



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

Mahdi Rahmani Malek Abad¹ , Mohammad Reza Falsafinejad^{1*} ,
Asghar Minaei¹ , Noor Ali Farrokhi¹ 

1. Department of Educational Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabataba'i University, Tehran, Iran.

*Corresponding Author: Associate Professor, Department of Educational Measurement, Faculty of Psychology and Educational Sciences, Allameh Tabataba'i University, Tehran, Iran. Email: falsafinejad@atu.ac.ir

ARTICLE INFO

Article type:
Research Article

Article History:
Received: 08 Aug 2023
Revised: 27 Oct 2023
Accepted: 19 Nov 2023
Published: 02 Oct 2026

Keywords:
Depression,
Differential Item Functioning,
Mantel-Haenszel Method,
Rasch Model,
Recruitment Candidates.

ABSTRACT

The purpose of the present study is to identify the questions in the depression component of the Minnesota Multifaceted Personality Questionnaire-2, which are prejudicial to a specific demographic of job applicants on the basis of gender and age. This research is a survey, and the statistical population consisted of all recruitment candidates in Isfahan city during the 2022 and spring 2023. The one-step cluster method was used to select 5997 individuals, with 57.31% being male and 42.69% being female. The tool used was MMPI-2 questionnaire with 370 questions. Winsteps software version 3.64 was used to analyze the data. 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 findings indicated that over half of the items (64.9%) were biased against women and young people. In addition to Rasch's model, Mantel-Haenszel also found a comparable result, with the exception of two items. The use of depression component questions simultaneously can result in non-bias. However, the use of Harris-Lingoes subscales or selective questions will result in severe DIF that is heavily biased toward women and young people. Additionally, the Rasch model exhibits a relative advantage over the Mantel-Haenszel model in detecting DIF.

Cite this article: Rahmani Malek Abad, M., Falsafinejad, M. R., Minaei, A., & Farrokhi, N. A. (2026). Differential Item Functioning of Minnesota Multiphasic Personality Inventory-2 (Depression) Items in Recruitment Candidates: A Comparison Between the Rasch Model and Mantel-Haenszel. *Journal of Applied Psychological Research*, 17(3), 323-338. <https://doi.org/10.22059/japr.2024.363263.644703>



© Author(s) retain the copyright.

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

Publisher: University of Tehran Press.

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

Extended Abstract

Aim

One of the important requirements of psychological measurement is fairness in testing, particularly when an assessment is used to evaluate the mental health of recruitment candidates. Accurate and fair assessment is essential because biased items may lead to inappropriate decisions in employment. One feature that underlines the fairness of assessment is the absence of item bias (Baker et al., 2023).

One way to identify biased items is through Differential Item Functioning (DIF). DIF makes it possible to identify items that function differently across groups of individuals with the same level of the underlying trait and to determine whether an item favors or disadvantages a particular group (Wiberg, 2007). Gender and age may influence individuals' responses to psychological assessment items and consequently affect the fairness and accuracy of employment decisions. This issue is particularly important for the depression component of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2), which is used to assess the mental health of recruitment candidates.

Several methods have been proposed to detect DIF, which may produce different results because they differ in their statistical foundations and techniques (Karami, 2013). These methods include approaches based on Classical Test Theory, such as the Mantel-Haenszel (MH) method, and Item Response Theory, including the Rasch model. The MH method is a widely used approach for detecting DIF, whereas the Rasch model provides an IRT-based framework for examining item functioning and fit. Previous research has reported differences in the identification and fit of MMPI-2 items using different methodological approaches (Gori & Marin, 2015; Childs et al., 2000). However, the differential functioning of MMPI-2 depression items in relation to fairness in the selection of recruitment candidates has not been sufficiently examined.

Therefore, the present study aims to examine the depression component of the MMPI-2 among recruitment candidates by determining whether its items fit the Rasch model, identifying items with DIF using the Rasch and MH methods, and examining whether the two methods differ in identifying items with DIF.

Methodology

This applied, quantitative study used a descriptive survey design. The statistical population consisted of approximately 10,000 recruitment candidates for teaching positions (or student-teacher positions) in Isfahan who attended occupational medicine centers during the 2022 recruitment examinations and spring 2023. A one-stage cluster sampling method was used. Considering the data analysis procedure and the use of the Rasch model, a sample of more than 4,000 participants was considered to increase the reliability of the results (Şahin & Anıl, 2017). Initially, 6,213 candidates were selected from three occupational medicine centers, and after excluding incomplete questionnaires, 5,997 participants (57.31% men and 42.69% women) were included in the analysis. Inclusion criteria were being a recruitment candidate for the Ministry of Education, completing the occupational medicine screening process through the psychological assessment stage, and responding accurately to the questionnaire. Participants were excluded if more than 35 items were unanswered, they had an active treatment record, or their test validity indices were invalid.

The instrument was the depression scale of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) (Groth-Marnat & Wright, 2016). The scale contains 57 items with dichotomous responses (Yes = 1, No = 0). In the present study, the depression component was examined because of the number of items in the full MMPI-2. The scale showed acceptable reliability in the present study, with Cronbach's alpha of .78 and split-half reliability of .73.

For data analysis, descriptive statistics and Cronbach's alpha were calculated using SPSS version 27. Item fit was examined using the Infit and Outfit mean-square (MnSq) statistics in Winsteps version 3.64. Given the large sample size and the sensitivity of ZSTD to sample size, item fit was evaluated primarily using MnSq. Differential item functioning (DIF) by gender and age was examined using the Mantel-Haenszel (MH) method and the Rasch model, and the results of the two approaches were compared.

Findings

Among the 5,997 participants, 57.31% were men and 42.69% were women. The mean age was 28.97 years ($SD = 4.15$). Regarding educational level, 37.17% had a diploma, 42.87% had a bachelor's degree, 15.81% had a master's or doctoral degree, and 4.15% had religious education.

First, the fit of the 57 depression items of the MMPI-2 to the Rasch model was examined. The most difficult items were 45, 75, and 39, whereas items 68, 134, 37, and 238 received positive responses most frequently. The fit indices of 53 items were within the desirable range of 0.5–1.5. Four items (37, 134, 142, and 178) had Outfit MnSq values above 1.5 but below 2, which were considered acceptable. Overall, all items showed an acceptable fit to the Rasch model.

DIF was then examined by gender and age using both the Mantel–Haenszel (MH) method and the Rasch model. Based on the Rasch results, 28 of the 57 items showed DIF by gender, including 18 items with moderate DIF and 10 with large DIF. By age group (18–25 years vs. 25–60 years), 14 items showed DIF, including 9 with moderate DIF and 5 with large DIF. According to the direction of DIF, 13 gender-related items favored women and 15 favored men; regarding age, 8 items favored the older group and 5 favored the younger group.

The MH method yielded findings highly consistent with those of the Rasch model. All items identified as having DIF by the Rasch model were also identified by the MH method, except for items 165 and 233, which were detected only by the Rasch model. Thus, the two methods showed largely similar performance in identifying DIF.

Overall, 49.1% of the depression items showed DIF by gender and 24.6% by age, indicating that a substantial proportion of the items may be influenced by characteristics unrelated to the depression trait. These findings highlight the importance of considering DIF when interpreting the depression scale of the MMPI-2 in recruitment assessments.

Conclusion

The findings showed that almost all 57 depression items of the MMPI-2 demonstrated acceptable fit to the Rasch model, with only four items showing minor and acceptable deviations. This indicates that the data were adequately fitted to the Rasch model and that the identified DIF findings were sufficiently reliable. Regarding gender, 28 items showed DIF, including 18 with moderate and 10 with large DIF. Thirteen items showed DIF in favor of women and 15 in favor of men, indicating a relatively balanced direction of gender-related DIF. Regarding age, 14 items showed DIF, with 9 items favoring the older group (25–60 years) and 5 favoring the younger group (18–25 years). Approximately 75% of the items were free from age-related DIF, indicating relatively fair functioning across age groups. These findings suggest that age and gender should be considered when interpreting and scoring the depression component of the MMPI-2, particularly when individual items or Harris–Lingoes subscales are interpreted separately. Differences in MMPI-2 item functioning reported in previous studies may partly be related to the use of Harris–Lingoes subscales (Gori & Marin, 2015; McNulty et al., 2003; Childs et al., 2000).

The comparison of DIF detection methods showed that the Mantel–Haenszel method performed well in identifying items with moderate or large DIF and produced findings largely consistent with those of the Rasch model. All items identified by the Rasch model were also identified by the Mantel–Haenszel method, except for items 165 and 233. Thus, the Mantel–Haenszel method showed relatively comparable performance to the Rasch model in detecting DIF, consistent with previous findings (Karami & Khodi, 2021; Wilson & Hoskens, 2005). The Rasch model also provides an appropriate approach for transforming nonlinear raw rank data into interval measures, thereby addressing the issue of specific objectivity or invariance of measurement scales in psychological and educational assessment (Minaei, 2015; Tennant & Conaghan, 2007). However, the generalizability of these findings should be considered cautiously because the depression component of the MMPI-2 represents a unidimensional construct and the analysis was conducted based on the assumptions of Item Response Theory.

From a practical perspective, psychologists and psychometricians involved in the psychological assessment of recruitment candidates should consider age and gender when interpreting MMPI-2 depression scores. Future studies should examine DIF in other MMPI-2 components, use additional DIF detection methods, and investigate age- and gender-specific cut-off points to contribute to fairer interpretation of MMPI-2 scores in recruitment assessments.

Keywords: Depression, Differential Item Functioning, Mantel-Haensze Method, Rasch Model, Recruitment Candidates.

Ethical Considerations

Confidentiality of the participants' identities, personal information, and test results was maintained throughout the study. The results obtained from the participants' tests were used solely for the purposes of this research and were not used against participants in the recruitment or employment process. The data were obtained and used with the knowledge and coordination of the relevant occupational medicine centers.

Acknowledgments and Funding

This article is based on a doctoral dissertation in Educational Measurement at Allameh Tabataba'i University, entitled "Application of the Rasch Model and Its Comparison with Classical Measurement Theory in Increasing the Accuracy and Validity of the MMPI-2 Questionnaire in the Job Recruitment Process." The authors received no financial or non-financial support from any organization or company for conducting or preparing this research.

Conflict of Interest

The authors declare that there is no conflict of interest with respect to this article, its publication, or any organization.

Cite this article: Rahmani Malek Abad, M., Falsafinejad, M. R., Minaei, A., & Farrokhi, N. A. (2026). Differential Item Functioning of Minnesota Multiphasic Personality Inventory-2 (Depression) Items in Recruitment Candidates: A Comparison Between the Rasch Model and Mantel-Haenszel. *Journal of Applied Psychological Research*, 17(3), 323-338. <https://doi.org/10.22059/japr.2024.363263.644703>



© Author(s) retain the copyright.

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

Publisher: University of Tehran Press.



بررسی کنش افتراقی سؤالات پرسشنامه شخصیت چندوجهی مینه‌سوتا-۲ (مؤلفه افسردگی) در بین داوطلبان استخدامی: مقایسه مدل راش و منتل‌هنزل^۱

مهدی رحمانی ملک‌آباد^۱✉، محمدرضا فلسفی‌نژاد^{۱*}✉، اصغر مینائی^۱✉، نورعلی فرخی^۱✉

۱. گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران.

*نویسنده مسئول: دانشیار، گروه سنجش و اندازه‌گیری، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه علامه طباطبائی، تهران، ایران.
رایانامه: falsafinejad@atu.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: پژوهشی تاریخ‌های مقاله: تاریخ دریافت: ۱۴۰۲/۰۵/۱۷ تاریخ بازنگری: ۱۴۰۲/۰۸/۰۵ تاریخ پذیرش: ۱۴۰۲/۰۸/۲۸ تاریخ انتشار: ۱۴۰۵/۰۷/۱۰ کلیدواژه‌ها: افسردگی، داوطلبان استخدامی، روش منتل‌هنزل، کنش افتراقی، مدل راش.	هدف پژوهش حاضر، شناسایی سؤالات مؤلفه افسردگی پرسشنامه شخصیت چندوجهی مینه‌سوتا-۲ بود که برحسب جنسیت و سن در میان داوطلبان استخدامی به‌سمت گروه خاصی تورش دارند. روش پژوهش پیمایشی و جامعه آماری کلیه داوطلبان استخدامی شهر اصفهان در سال ۱۴۰۱ و بهار ۱۴۰۲ بودند. حجم نمونه ۵۹۹۷ نفر (۵۷/۳۱ درصد مرد و ۴۲/۶۹ درصد زن) بود که به روش خوشه‌ای یک‌مرحله‌ای انتخاب شدند. ابزار مورد استفاده پرسشنامه MMPI-2 با ۳۷۰ سؤال بود. برای تجزیه و تحلیل اطلاعات از نرم‌افزار Winsteps نسخه ۳.۶۴ بهره گرفته شد. برازش سؤالات با مدل با استفاده از شاخص‌های outfit MnSq و infit MnSq تأیید شد و کنش افتراقی سؤالات برحسب جنسیت و سن به دو روش منتل‌هنزل و مدل راش با نرم‌افزار Winsteps نسخه ۳.۶۴ بررسی و مقایسه شد. پس از تأیید برازش داده‌ها با مدل، تحلیل کنش افتراقی انجام گرفت. نتایج نشان داد بیش از نیمی سؤالات (۶۴/۹ درصد) در عملکرد خود دارای تورش به ضرر زنان و جوانان بوده‌اند. روش منتل‌هنزل نیز به‌جز اختلاف در دو سؤال، یافته مشابهی را نسبت به مدل راش ارائه داد. به‌کارگیری سؤالات انتخابی یا استفاده از خرده‌مقیاس‌های هریس-لینگز موجب تورش شدید سؤالات به جهت زنان و افراد جوان شد، اما استفاده از سؤالات مؤلفه افسردگی به‌صورت یک‌جا می‌تواند موجب عدم تورش شود. همچنین مدل راش برتری نسبی به روش منتل‌هنزل در تشخیص DIF داشته است.

استناد: رحمانی ملک‌آباد، م.، فلسفی‌نژاد، م.، مینائی، ا.، و فرخی، ن. ع. (۱۴۰۵). بررسی کنش افتراقی سؤالات پرسشنامه شخصیت چندوجهی مینه‌سوتا-۲ (مؤلفه افسردگی) در بین داوطلبان استخدامی: مقایسه مدل راش و منتل‌هنزل. فصل‌نامه پژوهش‌های کاربردی روانشناختی، ۱۷(۳)، ۳۳۳-۳۳۸.

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

ناشر: انتشارات دانشگاه تهران

© نویسندگان.

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



۱. مقدمه

گزینش و استخدام نیروی کارآمد، مؤثر و سالم یکی از مهم‌ترین نیازهای یک سازمان و مؤثر بر عملکرد و بازده کاری^۱ فرد و سازمان است. به همین سبب همواره ابتدا سلامت و توانمندی‌های داوطلبان مشاغل سنجش می‌شود و پس از آن افراد به خدمت گرفته می‌شوند. به‌منظور سنجش سلامت جسمی و روانی داوطلبان و به‌کارگیری آنان در جایگاه متناسب با توانمندی‌هایشان، معایناتی با عنوان معاینات بدو استخدام صورت می‌پذیرد (سلطانی شال و همکاران، ۱۳۹۹). براساس دستورالعمل سنجش و ارزیابی سلامت روان کارکنان (۱۳۹۶) نتایج معاینات بدو استخدام به‌عنوان سندی قانونی برای ثبت وضعیت سلامت فرد در زمان استخدام، تشخیص بیماری‌های مؤثر در محیط کار و رد یا تأیید افراد به‌کار می‌رود. در این سند، سلامت افراد به معنای نداشتن اختلال روانی و جسمی درنظر گرفته شده است. یکی از مهم‌ترین شاخص‌های سلامت روان داوطلبان در بدو استخدام، تشخیص نداشتن افسردگی^۲ است که به‌وسیله پرسشنامه چندوجهی مینه‌سوتا^۳ ویرایش دوم (MMPI-2) ارزیابی می‌شود. پرسشنامه شخصیت چندوجهی مینه‌سوتا-۲ از مهم‌ترین و معتبرترین پرسشنامه‌ها در زمینه بررسی شخصیت و برخی اختلالات روانی است (چری^۴، ۲۰۲۲). از طرفی افسردگی یکی از شایع‌ترین اختلال‌های روانی و در حال حاضر چهارمین بیماری شایع در جهان محسوب می‌شود و یکی از مسائل اساسی بهداشت همگانی است که مانند مشکلات مزمن دیگر، حالات عود و بازگشت دارد (اون ولی و همکاران^۵، ۲۰۱۵) و عملکرد اجتماعی و شغلی را تضعیف می‌کند؛ به‌نحوی که زمینه‌ساز اختلالات دیگر شغلی مانند استرس و اضطراب شغلی نیز می‌شود. از آنجا که افسردگی می‌تواند ریشه و یکی از ویژگی‌های هر اختلال خلقی و روانی باشد (کاپلان و سادوک^۶، ۲۰۰۹ به نقل از بولاند و همکاران^۷، ۲۰۲۱)، این پژوهش فقط به بررسی مؤلفه افسردگی می‌پردازد.

در کنار داشتن ابزاری استاندارد، یکی دیگر از مهم‌ترین مسائل در سنجش داوطلبان استخدام، ارزیابی صحیح و دقیق، داشتن انصاف در گزینش و عدم سوگیری سؤالات در اعلام نتایج است. گاهی برخی سؤالات در تعیین سطح و مشخص کردن صفتی در افراد یک گروه نسبت به گروهی دیگر با همان سطح از صفت دچار سوگیری می‌شوند و نتایج را به‌صورت جهت‌دار گزارش می‌کنند که این امر سبب تبعیض در استخدام و حتی گزینش اشتباه افراد می‌شود. این امر سبب می‌شود تا افرادی که سالم هستند در مرحله ارزیابی رد شوند و افرادی که دارای افسردگی یا بیمار هستند تأیید و استخدام شوند. درنتیجه این سوگیری، سازمان‌ها دچار مشکل و آسیب و حتی گاهی به آزمون‌های روان‌شناختی بدبین می‌شوند. برای مشخص کردن سوگیری سؤالات پرسشنامه چندوجهی مینه‌سوتا باید کنش افتراقی سؤالات را بررسی کرد تا مشخص شود چه سؤالاتی در تشخیص صفت دارای سوگیری است و به چه سمت جهت دارد و کدام سؤال بدون تورش به ارزیابی صفات روان‌شناختی داوطلبان استخدام می‌پردازد.

یکی از مسائلی که موجب تبعیض در انتخاب می‌شود، تبعیض‌های جنسیتی است. بکر و همکاران^۸ (۲۰۲۳) تفاوت‌های جنسیتی در نحوه پاسخگویی افراد به سؤالات آزمون‌های روان‌شناختی را یکی از مسائل مهم تأثیرگذار در ارزیابی‌ها عنوان کرده‌اند. این تأثیرگذاری تا حدی است که پزشکان و محققان پس از انتشار نسخه اول پرسشنامه چندوجهی مینه‌سوتا شروع به زیرسؤال بردن صحت این پرسشنامه کردند. منتقدان اشاره کردند که نشانه‌هایی از سوگیری در این پرسشنامه وجود دارد. گروهی نیز احساس کردند که پرسشنامه حاوی سؤالات جنسیتی و نژادپرستانه است که می‌تواند موجب عدم انصاف در نمره‌گذاری و اعلام نادرست نتایج شود (تالاریکو و همکاران^۹، ۲۰۲۳). تبعیض جنسیتی یکی از گسترده‌ترین شکل تبعیض‌ها است؛ زیرا زنان در

1. work efficiency

2. depression

3. Minnesota Multiphasic Personality Inventory

4. Cherry

5. Avenevoli et al.

6. Kaplan & Sadock

7. Boland et al.

8. Baker et al.

9. Talerico et al.

هر سنی، اگر رقیبشان مرد باشد، شانس بسیار کمتری برای دریافت شغل دارند، حتی اگر مردان دارای صلاحیت کمتر یا بی‌تجربه باشند. اگر در کنار تبعیض‌های جامعه، سؤالات نیز دارای تبعیض و سوگیری باشند، اطمینان کردن به نتایج بسیار دشوار می‌شود و افراد و سازمان‌ها متضرر می‌شوند؛ بنابراین تا حد امکان باید سؤالات دارای تورش را شناسایی کرد و راهکارهایی را در جهت رفع سوگیری سؤالات به کار گرفت.

در کنار جنسیت، تعصبات تبعیض‌آمیزی نیز در مورد سن نیز دیده شده که در هردو مورد، به افراد در جهت حذف از انتخاب منصفانه فشار آورده می‌شود (بوربای و خامزینا^۱، ۲۰۱۹). سن افراد یکی از عواملی است که می‌تواند موجب تغییر در نگرش یا نوع پاسخ‌دهی افراد شود؛ به‌نوعی که افراد در سنین بالا ممکن است برحسب نگرش و جهان‌بینی‌شان دارای رفتارها و پاسخ‌هایی باشد که نزدیک به علائم افسردگی باشد، اما درواقع افسرده نیستند؛ یا افراد در سنین کم نشانه‌هایی از شیدایی را نشان می‌دهند، ولی درواقع شیدا نیستند. به این ترتیب سن افراد می‌تواند در نحوه پاسخ‌دهی آنان مؤثر باشد و این فراوانی در افراد کمتر از ۲۵ سال و بالای ۲۵ سال نشان داده شده است (کاپلان و سادوک، ۲۰۰۹ به نقل از بولاند و همکاران، ۲۰۲۲)؛ بنابراین جنسیت و سن در مؤلفه‌های روان‌شناختی و به‌تبع آن عملکرد شغلی مؤثر است و تبعیض جنسیتی و سنی می‌تواند به صورت‌های مختلف در آموزش، اشتغال، شرایط اجتماعی و توانمندسازی افراد بروز یابد (زاهد و همکاران، ۱۳۹۹). از طرفی این تفاوت می‌تواند عاملی مؤثر بر نحوه پاسخ‌دهی باشد، اما تاکنون تأثیر کنش افتراقی سؤالات در انصاف و دقت‌گزینی برای داوطلبان استخدام بررسی نشده است. به‌علاوه، یکی از موانع و تهدیدات اصلی علیه اعتبار^۲ آزمون، وجود واریانس برگرفته از عوامل بی‌ربط و سازه‌های نامرتبط^۳ است. به این معنا که عملکرد فرد در آزمون نباید تحت تأثیر عواملی قرار گیرد که متفاوت با سازه مورد نظر آزمون است. در غیر این صورت، آزمون دارای کنش افتراقی^۴ (DIF) خواهد بود (کرمی^۵، ۲۰۱۳).

یکی از راه‌ها برای پی‌بردن به غیرمنصفانه‌بودن آزمون، کشف پرسش‌های دارای کنش افتراقی است. دورانس و هالند^۶ (۱۹۹۳) به نقل از وایبرگ^۷، ۲۰۰۷) کنش افتراقی را چنین تعریف می‌کنند: «منظور از کنش افتراقی (DIF)، تفاوت‌ها در عملکرد هر پرسش از گروه‌ها با توجه به توانایی یا ویژگی‌ای است که پرسش‌ها از افراد گروه به‌طور مفهومی می‌خواهند. کنش افتراقی تفاوتی ناخواسته بین گروه‌های امتحان‌دهنده است که قرار است براساس تفاوت در توانایی‌هایشان در هر پرسش و هر آزمون قیاس شوند.» کمیلی و گنکدون^۸ (۱۹۹۹) کنش افتراقی را در قالب مفهوم بی‌طرفی آزمون^۹ توضیح داده‌اند. روش‌های آماری آشکارسازی کنش افتراقی (DIF) را می‌توان در دو دسته کلی قرار داد. ماهیت دسته اول برگرفته از نظریه کلاسیک آزمون^{۱۰} (CTT) است. نمرات در این روش به‌عنوان متغیر هم‌تا برای پیش‌بینی نمرات واقعی در نظر گرفته می‌شوند (لرد و نایک^{۱۱}، ۱۹۶۸). از میان روش‌هایی که در این دسته قرار می‌گیرند، می‌توان از روش منتل‌هنزل^{۱۲} (MH)، روش منتل‌هنزل تعمیم‌یافته^{۱۳} (GMH) (وانگ^{۱۴}، ۲۰۰۳)، روش اختلاف میانگین استاندارد شده^{۱۵} و روش رگرسیون لجستیک (زومبو^{۱۶}، ۲۰۰۳).

1. Buribayev & Khamzina
2. Reliability
3. factors irrelevant-construct
4. differential item functioning
5. Karami
6. Dorans & Holland
7. Wiberg
8. Camilli & Congdon
9. test fairness
10. classic test theory
11. Lord & Novick
12. Mantel-Haenszel
13. generalized Mantel-Haenszel
14. Wang
15. standadized mean difference
16. Zumbo

دسته دوم شامل روش‌های مبتنی بر نظریه سؤال-پاسخ^۱ است؛ زیرا نمره متغیر مکنون به‌عنوان متغیر هم‌تا برای برآورد نمرات متغیر مکنون مورد استفاده قرار می‌گیرد. روش‌های این دسته شامل روش‌های مجذور کای-دو لرد^۲، راجو و روش‌های آزمون نسبت درست‌نمایی^۳ است (امبرتسون و ریس^۴، ۱۳۸۸؛ کرمی، گرامی‌پور و مینائی، ۱۴۰۰). روش‌های بررسی کنش افتراقی مبتنی بر نمرات مشاهده‌شده و متغیرهای مکنون می‌توانند در چارچوب مدل‌های مختلط خطی و غیرخطی تعمیم‌یافته ترکیب شوند و مدل‌های سلسله‌مراتبی^۵ نظریه سؤال-پاسخ و رگرسیون لجستیک^۶ را به‌وجود بیاورند. در این مدل‌ها اندازه اثر کنش افتراقی (DIF) و جایگزینی داده‌های گمشده امکان‌پذیر است (اسکروندال و رابه-هیسکت^۷، ۲۰۰۴). از آنجا که روش متل‌هنزل از نظر کاربرد و اعتبار بالاتر و پرکاربردتر از روش‌های دیگر کلاسیک بود و همچنین مدل راش با توجه به ساختار دوارزشی مدل‌های پارامتری دارای برازش مناسب سؤالات و اعتبار بالای شاخص‌ها است، در این پژوهش، این دو روش مقایسه می‌شوند. گوری و مارین^۸ (۲۰۱۵) در پژوهشی به بررسی کنش افتراقی سؤالات اسکیزوفرنی، اضطراب و پرخاشگری پرسشنامه MMPI-2 در میان دانشجویان تازه‌وارد دانشگاه با مدل راش پرداختند و نشان دادند برخی تفاوت‌ها در نحوه برازش و تشخیص سؤالات وجود دارد. مک‌نالتی و همکاران^۹ (۲۰۰۳) به‌منظور بررسی شاخص‌های روایی مقیاس MMPI-2 در نمونه‌ای مشخص‌شده از بین اتباع خارجی زندانی در ایالت‌های امریکا را با استفاده از نظریه کلاسیک مطالعه کردند. نتایج نشان داد زنان بیشتر از مردان و قفقازی‌ها بیشتر از آمریکایی‌های آفریقایی‌تبار محتوای جعلی و غیرمعتبر تولید کرده‌اند. چیلدز و همکاران^{۱۰} (۲۰۰۰) ارزیابی شخصیت را با استفاده از نظریه سؤال-پاسخ و مقیاس افسردگی را در آزمون MMPI-2 مطالعه کردند. در این تحقیق از مدل نظریه سؤال-پاسخ دویارامتری استفاده شد. نتایج نشان داد مقیاس افسردگی دارای سؤالات بسیار تبعیض‌آمیز است. با توجه به اهمیت رعایت انصاف در گزینش نیروی انسانی سالم و مناسب در آزمون‌های بدو ورود و انتخاب بدون سوگیری افراد، سؤال اول این پژوهش این است که آیا سؤالات مؤلفه افسردگی پرسشنامه MMPI-2 با مدل راش برازش دارد. دوم اینکه کدام سؤالات مؤلفه افسردگی با مدل راش و روش متل‌هنزل دارای کنش افتراقی هستند و سوم اینکه آیا تفاوتی در تشخیص سؤالات دارای سوگیری بین دو روش متل‌هنزل و مدل راش وجود دارد.

۲. روش

۲-۱. جامعه، نمونه و روش اجرا

پژوهش از نوع کاربردی و همچنین به‌دلیل استفاده از ابزار، کمی است. به‌علاوه در گروه تحقیقات توصیفی و پیمایشی نیز قرار می‌گیرد. جامعه آماری پژوهش شامل کلیه داوطلبان استخدامی متقاضی شغل معلمی (یا دانشجو-معلمی) شهر اصفهان به تعداد ۱۰ هزار نفر هستند که در آزمون‌های سال ۱۴۰۱ و بهار ۱۴۰۲ به مراکز طب کار در شهر اصفهان مراجعه کرده بودند. روش نمونه‌گیری در این پژوهش، خوشه‌ای یک‌مرحله‌ای بود. با توجه به شیوه تجزیه و تحلیل داده‌ها و استفاده از مدل راش، حجم نمونه بیشتر از ۴۰۰۰ نفر در نظر گرفته شد تا نتایج قابل‌اعتمادتر باشد (شاین و آنیل^{۱۱}، ۲۰۱۷). با توجه به نمونه‌گیری خوشه‌ای، ابتدا ۶۲۱۳ نفر از میان سه مرکز طب کار انتخاب شدند که پس از حذف پرسشنامه‌های ناقص، ۵۹۹۷ نفر در تحلیل قرار گرفتند. معیارهای ورود به پژوهش عبارت‌اند از: فرد داوطلب استخدام در آموزش و پرورش باشد، مراحل غربالگری طب کار را تا مرحله ارزیابی روان‌شناختی سپری کرده باشد و به پرسشنامه با دقت و صحت پاسخ داده باشد. معیارهای خروج نیز عبارت‌اند از: بیش از ۳۵ سؤال پرسشنامه بدون پاسخ باشد، فرد دارای پرونده درمانی فعال باشد و شاخص‌های روایی آزمون نامعتبر باشد.

1. item response theory
2. Lord's chi-square
3. likelihood ratio test
4. Embretson & Reise
5. hierarchical models
6. logistic regression
7. Skron dal & Rabe-Hesketh
8. Gori & Marin
9. McNulty et al.
10. Childs et al.
11. Şahin & Aml

هدف این پژوهش بررسی و مقایسه کنش افتراقی سؤالات مؤلفه افسردگی پرسشنامه چندوجهی شخصیتی مینه‌سوتا به روش متل‌هنزل و مدل راش برای داوطلبان استخدامی بود. ابتدا فهرست مراکز طب کار شهر اصفهان استخراج شدند. از بین ۱۷ مرکز ۹ مرکز برای همکاری اعلام آمادگی کردند و از این بین، سه مرکز به تصادف و با توجه به حجم مراجعان استخدامی انتخاب شدند. پس از صحبت با مسئول مراکز مقرر شد با حفظ محرمانه‌بودن اسامی داوطلبان و محرمانه‌ماندن نتایج آزمون‌های هر فرد، آزمون MMPI-2 در مرکز گرفته شود و نتایج به‌صورت هفتگی در اختیار دستیار پژوهشگر قرار گیرد. دستیار پژوهشگر نیز پاسخنامه‌های افراد را وارد نرم‌افزار کرد و پس از تجمیع در اختیار پژوهشگر قرار داد.

۲-۲. ابزار پژوهش

۲-۲-۱. پرسشنامه چندوجهی شخصیتی مینه‌سوتا-۲ (MMPI-2)

این پرسشنامه توسط [هاتاوی و مک‌کینلی^۲ \(۱۹۴۳\)](#) ساخته شده و در سال ۱۹۷۰ و در نهایت در سال ۲۰۰۸ دانشگاه مینه‌سوتا آخرین بازنگری آن را انجام داد و منتشر کرد. ابزار پژوهش نسخه MMPI-2 است که براساس ویرایش سال ۲۰۰۸ توسط هاتاوی و مک‌کینلی و با سرپرستی دانشگاه مینه‌سوتا دارای ۳۷۰ ماده است ([گروث-مارنات و رایت^۳، ۲۰۱۶](#)). این مقیاس سه شاخص روایی دروغ^۴ (L)، غیرمعمولی^۵ (IF) و انکار^۶ (K) و ده شاخص بالینی هیپوکندریا^۷ (HS)، افسردگی^۸ (D)، هسیتی^۹ (Hy)، انحراف روانی-اجتماعی^{۱۰} (Pd)، زنانگی-مردانگی^{۱۱} (MF)، پارانویا^{۱۲} (Pa)، ضعف روانی^{۱۳} (Pt)، اسکیزوفرنی^{۱۴} (Sc)، هیپومانیا^{۱۵} (Ma) و درون‌گرایی اجتماعی^{۱۶} (Si) را ارزیابی می‌کند. در این پژوهش با توجه به حجم سؤالات و عدم امکان بررسی تمام مؤلفه‌ها، فقط به مؤلفه افسردگی از پرسشنامه MMPI-2 با ۵۷ سؤال پرداخته شد. پاسخ به این ماده‌ها به‌صورت دوگزینه‌ای بلی/درست (۱) و خیر/غلط (صفر) است. از میان ۵۷ سؤال، ۲۰ سؤال به‌صورت مستقیم و ۳۷ سؤال به‌صورت معکوس نمره‌گذاری شدند (بلی= صفر و خیر= ۱). روایی این مقیاس در مطالعات متعدد با استفاده از تحلیل عاملی متمایل و چرخش ابلیمین و همچنین تشخیص افتراقی روان‌شناسان بالینی بررسی و تأیید شد و نشان داده شد که مقیاس افسردگی با افرادی که دارای نشانگان بالینی افسردگی دارند همخوان و دارای روایی قابل قبول است ([گروث-ارنات و رایت^۳، ۲۰۱۶](#)). در مطالعه [دیسکویتس و همکاران^{۱۷} \(۲۰۱۶\)](#) اعتبار مقیاس‌های آزمون اصلی با روش دونیمه‌کردن^{۱۸} از ۵۱/۰ به بالا بوده است که بعضی از این ضرایب برابر با ۹۵/۰ به‌دست آمد. ضرایب همبستگی مقیاس‌ها از راه بازآزمایی نیز از ۵۰/۰ تا ۹۰/۰ بود. [ولیانپور و همکاران^{۱۹} \(۲۰۲۰\)](#) در مطالعه‌ای که روی جامعه سالم و جامعه بالینی ایرانی انجام دادند، اعتبار مقیاس افسردگی پرسشنامه MMPI-2 را به روش آلفای کرونباخ^{۲۰} ۸۰/۰ و ضریب اعتبار مقیاس از راه بازآزمایی را ۸۲/۰ گزارش کردند. اعتبار این مقیاس در این پژوهش به روش آلفای کرونباخ ۷۸/۰ و به روش دونیمه‌کردن ۷۳/۰ به‌دست آمد. این سطح از اعتبار در مقایسه با منابع اولیه کمی پایین‌تر است

1. Minnesota Multiphasic Personality Inventory-2 (MMPI-2)

2. Hathaway & Mckinely

3. Groth-Marnat & Wright

4. Lie (L)

5. Infrequency (F)

6. Defensiveness (K)

7. Hypochondriasis (Hs)

8. Depression (D)

9. Hysteria (Hy)

10. Psychopathic 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 et al.

20. cronbach alpha reliability

که علت آن را می‌توان کم‌بودن واریانس صفت در بین اعضای جامعه دانست؛ زیرا افراد برای گزینش‌شدن در فرایندهای اولیهٔ آزمون‌های استخدامی فرهنگیان انتخاب شده‌اند و همگن هستند. باین‌حال این ضرایب اعتبار در سطح مطلوبی قرار دارد.

۲-۳. روش تجزیه و تحلیل داده‌ها

برای محاسبهٔ آماره‌های توصیفی میانگین و انحراف استاندارد و ضریب آلفای کرونباخ و اعتبار مقیاس از طریق نظریهٔ کلاسیک از نرم‌افزار SPSS نسخهٔ ۲۷ استفاده شد و بررسی برازش سؤال به کمک آمارهٔ Mnsq infit و Mnsq outfit و با استفاده از نرم‌افزار Winsteps نسخهٔ ۳/۶۴ صورت گرفت. ضریب مطلوبیت سؤالات نیز با شاخص Measure بیان شد. در بررسی تناسب شاخص‌ها ابتدا به شاخص Mnsq ^۱ مراجعه می‌شود و در صورت عدم تأیید باید به شاخص ZSTD ^۲ رجوع کرد. مقادیر MNSQ از صفر تا مثبت بی‌نهایت قرار دارد و مقدار مطلوب ۱ است. هرچه این مقدار به صفر نزدیک شود، نشان از بیش‌برازش الگوهای پاسخ با مدل دارد که به معنای وابستگی در میان پاسخ‌ها یا سؤالات است. مقادیر بالاتر از ۱ نیز بیانگر انحراف از تک‌بعدی‌بودن است. در کل مقادیر بیشتر از ۲ غیرمفید و حاکی از اعتبار و برازش اندک است. مقادیر ۱/۵ تا ۲ با اینکه زیاد است، اما گمراه‌کننده نیست و قابل‌اغماض است. مقادیر ۰/۵ تا ۱/۵ برای برازش مقیاس اندازه‌گیری مناسب است و میزان اعتبار مقیاس را افزایش می‌دهد؛ و درنهایت مقادیر کمتر از ۰/۵ برای برازش مقیاس اندازه‌گیری بهره‌پایینی دارد و میزان اعتبار را به‌صورت غیرواقعی بالا می‌برد. مقادیر ZSTD نیز بهتر است بین ۲- تا ۲+ گزارش شود. البته با توجه به اینکه این آماره به حجم نمونه بسیار حساس است (لیناکر^۳، ۲۰۱۲) و حجم نمونهٔ این پژوهش بالا است، در این پژوهش برای بررسی برازش سؤال با مدل، از آمارهٔ MnSq که به دو شکل infit و outfit آمده استفاده شد.

۳. یافته‌ها

۳-۱. توصیف جمعیت‌شناختی

از میان مشارکت‌کنندگان، ۵۷/۳۱ درصد مرد و ۴۲/۶۹ درصد زن بودند. میانگین سنی مشارکت‌کنندگان ۲۸/۹۷ سال و با انحراف استاندارد ۴/۱۵ سال بود. تحصیلات ۳۷/۱۷ درصد افراد دیپلم، ۴۲/۸۷ درصد کارشناسی، ۱۵/۸۱ درصد کارشناسی ارشد و یا دکتری و ۴/۱۵ درصد حوزوی بود.

۳-۲. بررسی برازش مدل با داده‌ها

با توجه به اینکه هدف این پژوهش بررسی کنش افتراقی سؤالات پرسشنامهٔ MMPI-2 در بین داوطلبان استخدامی با مدل راش در کنار روش متل‌هنزل بود، ابتدا تناسب مدل و داده‌ها و مفروضات نظریهٔ راش تبیین شدند، سپس شاخص‌های کمی سنجش سؤالات به‌دست‌آمده ارائه شد. شاخص‌های کمی در جدول ۱ آمده است.

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

شماره سؤال	Measure	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
۴۵	۲/۵۵	۰/۹۹	-۰/۲	۰/۹۳	-۰/۹
۷۵	۲/۴۴	۰/۹۴	-۱/۰	۰/۶۸	-۵/۱
۳۹	۲/۳۸	۰/۹۵	-۰/۹	۰/۷۹	-۳/۳
۱۰	۲/۱۹	۰/۹۶	-۰/۷	۰/۹۱	-۱/۴
۲	۲/۱۴	۰/۹۸	-۰/۴	۰/۸۷	-۲/۳
۱۷۵	۲/۱۲	۰/۹۴	-۱/۱	۰/۷۶	-۴/۴
۱۵	۲/۰۳	۰/۹۶	-۰/۸	۰/۸۸	-۲/۳
۹۲	۲/۰۳	۱/۰۰	-۰/۰	۱/۰۴	۰/۸
۱۰۹	۲/۰۰	۰/۹۳	-۱/۶	۰/۷۶	-۴/۸
۳۱	۱/۹۹	۰/۹۵	-۱/۰	۰/۹۰	-۱/۹

1. mean-square t standardized fit

2. standardized as a z-score

3. Linacre

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

در ستون دوم (Measure) جدول ۱ درجه سختی هر پرسش در واحد لجیته^۱ اشاره دارد که در پرسشنامه‌های شخصیتی به‌عنوان ضریب دشواری یا مطلوبیت معرفی می‌شود (لیناکر، ۲۰۱۲). در این مؤلفه دشوارترین سؤالات ۷۵، ۴۵ و ۳۹ بودند و به‌ترتیب سؤالات ۶۸، ۱۳۴، ۳۷ و ۲۳۸ بیش از همه پاسخ مثبت گرفته‌اند. ستون سوم و چهارم مربوط به شاخص‌های تناسب Infit و ستون پنجم و ششم متعلق به شاخص‌های تناسب Outfit است. با توجه به حجم زیاد نمونه در این پژوهش، برای بررسی برازش سؤال با مدل از آماره MnSq استفاده شد. با توجه به نتایج جدول ۱ می‌توان گفت از میان ۵۷ سؤال مقیاس افسردگی، شاخص‌های برازش ۵۳ سؤال در سطح مناسب (بین ۰/۵ تا ۱/۵) قرار دارد. فقط چهار سؤال (۳۷، ۱۳۴، ۱۴۲ و ۱۷۸) دارای مقدار Outfit MnSq بالاتر از ۱/۵ بوده‌اند و به‌علت اینکه این مقادیر از ۲ کمتر است، همراه‌کننده نیست و قابل‌اغماض است. درمجموع می‌توان گفت شاخص‌های برازش تمامی سؤالات در دامنه مطلوبی قرار دارد و سؤالات برازش خوبی با مدل را دارند.

۳-۳. بررسی کنش افتراقی سؤالات

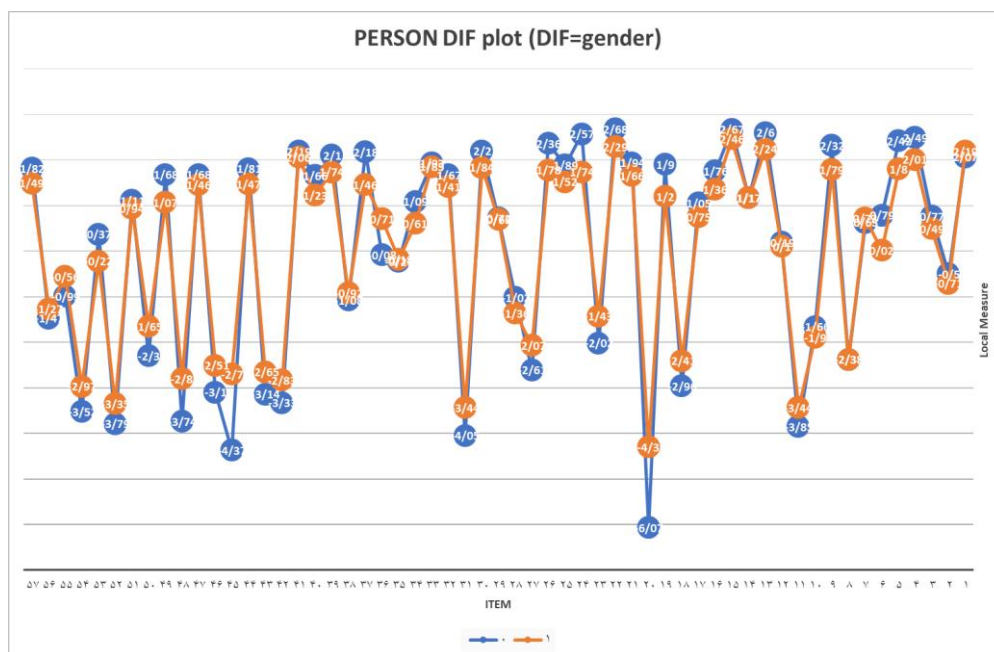
پس از تأیید برازش داده‌ها با مدل، در ادامه برای تشخیص و مقایسه سؤالات دارای کنش افتراقی برحسب جنسیت و سن از دو شیوه متل‌هنزل و مدل راش استفاده شد و نتایج آن در جدول ۲ آمده است. در رویکرد متل‌هنزل مقادیر کنش افتراقی با Size و سطح معناداری آن با Prob. آمده است. پس از آن، نتایج کنش افتراقی مدل راش گزارش 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، نشانه عدم کنش افتراقی (A) است. اگر از ۰/۴۳ تا ۰/۶۴ باشد دارای کنش افتراقی متوسط (B) و اگر بالاتر از ۰/۶۴ باشد، دارای کنش افتراقی بزرگ (C) است (لیناکر، ۲۰۱۲). کنش افتراقی بزرگ گواهی بر ناعادلانه‌بودن سؤال است. به این ترتیب، اگر پاسخ به سؤال‌های آزمون با کنترل توانایی آزمودنی‌ها تابعی از ویژگی‌های آزمودنی‌ها مانند جنسیت و سن باشد، آن سؤال‌ها نسبت به آن افراد دارای کنش افتراقی و سوگیری است.

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

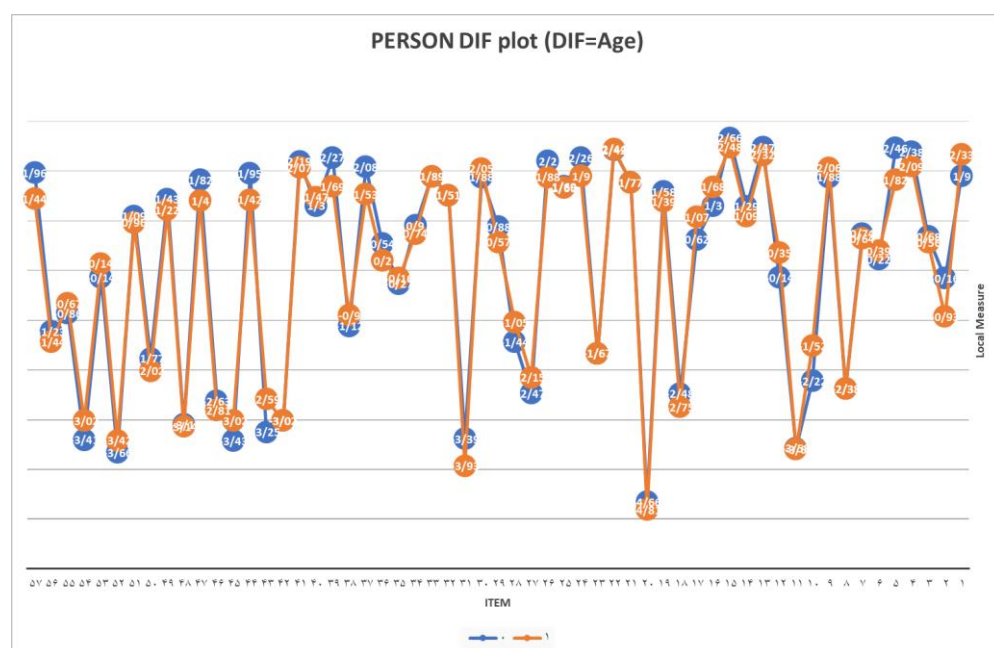
شماره سؤال	متل‌هنزل			مدل راش			جهت تورش
	نتیجه	Prob.	Size	DIF Contrast	DIF Size	Welch Prob.	نتیجه
۲	C	-۰/۰۰۰۰	۰/۷۹	۰/۷۶	۰/۴۵	-۰/۰۰۰۰	C
۹	B	-۰/۰۰۰۲	۰/۳۸	۰/۴۹	۰/۳۰	-۰/۰۰۰۱	B
۱۰	C	-۰/۰۰۰۰	۰/۶۴	۰/۶۲	۰/۴۰	-۰/۰۰۰۰	C
۱۵	C	-۰/۰۰۰۰	۰/۷۵	۰/۷۷	۰/۴۷	-۰/۰۰۰۰	C
۲۹	B	-۰/۰۰۰۰	۰/۵۲	۰/۵۳	۰/۳۳	-۰/۰۰۰۰	B
۳۱	C	-۰/۰۰۰۰	-۰/۶۶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

شماره سؤال	متل‌هنزل			مدل راش			جهت تورش
	نتیجه	Prob.	Size	DIF Contrast	DIF Size	Welch Prob.	نتیجه
۹۲	B	-/.۰۰۰۱	۰/۴۴	۰/۳۸	۰/۲۳	۰/۰۰۰۲	F
۹۵	B	-/.۰۰۰۰	۰/۵۷	۰/۵۸	۰/۳۶	۰/۰۰۰۰	F
۱۰۹	B	-/.۰۰۰۰	-۰/۵۰	-۰/۵۴	-۰/۳۳	۰/۰۰۰۰	M
۱۳۰	B	-/.۰۰۰۰	-۰/۵۰	-۰/۶۲	-۰/۳۹	۰/۰۰۰۰	M / U
۱۴۱	B	-/.۰۰۰۰	۰/۴۴	۰/۴۸	۰/۲۹	۰/۰۰۰۰	F
۱۴۳	C	-/.۰۰۰۰	-۰/۸۵	-۰/۷۹	-۰/۴۱	۰/۰۰۰۰	M
۱۴۶	C	-/.۰۰۰۰	۰/۷۱	۰/۷۲	۰/۴۶	۰/۰۰۰۰	F / U
۱۴۸	B	-/.۰۰۰۰	۰/۴۴	۰/۵۸	۰/۳۸	۰/۰۰۰۰	U
۱۶۵	A	-/.۰۰۰۰	۰/۳۸	۰/۴۳	۰/۲۶	۰/۰۰۰۰	F
۱۷۵	B	-/.۰۰۰۰	-۰/۴۳	-۰/۵۰	-۰/۳۱	۰/۰۰۰۰	M
۱۷۸	B	-/.۰۰۰۰	-۰/۴۴	-۰/۴۹	-۰/۳۰	۰/۰۰۰۰	M / L
۱۸۱	B	-/.۰۰۰۰	۰/۴۹	۰/۵۳	۰/۳۵	۰/۰۰۰۰	U
۱۸۸	C	-/.۰۰۰۰	۱/۵۷	-۱۰/۶۷	-۱/۲۰	۰/۰۰۰۰	M
۱۸۹	B	-/.۰۰۰۰	-۰/۵۷	-۰/۵۸	-۰/۳۷	۰/۰۰۰۰	M
۲۱۲	B	-/.۰۰۰۱	۰/۴۳	۰/۴۳	۰/۲۷	۰/۰۰۰۰	U
۲۱۵	C	-/.۰۰۰۰	-۰/۸۵	-۰/۹۴	-۰/۶۲	۰/۰۰۰۰	M
۲۲۱	B	-/.۰۰۰۰	۰/۵۵	۰/۶۱	۰/۳۸	۰/۰۰۰۰	F
۲۲۳	C	-/.۰۰۰۰	-۰/۶۱	-۰/۶۶	-۰/۴۰	۰/۰۰۰۰	M
۲۳۳	A	-/.۰۰۱۸	-۰/۲۹	-۰/۴۳	-۰/۲۷	۰/۰۰۰۱	M
۲۳۸	B	-/.۰۰۰۰	۰/۵۷	۰/۵۸	۰/۳۵	۰/۰۰۰۰	F
۲۴۵	B	-/.۰۰۰۰	-۰/۴۳	-۰/۵۵	-۰/۳۵	۰/۰۰۰۰	M
۲۴۸	B	-/.۰۰۰۰	-۰/۴۰	-۰/۴۳	-۰/۲۴	۰/۰۰۰۰	M
۲۶۷	B	-/.۰۰۰۰	۰/۴۸	۰/۵۲	۰/۳۴	۰/۰۰۰۰	U
۲۷۸	C	-/.۰۰۰۰	۰/۷۹	۰/۷۶	۰/۴۵	۰/۰۰۰۰	U

M=مردان، F=زنان، U=۲۵ تا ۶۰ سال، L=۱۸ تا ۲۵ سال



شکل ۱. نمودار کنش افتراقی سؤالات افسردگی برحسب ویژگی جنسیت (صفر= زن و ۱= مرد)



شکل ۲. نمودار کنش افتراقی سؤالات افسردگی برحسب ویژگی سن (صفر= ۱۸ تا ۲۵ سال و ۱= ۲۵ تا ۶۰ سال)

با توجه به نتایج جدول ۲ می‌توان گفت از بین ۵۷ سؤال مؤلفه افسردگی پرسشنامه MMPI-2 که روی داوطلبان استخدامی اجرا شده، ۲۸ سؤال دارای کنش افتراقی است و مابقی سؤالات کنشی افتراقی بین زنان و مردان نداشته است. از بین سؤالات دارای کنش افتراقی ۱۸ مورد (سؤالات ۹، ۲۹، ۴۹، ۷۵، ۹۲، ۹۵، ۱۰۹، ۱۳۰، ۱۴۱، ۱۶۵، ۱۷۵، ۱۷۸، ۱۸۹، ۲۲۱، ۲۳۳، ۲۴۵، ۲۳۸ و ۲۴۸) کنش افتراقی متوسط و ۱۰ مورد (سؤالات ۱۰، ۱۵، ۵۵، ۵۶، ۷۶، ۱۴۳، ۱۴۶، ۱۸۸، ۲۱۵ و ۲۲۳) کنش افتراقی بزرگی را برحسب ویژگی جنسیت گزارش داده‌اند. از طرفی با توجه به پیشینه پژوهش‌های انجام‌شده در حوزه افسردگی گفته شده که فراوانی افسردگی در افراد کمتر از ۲۵ سال و بالای ۲۵ سال متفاوت است (کاپلان و سادوک، ۲۰۰۹ به نقل از بولاند و

همکاران، ۲۰۲۲). از این جهت، داوطلبان استخدامی از نظر سنی به طبقات ۱۸ تا ۲۵ سال (L) و بالای ۲۵ سال (U) تقسیم شدند. در بررسی کنش افتراقی برحسب سن نیز از بین سؤالات مؤلفه افسردگی ۹ مورد (سؤالات ۳۷، ۴۵، ۴۶، ۱۳۰، ۱۴۸، ۱۷۸، ۱۸۱، ۲۱۲ و ۲۶۷) دارای کنش افتراقی متوسط و ۵ مورد (سؤالات ۲، ۱۰، ۳۱، ۱۴۶، ۲۷۸) دارای کنش افتراقی بزرگ بودند. در مقایسه بین دو رویکرد متل‌هنزل و مدل راش در بررسی کنش افتراقی نیز می‌توان گفت همه سؤالاتی نیز که در مدل راش دارای کنش افتراقی تشخیص داده شدند، در مدل متل‌هنزل نیز دارای کنش افتراقی تشخیص داده شدند به‌جز سؤال ۱۶۵ و ۲۳۳ که در روش راش دارای کنش افتراقی بود، اما در روش متل‌هنزل شناسایی نشد. با توجه به حجم سؤالات و تعداد نسبتاً یکسان در شناسایی سؤالات دارای کنش افتراقی توسط دو روش، می‌توان نتیجه گرفت که دو روش متل‌هنزل و مدل راش عملکرد یکسانی در شناسایی سؤالات دارای کنش افتراقی داشتند. همچنین این نتیجه نشانه دیگری از برازش مطلوب داده‌ها با مدل راش دارد (لیناکر، ۲۰۱۲). در شکل‌های ۱ و ۲ نتایج کنش افتراقی سؤالات افسردگی برحسب ویژگی جنسیت و سن به‌صورت نمودار آمده است.

در ستون آخر جدول، جهت تورش سؤالات دارای کنش افتراقی نشان داده شده است. اگر به سؤال تورش‌دار، مردان بیشتر جواب مثبت داده باشند، آن سؤال با M نشان داده شده و نشانگر این است که برای مردان دارای کنش افتراقی بیشتری است و اگر در سؤال تورش‌دار، زنان جواب مثبت بیشتر داده باشند، آن سؤال با F نشان داده شده و نشان می‌دهد این سؤال برای زنان کنش افتراقی بیشتری دارد. با توجه به ستون آخر جدول ۲، ۱۳ سؤال دارای تورش به جهت زنان و ۱۵ سؤال دارای تورش به جهت مردان بوده‌اند و سؤالات دارای کنش افتراقی، تقریباً به یک اندازه به جهت دو جنس تورش داشته‌اند. همچنین از این بین، هشت سؤال در جهت اعمال نمره بالا به گروه بالای سنی و پنج سؤال به ضرر گروه پایین سنی تشخیص داده شد.

تأثیر کنش افتراقی بر انصاف در آزمون را می‌توان از منظر تفاوت‌های عملکردی شرکت‌کنندگان در پرسشنامه مشخص کرد. وجود پرسش‌هایی با ویژگی کنش افتراقی به این مفهوم است که سازه‌های غیرمرتبطی به پاسخ افراد تأثیر می‌گذارد و ملاک توانایی شرکت‌کنندگان در آن آزمون نیست. در مقیاس افسردگی پرسشنامه MMPI-2، به دلیل تعداد پرسش‌ها با ویژگی کنش افتراقی می‌توان گفت ۴۹/۱ درصد سؤالات، تأثیرپذیر از عامل نامرتبط جنسیت و ۲۲/۸ درصد سؤالات، متأثر از سن شرکت‌کنندگان بود که در برخی از موارد به نفع آنان و گاهی به ضرر آنان بوده است. به‌طور کلی بخش بسیار اعظمی از سؤالات پرسشنامه حاضر دارای تبعیضی برای شرکت‌کنندگان بوده و انصاف در آزمون را تا حد زیادی نقض کرده است و فقط در ۲۰ سؤال (۶۴/۹ درصد) از نظر تبعیض جنسیتی، انصاف در تشخیص برای مؤلفه افسردگی رعایت شده است. باین‌حال باید در نظر داشت که مفهوم انصاف در آزمون چندبعدی است و وجود یا نبود آن متأثر از عوامل مختلفی است که از زمان طراحی آزمون باید آن‌ها را کنترل کرد و سرانجام در زمان تفسیر نتایج، اثرگذاری آن عوامل را در نظر گرفت.

۴. بحث و نتیجه‌گیری

هدف این پژوهش، بررسی و تبیین قدرت هریک از شیوه‌های سنجش و تشخیص کنش افتراقی سؤالات مؤلفه افسردگی پرسشنامه چندوجهی مینه‌سوتا-۲ براساس مدل راش و رویکرد متل‌هنزل بود که پیش‌تر معرفی شد. هدف دیگر، بررسی اثرگذاری این شیوه‌ها بر روایی و صحت تحلیل‌های آماری انجام‌شده براساس نحوه پاسخگویی داوطلبان استخدامی و با توجه به خصیصه جنسیت و سن بود. براساس نتایج، تقریباً تمامی سؤالات مؤلفه افسردگی با مدل راش یک‌پارامتری دارای برازش مناسب و فقط چهار سؤال از مدل اندکی انحراف داشته‌اند که این انحراف‌ها قابل‌اغماض‌اند. در نتیجه داده‌ها با مدل راش برازش دارند. به این ترتیب نتایج به‌دست‌آمده از کنش‌های افتراقی قابل‌اتکا و قابل‌استفاده است.

از بین ۵۷ سؤال مقیاس افسردگی آزمون MMPI-2 روی داوطلبان استخدامی، ۲۸ سؤال دارای کنش افتراقی بین زنان و مردان بود که از این بین ۱۳ سؤال به ضرر مردان و ۱۵ سؤال به ضرر زنان بوده است. با بررسی سؤالات دارای تورش مشخص شد سؤالات مربوط به سنجش ملال ذهنی (۹، ۱۰، ۱۵، ۸۲ و...) بیشتر دارای کنش افتراقی به ضرر مردان بود و سؤالات سنجش افسردگی ذهنی و آشفته‌گی فکری (۴۹، ۵۶، ۷۵، ۹۵، ۱۳۰ و...) بیشتر به ضرر زنان تورش داشت. به عبارت دقیق‌تر، سؤالات

دارای کنش افتراقی بیشتر به ضرر زنان است؛ زیرا این سؤالات به سنجش حالات هیجانی و عاطفی افسردگی پرداخته است که به‌طور طبیعی زنان در این سؤالات، اقبال بیشتری از خود نشان می‌دهند. باین‌حال اگر با کمی اغماض به سؤالات دارای کنش افتراقی نگاه کنیم، می‌توان گفت سؤالات به یک اندازه به ضرر مردان و زنان دارای تورش است؛ بنابراین اگر کل سؤالات مقیاس افسردگی با هم اجرا شود، تا حدی جلوی تورش را در کل آزمون می‌گیرد، اما اگر به‌صورت گزینشی اجرا شود یا اینکه از خرده‌مقیاس‌های هریس-لینگوز^۱ استفاده شود، امکان تورش شدید در نتایج بیشتر است و نتایج تا حد زیادی ناعادلانه و غیرقابل‌اتکا خواهد بود. این دلیل تفاوت در یافته‌های گوری و مارین (۲۰۱۵)، مک‌نالتی و همکاران (۲۰۰۳) و چیلدز و همکاران (۲۰۰۰) با یافته‌های پژوهش حاضر است؛ چرا که این پژوهش‌ها از خرده‌مقیاس‌های هریس-لینگوز استفاده کرده‌اند. البته در تبیین این نکته که چرا در مقیاس افسردگی سؤالات زیادی دارای کنش افتراقی تشخیص داده شده می‌توان گفت نمونه این پژوهش، داوطلبین استخدامی بودند و این گروه نمونه با توجه به اینکه در حیطه غیربالینی و جامعه نرمال قرار گرفته‌اند، شاخص‌های خصیصه‌ای را بیشتر از شاخص‌های بالینی بروز می‌دهند و به همین سبب تعداد سؤالات دارای کنش افتراقی در این مقیاس بیشتر بوده است. البته سؤالاتی را نیز که دارای کنش افتراقی نبودند، می‌توان منصفانه دانست؛ زیرا گروه سنی یا جنسیت تأثیری در نتایج آن‌ها ندارد و این سؤالات به دور از هرگونه سوگیری، نتایج را گزارش می‌کنند.

همچنین درمورد سؤالات دارای کنش افتراقی براساس سن می‌توان گفت از مجموع ۵۷ سؤال مقیاس افسردگی، فقط ۱۴ سؤال دارای کنش افتراقی تشخیص داده شدند که از این بین، ۹ سؤال دارای کنش افتراقی به نفع افراد بالای ۲۵ سال و ۵ سؤال به نفع افراد ۱۸ تا ۲۵ سال بود. در نتیجه سؤالات بیشتری به ضرر جوانان دارای سوگیری است. از بین این سؤالات، مواردی که به بررسی بی‌ثباتی خلقی در مقیاس افسردگی می‌پردازند (برای مثال سؤالات ۲، ۱۰، ۱۳۰، ۱۴۶ و...)، به ضرر گروه جوان دارای تورش است و سؤالاتی که مربوط به عملکرد جسمی و تمرکزند (برای مثال سؤالات ۳، ۳۷، ۴۵ و...)، به ضرر گروه بزرگ‌تر دارای کنش افتراقی هستند. البته بیشتر سؤالات (۷۵ درصد) مقیاس افسردگی به دور از کنش افتراقی بودند و مشخص شد که به‌طور منصفانه به بررسی مؤلفه افسردگی در بین سنین مختلف پرداخته‌اند. با مقایسه شیوه‌های سنجش کنش افتراقی، در تشخیص پرسش‌هایی با سطح کنش افتراقی متوسط یا بالا (۳۷ سؤال از ۵۷ سؤال) مدل متل‌هنزل کارآمدی مناسبی داشت و مشخص شد مدل متل‌هنزل نسبتاً به اندازه مدل راش در تشخیص کنش افتراقی توانمند است. البته این نکته را باید در نظر داشت که روش متل‌هنزل برای سؤالات با تمام سطوح توانایی، نسبتاً در یک حد توانمند بوده که این نتیجه ناهمسو با یافته کرمی و خودی (۱۳۹۹) است که توضیح داده بودند این روش برای سؤالات بسیار ساده یا بسیار دشوار کارآمدتر از سؤالات متوسط است. شایان ذکر است که آزمون مورد استفاده در پژوهش کرمی و خودی (۱۳۹۹) از نوع شناختی و مربوط به درس زبان بوده که محتوای آن کاملاً متفاوت با آزمون‌های غیرشناختی است.

یافته‌های این پژوهش برای روان‌سنجان و متخصصان طراحی آزمون و تحلیلگران نتایج آزمون، رویکردهای جدید سنجش کنش افتراقی را توصیه می‌کند؛ چرا که مدل‌های راش برتری نسبی خوبی از نظریه کلاسیک اندازه‌گیری دارند (ویلسون و هاسکنس^۲، ۲۰۰۵). به نظر لی و همکاران^۳ (۲۰۱۰) دقت همبستگی‌های بین خرده‌مقیاس‌ها و اعتبار آزمون با استفاده از مدل راش بهبود می‌یابد؛ زیرا این مدل خطای اندازه‌گیری را در برآورد به حساب می‌آورد. مدل راش با فراهم‌آوردن روشی مناسب برای تبدیل غیرخطی داده‌های خام رتبه‌ای به اندازه‌های فاصله‌ای، بر یکی از دغدغه‌ها در ساخت ابزارهای اندازه‌گیری روانی و تربیتی، یعنی دستیابی به عینیت خاص یا تغییرناپذیری مقیاس‌ها، غلبه کرده است (مینائی، ۱۳۹۳؛ تنانت و کانگان^۴، ۲۰۰۷). البته در تعمیم‌پذیری نتایج به‌دست‌آمده از مدل راش باید دقت داشت؛ چرا که مؤلفه افسردگی پرسشنامه چندوجهی شخصیتی مینه‌سوتا دارای سازه‌ای تک‌بعدی^۵ است و مفروضات لازم نظریه سؤال-پاسخ را رعایت کرده است. اما اگر مفروضه عدم وابستگی مکانی^۶

1. Harris-Lingoes

2. Wilson & Hoskens

3. Lee et al.

4. Tennant & Conaghan

5. unidimensional

6. local independence

در سازه های رعایت نشده باشد، ممکن است در سنجش با شیوه های ذکر شده نتایج متفاوتی را به دست آورد. همچنین باید توجه کرد که اگر در مدل های نظریه جدید اندازه گیری فراتر از مدل راش برویم و مدل های دو یا سه پارامتری را استفاده کنیم، احتمال دارد نتایج تغییر کند و حساسیت نظریه سؤال-پاسخ افزایش یابد.

با توجه به نتایج این پژوهش، به روان سنجان و روان شناسانی که برای بررسی سلامت روان داوطلبان استخدام، با مراکز طب کار همکاری دارند و به ویژه مجریان آزمون MMPI-2 توصیه می شود به نتایج کنش های افتراقی سوالات براساس سن و جنسیت توجه داشته باشند و در تفسیر و نمره گذاری به این ویژگی های دموگرافیک داوطلبان توجه بیشتری نشان دهند. در نهایت به پژوهشگران پیشنهاد می شود تا کنش افتراقی مؤلفه های دیگر پرسشنامه چندوجهی مینه سوتا مانند پرخاشگری، اضطراب، پارانوئیا و اسکیزوفرنی را در میان داوطلبان استخدام بررسی کنند و از روش های دیگر بررسی کنش های افتراقی نیز در کنار مدل راش بهره ببرند. به پژوهشگران نیز توصیه می شود به بررسی نقاط برش مؤلفه افسردگی و دگر مؤلفه های این پرسشنامه به تفکیک جنسیت و سن بپردازند تا نهایت انصاف در آزمون رعایت شود. در نهایت اینکه پیشنهاد می شود تا از مسیرهای دیگر کنش افتراقی، سوالات افسردگی بررسی و نتایج آن با نتایج پژوهش حاضر مقایسه شود.

۵. ملاحظات اخلاقی

در این پژوهش نام و مشخصات مشارکت کنندگان و نتایج آزمون ها محرمانه باقی ماند. همچنین نتایج آزمون ها صرفاً در این پژوهش استفاده شد و در فرایند شغلی و استخدامی، به سود یا زیان افراد ترتیب اثر داده نشد. داده ها با اطلاع و هماهنگی با مراکز طب کار استفاده شدند.

۶. سیاست گذاری و حمایت مالی

این مقاله برگرفته از رساله دکتری نویسنده اول در رشته سنجش و اندازه گیری دانشگاه علامه طباطبائی با عنوان «کاربرد مدل راش و مقایسه آن با نظریه کلاسیک اندازه گیری در افزایش دقت و روایی پرسشنامه شخصیت چندوجهی مینه سوتا (MMPI) در فرایند استخدام های شغلی» بوده است و برای نگارش این مقاله، حمایت مالی و معنوی از سازمان یا شرکتی دریافت نشده است.

۷. تعارض منافع

در این مقاله هیچ گونه تعارض منافع وجود ندارد.

منابع

- امبرتسون، س.، و رایس، آس. پی. (۱۳۸۸). *نظریه های جدید روان سنجی برای روان شناسان*. ترجمه حسن پاشا شریفی، ولی اله فرزاد، مجتبی حبیبی عسگرآباد، و بلال ایزانلو. تهران: رشد.
- دستورالعمل سنجش و ارزیابی سلامت روانی کارکنان وظیفه در مراکز آموزشی نیروهای مسلح. (۱۳۹۶). *پایگاه اطلاع رسانی قوانین و مقررات کشور*.
<https://dotic.ir/news/2979>
- زاهد، ز.، نصراللهی، ط.، و مهینی زاده، م. (۱۳۹۹). از تبعیض جنسیتی به سوی برابری و رشد اقتصادی (مطالعه کشورهای در حال توسعه). *زن در توسعه و سیاست*، ۱۸ (۴)، ۶۶۲-۶۴۳.
<https://doi.org/10.22059/jwdp.2021.311205.1007897>
- سلطانی شال، ر.، سعادت بین جواهری، ف.، و زبردست، ع. (۱۳۹۹). بررسی سطح بهزیستی و ویژگی های روان سنجی مقیاس بهزیستی پرستاران در محیط کار. *طب کار*، ۱۲ (۱)، ۶۸-۵۵.
<http://dx.doi.org/10.18502/tkj.v12i1.3659>
- کریمی، ح. ر.، گرامی پور، م.، و مینائی، ا. (۱۴۰۰). نرخ آشکار سازی کنش افتراقی سؤال با استفاده از مدل درخت های راش یک مطالعه شبیه سازی شده و داده های واقعی آزمون های سرنوشت ساز ناچا. *اندازه گیری تربیتی*، ۱۱ (۴۴)، ۳۰-۱.
<https://doi.org/10.22054/jem.2021.61694.2190>

کرمی، ح. و خودی، ع. (۱۳۹۹). بررسی کنش افتراقی پرسش‌ها و عملکرد در آزمون: مقایسه رگرسیون لجستیک، مدل رش و متل-هنزل. پژوهش‌های زبان‌شناختی در زبان‌های خارجی، ۱۰(۴)، ۸۵۳-۸۴۲. <https://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، ۵(۱۸)، ۷۷-۱۱۴. <https://doi.org/10.22054/jem.2015.790>

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. <https://doi.org/10.1016/j.jaac.2014.10.010>
- Baker, C. A., Baum, L. J., Francis, J. P., Shura, R. D., & Ord, A. S. (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. <https://doi.org/10.1037/pro0000517>
- 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. <https://doi.org/10.1177/1358229119846784>
- 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. <https://doi.org/10.3102/10769986024004323>
- Cherry, K. (2022) *The Minnesota Multiphasic Personality Inventory (MMPI)*. Updated on September 02 Medically reviewed by Amy Morin. <https://www.verywellmind.com/what-is-the-minnesota-multiphasic-personality-inventory-2795582>
- 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. *Assessment*, 7(1), 37-54. <https://doi.org/10.1177/107319110000700103>
- Deskovitz, M. A., Weed, N. C., McLaughlan, J. K., & Williams, J. E. (2016). Interpretive reliability of six computer-based test interpretation programs for the Minnesota Multiphasic Personality Inventory-2. *Assessment*, 23(2), 250-261. <https://doi.org/10.1177/1073191115584970>
- Embretson, S. E., & Reise, S. P. (2009). *Item response theory for psychologists*. Translated by Hassan Pasha Sharifi, Vali Allah Farzad, Mojtaba Habibi Asgar Abadi & Balal Ezanlow, Tehran: Roshd (Original work published 2000). (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>
- Groth-Marnat, G. G. & Wright, A. J. (2016). *Handbook of Psychological Assessment*. (6th ed.). Wiley.
- Boland, R. J., Verduin, M. L., & Ruiz, P. (2021). *Kaplan & Sadock's synopsis of psychiatry*, (12th ed). Lippincott Williams & Wilkins.
- 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. <https://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. *Quarterly of Educational Measurement*, 11(44), 1-30. <https://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.
- McNulty, J. L., Forbey, J. D., Graham, J. R., ... & Burlew, A. K. (2003). MMPI-2 validity scale characteristics in a correctional sample. *Assessment*, 10(3), 288-298. <https://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. <https://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. [EJ1130806.pdf](http://ejournal.ekitap.com.tr/index.php/estp/article/view/1130806)
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9780203489437>
- 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. <https://doi.org/10.18502/tkj.v12i1.3659> (In Persian)
- Talerico, G. M., McCallum, J. J., Whitman, M. R., ... & 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. <https://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. <https://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>
- Wang N. (2003). Use of the Rasch IRT model in standard setting: An item- mapping method. *Journal of Educational Measurement*, 40(3), 231-253. <https://doi.org/10.1111/j.1745-3984.2003.tb01106.x>
- Wiberg, M. (2007). *Measuring and detecting differential item functioning in criterion-referenced licensing test: A theoretic comparison of methods*. Umea Universitet. <https://trid.trb.org/View/1152612>
- Wilson, M., Hoskens, M. (2005). Multidimensional Item Responses: Multimethod-Multitrait Perspectives. In R. Maclean, R. Watanabe, R. Bake...& N. Hungi (eds.). *Applied Rasch Measurement: A Book of Exemplars* (pp 287-307). Education in the Asia-Pacific Region: Issues, Concerns and Prospects, vol 4. Springer, Dordrecht. https://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*, 18(4), 662-643. <https://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. <https://doi.org/10.1191/0265532203lt248oa>