z. A. (2019). Gender equality in employment: The experience of Kazakhstan. *International Journal of Discrimination and the Law*, 19(2), 110-124. <http://dx.doi.org/10.1177/1358229119846784>
- Camilli, G. (2006). Test fairness. In R. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 221-256). New York: American Council on Education & Praeger series on higher education. <http://dx.doi.org/10.1037/14049-027>
- Camilli, G., & Congdon, P. (1999). Application of a method of estimating DIF for polytomous test items. *Journal of Educational and Behavioral Statistics*, 24(4), 323-341. <http://dx.doi.org/10.3102/10769986024004323>
- Camilli, G., & Shepard, L. (1994). Methods for identifying biased test items. Newbury Park, CA Sage.
- Chapelle, C. A. (2020). *Validity in language assessment*. The Routledge Handbook of Second Language Acquisition and Language Testing, 11. <http://dx.doi.org/10.4324/9781351034784-3>



- Chen, M. Y., Liu, Y., & Zumbo, B. D. (2020). A propensity score method for investigating differential item functioning in performance assessment. *Educational and Psychological Measurement*, 80(3), 476-498. <http://dx.doi.org/10.1177/0013164419878861>
- Cherry, K. (2022) The Minnesota Multiphasic Personality Inventory (MMPI). Updated on September 02 Medically reviewed by Amy Morin. [MMPI Test: The Minnesota Multiphasic Personality Inventory \(verywellmind.com\)](https://www.verywellmind.com/mmpi-test-the-minnesota-multiphasic-personality-inventory)
- Childs R. A., Dahlstrom W. G., Kemp S. M. & Panter. A. T. (2000). Item Response Theory in personality assessment: A demonstration using the MMPI-2 depression scale. *Psychological Assessment*. 7(1), 37-54. <http://dx.doi.org/10.1177/107319110000700103>
- Deskovitz MA, Weed NC, McLaughlan JK, Williams JE. (2016). Interpretive Reliability of Six Computer-Based Test Interpretation Programs for the Minnesota Multiphasic Personality Inventory-2. *Assessment*. 23(2), 250-261. <http://dx.doi.org/10.1177/1073191115584970>
- Embretson, S. E., and Reise, S. P. (2000). Item Response Theory for Psychologists. Translated by Hassan Pasha Sharifi, Vali Allah Farzad, Mojtaba Habibi Asgar Abadi & Balal Ezanlow, Tehran: Roshd. (In Persian)
- Gori. E., & Marin. R. F. (2015). Rasch Analysis of some MMPI-2 scales in a sample of university freshmen. *International Journal of Arts & Sciences*. 08(03), 107-150. <https://www.universitypublications.net/ijas/0803/pdf/B5R247.pdf>
- Guidelines for measuring and evaluating the mental health of employees on duty in the training centers of the armed forces. (2016). 2215/1/3212/703. <https://dotic.ir/news/2979> (In Persian)
- Hambleton, R. K., & Rogers, H. J. (1989). Detecting potentially biased test items: Comparison of IRT area and Mantel-Haenszel methods. *Applied Measurement in Education*, 2(4), 313-334. [https://doi.org/10.1207/s15324818ame0204\\_4](https://doi.org/10.1207/s15324818ame0204_4)
- Hambleton, Ronald K.; Saminathan, H.; Rogers, H. Jane (1991). Fundamentals of Question-Answer Theory. Translated by Mohammad Reza Filsafinejad (2010). Tehran: Allameh Tabatabai University. (In Persian)
- Kaplan & Sadock's Synopsis of Psychiatry, (2022). 12th Edition is designed to replicate how to approach clinical cases and is comprehensive including DSM-5 criteria and ICD-10 diagnoses.
- Karami, H., & Khodi, A. (2021). Differential Item Functioning and Test Performance: a Comparison Between the Rasch Model, Logistic Regression and Mantel-Haenszel. *Journal of Foreign Language Research*, 10 (4), 842-853. <http://dx.doi.org/10.22059/jflr.2021.315079.783> (In Persian)
- Karami, H. (2013) The quest for fairness in language testing. *Educational Research and Evaluation*, 19(2&3), 158-169. <http://dx.doi.org/10.1080/13803611.2013.767618>
- Karami. H. R., Gramipour. M., & Minaei. A. (2021). Differential Item Functioning (DIF) Detection Rate Using Rasch Trees Model: A Simulated and Real Data Study of the NAJA High Stakes Tests. *Educational Measurement*. 11(44), 1-30. <http://dx.doi.org/10.22054/jem.2021.61694.2190> (In Persian)
- Lee, J.C., Zhang, Z., Yin, H. (2010). Using multidimensional Rasch analysis to validate the Chinese version of the Motivated Strategies for Learning Questionnaire (MSLQ-CV). *European Journal of Psychology of Education*. 25, 141-155. <http://dx.doi.org/10.1007/s10212-009-0009-6>
- Linacre, J. M. (2012). A User's Guide to WINSTEPS®. Retrieved May 2, from <http://www.winsteps.com>
- Hathaway, S. R., & McKinley, J. C. (1943). MMPI. Manual for administration and scoring. Minneapolis: University of Minnesota Press.
- Marnat, G. G. & Wright, A. J. (2016). *Handbook of Psychological Assessment*. 6<sup>th</sup> ed. by John Wiley & Sons, Inc. ISBN: 978-1-118-96064-6.
- McNulty J. L., Forbey J. D., Graham J. R., Ben-Porath Y. S., Black M. S., Anderson S. V. & Burlew A. K. (2015). MMPI-2 Validity Scale Characteristics in a Correctional Sample. *Assessment*, 10(3), 288-298. <http://dx.doi.org/10.1177/1073191103255623>
- Minaei. A. (2015). Application of Rasch measurement model to evaluate measurement properties of the Test of Visual- Motor Skills-Revised. *Journal of Educational Measurement*, 5(18), 77-114. <http://dx.doi.org/10.22054/jem.2015.790> (In Persian)

- Şahin, A. & Anıl, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory & Practice*, 17, 321–335. <http://dx.doi.org/10.12738/estp.2017.1.0270>
- Siebert, R. J., Tennant, A., & Turner-Stokes, L. (2010). Rasch analysis of the Beck Depression Inventory-II in a neurological rehabilitation sample. *Disability and Rehabilitation*, 32(1), 8–17. <http://dx.doi.org/10.3109/09638280902971398>
- Skrondal, A., Rabe-Hesketh, S., and Boca Raton, F. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Chapman & Hall/ CRC Press. ISBN:9781584880004
- Soltani Shal R, Saadatbin Javaheri F, Zebardast F. (2020). Survey the level of well-being and Psychometric characteristics of hospital nurses' well-being at work scale. *Occupational Medicine Quarterly Journal*, 12(1), 55-68. <http://dx.doi.org/10.18502/tkj.v12i1.3659> (In Persian)
- Su. Y & Wang N. (2005). Use of the Rasch IRT model in standard setting: An item- mapping method. *Journal of Educational Measurement*. 40 (3), 231- 53. <http://dx.doi.org/10.1111/j.1745-3984.2003.tb01106.x>
- Talerico G. M., McCallum. J. J., Whitman M. R., Tarescavage. M., Corey D. M., & Ben-Porath. Y. B. (2023). Comparing the Validity of MMPI-3 Scores in Prehire Psychological Screenings of Male and Female Police Officer Candidates, *Journal of Personality Assessment*, 106(1), 27-36. <http://dx.doi.org/10.1080/00223891.2023.2191278>
- Tennant, A., & Conaghan, P.G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism* 57, 1358– 1362. <http://dx.doi.org/10.1002/art.23108>
- Valianpour. Z., Modarres Gharavi. M., & Mahram. B. (2020). Validation of the Minnesota Multiphasic Personality Inventory (MMPI 2) in psychiatric patients and non-patient individuals in Mashhad city, Iran. *Journal of Fundamentals of Mental Health*, 22(6), 399-407. <https://doi.org/10.22038/jfmh.2020.17815>.
- Wiberg, M. (2007). *measuring and detecting Differential Item Functioning in criterion-referenced licensing test*. Umea Universitet. ISSN 1103-2685.
- Wilson, M., & Hoskens, M. (2005). *Multidimensional Item Responses: Multimethod-Multitrait Perspectives Applied Rasch Measurement: A Book Of Exemplars*. California: Springer. [http://dx.doi.org/10.1007/1-4020-3076-2\\_16](http://dx.doi.org/10.1007/1-4020-3076-2_16)
- Zahid. Z., nasrollahi. Z., & Mahinizadeh. M. (2020). From Gender Discrimination to Equality and Economic Growth (Study of Developing Countries). *Journal of Woman in Development and Politics*.4(18), 662- 643. <http://dx.doi.org/10.22059/jwdp.2021.311205.1007897> (In Persian)
- Zumbo, B. D. (2003). Does item-level DIF manifest itself in scale-level analyses? Implications for translating language tests. *Language testing*, 20(2), 136-147. <http://dx.doi.org/10.1191/0265532203lt248oa>

## Differential Item Functioning of Minnesota Multiphasic Personality Inventory-2 (Depression) Items in among Recruitment Candidates: A Comparison Between the Rasch Model and Mantel-Haenszel

Mahdi Rahmani Malek Abad<sup>1</sup> , Mohammad Reza Falsafinejad<sup>2\*</sup> , Asghar Minaei<sup>3</sup>   
, Noor Ali Farrokhi<sup>4</sup> 

1. Department of Educational Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabataba'i University, Tehran, Iran. E-mail: [m.rahmani\\_psy@yahoo.com](mailto:m.rahmani_psy@yahoo.com)
2. Corresponding Author, Department of Educational Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabataba'i University, Tehran, Iran. E-mail: [falsafinejad@atu.ac.ir](mailto:falsafinejad@atu.ac.ir)
3. Department of Educational Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabataba'i University, Tehran, Iran. E-mail: [as\\_minaei@atu.ac.ir](mailto:as_minaei@atu.ac.ir)
4. Department of Educational Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabataba'i University, Tehran, Iran. E-mail: [farrokhi@atu.ac.ir](mailto:farrokhi@atu.ac.ir)

### Extended Abstract

**Aim:** The purpose of the present study is to identify the questions of the depression component of the Minnesota Multifaceted Personality Questionnaire-2, which are biased toward a certain group among job candidates based on gender and age. One of the tasks of measurement is to have fairness in the test, especially if the test is important for selecting healthy people for selection in jobs. One of the features that underlines the fairness of the assessment is the bias of the questionnaire items (Baker, Baum & Francis, 2023). One of the ways to identify biased items is the Differential Item Functioning. By using DIF, it is possible to identify the items that have a bias in favor of or against a certain group in order to prevent the occurrence of bias in the test results (Wiberg, 2007). So far, several methods have been proposed to detect DIF, which has caused some confusion to researchers, because they are sometimes different in terms of statistical basis and techniques, and the results of detecting biased items have had difference (Karami, 2013). Due to the importance of identifying biased depression scale items from MMPI-2, which is used to diagnose the mental health of Recruitment candidates, it is possible to increase the fairness in the correct selection.

**Method:** The present study investigated and compared the results obtained from the two methods of differential item functioning of classical test theory (Mantel-Haenszel) and Rasch model for the depression scale items of Minnesota Multifaceted Personality Questionnaire-2 according to gender and age. This research is survey research, and the statistical population was all the Recruitment candidates of Isfahan city in 2022 and spring 2023. The sample size was 5997 people (57.31% men and 42.69% women) who were selected by one-step cluster method. The tool used was the depression scale of Minnesota Multiphasic Personality Inventory-2. (Marnat & Wright, 2016). This scale included 57 items that have two options: Yes (1) and No (0). The fit of the Items with the model was confirmed using outfit MnSq and infit MnSq, and DIF according to gender and age was checked and compared with Mantel-Haenszel and Rasch model by Winsteps version 3.64.

**Results:** At first, data fit index was checked with Rasch model. In this component, the most difficult item was related to items 45, 75 and 39, and items to which most people answered positively were items 68, 134, 37 and 238. that among the 57 depression items, the fit indices of 53 items were at the appropriate level (between 0.5 and 1.5) and only 4 items had MnSq Outfit values higher than 1.5, in total the results showed that the model had a good fit with the data. Then uniform differential item functioning analysis was performed with Rasch model and Mantel-Haenszel method. Among the 57 items of the depression scale, 28 items had a differential action between men and women. Among the items with differential action, 18 items showed moderate differential action and 10 items showed large differential action according to gender and the remaining items had no bias. Also, among all the items of the depression scale, 13 items had DIF in terms of age category (people under 25 years and people over 25 years old), 9 items had moderate differential action and 5 items had large differential action. In general, the results showed that more than half of the items (64.9%) in their performance had a bias

towards men and people 25 years old and above, and these people naturally get a higher score on the depression scale. Mantel-Henzel's method also provided a similar finding to Rasch's model, except for the difference in two items (165 and 233).

**Conclusion:** The results showed that scale Depression items have been at the bias of women and older people, and in the interpretation and scoring, age should be considered in addition to gender, because that men and older people naturally score higher on the depression scale of the MMPI-2. Also, in the discussion of the comparison of DIF measurement methods, it was found that the Mantel-Henzel model performed well in detecting questions with a medium or high level of differential action (37 questions out of 57 questions) and the results showed that the Mantel-Henzel model is comparable to the Rasch model in detecting Differential action is powerful (Karami & Khodi, 2021). Also, in the discussion of the comparison of DIF detection methods, it was found that in detecting items with a medium or high level of Differential Item Functioning (37 questions out of 57 questions), the Mantel-Henzel model performed well, and the results showed that the Mantel-Henzel model is relatively as good as the Rasch model in detecting Differential action is powerful (Wilson & Hoskens, 2005), because the Rasch model has overcome one of the common concerns and concerns in the construction of psychological and educational measurement tools by providing a suitable method for non-linear transformation of raw rank data into distance measures, i.e. achieving specific objectivity or immutability of scales (Tennant & Conaghan, 2007).

**Keyword:** *Depression, Recruitment candidates, Differential Item Functioning, Rasch Model, Mantel-Haenszel Method.*

**Ethical Considerations:** In this research, it was tried to keep the names and details of the participants of the research project and the results obtained from the tests confidential. Also, the results obtained from the participants' tests were only used in this research, and they were not used against people in the job and employment process. The obtained data were used with information and coordination with medical centers.

**Acknowledgments:** This article is based on a doctoral thesis in the field of Educational Measurement of Allameh Tabatabai University, entitled the application of the Rasch model and its comparison with the classical theory of measurement in increasing the accuracy and validity of the MMPI-2 questionnaire in the job recruitment process. No financial or moral support was received from any organization or company for writing this article.

**Conflict of Interest:** This article has no conflict of interest with the journal or any other organization.

## References

- Baker. C. A, Baum. L. J & Francis. J. C. (2023). Assessment of test bias on the MMPI-2-RF higher order and restructured clinical scales as a function of gender and race. *Professional Psychology Research and Practice* 54 (4) Follow journal. <http://dx.doi.org/10.1037/pro0000517>
- Karami, H & Khodi, A. (2021). Differential Item Functioning and Test Performance: a Comparison Between the Rasch Model, Logistic Regression and Mantel-Haenszel. *Journal of Foreign Language Research*, 10 (4), 842-853. (in Persian). <http://dx.doi.org/10.22059/jflr.2021.315079.783>
- Karami, H. (2013) The quest for fairness in language testing. *Educational Research and Evaluation*, 19(2&3), 158-169. (in Persian). <http://dx.doi.org/10.1080/13803611.2013.767618>
- Marnat, G. G. & Wright, A. J. (2016). *Handbook of Psychological Assessment*. 6<sup>th</sup> ed. by John Wiley & Sons, Inc. ISBN: 978-1-118-96064-6.

- Tennant, A., & Conaghan, P.G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism* 57, 1358– 1362. <http://dx.doi.org/10.1002/art.23108>
- Wiberg, M. (2007). measuring and detecting Differential Item Functioning in criterion-referenced licensing test. Umea Universitet. ISSN 1103-2685.
- Wilson, M., & Hoskens, M. (2005). *Multidimensional Item Responses: Multimethod-Multitrait Perspectives Applied Rasch Measurement: A Book Of Exemplars*. California: Springer. [http://dx.doi.org/10.1007/1-4020-3076-2\\_16](http://dx.doi.org/10.1007/1-4020-3076-2_16)

in Press / Accepted Manuscript