20(73), Autumn 2024, 7-38 ISSN: 2538-3183
eISSN: 2538-3191
Article Type: Reserch Article
Received in revised form February 21, 2020
Accepted September 19, 2024
Received Apri 25, 2019
Published Online October 06, 2024

Educational Psychology
https://jep.atu.ac.ir/

ATU PRESS

# Comparison of the Accuracy of Person Fit Statistics Based on the Rasch Model in Identifying Acquiescence Response Styles

**Tayebe Dehghan nayeri** — PhD in Measurement and Measurement, Allameh Tabataba'i University, Tehran, Iran. E-mail: dehniri@gmail.com

**Noorali Farokhi** * — *Corresponding Author*, Professor, Department of Educational Psychology, Allameh Tabatabae'i University, Tehran, Iran n. E-mail: farrokhinoorali@gmail.com

**Ahmad Borjali** — Professor, Department of Clinical Psychology, Allameh Tabatabaei University, Tehran, Iran. E-mail: borjali@atu.ac.ir

**Ebrahim Alizade** — Assistant Professor, Department of Counseling, Shahid Beheshti University, Tehran, Iran. E-mail: ebalizade@gmail.com

## Abstract

The aim of the present study is to evaluate the accuracy of person fit statistics based on the Rasch model in identifying acquiescence response styles. To this end, the acquiescence response style was simulated in four different scenarios, including sample size, test length, the percentage of aberrant items, and the percentage of respondents with aberrant response patterns, with 25 repetitions. The findings of the study indicated that the statistics Up and Wp performed better in identifying the acquiescence response style in polytomous data compared to the statistics lp and lzp. However, the accuracy of these statistics decreases as the percentage of respondents with aberrant response patterns increases. Conversely, the lzp and lp statistics were unable to identify this response style in any scenario. Additionally, the examination of the factors affecting the accuracy of person fit statistics revealed that, in general, the accuracy of Up and Wp statistics increases with an increase in sample size, the number of questions, and the percentage of aberrant items. Therefore, in conditions where the likelihood of acquiescence response style exists, researchers can utilize Wp and Up statistics by considering the sample size and questionnaire length. However, given that these statistics fail to identify the acquiescence response style as the percentage of respondents with aberrant response patterns increases, it is recommended to examine the accuracy of other person fit statistics in identifying acquiescence response styles in polytomous data.

**Keywords:** Person Fit Statistics, aberrant Response Patterns, Rasch Model, Acquiescence Response Style

**Cite this Article:** Dehghan nayeri, T., Farokhi, N., Borjali, A., & Alizade, E. (2024). Comparison of the Accuracy of Person Fit Statistics Based on the Rasch Model in Identifying Acquiescence Response Styles. *Educational Psychology*, *20*(73), 135-152. https://doi.org/10.22054/jep.2025.42356.2678

© 2016 by Allameh Tabataba'i University Press
**Publisher:** Allameh Tabataba'i University Press
**DOI**: https://doi.org/10.22054/jep.2025.42356.2678

## Introduction

Psychological questionnaires play a significant role in the selection of candidates and psychological assessments. Typically, individuals respond to questions in a paper-and-pencil format or, preferably, via computer-based methods, and decisions about individuals are made based on the results of these tests. Therefore, respondents may distort their answers to present a favorable image of themselves or to fake results, deceive (Emons, 2009), display non-cooperation with the interviewer (Woods, Altman, & Tarquini, 2008), experience fatigue during responding, require specific knowledge to answer (Armstrong & Shav, 2009), misunderstand the questionnaire instructions, or due to the disclosure and leakage of questions (Mayer, 2003). Distortion in responding, regardless of the content of the questions (such as acquiescing to the questionnaire items), is termed response style (Schwarz, Weldon, & Groenendijk, 2015; Tipton, 2011).

Acquiescence and non-acquiescence response styles (Schwarz et al., 2015; Cabooter, 2010; Wetters, 2006; Razavi, 2001; Baumgartner & Steenkamp, 2001), extreme response style (Rapp, 2013; Cabooter, 2010; Emons et al., 2008; Wetters, 2006; Razavi, 2001; Baumgartner & Steenkamp, 2001), and middle category response style (Schwarz et al., 2015; Cabooter, 2010; Wetters, 2006; Baumgartner & Steenkamp, 2001; Razavi, 2001) are commonly used in non-cognitive questionnaires.

Aberrant responses, regardless of their cause, must be identified prior to analyzing respondents' scores, as a questionnaire might possess acceptable and satisfactory psychometric characteristics, yet a respondent's score may be invalid due to the presence of aberrant responses, leading to adverse consequences in decision-making (Kanighen et al., 2015; Kriths, 2004).

Aberrant response patterns can be identified based on two approaches: content-based and statistical. The first approach, with a history of nearly 80 years, utilizes content scales to identify response styles. Scales such as K (correction), F (faking), and L (lying) in the Minnesota Multiphasic Personality Inventory (Butcher, Graham, & Ben-Porath, 2001), social desirability scales, and generally, forced-choice items (Chen, Lee, & Yen, 2004) use this approach. Conversely, the statistical approach identifies aberrant response patterns by comparing an individual's response pattern with the expected pattern. Person fit statistics within the framework of item response theory fall

under this approach (Emons, 2008; Mejier & Sijtsma, 2001), estimating the logical consistency of a respondent's response pattern to a questionnaire based on their actual latent trait (De Ayala, 2009; Emons, 2008).

The accuracy of person fit statistics depends on various factors such as sample size, the number of simulated questions, the statistical model chosen for data generation, and values for person and item parameters (Rapp, 2013), the type of parameter estimators, and the type of aberrant responses (De la Torre & Deng, 2008; Steyn, Wallos, Abdos, & Germain, 2011), as well as the percentage of aberrant response patterns (Armstrong & Shav, 2009; Emons, 2009), which must be considered when analyzing person fit statistics.

Various person fit statistics exist within different frameworks of item response theory. For example, the D, lzm, and ECI statistics are used in two- and three-parameter models, and the U, W, l, and lz statistics are used in the Rasch model; however, many person fit statistics, such as lz and U, can also be used in other models of item response theory (Mejier & Sijtsma, 2001). To date, more than thirty-six person fit statistics (Karabatsos, 2003; Mejier & Sijtsma, 2001) have been introduced, mostly studied in dichotomous data (Kanighen et al., 2014; Emons, 2008). In the context of polytomous data, among parametric person fit statistics, the lzp statistic has been most studied (Kanighen et al., 2014; C-You, 2013; Steyn et al., 2011; Emons, 2008), while the accuracy of other parametric statistics like W and U, which have shown satisfactory performance in dichotomous data (Rapp, 2013; Karabatsos, 2003; Glass & Mejier, 2003; Mejier & Sijtsma, 2001), has rarely been studied. Moreover, given that there are more than thirty-six person fit statistics for identifying response styles (Karabatsos, 2003), this approach still requires research to identify specific fit statistics that can be used to detect response styles present in non-cognitive questionnaires.

## Literature Review

A significant application of person fit statistics in the last ten years has been in personality (C-Yun, 2013; Ferrando, 2012; Ferrando & Chico, 2001), attitude (Curtis, 2004), and health outcome assessments (Mejier, Nissen, & Tandreau, 2016; Kanighen et al., 2015; Konard et al., 2010), where researchers aim to identify the most accurate person fit statistic for detecting aberrant responses in each of these fields.

Research on person fit statistics can be categorized into two types. In the first type, researchers seek solely a statistic that can identify aberrant response patterns (Kanighen et al., 2014; Ferrando, 2012; Ferrando & Chico, 2001). In the second type, researchers (Hang, 2012; Steyn et al., 2011; Emons, 2008; Meijer, 2003; Glass & Meijer, 2003; Embretson & Riese, 2000, translated by Sharifi et al., 1384) believe that a particular fit index can be employed to identify each response style. Therefore, the present study attempts to improve measurement at the individual level to select competent and qualified individuals. Among item response theory models, the Rasch model was chosen due to its distinct objectivity and invariance (Embretson & Riese, 2000, translated by Sharifi et al., 1384). This study examined the accuracy and sensitivity of four-person fit statistics based on this model (Wp, Up, lzp, and lp) in identifying the most common response style, acquiescence, in polytomous data. Furthermore, to better conclude, factors such as sample size, questionnaire length, the percentage of aberrant items, and the percentage of respondents with aberrant response patterns were examined.

## Methodology

The accuracy of four person fit statistics ($W^p$, $U^p$, $l^p$, $lz^p$) was compared in identifying acquiescence response styles within the framework of the Rasch model. The Rasch model is one of the most robust models in the family of item response theory. Key characteristics of the Rasch model include specific objectivity (equal measurement units) and invariance. The Rasch model comprises a family of models, including the Partial Credit Model (PCM), the Generalized Partial Credit Model (GPCM), the Rating Scale Model (RSM), and the Simple Logistic Model (SLM). The Partial Credit Model is one of the widely used Rasch models, particularly suitable for analyzing responses to attitudinal and personality scales where respondents rate their opinions or respond to items on a multipoint scale (Embretson & Reise, 2000, translated by Sharifi et al., 1384).

### *Data Simulation Design*

Data simulation was conducted using the WinGen software version 3 (Han, 2007). Initially, the assumptions of unidimensionality and local independence under the Rasch Partial Credit Model (PCM) were selected in the software. The next step involved data simulation according to the specified values for each of the influencing factors (500

and 2000 subjects, 30 and 60 items, 15% and 30% non-uniform items, 15% and 30% respondents with response style) as noted by sources such as Rupp (2013), Armstrong and Shaye (2009), Emons (2008, 2009), Karabatsos (2003), and Ferrando and Chico (2001).

Scores for 2000 subjects were generated with a normal ability distribution in response to a 60-item questionnaire, repeated 25 times. Additionally, the distribution of item locations was uniformly set between -3 and 3. Subsequently, a random sample of 500 from each dataset was selected. Similarly, a shorter 30-item test was randomly sampled from the 60-item test. Responses from 15% and 30% of subjects with low ability θ ($\partial \leq \cdot$) were changed to "Agree" (option 4) for 15% and 30% of items whose location exceeded the subject's ability ($1.5 \geq i(\partial)$), with no change if their response was already option 5.

Four person fit statistics ($W^p$), ($U^p$), ($l^p$), and ($lz^p$) were estimated for each subject using Excel for the first three and the Perfit package (Tendeiro, 2015) in R for ($lz^p$). Using ROC curves (SPSS version 25), it was determined which of the four person fit statistics exhibited greater accuracy and sensitivity in identifying the acquiescence response style under various conditions (sample size, test length, percentage of non-uniform items, and percentage of respondents with response style). The mean detection rate of each person fit statistic across 25 data sets was then reported (Table 1).

To determine the sensitivity and accuracy of person fit statistics in various situations, the area under the ROC curve was used. The area represents the accuracy of the person fit statistic in distinguishing between consistent and inconsistent response vectors. Diagnostically, an area less than 0.69 is considered unacceptable, between 0.70 and 0.79 acceptable, between 0.80 and 0.89 excellent, and above 0.90 outstanding, indicating that the statistic can effectively differentiate between the two groups. The point with the highest sensitivity and specificity was selected as the optimal cutoff point (Myers, Gamst, Guarino, 2016), as shown in the table in the appendix.

## Results

Accuracy in Identifying Acquiescence Response Style: Table 1 illustrates the accuracy of identifying the acquiescence response style using four person fit statistics ($W^p$), ($U^p$), ($lz^p$), and ($l^p$) under different conditions when the percentage of non-uniform items is 15% and 30%.

**Table 1**. Accuracy of response style identification, confirming the aberrant Response Patterns four person fit statistics (25 repetitions)

| **Percentage of questions with aberrant response patterns = 15% (α=0.05)** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Sample size** | **Test length** | **Percentage of people with aberrant response pattern = 15%** | | | | **Percentage of people with aberrant response pattern = 30%** | | | |
| | | **$lz^p$** | **$l^p$** | **$U^p$** | **$W^p$** | **$lz^p$** | **$l^p$** | **$U^p$** | **$W^p$** |
| 500 | 30 | 0.03 | 0.00 | 0.71 | 0.70 | 0.00 | 0.00 | 0.58 | 0.41 |
| | 60 | 0.03 | 0.00 | 0.71 | 0.73 | 0.00 | 0.00 | 0.59 | 0.42 |
| 2000 | 30 | 0.03 | 0.00 | 0.78 | 0.77 | 0.00 | 0.00 | 0.67 | 0.52 |
| | 60 | 0.03 | 0.00 | 0.79 | 0.78 | 0.00 | 0.00 | 0.68 | 0.53 |
| **Percentage of questions with aberrant response patterns = 30% (α=0.05)** | | | | | | | | | |
| **Sample size** | **Test length** | **Percentage of people with aberrant response pattern = 15%** | | | | **Percentage of people with aberrant response pattern = 30%** | | | |
| | | **$lz^p$** | **$l^p$** | **$U^p$** | **$W^p$** | **$lz^p$** | **$l^p$** | **$U^p$** | **$W^p$** |
| 500 | 30 | 0.02 | 0.00 | 0.72 | 0.72 | 0.00 | 0.00 | 0.58 | 0.40 |
| | 60 | 0.01 | 0.00 | 0.74 | 0.75 | 0.00 | 0.00 | 0.57 | 0.43 |
| 2000 | 30 | 0.00 | 0.00 | 0.79 | 0.78 | 0.00 | 0.00 | 0.69 | 0.53 |
| | 60 | 0.03 | 0.00 | 0.80 | 0.79 | 0.02 | 0.02 | 0.68 | 0.52 |

The table details how these statistics perform across various scenarios, including sample size, test length, and the percentage of respondents exhibiting inconsistent response patterns. The findings presented in the table above illustrate the accuracy of identifying the four person fit statistics under different conditions, including sample size, questionnaire length, the percentage of respondents, and the percentage of items reflecting the acquiescence response style. In the scenario where 15% of respondents in a 500-person sample endorsed 15% of the items, the accuracy of the fit statistics for 30-item and 60-item questionnaires was estimated as follows: ($lz^p$: (0.03), $l^p$: (0), $U^p$: (0.71), $W^p$: (0.73)) for the 30-item questionnaire, and ($lz^p$: (0.03), $l^p$: (0), $U^p$: (0.71), $W^p$: (0.70)) for the 60-item questionnaire. In this condition, the Up and $W^p$ statistics showed acceptable identification accuracy, while the ($lz^p$) and ($l^p$) statistics failed to identify the acquiescence response style. In the same context, if the percentage of respondents exhibiting the acquiescence response pattern doubles, the accuracy of the fit statistics for the 30-item and 60-item questionnaires changed to ($lz^p$ (0), $l^p$ (0), $U^p$ (0.59), $W^p$ (0.42)) and ($lz^p$ (0), $l^p$ (0), $U^p$ (0.58), Wp (0.41)), respectively. In this scenario, none of the statistics were able to identify the acquiescence response style.

Now, if the percentage of items exhibiting non-uniform response patterns doubles while 15% of respondents still exhibit such patterns, the identification accuracy of the fit statistics for the 30-item and 60-item questionnaires was estimated as follows: ($lz^p$ (0.12), $l^p$ (0), $U^p$ (0.74), $W^p$ (0.75)) for the 30-item questionnaire, and ($lz^p$ (0.02), $l^p$ (0), $U^p$ (0.72), $W^p$ (0.72)) for the 60-item questionnaire. The Up and $W^p$ statistics demonstrated the highest detection accuracy, while the ($lz^p$) and ($l^p$) statistics exhibited the lowest.

When the sample size increased to 2000 respondents, the accuracy of the fit statistics was examined. When 15% of respondents endorsed 15% of the items, the accuracy of the fit statistics for a 30-item or 60-item questionnaire was (lzp (0.03), $l^p$ (0), $U^p$ (0.79), $W^p$ (0.78)) and ($lz^p$ (0.03), $l^p$ (0), Up (0.78), $W^p$ (0.77)), respectively. The ($lz^p$) and ($l^p$) statistics were unable to identify this response style. Under the same conditions, if the percentage of items doubled, the accuracy of the fit statistics for both the 30-item and 60-item questionnaires changed to ($lz^p$: (0.03), $l^p$: (0), $U^p$: (0.80), $W^p$: (0.79)) and ($lz^p$: (0), $l^p$: (0), $U^p$: (0.79), $W^p$: (0.78)), respectively. In this situation, the accuracy of the ($lz^p$) and ($l^p$) statistics was deemed unacceptable, whereas the Up and $W^p$ statistics displayed acceptable accuracy in identifying the acquiescence response style. Overall, when the percentage of respondents exhibiting acquiescence response style doubled in sample sizes of 500 and 2000, none of the fit statistics were capable of identifying this response style, as they fell below the acceptable threshold of 0.70 as classified by Mirza et al .(2016).

## Discussion

The aim of the present study was to identify the best fit statistic for detecting acquiescence response style in multidimensional data. To this end, four-person fit statistics based on the Rasch model ($W^p$, $U^p$, $lz^p$, and $l^p$) were utilized. To facilitate a more comprehensive analysis, different conditions including sample size, test length, the percentage of respondents exhibiting acquiescence response style, and the percentage of non-uniform items were considered. The results yielded from the simulation data analysis indicated that the accuracy of $U^p$ and $W^p$ statistics improved in identifying acquiescence response style as sample size increased; in contrast, the $l^p$ and $lz^p$ statistics were unaffected by the increase in sample size. Ciu (2013) concluded in examining person fit statistics that as the sample size exceeds 500, the

average identification rate of the Guttman error index increases, which aligns with the findings associated with the $U^p$ and $W^p$ statistics. Statistically, it is expected that in larger samples, the accuracy of comparisons between observed and expected response vectors will increase, leading to a more precise estimation of item parameters; thus, the identification accuracy of fit statistics will also improve with an increasing sample size (Rapp, 2013). Conversely, the ($lz^p$) statistic (Ciu, 2013) and the Breusch-Wallis statistics (BW) (Hang, 2012) demonstrated higher identification accuracy in smaller sample sizes. However, the methodologies utilized in these studies (Ciu, 2013; Hang, 2012) differ from the current study regarding data simulation models, response styles, and the number of questionnaire items. For instance, in Ciu's (2013) study, data was simulated within the framework of item response theory, and the effect of the number of items on the accuracy of person fit statistics was not examined.

The number of questions is another influencing factor in the identification accuracy of fit statistics, with $U^p$ and $W^p$ being more effective in correctly identifying the acquiescence response style as the number of items increased when compared to the other two statistics. The results of previous studies (Kanygin et al., 2014; Stieng et al., 2011; Emons, 2008, 2009; Karabatsos, 2003) support this finding. Woods et al. (2008) also demonstrated that non-uniform responses are not always identifiable and that their detection is more effective in longer scales with highly correlated questions and a relatively varied difficulty range. Scales with lower measurement characteristics and shorter lengths do not show person fit misalignment. Consequently, as the number of items in a questionnaire increases, the range of item placement will also expand, and the likelihood of identifying non-uniform response patterns will be enhanced (Umbertson & Rice, 2000, translated by Sharifi et al., 2005).

Another influential factor on the accuracy of person fit statistics is the percentage of respondents exhibiting an acquiescence response style; an increase in this percentage made it more challenging to identify fit statistics such as $U^p$ and $W^p$. Statistically, as the number of respondents with this response style increases, it becomes more difficult to identify them (Rapp, 2013; Karabatsos, 2003). Contrary to the findings of the current research, Hang (2012) found that the proportion of respondents with a non-uniform response pattern had a limited effect on the accuracy and sensitivity of person fit statistics based on response

theory, as well as on group-based statistics and BW statistics. In that study, three different percentages (10%, 20%, and 30%) of respondents with non-uniform response patterns were considered, and it is possible that increasing the percentage of such respondents may influence the accuracy of fit statistics. Additionally, other conditions like test length and sample size that might affect the accuracy of fit statistics were not examined in this research, which could simultaneously impact the identification accuracy of the statistics.

The impact of the percentage of non-uniformly responding items on the identification accuracy of the four person fit statistics was also investigated. The accuracy of $U^p$ and $W^p$ statistics significantly improved with an increase in the percentage of items reflecting response styles. Statistically, as the percentage of non-uniform items increases, the detection of expected patterns becomes easier; thus, the identification accuracy of person fit statistics will also increase.

## Conclusion

The findings of the current study and the literature review indicate that, in order to assess the accuracy of a person fit statistic, various factors must be taken into consideration, including sample size, test length, the percentage of non-uniform items, and the proportion of individuals exhibiting response styles. In situations where an acquiescence response style is prevalent, such as in employment contexts where candidates may seek to appear favorable, they tend to endorse items that are socially accepted. In this situation, among the four person fit statistics based on the Rasch model, $U^p$ and $W^p$ demonstrated higher identification accuracy in multidimensional data; however, as the percentage of respondents exhibiting this response style increased, these statistics were unable to identify the acquiescent response style. It is therefore suggested that the accuracy of other person fit statistics in identifying acquiescence response style in non-cognitive multidimensional questionnaires should be studied. Moreover, the two statistics ($lz^p$) and ($l^p$), frequently employed in multidimensional data and particularly in two-dimensional data, are deemed unsuitable for identifying acquiescence response style, as they do not have sufficient accuracy for this purpose.

روان‌شناسی تربیتی
https://jep.atu.ac.ir/

دوره ۲۰، شماره ۷۳، پاییز ۱۴۰۳، ۱۰-۱۰
eISSN: 2538-3191 ISSN: 2538-3183
نوع مقاله: مقاله پژوهشی
تاریخ دریافت: ۱۳۹۸/۱۲/۰۲ تاریخ بازنگری: ۱۳۹۸/۰۲/۰۶؛ تاریخ پذیرش: ۱۴۰۳/۰۷/۱۵ تاریخ انتشار: ۱۴۰۳/۰۶/۲۹؛

# مقایسه دقت آماره‌های برازش شخص مبتنی بر مدل راش در شناسایی سبک پاسخ‌دهی تصدیق

**طیبه دهقان‌نیری** دکتری رشته سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: dehniri@gmail.com

**نورعلی فرخی*** *نویسنده مسئول*، استاد گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: dr.delavarali@gmail.com

**احمد برجعلی** استاد گروه روان‌شناسی بالینی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: borjali@atu.ac.ir

**ابراهیم علیزاده** استادیار گروه مشاوره، دانشگاه شهید بهشتی، تهران، ایران ebalizade@gmail.com

## چکیده

هدف پژوهش حاضر، بررسی دقت آماره‌های برازش شخص مبتنی بر مدل راش در شناسایی سبک پاسخ‌دهی تصدیق است. به این منظور، سبک پاسخ‌دهی تصدیق در چهار موقعیت متفاوت شامل حجم نمونه، طول آزمون، درصد سؤال‌های ناهمسان و درصد پاسخ‌دهندگان با الگوی پاسخ ناهمسان با ۲۵ تکرار، شبیه‌سازی شد. یافته‌های پژوهش نشان داد، آماره‌های Up و Wp در مقایسه با آماره‌های lp و lzp در شناسایی سبک پاسخ‌دهی تصدیق در داده‌های چند ارزشی، عملکرد بالاتری دارند؛ اما با افزایش درصد پاسخ‌دهندگان با الگوی پاسخ ناهمسان از دقت شناسایی این آماره‌ها کاسته می‌شود؛ در مقابل، آماره‌هایlzp و lp در هیچ موقعیتی نمی‌توانند این سبک پاسخ‌دهی را شناسایی نمایند. همچنین، بررسی عوامل مؤثر بر دقت آماره‌های برازش شخص نشان داد که به‌طورکلی، دقت آماره‌های Up و Wp با افزایش حجم نمونه، تعداد سؤالات پرسشنامه و درصد سؤال‌های ناهمسان بیشتر می‌شود؛ بنابراین در شرایطی که احتمال سبک پاسخ‌دهی تصدیق وجود دارد، پژوهشگران می‌توانند با در نظر گرفتن حجم نمونه و طول پرسشنامه از آماره‌های Wp و Up استفاده کنند؛ اما با توجه به آنکه این آماره‌ها با افزایش درصد پاسخ‌دهندگان با الگوی پاسخ ناهمسان، قادر به شناسایی سبک پاسخ‌دهی تصدیق نیستند؛ پیشنهاد می‌شود دقت سایر آماره‌های برازش شخص در شناسایی سبک پاسخ‌دهی تصدیق در داده‌های چند ارزشی موردبررسی قرار بگیرد.

**کلیدواژه‌ها:** آماره برازش شخص، الگوهای پاسخ ناهمسان، مدل راش، سبک پاسخ‌دهی تصدیق

**استناد به این مقاله:** دهقان‌نیری، طیبه، فرخی، نورعلی، برجعلی، احمد، و علیزاده، ابراهیم. (۱۴۰۳). مقایسه دقت آماره‌های برازش شخص مبتنی بر مدل راش در شناسایی سبک پاسخ‌دهی تصدیق. *فصلنامه روان‌شناسی تربیتی*، ۲۰(۷۳)، ۱۰-۱۰. https://doi.org/10.22054/jep.2025.42356.2678

© ۲۰۱۶ دانشگاه علامه طباطبائی
**ناشر:** دانشگاه علامه طباطبائی

## مقدمه

پرسشنامه‌های روان‌شناختی در زمینه گزینش داوطلبان و بررسی‌های روان‌شناختی نقش بسزایی دارد. افراد معمولاً به‌صورت مداد- کاغذی و یا ترجیحاً به روش کامپیوتری به سؤالات پاسخ می‌دهند و با توجه به نتایج این آزمون‌ها درباره افراد تصمیم‌گیری می‌شود؛ بنابراین پاسخ‌دهندگان به سؤال‌های یک پرسشنامه ممکن است به‌منظور نشان دادن یک تصویر خوب از خودشان یا تمارض، بی دقّتی (Emons, 2009)، عدم همکاری با پرسشگر (Woods et al., 2008)، خستگی در حین پاسخ‌دهی، نیاز به داشتن دانش خاص برای پاسخ‌دهی (Armstrong & Shi, 2009)، فهم نادرست دستورالعمل پرسشنامه و یا افشا شدن و لو رفتن سؤالات (Meyer, 2003) پاسخ‌هایشان را تحریف کنند. تحریف در پاسخ بدون توجه به محتوای سؤال (مانند تصدیق سؤالات پرسشنامه) سبک پاسخ‌دهی[1] نامیده می‌شود (Tippins, 2011؛ Schoonees et al., 2015).

سبک‌های پاسخ‌دهی تصدیق[2] و عدم تصدیق[3] (Schoonees et al., 2015؛ Cabooter, 2010؛ Weijters, 2006؛ Razavi, 2001؛ Baumgartner & Steenkamp, 2001)، انتخاب گزینه‌های انتهایی[4] (Rupp, 2013؛ Cabooter, 2010؛ Emons, 2008؛ Weijters, 2006؛ Razavi, 2001؛ Baumgartner & Steenkamp, 2001) و انتخاب گزینه میانی[5] (Schoonees et al., 2015؛ Cabooter, 2010؛ Weijters, 2006؛ Baumgartner & Steenkamp, 2001؛ Razavi, 2001) به ترتیب در پرسشنامه‌های غیر شناختی، کاربرد بیشتری دارند.

پاسخ‌های ناهمسان، به هر دلیلی که رخ داده باشند، باید پیش از تحلیل نمره پاسخ‌دهندگان، شناسایی شوند؛ زیرا یک پرسشنامه علی‌رغم آنکه دارای ویژگی‌های روان‌سنجی قابل‌قبول و رضایت بخشی است، ممکن است، نمره یک پاسخ‌دهنده به دلیل وجود پاسخ‌های ناهمسان معتبر نباشد و تصمیم‌گیری بر اساس آن، پیامدهای نامطلوبی به دنبال داشته باشد (Curtis, 2004؛ Kanighen et al., 2015).

---

1. response style
2. acquiescence
3. disacquiescence
4. extreme
5. mid-point

الگوهای پاسخ ناهمسان بر اساس دو رویکرد محتوایی و آماری قابل‌شناسایی می‌باشند. در رویکرد اول که سابقه‌ای نزدیک به ۸۰ سال دارد، از مقیاس‌های محتوایی برای شناسایی سبک‌های پاسخ‌دهی استفاده می‌شود. مقیاس‌های (تصحیح)K، (وانمود)F و (دروغ‌سنجی) L در پرسشنامه شخصیّت چندوجهی مینه‌سوتا[1] (Butcher et al., 2001)، مقیاس مطلوبیت اجتماعی و به‌طورکلی، سؤالات گزینه بایست (Chen et al., 2004) از این رویکرد استفاده می‌کنند؛ در مقابل، رویکرد آماری با مقایسه الگوی پاسخ فرد با الگوی مورد انتظار، ناهمسانی الگوهای پاسخ را شناسایی می‌نماید. آماره‌های برازش شخص موجود در چارچوب نظریه سؤال پاسخ در این رویکرد قرار می‌گیرند (Emons, 2008؛ Meijer & Sijtsma, 2001) که میزان منطقی بودن الگوی پاسخ یک آزمودنی به مجموعه‌ای از سؤال‌های یک پرسشنامه را با توجه به میزان واقعی خصیصه مکنون[2] او برآورد می‌کنند (De Ayala, 2009؛ Emons, 2008).

دقت آماره‌های برازش شخص به عوامل مختلفی مانند حجم نمونه، تعداد سؤالات شبیه‌سازی‌شده، مدل آماری انتخاب‌شده برای تولید داده‌ها و مقادیر برای پارامترهای شخص و سؤال (Rupp, 2013) نوع برآوردکننده‌های پارامتر و نوع پاسخ‌های ناهمسان (De la Torre & Deng, 2008؛ St-Onge et al., 2011) و درصد الگوهای پاسخ ناهمسان (Armstrong & Shi, 2009؛ Emons, 2009) بستگی دارد که بایستی در تحلیل آماره‌های برازش شخص به آن‌ها توجه کرد.

آماره‌های برازش شخص متفاوتی در چارچوب مدل‌های مختلف نظریه سؤال پاسخ وجود دارند. برای نمونه در مدل‌های دو و سه پارامتری از آماره‌های برازش D، $l_{zm}$ و ECI و در مدل راش از آماره‌های U، W، l و lz استفاده می‌شود، البته بیشتر آماره‌های برازش شخص مانند lz و U را می‌توان در سایر مدل‌های نظریه سؤال پاسخ نیز به کار گرفت (Meijer & Sijtsma, 2001). تاکنون، بیش از سی‌وشش آماره برازش شخص (Karabatsos, 2003؛ Meijer& Sijtsma, 2001) معرفی‌شده که بیشتر در داده‌های دو ارزشی مطالعه شده است (Emons, 2008؛ Kanygin et al., 2014). در بافت داده‌های چندارزشی از میان آماره‌های برازش شخص پارامتریک، بیشتر آماره $lz^p$، موردپژوهش قرار گرفته است (Emons, ؛St-Onge et al., 2011 ؛Ciu's , 2013 ؛Kanygin et al., 2014)

1. Minnesota Multiphasic Personality Inventory (MMPI)
2. latent trait

2008) و دقت آماره‌های پارامتریک دیگری مانند W و U (Rupp, 2013؛ Karabatsos, 2003؛ Glass & Meijer, 2003؛ Meijer & Sijtsma, 2001)که دقت قابل قبولی در داده‌های دوارزشی داشته‌اند، به‌ندرت موردپژوهش قرار گرفته است. همچنین، با توجه به آنکه برای شناسایی سبک‌های پاسخ‌دهی بیش از سی‌وشش آماره برازش شخص وجود دارد (Karabatsos, 2003)، این رویکرد همچنان نیازمند پژوهش است که آماره‌های برازش خاصی شناسایی شود که بتوانند برای تشخیص سبک‌های پاسخ‌دهی موجود در پرسشنامه‌های غیرشناختی مورداستفاده قرار بگیرند.

کاربرد قابل‌توجهی از آماره‌های برازش شخص در ده سال اخیر در حوزه‌های شخصیت (Ciu's, 2013؛ Ferrando, 2012؛ Ferrando & Chico, 2001)، نگرش (Curtis, 2004) و ارزیابی پیامدهای سلامت (Meijer et al., 2016؛Kanighen et al., 2015؛ Konard et al., 2010) ارائه شده است که در هر یک از این حوزه‌ها هدف پژوهشگران این بوده است که دقیق‌ترین آماره برازش شخص را برای شناسایی پاسخ‌های ناهمسان شناسایی کنند.

پژوهش‌های انجام‌شده در حوزه آماره‌های برازش شخص را می‌توان در دو دسته جای داد. در دسته نخست، پژوهشگران فقط در جستجوی آماره‌ای هستند که بتواند الگوهای پاسخ ناهمسان را شناسایی نماید (Conijn et al., 2014؛ Ferrando, 2012؛ Ferrando & Chico, 2001) در دسته دوم، پژوهشگران (Hang, 2012؛ St-Onge et al., 2011؛ Emons, 2008؛ Meijer, 2003؛ Glas & Meijer, 2003؛ Umbertson & Rice, 2000، ترجمه شریفی و همکاران، ۱۳۸۴) بر این باورند که برای شناسایی هر سبک پاسخ‌دهی این امکان وجود دارد که از یک نوع شاخص برازش خاص استفاده کرد؛ بنابراین در پژوهش حاضر، سعی بر بهبود سیستم اندازه‌گیری در سطح فردی هست تا افراد شایسته و دارای صلاحیت انتخاب شوند. به این منظور از میان مدل‌های نظریه سؤال پاسخ، با توجه به ویژگی‌های عینیت خاص[1] و نامتغیر بودن[2] (Umbertson & Rice, 2000، ترجمه شریفی و همکاران، ۱۳۸۴) مدل راش انتخاب شد و دقت و حساسیت چهار آماره برازش شخص مبتنی بر این مدل ($l^p$ و $lz^p$، $U^p$،$W^p$) در شناسایی شایع‌ترین سبک پاسخ‌دهی یعنی تصدیق در داده‌های چند ارزشی موردمطالعه قرار گرفت. همچنین، برای نتیجه‌گیری بهتر اثر عوامل

1. specific objectivity
2. invariant

مؤثری مانند حجم نمونه، تعداد سؤالات پرسشنامه، درصد سؤالات ناهمسان و درصد پاسخ‌دهندگان با الگوی پاسخ ناهمسان موردبررسی قرار گرفت.

## روش

دقت چهار آماره برازش شخص ($W^p$، $U^p$، $l^p$، $lz^p$) در شناسایی سبک پاسخ‌دهی تصدیق در چارچوب مدل راش با یکدیگر مقایسه شد. مدل راش یکی از توانمندترین مدل‌های خانواده نظریه سؤال پاسخ است. از ویژگی‌های شاخص مدل راش، ویژگی عینیت خاص (واحدهای اندازه‌گیری یکسان) و نامتغیر بودن است. مدل راش شامل خانواده‌ای از مدل‌ها است که ازجمله آن‌ها می‌توان به مدل امتیاز پاره‌ای[1] (PCM)، مدل امتیاز پاره‌ای تعمیم‌یافته[2] (GPCM)، مدل مقیاس درجه‌بندی[3] (RSM) و مدل لوجستیک ساده[4] (SLM) اشاره کرد. مدل امتیاز پاره‌ای از مدل‌های پرکاربرد راش است که برای تحلیل پاسخ‌های مقیاس‌های نگرشی و شخصیتی که پاسخ‌دهنده عقاید خود را درجه‌بندی می‌کند و یا به عبارت‌ها در یک مقیاس چندنقطه‌ای پاسخ می‌دهد کاملاً مناسب است (Umbertson & Rice, 2000، ترجمه شریفی و همکاران، ۱۳۸۴).

### طرح شبیه‌سازی داده‌ها

شبیه‌سازی داده‌ها بر اساس نسخه سه نرم‌افزار وینجن (Han, 2007) انجام شد. نخست، پیش‌فرض تک‌بعدی بودن و استقلال موضعی تحت مدل امتیاز پاره‌ای راش[5] (PCM) در نرم‌افزار انتخاب شد. در مرحله بعد، شبیه‌سازی داده‌ها با توجه به مقادیر تعیین‌شده برای هر یک از عوامل مؤثر (۵۰۰ و ۲۰۰۰ آزمودنی، ۳۰ و ۶۰ سؤال، ۱۵ و ۳۰ درصد سؤال ناهمسان، ۱۵ و ۳۰ درصد پاسخ‌دهندگان با سبک پاسخ‌دهی) (Rupp, 2013؛ Armstrong & Shi, 2009؛ Emons, 2008, 2009؛ Karabatsos, 2003، Ferrando & Chico, 2001) به روش زیر انجام شد:

نمره ۲۰۰۰ آزمودنی با یک توزیع توانایی نرمال در پاسخ به یک پرسشنامه ۶۰ سؤالی با ۲۵ تکرار تولید شد. همچنین، توزیع جایگاه سؤالات، یکنواخت (۳، ۳-) در نظر گرفته شد. در مرحله بعد از هر مجموعه داده، یک نمونه ۵۰۰ تایی به روش تصادفی انتخاب شد. به

1. partial credit model
2. generalized partial credit model
3. Rating scale model
4. simple logistic model
5. partial credit model

همین ترتیب، آزمون کوتاه‌تر (۳۰ سؤالی) از آزمون ۶۰ سؤالی به‌صورت تصادفی نمونه‌گیری شد. پاسخ ۱۵ و ۳۰ درصد آزمودنی‌ها با توانایی پایین ($\theta \leq$ ۰) به ۱۵ و ۳۰ درصد سؤالاتی که جایگاهشان بالاتر از توانایی آزمودنی ($\partial i \geq$ ۱/۵) بود به گزینه ۴ (موافقم) تغییر یافت و درصورتی‌که پاسخ آن‌ها گزینه ۵ بود تغییری نکرد.

چهار آماره برازش شخص؛ $W^p$، $U^p$، $l^p$ (با نرم‌افزار آماری excel) و $lz^p$ (با پکیج Perfit (Tendeiro, 2015) در نرم‌افزار R) برای هر یک از آزمودنی‌ها برآورد شد. سپس به‌وسیله منحنی راک (نرم‌افزار spss نسخه ۲۵) مشخص شد که تحت شرایط مختلف (حجم نمونه، طول آزمون، درصد سؤال‌های ناهمسان و درصد پاسخ‌دهندگان با سبک پاسخ‌دهی) دقت و حساسیت کدام‌یک از چهار آماره برازش شخص در شناسایی سبک پاسخ‌دهی تصدیق بالاتر است که درنهایت، متوسط میانگین شناسایی هر یک از آماره‌های برازش شخص در هر ۲۵ مجموعه داده گزارش شد (جدول یک).

برای تعیین حساسیت و دقت آماره برازش شخص در موقعیت‌های مختلف از فضای زیر منحنی راک استفاده شد. مقدار این فضا، نشانه دقت آماره برازش شخص در تمایز بین بردارهای پاسخ همسان و ناهمسان هست. ازنظر تشخیصی مقدار سطح زیر منحنی کمتر از ۶۹/. غیرقابل‌قبول، بین ۰/۷۰ تا ۰/۷۹ قابل‌قبول، بین ۰/۸۰ تا ۰/۸۹ عالی، بالاتر از ۹/. بسیار عالی هست و نشانگر این است که آماره برازش موردنظر می‌تواند بین افراد دو گروه تمایز بسیار خوبی بگذارد. سپس نقطه‌ای که بالاترین مقدار حساسیت و ویژگی را داشت به‌عنوان بهترین نقطه برش انتخاب شد (Meyers et al., 2016) که در جدول ارائه شده در پیوست قابل‌ملاحظه است.

## یافته‌ها

دقت شناسایی سبک پاسخ‌دهی تصدیق کردن: جدول یک نشان می‌دهد هنگامی‌که درصد سؤالات ناهمسان ۱۵٪ و ۳۰٪ است، دقت شناسایی سبک پاسخ‌دهی تصدیق کردن چهار آماره برازش شخص $W^p$، $U^p$، $lz^p$ و $l^p$ در موقعیت‌های مختلف (حجم نمونه، طول آزمون و درصد پاسخ‌دهندگان با الگوهای پاسخ ناهمسان) چقدر است.

**جدول ۱.** دقت شناسایی سبک پاسخ‌دهی تصدیق کردن چهار آماره برازش شخص (۲۵ تکرار)

| درصد سؤالات با الگوی پاسخ ناهمسان=۱۵٪ (α=۰/۰۵) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| درصد افراد با الگوی پاسخ ناهمسان=۳۰٪ | | | | درصد افراد با الگوی پاسخ ناهمسان= ۱۵٪ | | | | طول آزمون | حجم نمونه |
| $W^p$ | $U^p$ | $l^p$ | $lz^p$ | $W^p$ | $U^p$ | $l^p$ | $lz^p$ | | |
| ۰/۴۱ | ۰/۵۸ | ۰ | ۰ | ۰/۷۰ | ۰/۷۱ | ۰ | ۰/۰۳ | ۳۰ | ۵۰۰ |
| ۰/۴۲ | ۰/۵۹ | ۰ | ۰ | ۰/۷۳ | ۰/۷۱ | ۰ | ۰/۰۳ | ۶۰ | |
| ۰/۵۲ | ۰/۶۷ | ۰ | ۰ | ۰/۷۷ | ۰/۷۸ | ۰ | ۰/۰۳ | ۳۰ | ۲۰۰۰ |
| ۰/۵۳ | ۰/۶۸ | ۰ | ۰ | ۰/۷۸ | ۰/۷۹ | ۰ | ۰/۰۳ | ۶۰ | |
| درصد سؤالات با الگوی پاسخ ناهمسان=۳۰٪ (α=۰/۰۵) | | | | | | | | | |
| درصد افراد با الگوی پاسخ ناهمسان= ۳۰٪ | | | | درصد افراد با الگوی پاسخ ناهمسان= ۱۵٪ | | | | طول آزمون | حجم نمونه |
| $W^p$ | $U^p$ | $l^p$ | $lz^p$ | $W^p$ | $U^p$ | $l^p$ | $lz^p$ | | |
| ۰/۴۰ | ۰/۵۸ | ۰ | ۰ | ۰/۷۲ | ۰/۷۲ | ۰ | ۰/۰۲ | ۳۰ | ۵۰۰ |
| ۰/۴۳ | ۰/۵۷ | ۰ | ۰ | ۰/۷۵ | ۰/۷۴ | ۰ | ۰/۰۱۲ | ۶۰ | |
| ۰/۵۳ | ۰/۶۹ | ۰ | ۰ | ۰/۷۸ | ۰/۷۹ | ۰ | ۰ | ۳۰ | ۲۰۰۰ |
| ۰/۵۲ | ۰/۶۸ | ۰/۰۲ | ۰/۰۲ | ۰/۷۹ | ۰/۸۰ | ۰ | ۰/۰۳ | ۶۰ | |

نتایج جدول بالا دقت شناسایی چهار آماره برازش شخص در شرایط مختلف از حجم نمونه، طول پرسشنامه، درصد پاسخ‌دهندگان و درصد سؤالات با سبک پاسخ‌دهی (تصدیق کردن) را نشان می‌دهد. در شرایطی که، ۱۵٪ پاسخ‌دهندگان در یک نمونه ۵۰۰ نفری، ۱۵٪ سؤالات را تصدیق کرده باشند، دقت آماره‌های برازش شخص در پرسشنامه‌های ۳۰ و ۶۰ سؤالی به ترتیب: (۰/۰۳)، ($lz^p$) (۰/۰۳)، ($l^p$)(۰)، ($l^p$) (۰/۷۱)، ($U^p$) (۰/۷۱) و (۰/۷۳)، ($W^p$)(۰/۷۰) برآورد شد. در این شرایط آماره‌های برازش $U^p$ و $W^p$، دقت شناسایی قابل‌قبولی داشتند و آماره‌های $lz^p$ و $l^p$، در این شرایط نتوانستند سبک پاسخ‌دهی تصدیق را شناسایی نمایند. در همین موقعیت، اگر درصد پاسخ‌دهندگان با الگوی پاسخ (تصدیق) دو برابر شود، در پرسشنامه ۳۰ و ۶۰ سؤالی به ترتیب؛ دقت آماره‌های برازش شخص به مقادیر (۰)، ($lz^p$) (۰)، ($l^p$)(۰)، (۰/۵۹)، ($U^p$)(۰/۵۸) و (۰/۴۲)، ($W^p$)(۰/۴۱) تغییر یافت در این موقعیت، هیچ‌کدام از آماره‌ها نتوانستند، سبک پاسخ‌دهی تصدیق را شناسایی نمایند.

اکنون اگر درصد سؤالات با الگوی پاسخ ناهمسان دو برابر شود و درصد افراد با الگوی پاسخ ناهمسان ۱۵٪ باشد در پرسشنامه‌های ۳۰ و ۶۰ سؤالی، دقت شناسایی آماره‌های برازش شخص به ترتیب: (۰/۱۲)، ($lz^p$) (۰/۰۲)، ($l^p$)(۰)، ($l^p$) (۰/۷۴)، ($U^p$)(۰/۷۲) و (۰/۷۵)، ($W^p$)(۰/۷۲)

برآورد شد. آماره‌های برازش $U^p$ و $W^p$، بالاترین و آماره $lz^p$ و lp پایین‌ترین دقت شناسایی را داشتند.

حجم نمونه تا ۲۰۰۰ نفر افزایش یافت و دقت آماره‌های برازش شخص بررسی شد. هنگامی‌که، ۱۵٪ پاسخ‌دهندگان، ۱۵٪ سؤالات را تصدیق کرده باشند، دقت آماره‌های برازش شخص در یک پرسشنامه ۳۰ و یا ۶۰ سؤالی به ترتیب: (۰/۰۳) ، (۰/۰۳)$lz^p$، (۰)، (۰)$l^p$، (۰/۷۹)، (۰/۷۸) $U^p$ و (۰/۷۸)، (۰/۷۷) $W^p$ برآورد شد. آماره‌های$lz^p$ و $l^p$ قادر به شناسایی این سبک پاسخ‌دهی نبودند. در همین شرایط، اگر درصد سؤالات دو برابر شود در یک پرسشنامه ۳۰ و ۶۰ سؤالی دقت آماره‌های برازش شخص به ترتیب به مقادیر (۰/۰۳) ، (۰)$lz^p$، (۰)، (۰)$l^p$، (۰/۸۰)، (۰/۷۹)$U^p$ و (۰/۷۹)،(۰/۷۸) $W^p$ تغییر یافت. در این موقعیت، دقت آماره‌های برازش $lz^p$ و $l^p$، غیرقابل‌قبول و در مقابل، آماره‌های $U^p$ و $W^p$، دقت قابل‌قبولی در شناسایی سبک پاسخ‌دهی تصدیق داشتند. به‌طورکلی، هنگامی‌که درصد پاسخ‌دهندگان با سبک پاسخ‌دهی دو برابر شد در دو حجم نمونه ۵۰۰ و ۲۰۰۰ آزمودنی، هیچ‌کدام از آماره‌های برازش شخص قادر به شناسایی این سبک پاسخ‌دهی نبودند زیرا بر طبق طبقه‌بندی Meyers و همکاران (2016)، دقت این آماره‌ها کمتر از ۰/۷۰ برآورد شد که غیرقابل‌قبول است.

## بحث و نتیجه‌گیری

هدف مطالعه حاضر یافتن بهترین آماره برازش شخص در شناسایی سبک پاسخ‌دهی تصدیق در داده‌های چند ارزشی بود. به این منظور از چهار آماره برازش شخص مبتنی بر مدل راش ($W^p$، $U^p$، $lz^p$و$l^p$) استفاده شد. برای بررسی دقیق‌تر، موقعیت‌های مختلف اعم از حجم نمونه، طول آزمون، درصد افراد با سبک پاسخ‌دهی و درصد سؤالات ناهمسان به کار گرفته شد. نتایج حاصل از تحلیل داده‌های شبیه‌سازی نشان داد که دقت آماره‌های $U^p$ و $W^p$ در شناسایی سبک پاسخ‌دهی تصدیق با افزایش حجم نمونه، بیشتر شد؛ در مقابل دو آماره $l^p$ و $lz^p$ تحت تأثیر افزایش حجم نمونه قرار نگرفتند. Ciu's (2013) در بررسی آماره‌های برازش شخص به این نتیجه دست یافت که با افزایش حجم نمونه (بیشتر از ۵۰۰ نفر)، میانگین شناسایی شاخص خطای گاتمن، افزایش می‌یابد که این یافته همسو با نتایج به‌دست‌آمده از آماره‌های برازش شخص $U^p$ و $W^p$ است. ازنظر آماری انتظار می‌رود در حجم نمونه بزرگ‌تر، دقت مقایسه‌های بین بردارهای پاسخ مشاهده‌شده و پیش‌بینی‌شده بیشتر شود و پارامترهای سؤال با

دقت بالاتری برآورد شوند؛ بنابراین با افزایش حجم نمونه، دقت شناسایی آماره‌ها نیز بیشتر خواهد شد (Rupp, 2013). برخلاف این یافته‌ها، آماره $lz^p$ (Ciu's, 2013) و آماره‌های (BW) (Hang, 2012) در حجم نمونه‌های کوچک‌تر، دقت شناسایی بالاتری داشتند. البته در این پژوهش‌ها (Ciu's, 2013؛ Hang, 2012)، مدل شبیه‌سازی داده‌ها، نوع سبک پاسخ‌دهی و تعداد سؤالات پرسشنامه با پژوهش حاضر متفاوت است. برای مثال در پژوهش Ciu's (2013) داده‌ها در چهارچوب مدل پاسخ مدرج شبیه‌سازی شدند و اثر تعداد سؤالات بر دقت آماره‌های برازش شخص موردبررسی قرار نگرفت.

تعداد سؤالات عامل مؤثر دیگر بر دقت شناسایی آماره‌های برازش شخص است که آماره‌های $U^p$ و $W^p$ در مقایسه با دو آماره دیگر با افزایش تعداد سؤالات، با دقت بالاتری توانسته‌اند، سبک پاسخ‌دهی تصدیق را شناسایی نمایند. نتایج پژوهش‌های انجام‌شده (Kanygin et al., 2014؛ St-Onge et al., 2011؛ Emons, 2008, 2009؛ Karabatsos, 2003) از این یافته حمایت می‌کند. Woods و همکاران (2008) نیز در پژوهش خود نشان دادند که پاسخ‌های ناهمسان همیشه قابل‌شناسایی نیستند و در مقیاس‌های طولانی‌تر با سؤالاتی که با یکدیگر و همچنین با مقدار توانایی، همبستگی بالایی داشته باشند و دامنه دشواری آن‌ها نسبتاً متفاوت باشد، قدرت شناسایی آن‌ها توسط آماره‌های برازش شخص بالاتر خواهد بود. مقیاس‌هایی که ویژگی‌های اندازه‌گیری پایینی دارند و کوتاه‌ترند، ناهمگونی در برازش شخص را نشان نمی‌دهند؛ بنابراین با افزایش تعداد سؤالات یک پرسشنامه، دامنه جایگاه سؤالات نیز بیشتر خواهد شد و احتمال شناسایی الگوهای پاسخ ناهمسان افزایش می‌یابد (Umbertson & Rice, 2000، ترجمه شریفی و همکاران، ۱۳۸۴).

عامل مؤثر دیگر بر دقت آماره‌های برازش شخص، درصد پاسخ‌دهندگان با سبک پاسخ‌دهی است که با افزایش این عامل دقت شناسایی آماره‌های برازش $U^p$ و $W^p$ کاهش یافت. ازنظر آماری نیز با افزایش پاسخ‌دهندگان با سبک پاسخ‌دهی میزان شناسایی آن‌ها دشوارتر خواهد شد (Rupp, 2013؛ Karabatsos, 2003). برخلاف یافته‌های حاضر، Hang (2012) در بررسی تأثیر نسبت پاسخ‌دهندگان با الگوی پاسخ ناهمسان بر دقت آماره‌های برازش شخص مبتنی بر نظریه سؤال پاسخ، آماره‌های مبتنی بر گروه و آماره‌های (BW) به این نتیجه دست یافت که نسبت افرادی که پاسخ‌های ناهمسان داده‌اند، اثر کمی بر دقت و حساسیت آماره‌های برازش شخص دارد. در این پژوهش سه درصد متفاوت (۱۰٪، ۲۰٪ و

۳۰٪) از پاسخ‌دهندگان با الگوی پاسخ ناهمسان در نظر گرفته شده بود، ممکن است با افزایش پاسخ‌دهندگان با الگوی پاسخ ناهمسان، دقت آماره‌های برازش شخص نیز تغییر یابد، همچنین موقعیت‌های دیگری مانند طول آزمون و حجم نمونه که بر دقت آماره‌های برازش شخص تأثیر خواهند گذاشت در این پژوهش بررسی نشده است که می‌تواند هم‌زمان با تغییر درصد پاسخ‌دهندگان با الگوی پاسخ ناهمسان بر دقت آماره‌های برازش شخص نیز اثر بگذارد.

تأثیر درصد سؤال‌هایی با الگوی پاسخ ناهمسان در دقت شناسایی چهار آماره برازش شخص نیز موردبررسی قرار گرفت. دقت آماره‌های برازش شخص $U^p$ و $W^p$ تقریباً با افزایش درصد سؤالات با سبک پاسخ‌دهی افزایش یافتند. ازنظر آماری نیز مطابق با نتایج حاضر با افزایش درصد سؤال‌های ناهمسان، تشخیص الگوهای مورد انتظار آسان‌تر خواهد شد؛ بنابراین، دقت شناسایی آماره‌های برازش شخص نیز افزایش خواهد یافت.

بررسی پیشینه پژوهشی در ارتباط با تأثیر درصد سؤال‌های ناهمسان، نتایج متفاوتی نشان داد؛ برای نمونه در پژوهش Ciu's (2013)، هنگامی‌که پاسخ‌دهندگان در تمامی سؤالات از سبک پاسخ‌دهی وانمود استفاده کردند، دقت شاخص خطای گاتمن برخلاف آماره $lz^p$ کاهش یافت و یا در پژوهش St-Onge و همکاران (2011) سه آماره برازش شخص ( lz, ECI4z،ECI2z ) در شناسایی سبک پاسخ‌دهی حدس زدن، عملکرد مشابهی داشتند و با افزایش سؤالات با احتمال حدس زدن بیشتر، دقت آماره‌های برازش شخص کاهش یافت. در پژوهش دیگری، دقت شناسایی آماره $l^p_z$ با افزایش سؤالاتی که تصادفی پاسخ داده شده‌اند، بیشتر شد؛ در مقابل با افزایش درصد سؤالات با سبک پاسخ‌دهی تصدیق، دقت این آماره کاهش یافت (Kanygin et al., 2014)؛ بنابراین، عوامل دیگری مانند نوع سبک پاسخ‌دهی نیز هم‌زمان با تعداد سؤالات ناهمسان بر دقت آماره‌های برازش شخص تأثیر می‌گذارد که باید در نظر گرفته شود.

یافته‌های پژوهش حاضر و بررسی پیشینه پژوهشی نشان داد که برای بررسی دقت یک آماره برازش شخص، باید عوامل مختلفی مانند حجم نمونه، طول آزمون، درصد سؤال‌های ناهمسان و درصد افراد با سبک پاسخ‌دهی را در نظر گرفت. در شرایطی که احتمال سبک پاسخ‌دهی تصدیق وجود دارد، مانند موقعیت‌های استخدام که داوطلبان سعی دارند خود را موجه جلوه دهند، بنابراین سؤالاتی که موردپسند جامعه است را تصدیق می‌کنند در این

شرایط از میان چهار آماره برازش شخص مبتنی بر مدل راش، آماره‌های $U^p$ و $W^p$ در داده‌های چندارزشی، دقت شناسایی بالاتری داشتند؛ البته با توجه به آنکه، با افزایش درصد پاسخ‌دهندگان با سبک پاسخ‌دهی، این آماره‌ها قادر به شناسایی سبک پاسخ‌دهی تصدیق نبودند؛ بنابراین پیشنهاد می‌شود، دقت سایر آماره‌های برازش شخص در شناسایی سبک پاسخ‌دهی تصدیق در پرسشنامه‌های غیرشناختی چندارزشی موردمطالعه قرار بگیرند. همچنین، دو آماره $lz^p$ و $l^p$ که به فراوانی در داده‌های چندارزشی و به‌ویژه داده‌های دوارزشی بکار گرفته شده‌اند در شناسایی سبک پاسخ‌دهی تصدیق در داده‌های چندارزشی مناسب نمی‌باشند و دقت کافی برای شناسایی این سبک پاسخ‌دهی را ندارند.

## تعارض منافع

نویسندگان هیچ‌گونه تعارض منافعی ندارند.

## پیوست ۱.

بهترین نقطه برش آماره‌های برازش شخص در شناسایی سبک پاسخ‌دهی تصدیق کردن (۲۵ تکرار)

| درصد سؤالات با الگوی پاسخ ناهمسان=۱۵٪ (α=۰/۰۵) | | | | | |
|---|---|---|---|---|---|
| درصد افراد با الگوی پاسخ ناهمسان= ۳۰٪ | | درصد افراد با الگوی پاسخ ناهمسان= ۱۵٪ | | طول پرسشنامه | حجم نمونه |
| $W^p$ | $U^p$ | $W^p$ | $U^p$ | | |
| – | – | ۸۸/۶۲ | ۳/۱۵ | ۳۰ | ۵۰۰ |
| – | – | ۹۵/۶۹ | ۲/۸۵ | ۶۰ | |
| – | – | ۹۶/۳۷ | ۲/۹۶ | ۳۰ | ۲۰۰۰ |
| – | – | ۹۳/۹۷ | ۲/۷۴ | ۶۰ | |
| درصد سؤالات با الگوی پاسخ ناهمسان=۳۰٪ (α=۰/۰۵) | | | | | |
| درصد افراد با الگوی پاسخ ناهمسان= ۳۰٪ | | درصد افراد با الگوی پاسخ ناهمسان= ۱۵٪ | | طول پرسشنامه | حجم نمونه |
| $W^p$ | $U^p$ | $W^p$ | $U^p$ | | |
| – | – | ۸۹/۸۶ | ۲/۹۸ | ۳۰ | ۵۰۰ |
| – | – | ۹۳/۹۴ | ۲/۶۵ | ۶۰ | |
| – | – | ۸۸/۵۲ | ۲/۹۵ | ۳۰ | ۲۰۰۰ |
| – | – | ۹۴/۲۳ | ۲/۸ | ۶۰ | |

## منابع

امبرتسون، سوزان ای؛ رایس، استیون پی (۲۰۰۰). *نظریه‌های جدید روان‌سنجی برای روان‌شناسان*، ترجمه حسن پاشا شریفی و همکاران. (۱۳۸۸). تهران: رشد.

## References

Armstrong, R. D. & M. Shi. (2009).A parametric cumulative sum statistic for person fit. *Appllied Psychological Measurement, 33*, 391-410.

Baumgartner, H. & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*, 143–156.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). MMPI–2 (Minnesota Multiphasic Personality Inventory–2): *Manual for administration and scoring* (Revised). Minneapolis, MN: University of Minnesota Press.

Conijn, J.M. Emons, W.H. De Jong, K. & Sijtsma, K. (2015). Using person-fit analysis to detect and explain aberrant responding to the Outcome Questionnaire-45. *Assessment, 22*, 5312-524.

Conijn, J. M. Emons W. H. M. & Sijtsma, K. (2014). Statistic lz-based person-fit methods for noncognitive multiscale measures. *Applied psychological Measurement, 38*, 122–136.

Cabooter, E. F. K. (2010). *The impact of situational and dispositional variables on response styles with respect to attitude measures*. Doctoral dissertation, Ghent University, Ghent, Belgium.

Conrad K. J. Bezruczko N. Chan Y. Riley B. Diamond G. Dennis M. L. (2010). Screening for atypical suicide risk with person fit statistics among people presenting to alcohol and other drug treatment. *Drug and Alcohol Dependence, 106*, 92-100.

Curtis, D. D. (2004). Person misfit in attitude surveys: Influences, impacts and implications. *International Education Journal, 5*, 125-143.

Chen, C. I., Lee, M. N., & Yen, C. L. (2004). Faking intention on the internet: Effects of test types and situational factors. *Chinese Journal of Psychology, 46*(4), 349- 359.

De Ayala, R. J. (2009) .The Theory and Practice of Item Response Theory.

De la Torre, J. & Deng, W. (2008). Improving person fit assessment by correcting the ability estimate and its reference distribution. *Journal of Educational Measurement, 45*, 159±1177.

Emons, W. H. M. (2009). Detection and diagnosis of person misfit from patterns of summed polytomous item. *Applied Psychological Measurement, 33*, 599-619.

Emons, W. H. M. (2008). Nonparametric person-fit analysis of polytomous item scores. *Applied Psychological Measurement, 32*, 224-247.

Ferrando, P. J. (2012). Assessing inconsistent responding in E and N measures: An application of person-fit analysis in personality. *Personality and Individual Differences, 52*, 718-722.

Ferrando, P.J. & Chico, E. (2001).Detecting dissimulation in personality test scores: A comparison between person-fit indices and detection scales. *Educational and Psychological Measurement, 61*, 997-1012.

Glas, C. A. W &Meijer, R. R. (2003). A Bayesian approach to person fit analysis in item response theory models, *Applied Psychological Measurement 27*(3), 217-233.

Huang, T.W. (2012). Aberrance Detection Powers of the BW and Person-Fit Indices. *Educational Technology & Society, 15*(1),128–37.

Han, K. T. (2007). WinGen: Windows software that generates IRT parameters and item responses. *Applied Psychological Measurement, 31*(5),457-459.

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education, 16*, 277-298.

Meyers, L. S. Gamst, G. & Guarino, A. J. (2016). Applied multivariate research: Design and interpretation. Sage publications.

Meijer, R. R., Niessen, A. S. M., & Tendeiro, J. N. (2016). A Practical Guide to Check the Consistency of Item Response Patterns in Clinical Research Through Person-Fit Statistics: Examples and a Computer Program. *Assessment, 23*(1), 52-62.

Meijer, R. R. (2003). Diagnosing item score patterns on a test using item response theorybased person-fit statistics. *Psychological Methods, 8*, 72-87.

Meyer IH. (2003). Prejudice, social stress and mental health in lesbian, gay, and bisexual populations: Conceptual issues and research evidence. *Psychological Bulletin, 129*, 674–697.

Meijer, R. R. & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement, 25*, 107-135.

Rupp, A. A. (2013). A systematic review of the methodology for person fit research in Item Response Theory: Lessons about generalizability of inferences from the design of simulation studies. *Psychological Test and Assessment Modeling*, 55.

Razavi, Tiffani. (2001). *Self-report measures: an overview of concerns and mitations of questionnaire use in occupational stress research*, Southampton, UK University of Southampton 23pp. (Discussion Papers in Accounting and Management Science, 01-175).

Schoonees, P. C., Van de Velden, M., Groenen, P. J.F. (2015). Constrained dual scaling for detecting response styles in categorical data, Econometric Institute Rep ort, *Psychometric Society, 804*, 968-994.

Syu, J. J. (2013). *Applying person – fit in faking detection the simulation and practice of non- parametric item response theory* (Unpublished doctoral dissertation). chengchi university.

St-Onge, C. Valois, P. Abdous, B. & Germain, S. (2011). Accuracy of person-fit statistics: A Monte Carlo study of the influence of aberrance rates. *Applied Psychological Measurement*, 35, 419-432.

Tendeiro, J. N. (2015). *Package 'PerFit*.Version 1.4.

Tippins, N. T. (2011). *Technology-Enhanced Assessment of Talent*. Hoboken, NJ, USA: Jossey-Bass, ProQuest ebrary. Web. 21 November 2015.

Woods, C. M. Oltmanns, T. F. &Turkheimer, E. (2008). Detection of aberrant responding on a personality scale in a military sample: An application of evaluating person fit with twolevel logistic regression. *Psychological assessment, 20*(2), 159.

Weijters, B. (2006). *Response styles in consumer research* Doctoral dissertation, Ghent University, Ghent, Belgium.

Zicker, M. J. & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriate measurement. *Applied Psychological Meagurement, 20*, 71,87.