



**INSAN**

Jurnal Psikologi dan Kesehatan Mental

<http://e-journal.unair.ac.id/index.php/JPKM>

p-ISSN 2528-0104 | e-ISSN 2528-5181



RESEARCH ARTICLE / ARTIKEL PENELITIAN

## **Psychometric Properties and Mixture Rasch Model Analysis of Body Regulation Exercise Questionnaire (BREQ-3)**

Bryan Andika, Ingrid Tunggowan, & Ananta Yudiarso  
Faculty of Psychology, Universitas Surabaya, Indonesia

### **ABSTRACT**

This study aimed to test the reliability and validity of Indonesian-translated Body Regulation Exercise Questionnaire (BREQ-3), originally developed by Wilson et al. (2006), using Rasch analysis. A total of 408 Indonesian participants aged 18–25 years completed the survey. The results supported the unidimensionality of each BREQ-3 subscale. The 5-point Likert scale functioned adequately, except for the amotivation subscale. Both person and item reliability, as well as separation indices, were within acceptable ranges. Most items effectively measured exercise regulation, although revision is required on six items. Three items from external and intrinsic regulation subscales showed gender bias. Further analysis using Mixture Rasch Model indicated that the identified classes were not distinguished by gender or age. Overall, the findings support the use of BREQ-3 Indonesian version for assessing physical activity regulation in young adults and provide recommendations for future research.

**Keywords:** *breq-3, exercise, exercise regulation, rasch analysis*

### **ABSTRAK**

Penelitian ini bertujuan untuk memvalidasi sifat psikometrik BREQ-3 Versi Bahasa Indonesia dengan menggunakan model Rasch. Studi melibatkan 408 partisipan warga negara Indonesia dalam kategori dewasa awal, dengan batasan usia antara 18 – 25 tahun. Hasil penelitian menunjukkan asumsi unidimensionalitas masing-masing subskala BREQ-3 terpenuhi. Skala Likert 5 poin instrumen ini memiliki fungsionalitas yang memadai, kecuali subskala amotivasi. Hasil tes reliabilitas individu, reliabilitas item, indeks pemisahan individu, dan pemisahan item cukup baik. Sebagian besar item instrumen ditemukan efektif dalam mengukur regulasi latihan, dengan pengecualian enam item yang membutuhkan revisi. Secara khusus, tiga item teridentifikasi memiliki bias gender pada subskala Regulasi Eksternal dan Regulasi Intrinsik. Analisis juga dilakukan dengan menggunakan Mixture Rasch Model dan hasilnya mengungkapkan bahwa ketiga kelas tidak dibedakan berdasarkan jenis kelamin dan usia. Temuan-temuan ini memiliki implikasi untuk penggunaan BREQ-3 Versi Indonesia untuk menilai regulasi olahraga pada masa dewasa awal dan rekomendasi untuk penelitian masa depan.

**Kata kunci:** *analisis rasch, breq-3, olahraga, regulasi olahraga*

INSAN Jurnal Psikologi dan Kesehatan Mental, 2025, Vol. 10(1), 21-52, doi: 10.20473/jpkm.v10i12025.21-52

Submitted: 18/12/2024 Accepted: 28/04/2025 Published: 27/06/2025

Editor: Rahkman Ardi

\*Corresponding Author's Email: bryandika43@gmail.com



This manuscript is under the open access policies and the Creative Common Attribution License (<http://creativecommons.org/licenses/by/4.0>); therefore, any use, distribution, and reproduction of this article, in any media, is not restricted as long as its original source is properly cited.

## INTRODUCTION

Motivation to exercise is the motives that exist within a person in bringing up exercise activities to achieve the desired goals, reflecting varying levels of autonomy (Wilson et al., 2002). This construct can be measured using Behavioral Regulation in Exercise Questionnaire-3 (BREQ-3), developed by Wilson, et al. (2002), based on Self-Determination Theory (SDT), which categorizes motivation along a continuum of autonomy. Understanding behavioral regulation in the context of exercise is a fundamental aspect of sports psychology and health promotion. SDT has emerged as the dominant theoretical framework for understanding exercise motivation, emphasizing that the quality of motivation—particularly its level of self-determination—is more crucial than its quantity in predicting sustained exercise behavior (Teixeira et al., 2012). BREQ-3 was developed to measure the spectrum of motivation regulation based on SDT, ranging from amotivation to intrinsic motivation. Nam et al. (2023) in their systematic review confirmed that BREQ-3 is the most widely used instrument to assess exercise motivation based on SDT, with good predictive ability for physical activity behavior. This is supported by a comprehensive bibliometric analysis of motivation measurement tools in sports, revealing an evolving landscape of measurement instruments (Clancy et al., 2017).

Hagger and Chatzisarantis (2008) highlighted the importance of developing psychometric instruments that can effectively capture the complex constructs of motivation to exercise. The main challenge is creating tools that can accurately differentiate between various types of motivation regulation, from external regulation to intrinsic motivation. Chung and Liu (2012) specifically examined the psychometric properties of translated instruments, emphasizing the challenges of cross-cultural validation. Several studies have analyzed BREQ, such as by Wilson et al. (2002), who critically reviewed the psychometric properties of the initial Behavioral Regulation in Exercise Questionnaire (BREQ), identifying strengths and weaknesses in the existing measurement approach. While BREQ-3 has been validated in various countries, there is still a gap in understanding its psychometric properties across different cultural contexts. Rogowska and Morouço (2024) emphasized the importance of cross-cultural validation to ensure that this instrument can accurately measure the exercise motivation construct in different contexts.

The results of the Portuguese version of BREQ-3 study have been applied using confirmatory factor analysis using AMOS 20.0 software and resulted that after removing 6 of the 24 items, the Portuguese version of BREQ-3 is valid and reliable (Cid et al., 2018). The confirmatory factor analysis of the Italian version of BREQ-3 using Mplus 8 software version 1.6 has proven it to be a reliable and valid instrument for measuring exercise behavior regulation in individuals with different demographic characteristics (Cavicchiolo et al., 2022). The results of the confirmatory factor analysis of BREQ-3M using Mplus 8 showed that Malaysian version of BREQ-3M is valid and reliable (Chai et al., 2022). The current version of the Chinese BREQ-3 has shown adequate evidence of validity and reliability based on its bi-factor structure but we recommend calculating amotivation, external regulation, introjected regulation, and single autonomous motivation scores (excluding item 19) when using the Chinese version of BREQ-3.

In relation to instrument validity, Standards for Educational and Psychological Testing, published by American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME), provide comprehensive guidelines for test development and emphasize the importance of evaluating the quality of testing practices (AERA et al.,

2004). According to the Standards (AERA et al., 1999), validity is defined as "the extent to which evidence and theory support the interpretation of test scores for proposed uses of tests." The standards also outline five sources of evidence that can be used to evaluate the validity of test score interpretations for specific purposes. The five sources consist of validity evidence based on: test content, response processes, internal structure, relations to other variables, and consequence of testing.

In Indonesia, no comprehensive study has yet evaluated the psychometric properties of the BREQ-3 using a modern measurement approach such as Rasch Model. Chemolli and Gagné (2014) showed that Rasch analysis provides a deeper understanding of the structure of motivation measures compared to the classical test theory. This aligns with the primary objective of the present study, which is to strengthen the internal structure validity evidence as recommended by AERA and other related professional organizations. The Rasch Model was selected for its ability to: (1) evaluate the unidimensionality of each BREQ-3 subscale, (2) identify misfitting items through item fit analysis, and (3) assess measurement invariance across demographic groups. Pietrabissa et al. (2020) emphasized that Rasch analysis provides stronger validity evidence for BREQ-3, especially in detecting item bias and ensuring measurement equivalence.

This study also employed Mixture Rasch Model (MRM) for further analysis. MRM is different from Latent Class Analysis (LCA) in its approach to handling heterogeneity. LCA divides the population into different latent groups based on similar response patterns, while MRM explains heterogeneity in Rasch data by identifying subpopulations that have different Rasch parameters (item difficulties) (McCreary et al., 2013). The application of MRM in psychometric analysis is well-justified under evidence-based principles for evaluating internal structure, as recommended by the International Test Commission (ITC) and AERA guidelines (AERA et al., 2014; Gregoire, 2018). These guidelines underscore the importance of establishing valid and reliable internal structures while accounting for population heterogeneity. Unlike Differential Item Functioning (DIF), MRM is used to identify latent subpopulations in the data without requiring prior group information, hence the focus is on heterogeneity in the population (Rost, 1990). In contrast, DIF aims to test for item bias toward a specific known group (e.g., gender or culture) by comparing item responses after controlling ability (Khine, 2020).

MRM enables a more in-depth analysis by identifying latent subgroups within the population that may exhibit distinct response parameters (Vermunt & Magidson, 2002). This approach aligns with ITC and AERA recommendations to evaluate whether test items function equivalently across different groups. MRM thus offers empirical evidence of the presence of subpopulations with divergent latent structures that may go undetected by conventional Rasch models (McCreary et al., 2013). MRM helps ensure that measurement is not only appropriate for the overall population, but also fair and accurate for each subgroup—supporting the principles of fairness and equity in psychological testing (AERA et al., 2004; Gregoire, 2018).

## METHOD

### *Research Design*

The psychometric validation study employed a quantitative approach to comprehensively evaluate the BREQ-3 within the Indonesian context using Rasch Model analysis. The study was designed to generate an in-depth understanding of the structure and quality of exercise of motivation measurement, with a primary focus on examining the instrument's internal validity and reliability.

### *Participants*

The sample size in Rasch modeling is influenced by the instrument calibration principle. When an instrument is calibrated on a different sample, slight variations in results are expected. Therefore, a

small sample size may lead to unstable calibration and reduced sensitivity in capturing the true measurement properties. According to Linacre (1994), for Rasch analysis with a 99% confidence level, the sample size that can be used ranges from 108 to 243 participants. This study involved 408 Indonesian participants, aged 18-25 years, representing the emerging adulthood stage, with a mean age of 20.11 and a standard deviation of 1.92. The participants consisted of 308 females and 100 males.

A non-random sampling was used, meaning that not all individuals in the population had an equal chance of being selected. This approach was intended to obtain a sample relevant to the research purpose based on specific criteria (Creswell, 2014). Specifically, accidental sampling was applied by recruiting individuals who were available and willing to participate at the time of data collection (Etikan, Musa, & Alkassim, 2016). Additionally, snowball sampling was employed, whereby initial participants referred others who met the inclusion criteria (Biernacki & Waldorf, 1981). Data collection was conducted both online via Google Forms and offline through physical questionnaires, enabling broader coverage and increased sample size through engagement.

### Measurement

BREQ-3 is based on its previous version, Behavioral Regulation in Exercise Questionnaire (BREQ) developed by Markland and Tobin (2004). Wilson et al. (2006) later revised the instruments by adding a new dimension—integrated motivation—resulting in BREQ-3. Prior to its use in this study, the scale was translated into Indonesian. BREQ-3 measures six dimensions of motivation: external regulation, introjected regulation, identified regulation, integrated regulation, intrinsic motivation, and amotivation.

Table 1. Blueprint Scale

| Subscale              | Item Coding | Item                                                                                                                          |
|-----------------------|-------------|-------------------------------------------------------------------------------------------------------------------------------|
| Identified Regulation | A1          | <i>Penting bagi saya untuk berolahraga secara teratur. / It is important for me to exercise regularly.</i>                    |
|                       | A7          | <i>Saya menghargai manfaat berolahraga. / I value the benefits of exercise.</i>                                               |
|                       | A13         | <i>Menurut saya penting untuk melakukan upaya berolahraga secara teratur. / I think it's important to exercise regularly.</i> |
|                       | A19         | <i>Saya menjadi gelisah jika saya tidak berolahraga secara teratur. / I get anxious if I don't exercise regularly.</i>        |
| Amotivation           | B2          | <i>Saya tidak mengerti mengapa saya harus berolahraga. / I don't understand why I have to exercise.</i>                       |
|                       | B8          | <i>Saya tidak mengerti mengapa saya harus repot-repot berolahraga. / I don't understand why I even bother to exercise.</i>    |
|                       | B14         | <i>Saya tidak melihat pentingnya berolahraga. / I don't see the importance of exercise.</i>                                   |
|                       | B20         | <i>Menurut saya, berolahraga hanya membuang-buang waktu. / In my opinion, exercising is a waste of time.</i>                  |

|                        |     |                                                                                                                                                                                    |
|------------------------|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Intrinsic Regulation   | C3  | <i>Saya berolahraga karena itu menyenangkan. / I exercise because it's fun.</i>                                                                                                    |
|                        | C9  | <i>Saya menikmati sesi-sesi berolahraga saya. / I enjoy my exercise sessions.</i>                                                                                                  |
|                        | C15 | <i>Menurut saya, olahraga merupakan aktivitas yang menyenangkan. / For me, doing sports is fun.</i>                                                                                |
|                        | C21 | <i>Saya mendapatkan kesenangan dan kepuasan dari berpartisipasi dalam olahraga. / I find joy and satisfaction in engaging in sports.</i>                                           |
| Introjected Regulation | D4  | <i>Saya merasa bersalah jika tidak berolahraga. / I feel guilty if I skip exercise.</i>                                                                                            |
|                        | D10 | <i>Saya merasa malu ketika melewatkan suatu sesi berolahraga. / I feel ashamed when I skip a workout session.</i>                                                                  |
|                        | D16 | <i>Saya merasa gagal jika sudah lama tidak berolahraga. / I feel like a failure if I haven't exercised in a long time.</i>                                                         |
|                        | D22 | <i>Saya akan merasa buruk tentang diri saya sendiri jika saya tidak meluangkan waktu untuk berolahraga. / I would feel bad about myself if I didn't take the time to exercise.</i> |
| Integrated Regulation  | E5  | <i>Saya berolahraga karena sesuai dengan tujuan hidup saya. / I exercise because it matches my life goals.</i>                                                                     |
|                        | E11 | <i>Saya menganggap olahraga sebagai bagian dari identitas diri saya. / I consider sports to be part of my identity.</i>                                                            |
|                        | E17 | <i>Saya menganggap olahraga sebagai bagian mendasar dari diri saya. / I consider sports to be a fundamental part of my identity.</i>                                               |
|                        | E23 | <i>Saya menganggap olahraga sesuai dengan nilai-nilai saya. / I consider sports to be in line with my values.</i>                                                                  |
| External Regulation    | F6  | <i>Saya berolahraga karena orang lain mengatakan saya harus berolahraga. / I exercise because others tell me I should do so.</i>                                                   |
|                        | F12 | <i>Saya ikut berolahraga karena teman/keluarga/pasangan saya mengatakan saya harus mengikutinya. / I exercise because my friends/family/partner tell me.</i>                       |
|                        | F18 | <i>Saya berolahraga karena orang lain tidak akan senang dengan saya jika saya tidak melakukannya. / I exercise because others won't approve of me if I don't.</i>                  |
|                        | F24 | <i>Saya merasa mendapat tekanan dari teman/keluarga saya untuk berolahraga. / I feel pressure from my friends/family to exercise.</i>                                              |

(Wilson et al., 2006)

The research scale was constructed using 24 items with a 5-point Likert scale; ranging from 0 "Not Applicable" to 4 "Very Applicable." BREQ-3 scale does not include unfavorable items.

#### *Data Analysis*

In this study, Rasch analysis was applied to BREQ-3 using Rasch Rating Scale Model (RSM), which was selected based on the Likert-type response format of the questionnaire. The data analysis was conducted using WINSTEP® version 5.1.0 and Jamovi. The psychometric properties examined included instrument reliability, person and item reliability, unidimensionality, rating scale functioning, bias detection through Differential Item Functioning (DIF). Rasch analysis was conducted separately for each BREQ-3 subscale, in accordance with its multidimensional structure.

## RESULTS

### *Unidimensionality*

Table 2. Unidimensionality of BREQ-3

| Subscale               | Raw Variance Explained by Measure | First Contrast | Conclusion     |
|------------------------|-----------------------------------|----------------|----------------|
| Amotivation            | 52.5%                             | 1.6            | Unidimensional |
| External Regulation    | 59.1%                             | 1.6            | Unidimensional |
| Introjected Regulation | 61.6%                             | 1.5            | Unidimensional |
| Identified Regulation  | 65.1                              | 1.7            | Unidimensional |
| Integrated Regulation  | 71%                               | 1.5            | Unidimensional |
| Intrinsic Regulation   | 66%                               | 1.5            | Unidimensional |

This table presents the unidimensionality analysis of the BREQ-3 subscales. Each subscale was evaluated based on the raw variance explained by the measure and the size of the first contrast. The results indicate that each subscale meets the criteria for unidimensionality, suggesting that each one measures a single underlying construct.

### *Rating Scale Diagnostics*

Table 3. Rating Scale Diagnostics of BREQ-3

| Subscale    | Rating Scale       | <i>f</i> | %  | Average Measure | Infit MNSQ | Outfit MNSQ | Threshold |
|-------------|--------------------|----------|----|-----------------|------------|-------------|-----------|
| Amotivation | Very inappropriate | 807      | 49 | -2.40           | 1.27       | 1.10        | NONE      |
|             | Inappropriate      | 559      | 34 | -1.36           | 0.70       | 0.87        | -2.24     |
|             | Neutral            | 132      | 8  | -0.44           | 0.63       | 0.58        | 0.56      |
|             | Appropriate        | 83       | 5  | 0.35            | 0.96       | 1.01        | 0.43*     |
|             | Very appropriate   | 51       | 3  | 1.01            | 1.40       | 1.88        | 1.25      |

|                        |                    |     |    |       |      |      |       |
|------------------------|--------------------|-----|----|-------|------|------|-------|
| External Regulation    | Very inappropriate | 473 | 29 | -2.16 | 1.09 | 1.05 | NONE  |
|                        | Inappropriate      | 463 | 28 | -1.15 | 0.86 | 0.88 | -2.03 |
|                        | Neutral            | 269 | 16 | -0.18 | 0.83 | 0.86 | -0.12 |
|                        | Appropriate        | 284 | 17 | -0.65 | 0.96 | 0.98 | 0.11  |
|                        | Very appropriate   | 143 | 9  | 1.45  | 1.28 | 1.33 | 2.04  |
| Introjected Regulation | Very inappropriate | 189 | 12 | -2.19 | 1.16 | 1.12 | NONE  |
|                        | Inappropriate      | 366 | 22 | -1.09 | 0.83 | 0.82 | -2.46 |
|                        | Neutral            | 433 | 27 | 0.10  | 0.83 | 0.79 | -0.60 |
|                        | Appropriate        | 393 | 24 | 1.29  | 0.86 | 0.89 | 0.69  |
|                        | Very appropriate   | 251 | 15 | 2.15  | 1.34 | 1.30 | 2.37  |
| Identified Regulation  | Very inappropriate | 67  | 4  | -1.90 | 1.21 | 1.13 | NONE  |
|                        | Inappropriate      | 124 | 8  | -1.07 | 0.87 | 0.97 | -2.08 |
|                        | Neutral            | 238 | 15 | 0.29  | 0.91 | 0.89 | -0.94 |
|                        | Appropriate        | 629 | 39 | 1.85  | 0.88 | 0.93 | 0.09  |
|                        | Very appropriate   | 574 | 35 | 3.40  | 1.11 | 1.05 | 2.93  |
| Integrated Regulation  | Very inappropriate | 106 | 6  | -3.29 | 0.98 | 1.02 | NONE  |
|                        | Inappropriate      | 221 | 14 | -1.86 | 0.87 | 0.92 | -3.39 |
|                        | Neutral            | 384 | 24 | -0.12 | 0.79 | 0.78 | -1.55 |
|                        | Appropriate        | 572 | 35 | 2.35  | 0.96 | 1.00 | 0.61  |
|                        | Very appropriate   | 349 | 21 | 4.10  | 1.33 | 1.21 | 4.34  |
| Intrinsic Regulation   | Very inappropriate | 18  | 1  | -3.88 | 1.63 | 1.82 | NONE  |
|                        | Inappropriate      | 77  | 5  | -1.89 | 0.94 | 0.99 | -4.49 |
|                        | Neutral            | 285 | 17 | 0.44  | 0.86 | 0.82 | -1.94 |
|                        | Appropriate        | 719 | 44 | 3.50  | 0.93 | 0.95 | 1.06  |
|                        | Very appropriate   | 533 | 33 | 5.33  | 1.10 | 1.05 | 5.37  |

Table 3 presents the rating scale diagnostics for each BREQ-3 subscale. It includes data on the frequency (f) and percentage (%) of responses for each rating scale category (Very inappropriate, Inappropriate,

Neutral, Appropriate, Very appropriate), as well as the average measure, Infit Mean Square (MNSQ), Outfit MNSQ, and thresholds for each subscale.

### Reliability

Table 4. Instrument Reliability

| Subscale                    | A    | B    | C    | D     | D    | E    |
|-----------------------------|------|------|------|-------|------|------|
| N                           | 6    | 6    | 6    | 6     | 6    | 6    |
| Person Reliability          | 0.57 | 0.75 | 0.79 | 0.67  | 0.86 | 0.81 |
| Item Separation Reliability | 2.47 | 7.82 | 6.81 | 14.90 | 6.82 | 1.62 |
| Alpha Cronbach              | 0.86 | 0.98 | 0.98 | 1.00  | 0.98 | 0.72 |
| Person Separation Index     | 1.15 | 1.73 | 1.72 | 1.54  | 2.47 | 2.08 |

Table 4 presents the reliability statistics for each subscale of the instrument, including key indicators such as Person Reliability, Item Separation Reliability, Cronbach's Alpha, and Person Separation Index. Person Reliability indicates the consistency with which the instrument distinguishes between different participants. Higher values suggest better differentiation. Item Separation Reliability reflects the ability of the items within each subscale to differentiate between participants at different levels of the underlying construct. Higher values indicate more reliable item separation. Cronbach's Alpha evaluates internal consistency, indicating how closely related the items within a subscale are. Higher values suggest better consistency among items. The Person Separation Index is similar to Person Reliability but provides a more detailed estimate of the instrument's ability to differentiate among levels of respondent traits.

### Item Fit

Table 5. Item Calibration

| Subscale                                        | Item | Measure | Infit MNSQ | Outfit MNSQ | Point Measure Correlation |
|-------------------------------------------------|------|---------|------------|-------------|---------------------------|
| <b>Subscale A<br/>Identified<br/>Regulation</b> | A1   | -0.32   | 0.84       | 0.87        | 0.71                      |
|                                                 | A7   | -1.25   | 1.08       | 1.07        | 0.56                      |
|                                                 | A13  | -0.55   | 0.82       | 0.87        | 0.69                      |
|                                                 | A19  | 2.13    | 1.13       | 1.10        | 0.79                      |
| <b>Subscale B<br/>Amotivation</b>               | B2   | -0.23   | 0.85       | 0.84        | 0.77                      |
|                                                 | B8   | -0.19   | 1.01       | 1.05        | 0.76                      |
|                                                 | B14  | 0.13    | 1.06       | 1.03        | 0.72                      |
|                                                 | B20  | 0.30    | 1.14       | 1.12        | 0.70                      |
| <b>Subscale C<br/>Intrinsic</b>                 | C3   | 0.26    | 1.23       | 1.21        | 0.83                      |
|                                                 | C9   | -0.27   | 0.97       | 0.98        | 0.82                      |

|                                                  |     |       |      |      |       |
|--------------------------------------------------|-----|-------|------|------|-------|
| <b>Regulation</b>                                | C15 | -0.12 | 0.70 | 0.68 | 0.89* |
|                                                  | C21 | 0.14  | 1.04 | 1.04 | 0.84  |
| <b>Subscale D<br/>Introjected<br/>Regulation</b> | D4  | -0.19 | 0.93 | 0.95 | 0.79  |
|                                                  | D10 | 0.78  | 1.08 | 1.01 | 0.79  |
|                                                  | D16 | -0.56 | 1.13 | 1.08 | 0.77  |
|                                                  | D22 | -0.03 | 0.82 | 0.84 | 0.82  |
| <b>Subscale E<br/>Integrated<br/>Regulation</b>  | E5  | -0.60 | 1.23 | 1.29 | 0.80  |
|                                                  | E11 | 0.98  | 0.85 | 0.83 | 0.90* |
|                                                  | E17 | 0.11  | 0.81 | 0.82 | 0.89* |
|                                                  | E23 | -0.49 | 0.99 | 0.98 | 0.84  |
| <b>Subscale F<br/>External<br/>Regulation</b>    | F6  | -0.59 | 0.94 | 0.97 | 0.80  |
|                                                  | F12 | -0.50 | 0.80 | 0.79 | 0.82  |
|                                                  | F18 | 0.61  | 1.16 | 1.15 | 0.74  |
|                                                  | F24 | 0.48  | 1.13 | 1.08 | 0.76  |

Table 5 provides the item calibration results for each subscale of the instrument. It includes key measures such as the item difficulty (Measure), the fit of the items to the Rasch model (Infit and Outfit MNSQ), and the relationship between each item and the overall subscale (Point Measure Correlation).

#### *Wright Map*

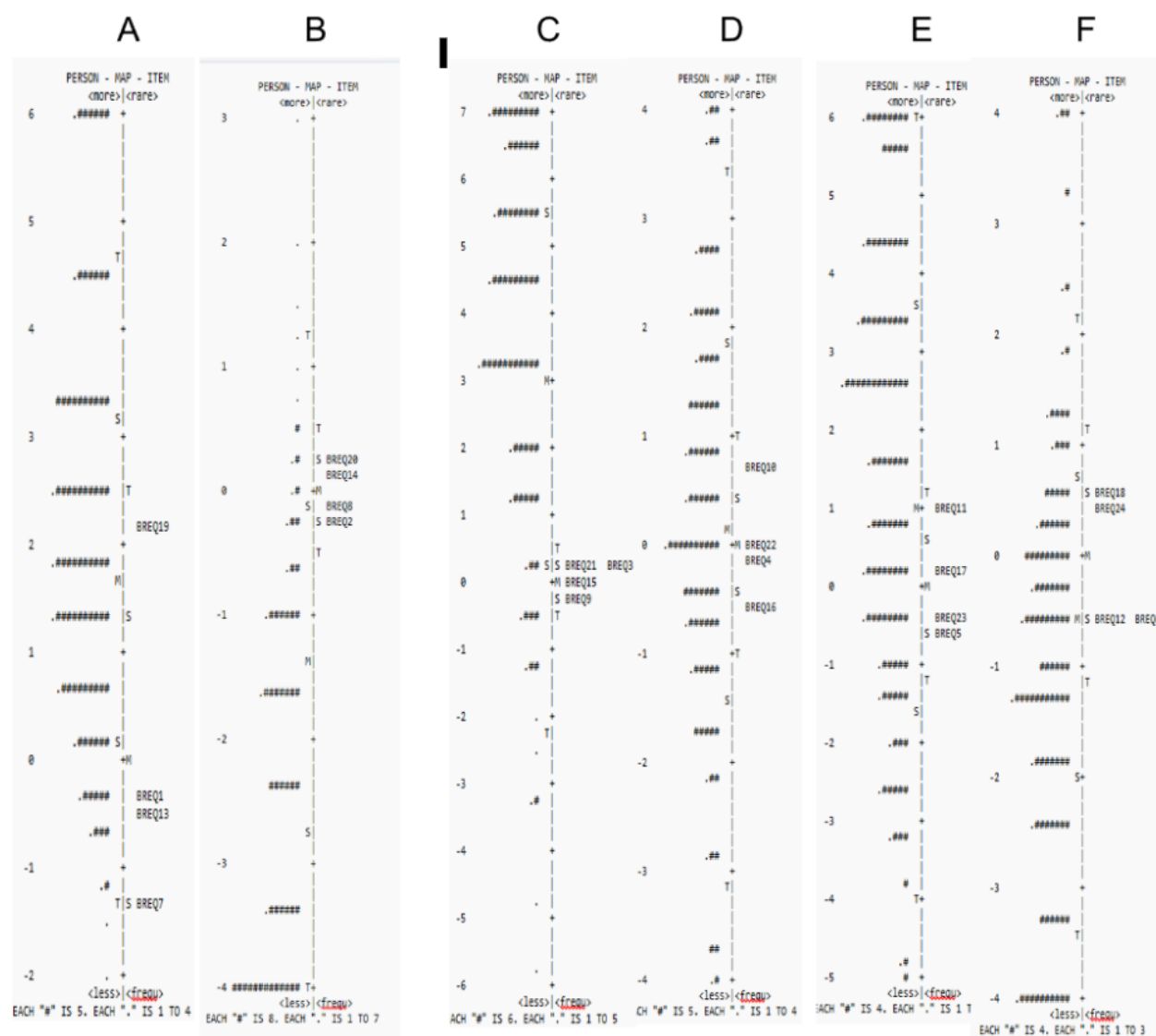


Figure 1. Wright Map

Wright Map is a graphical representation used in the Rasch measurement model to visualize the relationship between items and participants based on their responses. It shows how items and participants are distributed along a continuum of difficulty or ability.

#### Rasch Mixture Rasch Model

Table 6. Model Fit Information

| Class | AIC   | BIC   | CAIC  | Log-likelihood |
|-------|-------|-------|-------|----------------|
| 1     | 23786 | 24171 | 24267 | -11797         |
| 2     | 22151 | 22924 | 23117 | -10882         |
| 3     | 21774 | 22936 | 23226 | -10597         |

Table 6 presents the key model fit statistics for different classes in the Mixture Rasch Model (MRM). These statistics include the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Consistent Akaike Information Criterion (CAIC), and the log-likelihood value.

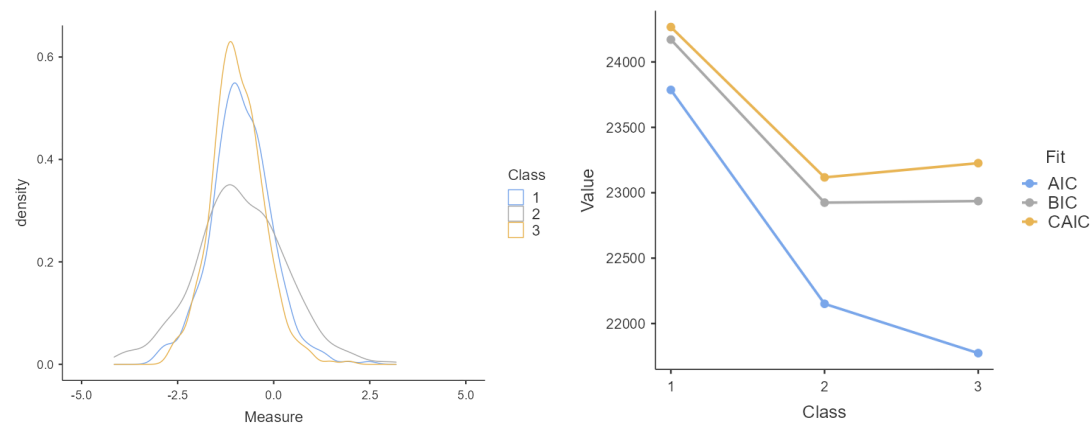


Figure 2. Person Distribution Across Class and Elbow Plot

Table 7. Frequencies of Genders

| Genders | Membership | Count | % of Total | Cumulative |
|---------|------------|-------|------------|------------|
| Male    | 1          | 18    | 4.4%       | 4.4%       |
|         | 2          | 7     | 1.7%       | 6.2%       |
|         | 3          | 74    | 18.1%      | 24.3%      |
| Female  | 1          | 103   | 25.2%      | 49.5%      |
|         | 2          | 69    | 16.9%      | 66.6%      |
|         | 3          | 136   | 33.4%      | 100%       |

This table presents the distribution of participants based on gender and their membership in different categories. The "Count" column indicates the number of participants in each category, while the "%" of Total column shows the percentage of participants within each gender category. The "Cumulative" column represents the cumulative percentage, summing up the percentages from the previous categories.

Table 8. Frequencies of Age

| Age | Membership | Count | % of Total | Cumulative |
|-----|------------|-------|------------|------------|
| 18  | 1          | 51    | 12.5%      | 12.5%      |
|     | 2          | 33    | 8.1%       | 20.6%      |
|     | 3          | 43    | 10.5%      | 31.1%      |
| 19  | 1          | 20    | 4.9%       | 36%        |
|     | 2          | 11    | 2.7%       | 38.7%      |

|    |   |    |       |       |
|----|---|----|-------|-------|
|    | 3 | 23 | 5.6%  | 44.3% |
| 20 | 1 | 16 | 3.9%  | 48.2% |
|    | 2 | 4  | 1.0%  | 49.2% |
|    | 3 | 28 | 6.9%  | 56.1% |
| 21 | 1 | 17 | 4.2%  | 60.3% |
|    | 2 | 21 | 5.1%  | 65.4% |
|    | 3 | 40 | 9.8%  | 75.2% |
| 22 | 1 | 8  | 2.0%  | 77.2% |
|    | 2 | 6  | 1.5%  | 78.7% |
|    | 3 | 47 | 11.5% | 90.2% |
| 23 | 1 | 3  | 0.7%  | 90.9% |
|    | 2 | 1  | 0.3%  | 91.2% |
|    | 3 | 14 | 3.3%  | 94.5% |
| 24 | 1 | 2  | 0.5%  | 95.0% |
|    | 2 | 0  | 0%    | 95.0% |
|    | 3 | 8  | 2.0%  | 97.0% |
| 25 | 1 | 4  | 1.0%  | 98.0% |
|    | 2 | 0  | 0%    | 98.0% |
|    | 3 | 8  | 2.0%  | 100%  |

This table presents the distribution of participants across different age groups and their corresponding membership categories. The "Count" column indicates the number of participants in each age category and membership combination, while the "% of Total" column represents the percentage of participants in that category. The "Cumulative" column tracks the cumulative percentage of participants, summing up the percentages across all categories.

## DISCUSSION

### *Unidimensionality*

The analysis in this study was performed separately for each dimension of exercise regulation, following the methodology outlined in the original study by Wilson et al. (2006). Unidimensionality for each subscale was evaluated based on the raw variance explained by the measure, determined through Rasch Principal Component Analysis of Residuals (PCAR). Holster and Lake (2016) argued that a raw variance explained greater than 40% is sufficient evidence of unidimensionality, while Linacre (2002) noted that a value exceeding 50% indicates an effective model. Furthermore, the eigenvalue of the first contrast should be below 2.0, as a value this low suggests that residuals reflect random noise rather than an additional dimension. The results of this study revealed that all subscales achieved a raw variance explained by measure above 50%, and the eigenvalue of the first contrast was less than 2.0. These

findings confirm the validity of the measurement model, demonstrating its ability to effectively explain data variation and confirming that each measurement dimension is unidimensional.

### *Rating Scale Diagnostics*

BREQ-3 employs a 5-point Likert scale, ranging from "very inappropriate" to "very appropriate." Diagnostic rating scale was used to evaluate how individuals make their choices and perceive the intervals between them. This approach allows for more accurate and interpretable assessment of the construct by capturing the actual distances applied when participants select among the available choices. Diagnostic results for each BREQ-3 subscale, as shown in Table 3, indicate that the responses categories functioned as intended. No response category had a frequency of 0 (zero) across any subscale. Furthermore, all subscales—except the amotivation subscale—exhibited properly ordered thresholds, progressing from negative to positive values across the five response options (Linacre, 2012). This suggests that participants utilize the response categories appropriately, despite one error in the amotivation subscale.

Rasch model also evaluates the reliability of both individuals (person reliability) and items (item reliability). Both person separation reliability (PSR) and item separation reliability (ISR) are interpreted similarly to Cronbach's  $\alpha$ , with a minimum value of approximately 0.80 indicating reliability. However, the person separation index (PSI), which uses a logit scale rather than raw scores, showed low reliability, with values ranging from 1.15 to 2.47. According to Linacre (2012), a PSI above 3.0 is necessary for high reliability.

As shown in Table 4, person reliability across the subscales ranged from 0.57 to 0.86. This indicates that subscales A and D, which measure identified regulation and introjected regulation, are only moderately effective in distinguishing individuals based on the constructs they assess. They may lack sensitivity in differentiating between high and low performers (Bond & Fox, 2013). In contrast, ISR ranged from 1.62 to 14.90, with Cronbach's  $\alpha$  values between 0.72 and 1.00. Separation indices ranged from 1.14 to 2.47 indicating low to moderate item-level separation. Overall, these results suggest that BREQ-3 demonstrates moderately high reliability and strong psychometric properties, particularly in terms of ISR values.

### *Item Fit*

The Rasch model evaluated the effectiveness of each item in measuring the construct by analyzing item infit, outfit, and point-measure correlation. Table 5 presented the results, sorted from the most difficult to the easiest items. For an item to be effective in measurement, the infit and outfit MNSQ values should fall within the range of 0.5 to 1.5, and the point-measure correlation should range from 0.40 to 0.85 (Fisher, 2007). This study found no misfitting items across all six subscales of the BREQ-3. The point-measure correlation values for the six subscales were positive, with most items meeting the criteria, except for C15, E11, and E17, which had values of 0.89, 0.89, and 0.90, respectively. In general, very high correlation between items (such as 0.89, 0.89, and 0.90) could indicate the potential that the item may be very similar to other items in the same subscale (redundancy), which can make the instrument less efficient. In addition, the items are prone to overfitting, where the items are too closely aligned with specific response patterns in the current sample, and may not generalize well across the broader population. The point-measure correlation ranges for each subscale are as follows: A (0.56–0.79), B (0.70–0.77), C (0.82–0.84), D (0.77–0.82), E (0.80–0.84), and F (0.74–0.82). These results indicate that all items in the instrument perform well in measuring the constructs of the BREQ-3, except for items C15, E11, and E17. These specific items should be revised to enhance the accuracy of the measurements.

### *Wright Map*

The construct validity can be assessed using the hierarchy of items displayed on a Wright Map. This map presents item difficulty on the right panel and person ability on the left panel. Easier items are positioned at the bottom, items with average difficulty are in the middle (denoted by "M" on the right side), and more difficult items are at the top (Yu, 2020). The Wright Maps for each BREQ-3 subscale are shown in Figures 1.

The Identified Regulation subscale (A) indicates that its items are relatively easy, with most individuals having higher ability than item difficulty, thereby reducing the scale's ability to differentiate individuals with higher levels of identified regulation. The Amotivation subscale (B) shows that the items are relatively difficult, with most individuals having lower ability than the item's difficulty. This reduces the scale's ability to distinguish individuals with lower levels of amotivation. In the Intrinsic Regulation subscale (C), all items were relatively easy, as indicated by the placement of most individuals above the mean value. The Introjected Regulation (D) and Integrated Regulation (E) subscale shows that participants' abilities are evenly distributed, while item difficulties are at a moderate level. This indicates that the scale is able to effectively distinguish individuals with different levels of introjected regulation. The External Regulation (F) subscale indicates that participants' abilities are predominantly at the medium to low level, while the difficulty of the dominant items is at the mean. This suggests that many items may be too difficult for a significant portion of participants, potentially limiting the scale sensitivity for measuring low levels of external regulation.

For suggestions, adding items with higher difficulty levels to the Identified Regulation (A) and Intrinsic Regulation (C) subscales to be more discriminatory for individuals with high identified regulation and intrinsic regulation. In addition, adding items with lower difficulty levels to the Amotivation (B) and External Regulation (F) subscales to be more effective in measuring individuals with low levels of amotivation and external regulation. This adjustment aims to increase the discriminatory power of the scale in measuring various levels of motivational regulation, so that the measurement results are more accurate and reflect differences between individuals more optimally.

#### *Differential Item Functioning*

Differential Item Functioning (DIF) analysis was conducted to evaluate whether subgroups within the sample (based on gender) responded differently to specific items. In this study, gender was classified into two subgroups: male and female. The item-trait chi-square method (Linacre, 2007) was applied to assess Differential Item Functioning (DIF). An item was considered significantly biased if its probability value was below 0.05. According to Zwick et al. (1999), a DIF contrast value below 0.43 is deemed negligible, values between 0.43 and 0.64 suggest slight to moderate bias, and values above 0.64 indicate significant bias. The results revealed that item F6 was biased toward gender ( $p = 0.0171$ ), suggesting that males and females might interpret the item "*Saya berolahraga karena orang lain mengatakan saya harus berolahraga*" / "I exercise because other people told me to" differently. This may reflect a greater tendency among males to report externally motivated exercise behavior. Therefore, it is necessary to be cautious in using item F6, as gender differences may affect its interpretation. Then, items C3 and C9 have a fairly high DIF contrast with values of -0.45 and 0.78, indicating that these items are suspected to cause bias in judgment, thus disrupting the validity of the measuring instrument.

#### *Mixture Rasch Model*

In addition, the study employed the Mixture Rasch Model to identify the number of Latent Classes in the dataset. Model evaluation was conducted using the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Consistent Akaike Information Criterion (CAIC). The AIC is a method based on model fit used to estimate the plausibility of a model in predicting future values. The BIC is another model selection criterion that assesses the trade-off between model fit and model complexity

(Mohammed et al., 2015). CAIC is a tool used to evaluate the consistency of the balance between model fit and complexity. As shown in Table 6, the model with three classes yielded the lowest values across all three criteria (AIC = 21.774; BIC = 22.936; CAIC = 23.226), compared to models with fewer classes. According to Mohammed et al. (2015), lower AIC and BIC values indicate way better model fit. Therefore, the three-class solution is considered the most optimal among the models tested. Figure 4 illustrates the density distribution of participants across the three classes. Furthermore, a re-analysis was carried out to determine the division of these three classes. The analysis was carried out by testing whether these three classes were determined by gender or age. Based on table 7 and 8, the distribution of class membership does not follow patterns based on gender or age. Thus, it can be ascertained that the three existing groups are not divided based on age or gender. Further analysis using a multidimensional Rasch model is recommended to explore potential interactions between latent dimensions and to deepen the understanding of the underlying structure.

## CONCLUSIONS

Several conclusions can be drawn from this study. The unidimensionality tests indicated that each subscale or dimension of BREQ-3 met the assumption of unidimensionality. The use of five response categories in each subscale functioned appropriately, with category thresholds increasing from negative to positive values—suggesting that the response options worked as intended, except in Amotivation subscale. This exception indicates potential ambiguity in how participants interpreted and rated items within the Amotivation dimension. Among all six subscales, only items C15, E11, and E17 were identified as misfitting, as their high point measure correlation suggests that these items may not be functioning optimally within the Rasch model framework. Following DIF analysis, we found that items F6, C3, and C9 are suspected of causing bias in the assessment, thus disrupting the validity of the measuring instrument.

These findings suggest that, aside from items C3, C9, C15, E11, E17, and F6, the remaining items effectively measured the construct of exercise regulation. The problematic items should be revised to improve the accuracy of the measurement. In terms of difficulty, all items in subscale A (Identified Regulation) were generally easy to moderate, while item B20 in subscale B was the most difficult. The distribution of responses in subscale C revealed that this subscale was relatively easy for participants, as few chose extreme responses. Another notable finding from DIF analysis was that item F6 demonstrated gender-related bias. Men tended to select higher ratings than women on this item, which aligns with the observation that men often exercise due to external motivation.

A limitation of this study lies in the data collection process, which may not fully represent emerging adults in Indonesia due to the use of a non-probability sampling method. However, because Rasch analysis does not rely heavily on sample characteristics, the findings regarding the measurement properties across the six BREQ-3 subscales may still be generalized with caution. Additionally, analysis using MRM indicated that three identified classes were not differentiated by gender or age. Future research could explore the characteristics and meaningful classification of these groups. Moreover, subsequent studies are encouraged to improve person reliability by revising problematic items and developing new ones—particularly subscale C—to enhance both person and item separation indices.

Further research is recommended to evaluate the validity and reliability of this instrument within specific populations or demographic groups. Such studies could provide a more comprehensive understanding of exercise regulation measurement in different contexts. Overall, the Indonesian version of BREQ-3 can be applied in future research and practical applications in organizations, with the considerations mentioned earlier.

## Properti Psikometri dan Analisis Campuran Model Rasch dari Body Regulation Exercise Questionnaire (BREQ-3)

Motivasi berolahraga adalah dorongan yang ada dalam diri seseorang dalam melakukan aktivitas olahraga untuk mencapai tujuan yang diinginkan yang merepresentasikan tingkat otonominya (Wilson dkk., 2002). Variabel motivasi berolahraga dapat diukur dengan skala *Behavioral Regulation in Exercise Questionnaire-3* (BREQ-3) oleh Wilson dkk. (2002) berdasarkan *Self-Determination Theory* (SDT) yang menggambarkan tingkat motivasi otonomi individu. Memahami regulasi perilaku dalam konteks olahraga merupakan aspek fundamental dalam psikologi olahraga dan promosi kesehatan. SDT telah menjadi kerangka teori yang dominan dalam memahami motivasi olahraga, dengan landasan bahwa kualitas motivasi lebih penting daripada kuantitasnya dalam memprediksi keberlanjutan perilaku olahraga (Teixeira dkk., 2012). BREQ-3 dikembangkan untuk mengukur spektrum regulasi motivasi berdasarkan SDT, mulai dari amotivasi hingga motivasi intrinsik. Nam dkk. (2023) dalam tinjauan sistematisnya menegaskan bahwa BREQ-3 merupakan instrumen yang paling banyak digunakan untuk menilai motivasi olahraga berdasarkan SDT dengan kemampuan prediksi perilaku aktivitas fisik yang baik. Hal ini didukung oleh analisis bibliometrik yang komprehensif tentang alat pengukuran motivasi dalam olahraga yang mengungkapkan lanskap instrumen pengukuran yang terus berkembang (Clancy dkk., 2017).

Hagger dan Chatzisarantis (2008) menyoroti pentingnya mengembangkan instrumen psikometrik yang dapat menangkap konstruk motivasi yang kompleks untuk berolahraga secara efektif. Tantangan utamanya adalah menciptakan alat yang dapat secara akurat membedakan antara berbagai jenis regulasi motivasi, dari regulasi eksternal hingga motivasi intrinsik. Chung dan Liu (2012) secara spesifik meneliti sifat psikometrik dari instrumen yang diterjemahkan dengan menekankan pada tantangan validasi lintas budaya. Beberapa penelitian telah menganalisis BREQ, seperti oleh Wilson dkk. (2002), yang secara kritis meninjau sifat psikometrik dari *Behavioral Regulation in Exercise Questionnaire* (BREQ) awal, mengidentifikasi kekuatan dan kelemahan dalam pendekatan pengukuran yang ada. Meskipun BREQ-3 telah divalidasi di berbagai negara, masih terdapat kesenjangan dalam memahami sifat-sifat psikometriksnya di berbagai konteks budaya yang berbeda. Rogowska dan Morouço (2024) menekankan pentingnya validasi lintas budaya untuk memastikan bahwa instrumen ini dapat secara akurat mengukur konstruk motivasi olahraga dalam konteks yang berbeda.

Hasil penelitian BREQ-3 versi Portugis telah diaplikasikan dengan analisis faktor konfirmatori menggunakan perangkat lunak AMOS 20.0 yang menunjukkan bahwa BREQ-3 versi Portugis valid dan reliabel setelah menghapus 6 dari 24 item (Cid dkk., 2018). Analisis faktor konfirmatori BREQ-3 versi Italia menggunakan perangkat lunak Mplus 8 versi 1.6 telah membuktikan bahwa BREQ-3 merupakan instrumen yang reliabel dan valid untuk mengukur regulasi perilaku olahraga pada individu dengan karakteristik demografi yang berbeda (Cavicchiolo dkk., 2022). Hasil analisis faktor konfirmatori BREQ-3M menggunakan Mplus 8 menunjukkan bahwa BREQ-3M versi Malaysia valid dan reliabel (Chai dkk., 2022). Versi BREQ-3 Cina saat ini telah menunjukkan bukti validitas dan reliabilitas yang memadai berdasarkan struktur dua faktornya, namun kami merekomendasikan untuk menghitung amotivasi, regulasi eksternal, regulasi yang dikenakan, dan skor motivasi otonom tunggal (tidak termasuk item 19) ketika menggunakan BREQ-3 Cina.

Terkait dengan validitas instrumen, Standar untuk Pengujian Pendidikan dan Psikologi yang diterbitkan oleh *American Educational Research Association* (AERA), *American Psychological Association* (APA), dan *National Council on Measurement in Education* (NCME), memberikan pedoman komprehensif untuk pengembangan tes dan menekankan pentingnya mengevaluasi kualitas praktik pengujian (AERA dkk., 2004). Menurut Standar AERA dkk. (1999), validitas didefinisikan sebagai “sejauh mana bukti dan teori mendukung interpretasi skor tes untuk penggunaan yang diusulkan dari tes tersebut.” Standar tersebut

juga menguraikan lima sumber bukti yang dapat digunakan untuk mengevaluasi validitas interpretasi skor tes untuk tujuan tertentu. Lima sumber tersebut terdiri dari bukti validitas berdasarkan: konten tes, proses respons, struktur internal, hubungan dengan variabel lain, dan konsekuensi dari pengujian.

Di Indonesia, belum ada studi komprehensif yang mengevaluasi sifat psikometrik BREQ-3 menggunakan pendekatan pengukuran modern seperti model Rasch. Chemolli dan Gagné (2014) menunjukkan bahwa analisis Rasch memberikan pemahaman yang lebih mendalam tentang struktur pengukuran motivasi dibandingkan dengan teori tes klasik. Hal ini sejalan dengan tujuan utama studi ini, yaitu memperkuat bukti validitas struktur internal sesuai rekomendasi AERA dan organisasi profesional terkait lainnya. Model Rasch dipilih karena kemampuannya untuk: (1) mengevaluasi unidimensionalitas setiap subskala BREQ-3, (2) mengidentifikasi item yang tidak sesuai melalui analisis kesesuaian item, dan (3) menilai invarian pengukuran di antara kelompok demografis. Pietrabissa dkk. (2020) menekankan bahwa analisis Rasch memberikan bukti validitas yang lebih kuat untuk BREQ-3, terutama dalam mendeteksi bias item dan memastikan kesetaraan pengukuran.

Studi ini juga menggunakan *Mixture Rasch Model* (MRM) untuk analisis lebih lanjut. MRM berbeda dengan *Latent Class Analysis* (LCA) dalam pendekatannya terhadap penanganan heterogenitas. LCA membagi populasi menjadi kelompok laten yang berbeda berdasarkan pola respons yang serupa, sementara MRM menjelaskan heterogenitas dalam data Rasch dengan mengidentifikasi subpopulasi yang memiliki parameter Rasch yang berbeda (kesulitan item) (McCreary dkk., 2013). Penerapan MRM dalam analisis psikometrik sangat dibenarkan berdasarkan prinsip-prinsip berbasis bukti untuk mengevaluasi struktur internal, sebagaimana direkomendasikan oleh Komisi Uji Internasional (ITC) dan pedoman AERA (AERA dkk., 2014; Gregoire, 2018). Pedoman ini menekankan pentingnya menetapkan struktur internal yang valid dan reliabel sambil memperhitungkan heterogenitas populasi. Berbeda dengan *Differential Item Functioning* (DIF), MRM digunakan untuk mengidentifikasi subpopulasi laten dalam data tanpa memerlukan informasi kelompok sebelumnya, sehingga fokusnya adalah pada heterogenitas dalam populasi (Rost, 1990). Di sisi lain, DIF bertujuan untuk menguji bias item terhadap kelompok tertentu yang diketahui (misalnya, gender atau budaya) dengan membandingkan respons item setelah mengontrol kemampuan (Khine, 2020).

MRM memungkinkan analisis yang lebih mendalam dengan mengidentifikasi subkelompok laten dalam populasi yang mungkin menunjukkan parameter respons yang berbeda (Vermunt & Magidson, 2002). Pendekatan ini sejalan dengan rekomendasi ITC dan AERA untuk mengevaluasi apakah item tes berfungsi secara ekuivalen di berbagai kelompok. Dengan demikian, MRM menyediakan bukti empiris tentang keberadaan subpopulasi dengan struktur laten yang berbeda yang mungkin tidak terdeteksi oleh model Rasch konvensional (McCreary dkk., 2013). MRM membantu memastikan bahwa pengukuran tidak hanya sesuai untuk populasi secara keseluruhan tetapi juga adil dan akurat untuk setiap subkelompok—mendukung prinsip keadilan dan kesetaraan dalam pengujian psikologis (AERA dkk., 2004; Gregoire, 2018).

## METODE

### *Desain Penelitian*

Penelitian ini merupakan studi validasi psikometrik dengan pendekatan kuantitatif, yang secara komprehensif bertujuan untuk mengevaluasi BREQ-3 dalam konteks Indonesia dengan menggunakan analisis model Rasch. Penelitian ini dirancang untuk menghasilkan pemahaman yang mendalam mengenai struktur dan kualitas pengukuran motivasi berolahraga, dengan fokus utama pada pengujian validitas internal dan reliabilitas instrumen.

### *Partisipan*

Ukuran sampel dalam model Rasch dipengaruhi oleh prinsip kalibrasi instrumen. Ketika instrumen dikalibrasi pada sampel yang berbeda, variasi kecil dalam hasil dapat diharapkan. Oleh karena itu, ukuran sampel yang kecil dapat menyebabkan kalibrasi yang tidak stabil dan sensitivitas yang berkurang dalam menangkap sifat pengukuran yang sebenarnya. Menurut Linacre (1994), untuk analisis Rasch dengan tingkat kepercayaan 99%, ukuran sampel yang dapat digunakan berkisar antara 108 hingga 243 partisipan. Studi ini melibatkan 408 partisipan dari Indonesia, berusia 18-25 tahun, yang mewakili tahap dewasa awal, dengan rata-rata usia 20,11 dan standar deviasi 1,92. Partisipan terdiri dari 308 perempuan dan 100 laki-laki.

Pengambilan sampel dilakukan secara tidak acak, yang berarti tidak semua individu dalam populasi memiliki peluang yang sama untuk dipilih. Pendekatan ini dimaksudkan untuk mendapatkan sampel yang relevan dengan tujuan penelitian berdasarkan kriteria tertentu (Creswell, 2014). Secara khusus, pengambilan sampel secara kebetulan dilakukan dengan merekrut individu yang bersedia berpartisipasi pada saat pengumpulan data (Etikan, Musa, & Alkassim, 2016). Selain itu, digunakan juga metode *snowball sampling*, di mana partisipan awal merekomendasikan orang lain yang memenuhi kriteria inklusi (Biernacki & Waldorf, 1981). Pengumpulan data dilakukan secara online melalui Google Forms dan secara offline melalui kuesioner fisik, yang memungkinkan cakupan yang lebih luas dan peningkatan ukuran sampel melalui partisipasi.

### Pengukuran

BREQ-3 dikembangkan berdasarkan versi sebelumnya yaitu *Behavioral Regulation in Exercise Questionnaire* (BREQ) oleh Markland dan Tobin (2004). Kemudian, Wilson dkk. (2006) menambahkan dimensi baru yaitu motivasi yang terintegrasi, sehingga menghasilkan versi terbaru. Skala ini telah diterjemahkan ke dalam bahasa Indonesia sebelum digunakan. BREQ-3 mengukur enam dimensi motivasi: regulasi eksternal, regulasi yang terinternalisasi, regulasi teridentifikasi, regulasi terintegrasi, motivasi intrinsik, dan amotivasi.

Tabel 1. *Blueprint* Skala

| Subskala                 | Koding Item | Item                                                                          |
|--------------------------|-------------|-------------------------------------------------------------------------------|
| Regulasi teridentifikasi | A1          | <i>Penting bagi saya untuk berolahraga secara teratur.</i>                    |
|                          | A7          | <i>Saya menghargai manfaat berolahraga.</i>                                   |
|                          | A13         | <i>Menurut saya penting untuk melakukan upaya berolahraga secara teratur.</i> |
|                          | A19         | <i>Saya menjadi gelisah jika saya tidak berolahraga secara teratur.</i>       |
| Amotivasi                | B2          | <i>Saya tidak mengerti mengapa saya harus berolahraga.</i>                    |
|                          | B8          | <i>Saya tidak mengerti mengapa saya harus repot-repot berolahraga.</i>        |
|                          | B14         | <i>Saya tidak melihat pentingnya berolahraga.</i>                             |
|                          | B20         | <i>Menurut saya, berolahraga hanya membuang-buang waktu.</i>                  |
| Regulasi Intrinsik       | C3          | <i>Saya berolahraga karena itu menyenangkan.</i>                              |
|                          | C9          | <i>Saya menikmati sesi-sesi berolahraga saya.</i>                             |
|                          | C15         | <i>Menurut saya, olahraga merupakan aktivitas yang menyenangkan.</i>          |

|                       |     |                                                                                                             |
|-----------------------|-----|-------------------------------------------------------------------------------------------------------------|
| Regulasi Introjeksi   | C21 | <i>Saya mendapatkan kesenangan dan kepuasan dari berpartisipasi dalam olahraga.</i>                         |
|                       | D4  | <i>Saya merasa bersalah jika tidak berolahraga.</i>                                                         |
|                       | D10 | <i>Saya merasa malu ketika melewatkan suatu sesi berolahraga.</i>                                           |
|                       | D16 | <i>Saya merasa gagal jika sudah lama tidak berolahraga.</i>                                                 |
|                       | D22 | <i>Saya akan merasa buruk tentang diri saya sendiri jika saya tidak meluangkan waktu untuk berolahraga.</i> |
| Regulasi Terintegrasi | E5  | <i>Saya berolahraga karena sesuai dengan tujuan hidup saya.</i>                                             |
|                       | E11 | <i>Saya menganggap olahraga sebagai bagian dari identitas diri saya.</i>                                    |
|                       | E17 | <i>Saya menganggap olahraga sebagai bagian mendasar dari diri saya.</i>                                     |
|                       | E23 | <i>Saya menganggap olahraga sesuai dengan nilai-nilai saya.</i>                                             |
| Regulasi Eksternal    | F6  | <i>Saya berolahraga karena orang lain mengatakan saya harus berolahraga.</i>                                |
|                       | F12 | <i>Saya ikut berolahraga karena teman/keluarga/pasangan saya mengatakan saya harus mengikutinya.</i>        |
|                       | F18 | <i>Saya berolahraga karena orang lain tidak akan senang dengan saya jika saya tidak melakukannya.</i>       |
|                       | F24 | <i>Saya merasa mendapat tekanan dari teman/keluarga saya untuk berolahraga.</i>                             |

(Wilson dkk., 2006)

Skala penelitian dibuat dengan menggunakan 24 item dengan skala Likert 5 poin; mulai dari 0 “Tidak sesuai” hingga 4 “Sangat sesuai.” Skala BREQ-3 tidak memuat item *unfavorable*.

#### Analisis Data

Dalam studi ini, analisis Rasch diterapkan pada BREQ-3 menggunakan Model Skala Penilaian Rasch (RSM), yang dipilih berdasarkan format respons tipe Likert pada kuesioner. Analisis data dilakukan menggunakan WINSTEP® versi 5.1.0 dan Jamovi. Sifat-sifat psikometrik yang dievaluasi meliputi reliabilitas instrumen, reliabilitas individu dan item, unidimensionalitas, fungsi skala penilaian, serta deteksi bias melalui *Differential Item Functioning* (DIF). Analisis Rasch dilakukan secara terpisah untuk setiap subskala BREQ-3, sesuai dengan struktur multidimensinya.

## HASIL PENELITIAN

### Unidimensionalitas

Tabel 2. Unidimensionalitas BREQ-3

| Subskala           | Raw Variance Explained by Measure | First Contrast | Simpulan   |
|--------------------|-----------------------------------|----------------|------------|
| Amotivasi          | 52,5%                             | 1,6            | Unidimensi |
| Regulasi Eksternal | 59,1%                             | 1,6            | Unidimensi |

|                          |       |     |            |
|--------------------------|-------|-----|------------|
| Regulasi Introjeksi      | 61,6% | 1,5 | Unidimensi |
| Regulasi Teridentifikasi | 65,1  | 1,7 | Unidimensi |
| Regulasi Terintegrasi    | 71%   | 1,5 | Unidimensi |
| Regulasi Intrinsik       | 66%   | 1,5 | Unidimensi |

Tabel ini menyajikan analisis unidimensionalitas subskala BREQ-3. Setiap subskala dinilai berdasarkan varians mentah yang dijelaskan oleh ukuran dan kontras pertama. Berdasarkan nilai yang diberikan, kesimpulan untuk setiap subskala diambil sebagai “Unidimensi,” yang menunjukkan bahwa setiap subskala mengukur satu konstruk yang mendasari.

#### *Diagnostik Skala Penilaian*

Tabel 3. Skala Penilaian Diagnostik BREQ-3

| Subskala            | Skala Penilaian     | <i>f</i> | %  | <i>Average Measure</i> | <i>Infit MNSQ</i> | <i>Outfit MNSQ</i> | <i>Threshold</i> |
|---------------------|---------------------|----------|----|------------------------|-------------------|--------------------|------------------|
| Amotivasi           | Sangat tidak sesuai | 807      | 49 | -2,40                  | 1,27              | 1,10               | NONE             |
|                     | Tidak sesuai        | 559      | 34 | -1,36                  | 0,70              | 0,87               | -2,24            |
|                     | Netral              | 132      | 8  | -0,44                  | 0,63              | 0,58               | 0,56             |
|                     | Sesuai              | 83       | 5  | 0,35                   | 0,96              | 1,01               | 0,43*            |
|                     | Sangat sesuai       | 51       | 3  | 1,01                   | 1,40              | 1,88               | 1,25             |
| Regulasi Eksternal  | Sangat tidak sesuai | 473      | 29 | -2,16                  | 1,09              | 1,05               | NONE             |
|                     | Tidak sesuai        | 463      | 28 | -1,15                  | 0,86              | 0,88               | -2,03            |
|                     | Netral              | 269      | 16 | -0,18                  | 0,83              | 0,86               | -0,12            |
|                     | Sesuai              | 284      | 17 | -0,65                  | 0,96              | 0,98               | 0,11             |
|                     | Sangat sesuai       | 143      | 9  | 1,45                   | 1,28              | 1,33               | 2,04             |
| Regulasi Introjeksi | Sangat tidak sesuai | 189      | 12 | -2,19                  | 1,16              | 1,12               | NONE             |
|                     | Tidak sesuai        | 366      | 22 | -1,09                  | 0,83              | 0,82               | -2,46            |
|                     | Netral              | 433      | 27 | 0,10                   | 0,83              | 0,79               | -0,60            |
|                     | Sesuai              | 393      | 24 | 1,29                   | 0,86              | 0,89               | 0,69             |

|                          |                     |     |    |       |      |      |       |
|--------------------------|---------------------|-----|----|-------|------|------|-------|
|                          | Sangat sesuai       | 251 | 15 | 2,15  | 1,34 | 1,30 | 2,37  |
| Regulasi Teridentifikasi | Sangat tidak sesuai | 67  | 4  | -1,90 | 1,21 | 1,13 | NONE  |
|                          | Tidak sesuai        | 124 | 8  | -1,07 | 0,87 | 0,97 | -2,08 |
|                          | Netral              | 238 | 15 | 0,29  | 0,91 | 0,89 | -0,94 |
|                          | Sesuai              | 629 | 39 | 1,85  | 0,88 | 0,93 | 0,09  |
|                          | Sangat sesuai       | 574 | 35 | 3,40  | 1,11 | 1,05 | 2,93  |
| Regulasi Terintegrasi    | Sangat tidak sesuai | 106 | 6  | -3,29 | 0,98 | 1,02 | NONE  |
|                          | Tidak sesuai        | 221 | 14 | -1,86 | 0,87 | 0,92 | -3,39 |
|                          | Netral              | 384 | 24 | -0,12 | 0,79 | 0,78 | -1,55 |
|                          | Sesuai              | 572 | 35 | 2,35  | 0,96 | 1,00 | 0,61  |
|                          | Sangat sesuai       | 349 | 21 | 4,10  | 1,33 | 1,21 | 4,34  |
| Regulasi Intrinsik       | Sangat tidak sesuai | 18  | 1  | -3,88 | 1,63 | 1,82 | NONE  |
|                          | Tidak sesuai        | 77  | 5  | -1,89 | 0,94 | 0,99 | -4,49 |
|                          | Netral              | 285 | 17 | 0,44  | 0,86 | 0,82 | -1,94 |
|                          | Sesuai              | 719 | 44 | 3,50  | 0,93 | 0,95 | 1,06  |
|                          | Sangat sesuai       | 533 | 33 | 5,33  | 1,10 | 1,05 | 5,37  |

Tabel 3 menyajikan diagnostik skala penilaian untuk setiap subskala BREQ-3. Tabel ini mencakup data frekuensi (f) dan persentase (%) tanggapan untuk setiap kategori skala penilaian (Sangat tidak sesuai, Tidak sesuai, Netral, Sesuai, Sangat sesuai), serta ukuran rata-rata, Infit Mean Square (MNSQ), Outfit MNSQ, dan ambang batas untuk setiap subskala.

#### Reliabilitas

Tabel 4. Reliabilitas Instrumen

| Subskala                  | A    | B    | C    | D    | D    | E    |
|---------------------------|------|------|------|------|------|------|
| <i>N</i>                  | 6    | 6    | 6    | 6    | 6    | 6    |
| <i>Person Reliability</i> | 0,57 | 0,75 | 0,79 | 0,67 | 0,86 | 0,81 |

|                                    |      |      |      |       |      |      |
|------------------------------------|------|------|------|-------|------|------|
| <i>Item Separation Reliability</i> | 2,47 | 7,82 | 6,81 | 14,90 | 6,82 | 1,62 |
| <i>Alpha Cronbach</i>              | 0,86 | 0,98 | 0,98 | 1,00  | 0,98 | 0,72 |
| <i>Person Separation Index</i>     | 1,15 | 1,73 | 1,72 | 1,54  | 2,47 | 2,08 |

Tabel 4 menyajikan statistik reliabilitas untuk setiap subskala instrumen, termasuk beberapa ukuran utama: Reliabilitas Individu, Reliabilitas Pemisahan Item, Alpha Cronbach, dan Indeks Pemisahan Individu. Reliabilitas Individu menunjukkan konsistensi instrumen dalam membedakan antara partisipan yang berbeda. Nilai yang lebih tinggi menunjukkan diferensiasi yang lebih baik. Reliabilitas Pemisahan Item mencerminkan kemampuan item dalam setiap subskala untuk membedakan antara partisipan pada level yang berbeda dari konstruk yang mendasarinya. Nilai yang lebih tinggi menunjukkan pemisahan item yang lebih andal. Alpha Cronbach mengukur konsistensi internal, yang menunjukkan seberapa erat hubungan antara item-item dalam subskala. Nilai yang lebih tinggi menunjukkan konsistensi yang lebih baik antar item. Person Separation Index mirip dengan Person Reliability, namun memberikan perkiraan yang lebih terperinci tentang kemampuan instrumen untuk membedakan tingkat sifat-sifat responden.

#### *Item Fit*

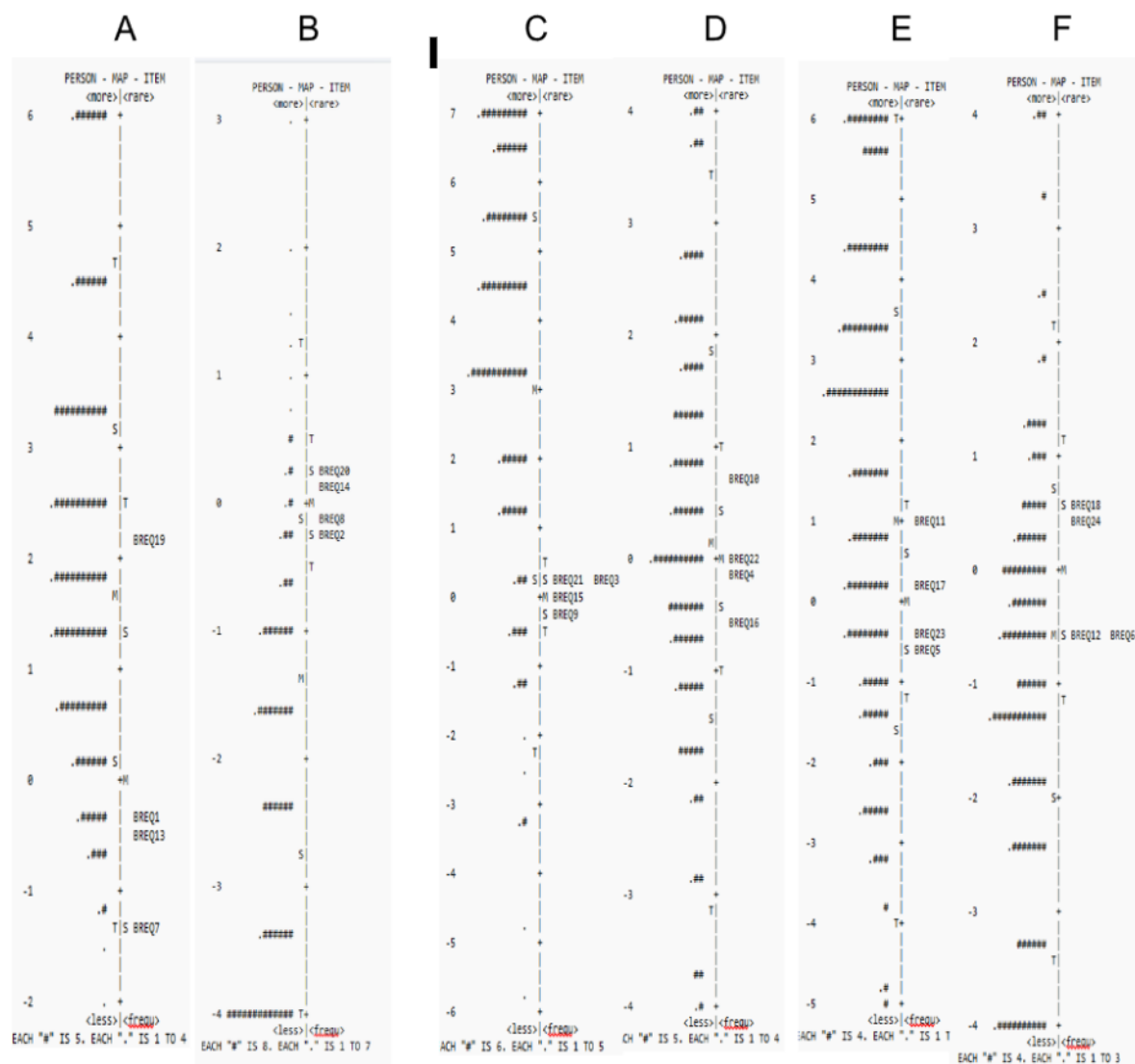
Table 5. Kalibrasi Item

| Subskala                                   | Item | Measure | Infit MNSQ | Outfit MNSQ | Point Measure Correlation |
|--------------------------------------------|------|---------|------------|-------------|---------------------------|
| <b>Subskala A Regulasi Teridentifikasi</b> | A1   | -0,32   | 0,84       | 0,87        | 0,71                      |
|                                            | A7   | -1,25   | 1,08       | 1,07        | 0,56                      |
|                                            | A13  | -0,55   | 0,82       | 0,87        | 0,69                      |
|                                            | A19  | 2,13    | 1,13       | 1,10        | 0,79                      |
| <b>Subskala B Amotivasi</b>                | B2   | -0,23   | 0,85       | 0,84        | 0,77                      |
|                                            | B8   | -0,19   | 1,01       | 1,05        | 0,76                      |
|                                            | B14  | 0,13    | 1,06       | 1,03        | 0,72                      |
|                                            | B20  | 0,30    | 1,14       | 1,12        | 0,70                      |
| <b>Subskala C Regulasi Intrinsik</b>       | C3   | 0,26    | 1,23       | 1,21        | 0,83                      |
|                                            | C9   | -0,27   | 0,97       | 0,98        | 0,82                      |
|                                            | C15  | -0,12   | 0,70       | 0,68        | 0,89*                     |
|                                            | C21  | 0,14    | 1,04       | 1,04        | 0,84                      |
| <b>Subskala D Regulasi Introjeksi</b>      | D4   | -0,19   | 0,93       | 0,95        | 0,79                      |
|                                            | D10  | 0,78    | 1,08       | 1,01        | 0,79                      |

|                                         |     |       |      |      |       |
|-----------------------------------------|-----|-------|------|------|-------|
|                                         | D16 | -0,56 | 1,13 | 1,08 | 0,77  |
|                                         | D22 | -0,03 | 0,82 | 0,84 | 0,82  |
| <b>Subskala E Regulasi Terintegrasi</b> | E5  | -0,60 | 1,23 | 1,29 | 0,80  |
|                                         | E11 | 0,98  | 0,85 | 0,83 | 0,90* |
|                                         | E17 | 0,11  | 0,81 | 0,82 | 0,89* |
|                                         | E23 | -0,49 | 0,99 | 0,98 | 0,84  |
| <b>Subskala F Regulasi Eksternal</b>    | F6  | -0,59 | 0,94 | 0,97 | 0,80  |
|                                         | F12 | -0,50 | 0,80 | 0,79 | 0,82  |
|                                         | F18 | 0,61  | 1,16 | 1,15 | 0,74  |
|                                         | F24 | 0,48  | 1,13 | 1,08 | 0,76  |

Tabel 5 menyajikan hasil kalibrasi item untuk setiap subskala instrumen. Tabel ini mencakup ukuran-ukuran kunci seperti tingkat kesulitan item (*Measure*), kecocokan item dengan model Rasch (*Infit and Outfit MNSQ*), dan hubungan antara setiap item dengan keseluruhan subskala (*Point Measure Correlation*).

*Wright Map*



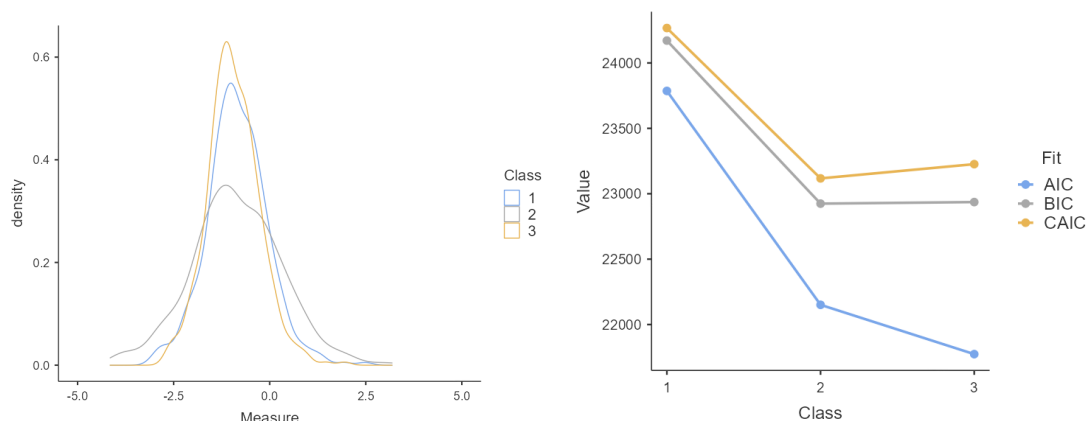
Gambar 1. Wright Map

*Rasch Mixture Rasch Model*

Tabel 6. Informasi Model Fit

| Class | AIC   | BIC   | CAIC  | Log-likelihood |
|-------|-------|-------|-------|----------------|
| 1     | 23786 | 24171 | 24267 | -11797         |
| 2     | 22151 | 22924 | 23117 | -10882         |
| 3     | 21774 | 22936 | 23226 | -10597         |

Tabel 6 menyajikan statistik kecocokan model utama untuk kelas-kelas yang berbeda dalam *Rasch Mixture Rasch Model (RM)*. Statistik ini termasuk *Akaike Information Criterion (AIC)*, *Bayesian Information Criterion (BIC)*, *Consistent Akaike Information Criterion (CAIC)*, dan *log-likelihood*.

Gambar 2. Distribusi Orang di Seluruh Kelas dan *Elbow Plot*

Tabel 7. Frekuensi Jenis Kelamin

| Jenis Kelamin | Keanggotaan | Hitungan | % dari Total | Kumulatif |
|---------------|-------------|----------|--------------|-----------|
| Laki-laki     | 1           | 18       | 4,4%         | 4,4%      |
|               | 2           | 7        | 1,7%         | 6,2%      |
|               | 3           | 74       | 18,1%        | 24,3%     |
| Perempuan     | 1           | 103      | 25,2%        | 49,5%     |
|               | 2           | 69       | 16,9%        | 66,6%     |
|               | 3           | 136      | 33,4%        | 100%      |

Tabel ini menyajikan distribusi partisipan berdasarkan jenis kelamin dan keanggotaan mereka dalam berbagai kategori. Kolom "Jumlah" menunjukkan jumlah partisipan dalam setiap kategori, sedangkan kolom "%" dari total menunjukkan persentase partisipan dalam setiap kategori jenis kelamin. Kolom "Kumulatif" menunjukkan persentase kumulatif, yang merupakan penjumlahan persentase dari kategori-kategori sebelumnya.

Tabel 8. Frekuensi Usia

| Age | Keanggotaan | Hitungan | % dari Total | Kumulatif |
|-----|-------------|----------|--------------|-----------|
| 18  | 1           | 51       | 12,5%        | 12,5%     |
|     | 2           | 33       | 8,1%         | 20,6%     |
|     | 3           | 43       | 10,5%        | 31,1%     |
| 19  | 1           | 20       | 4,9%         | 36%       |
|     | 2           | 11       | 2,7%         | 38,7%     |
|     | 3           | 23       | 5,6%         | 44,3%     |
| 20  | 1           | 16       | 3,9%         | 48,2%     |
|     | 2           | 4        | 1,0%         | 49,2%     |

|    |   |    |       |       |
|----|---|----|-------|-------|
|    | 3 | 28 | 6,9%  | 56,1% |
| 21 | 1 | 17 | 4,2%  | 60,3% |
|    | 2 | 21 | 5,1%  | 65,4% |
|    | 3 | 40 | 9,8%  | 75,2% |
| 22 | 1 | 8  | 2,0%  | 77,2% |
|    | 2 | 6  | 1,5%  | 78,7% |
|    | 3 | 47 | 11,5% | 90,2% |
| 23 | 1 | 3  | 0,7%  | 90,9% |
|    | 2 | 1  | 0,3%  | 91,2% |
|    | 3 | 14 | 3,3%  | 94,5% |
| 24 | 1 | 2  | 0,5%  | 95,0% |
|    | 2 | 0  | 0%    | 95,0% |
|    | 3 | 8  | 2,0%  | 97,0% |
| 25 | 1 | 4  | 1,0%  | 98,0% |
|    | 2 | 0  | 0%    | 98,0% |
|    | 3 | 8  | 2,0%  | 100%  |

Tabel ini menunjukkan distribusi partisipan di berbagai kelompok usia dan kategori keanggotaan yang sesuai. Kolom “Jumlah” menunjukkan jumlah partisipan di setiap kategori usia dan kombinasi keanggotaan, sedangkan kolom “% dari Total” menunjukkan persentase partisipan di kategori tersebut. Kolom “Kumulatif” melacak persentase kumulatif partisipan, dengan menjumlahkan persentase di semua kategori.

## DISKUSI

### *Unidimensionalitas*

Analisis dalam penelitian ini dilakukan secara terpisah untuk setiap dimensi regulasi olahraga, mengikuti metodologi yang diuraikan dalam penelitian oleh Wilson dkk. (2006). Unidimensionalitas untuk setiap subskala dievaluasi berdasarkan varians mentah yang dijelaskan oleh ukuran tersebut, yang ditentukan melalui *Rasch Principal Component Analysis of Residual (PCAR)*. Holster dan Lake (2016) berpendapat bahwa varians mentah yang dijelaskan lebih besar dari 40% adalah bukti yang cukup untuk unidimensionalitas, sementara Linacre (2002) mencatat bahwa nilai yang melebihi 50% mengindikasikan model yang efektif. Lebih lanjut, nilai eigen dari kontras pertama harus berada di bawah 2,0, karena nilai serendah ini menunjukkan bahwa residual mencerminkan *noise* acak daripada dimensi tambahan. Hasil dari penelitian ini menunjukkan bahwa semua subskala mencapai varians mentah yang dijelaskan oleh ukuran di atas 50%, dan nilai eigen dari kontras pertama kurang dari 2,0. Temuan ini mengonfirmasi validitas model pengukuran yang digunakan, menunjukkan kemampuannya untuk menjelaskan variasi data secara efektif dan mengonfirmasi unidimensionalitas setiap dimensi pengukuran.

### *Diagnostik Skala Penilaian*

BREQ-3 menggunakan skala Likert 5 poin, merepresentasikan respons dari “sangat tidak sesuai” hingga “sangat sesuai.” Skala penilaian diagnostik digunakan untuk mengevaluasi bagaimana individu membuat pilihan dan memandang interval di antara pilihan-pilihan tersebut. Pendekatan ini memungkinkan penilaian yang lebih akurat dan dapat diinterpretasikan terhadap konstruksi dengan menangkap jarak aktual yang diterapkan ketika partisipan memilih di antara pilihan-pilihan yang tersedia. Hasil diagnostik untuk setiap subskala BREQ-3, seperti yang ditunjukkan dalam Tabel 3, menunjukkan bahwa kategori respons berfungsi sesuai dengan yang diharapkan. Tidak ada kategori respons yang memiliki frekuensi 0 (nol) di subskala mana pun. Selain itu, semua subskala—kecuali subskala amotivasi—menunjukkan ambang batas yang teratur dengan baik, berkembang dari nilai negatif ke positif di lima opsi respons (Linacre, 2012). Hal ini menunjukkan bahwa partisipan menggunakan kategori respons dengan tepat, meskipun ada satu kesalahan dalam subskala amotivasi.

Model Rasch juga mengevaluasi reliabilitas individu dan item. Baik *person separation reliability* (PSR) maupun *item separation reliability* (ISR) diinterpretasikan secara serupa dengan Cronbach's  $\alpha$ , dengan nilai minimum sekitar 0,80 yang menunjukkan reliabilitas. Namun, *person separation index* (PSI), yang menggunakan skala logit daripada skor mentah, menunjukkan reliabilitas yang rendah, dengan nilai berkisar antara 1,15 hingga 2,47. Menurut Linacre (2012), PSI di atas 3,0 diperlukan untuk reliabilitas yang tinggi.

Sebagaimana ditunjukkan dalam Tabel 4, reliabilitas antar subskala berkisar antara 0,57 hingga 0,86. Hal ini menunjukkan bahwa subskala A dan D, yang mengukur regulasi yang diidentifikasi dan regulasi yang diinternalisasi, hanya memiliki efektivitas moderat dalam membedakan individu berdasarkan konstruk yang diukurnya. Mereka mungkin kurang sensitif dalam membedakan antara individu dengan kinerja tinggi dan rendah (Bond & Fox, 2013). Di sisi lain, ISR berkisar antara 1,62 hingga 14,90, dengan nilai Cronbach's  $\alpha$  antara 0,72 dan 1,00. Indeks pemisahan berkisar antara 1,14 hingga 2,47, menunjukkan pemisahan item pada tingkat rendah hingga moderat. Secara keseluruhan, hasil ini menunjukkan bahwa BREQ-3 memiliki reliabilitas yang cukup tinggi dan sifat psikometrik yang kuat, terutama dalam hal nilai ISR.

### *Item Fit*

Model Rasch mengevaluasi keefektifan setiap butir soal dalam mengukur konstruk dengan menganalisis *item infit*, *outfit*, dan korelasi *point-measure*. Hasil pada Tabel 5 diurutkan dari item yang paling sulit hingga yang paling mudah. Supaya sebuah item efektif dalam pengukuran, nilai *infit* dan *outfit* MNSQ harus berada dalam kisaran 0,5 hingga 1,5, dan korelasi *point-measure* harus berkisar antara 0,40 hingga 0,85 (Fisher, 2007). Penelitian ini tidak menemukan item yang tidak sesuai di keenam subskala BREQ-3. Nilai korelasi *point-measure* untuk keenam subskala adalah positif, dengan sebagian besar item memenuhi kriteria, kecuali C15, E11, dan E17, yang masing-masing memiliki nilai 0,89, 0,89, dan 0,90. Secara umum, korelasi yang sangat tinggi antar item (seperti 0,89, 0,89, dan 0,90) dapat mengindikasikan potensi item yang sangat mirip dengan item lain dalam subskala yang sama (redundansi), yang dapat membuat instrumen kurang efisien. Selain itu, item-item tersebut rentan terhadap *overfitting*, di mana item-item tersebut terlalu mirip dengan pola respons spesifik dalam sampel saat ini, dan mungkin tidak dapat digeneralisasi dengan baik pada populasi yang lebih luas. Rentang korelasi *point-measure* untuk setiap subskala adalah sebagai berikut: A (0,56-0,79), B (0,70-0,77), C (0,82-0,84), D (0,77-0,82), E (0,80-0,84), dan F (0,74-0,82). Hasil ini menunjukkan bahwa semua item dalam instrumen berkinerja baik dalam mengukur konstruk BREQ-3, kecuali untuk item C15, E11, dan E17. Item-item spesifik ini harus direvisi untuk meningkatkan akurasi pengukuran.

### *Wright Map*

Validitas konstruk dapat diukur dengan hierarki butir soal yang ditampilkan pada *Wright Map*. Peta ini menampilkan tingkat kesulitan butir soal pada panel kanan dan kemampuan seseorang pada panel kiri. Item yang lebih mudah diposisikan di bagian bawah, item dengan tingkat kesulitan rata-rata berada di tengah (dilambangkan dengan “M” di sisi kanan), dan item yang lebih sulit berada di bagian atas (Yu, 2020). Peta *Wright* untuk setiap subskala BREQ-3 ditunjukkan pada Gambar 1.

Subskala Regulasi Teridentifikasi (A) menunjukkan bahwa item-itemnya relatif mudah, dengan sebagian besar individu memiliki kemampuan yang lebih tinggi daripada tingkat kesulitan item, sehingga mengurangi kemampuan skala untuk membedakan individu dengan tingkat regulasi yang lebih tinggi. Subskala Amotivasi (B) menunjukkan bahwa item-itemnya relatif sulit dengan sebagian besar individu memiliki kemampuan yang lebih rendah daripada tingkat kesulitan item. Hal ini mengurangi kemampuan skala untuk membedakan individu dengan tingkat amotivasi yang lebih rendah. Pada subskala Regulasi Intrinsik (C), semua item relatif mudah, seperti yang ditunjukkan oleh penempatan sebagian besar individu di atas nilai rata-rata. Pada subskala Regulasi Introjeksi (D) dan Regulasi Terintegrasi (E) menunjukkan bahwa kemampuan partisipan terdistribusi secara merata, sementara tingkat kesulitan item berada pada level sedang. Hal ini mengindikasikan bahwa skala ini mampu membedakan individu dengan tingkat regulasi introjeksi yang berbeda secara efektif. Subskala Regulasi Eksternal (F) menunjukkan bahwa kemampuan partisipan dominan berada pada tingkat sedang hingga rendah, sedangkan tingkat kesulitan item dominan berada pada mean. Hal ini menunjukkan bahwa banyak item mungkin terlalu sulit bagi sebagian besar partisipan, sehingga berpotensi membatasi sensitivitas skala untuk mengukur tingkat regulasi eksternal yang rendah.

Sebagai saran, tambahkan item dengan tingkat kesulitan yang lebih tinggi pada subskala Regulasi Teridentifikasi (A) dan Regulasi Intrinsik (C) agar lebih diskriminatif untuk individu dengan regulasi teridentifikasi dan regulasi intrinsik yang tinggi. Selain itu, menambahkan item dengan tingkat kesulitan yang lebih rendah pada subskala Amotivasi (B) dan Regulasi Eksternal (F) agar lebih efektif dalam mengukur individu dengan tingkat amotivasi dan regulasi eksternal yang rendah. Penyesuaian ini bertujuan untuk meningkatkan daya diskriminasi skala dalam mengukur berbagai tingkat regulasi motivasi, sehingga hasil pengukuran lebih akurat dan merefleksikan perbedaan antar individu secara lebih optimal.

### *Differential Item Functioning*

Analisis *Differential Item Functioning* (DIF) dilakukan untuk mengevaluasi apakah subkelompok dalam sampel (berdasarkan jenis kelamin) memberikan respons yang berbeda terhadap butir-butir soal tertentu. Dalam penelitian ini, jenis kelamin diklasifikasikan menjadi dua subkelompok: laki-laki (L) dan perempuan (P). Metode *chi-square item-trait* (Linacre, 2007) digunakan untuk menilai *Differential Item Functioning* (DIF). Sebuah item dianggap bias secara signifikan jika nilai probabilitasnya di bawah 0,05. Menurut Zwick dkk. (1999), nilai kontras DIF di bawah 0,43 dianggap dapat diabaikan, nilai antara 0,43 dan 0,64 menunjukkan bias sedikit hingga sedang, dan nilai di atas 0,64 menunjukkan bias yang signifikan. Hasil penelitian menunjukkan bahwa item F6 bias terhadap gender ( $p = 0,0171$ ), yang menunjukkan bahwa laki-laki dan perempuan mungkin menginterpretasikan item “Saya berolahraga karena orang lain mengatakan saya harus berolahraga” secara berbeda. Hal ini mengindikasikan kecenderungan laki-laki untuk menunjukkan perilaku olahraga yang lebih proaktif karena adanya motivasi dari luar. Oleh karena itu, diperlukan kehati-hatian saat menggunakan item F6, karena perbedaan gender dapat memengaruhi interpretasinya. Kemudian, item C3 dan C9 memiliki kontras DIF yang cukup tinggi dengan nilai  $-0,45$  dan  $0,78$ , mengindikasikan bahwa item-item tersebut diduga dapat menimbulkan bias dalam penilaian, sehingga mengganggu validitas alat ukur.

### *Mixture Rasch Model*

Selain itu, analisis juga dilakukan dengan menggunakan *Mixture Rasch Model* untuk melihat jumlah *Latent Class* pada data yang telah diperoleh. Investigasi informasi dilakukan dengan menggunakan *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC), dan *Consistent Akaike Information Criterion* (CAIC). AIC adalah metode yang didasarkan pada kesesuaian model untuk memperkirakan kelayakan model dalam memprediksi nilai masa depan. BIC adalah kriteria seleksi model lain yang menilai keseimbangan antara kesesuaian model dan kompleksitas model (Mohammed dkk., 2015). CAIC adalah alat yang digunakan untuk mengevaluasi konsistensi keseimbangan antara kesesuaian model dan kompleksitas. Seperti yang ditunjukkan dalam Tabel 6, model dengan tiga kelas menghasilkan nilai terendah di semua tiga kriteria (AIC = 21,774; BIC = 22,936; CAIC = 23,226), dibandingkan dengan model dengan kelas yang lebih sedikit. Menurut Mohammed dkk. (2015), nilai AIC dan BIC yang lebih rendah menunjukkan kesesuaian model yang lebih baik. Oleh karena itu, solusi tiga kelas dianggap paling optimal di antara model-model yang diuji. Gambar 4 mengilustrasikan distribusi kepadatan partisipan di ketiga kelas. Selanjutnya, analisis ulang dilakukan untuk menentukan pembagian ketiga kelas tersebut. Analisis dilakukan dengan menguji apakah ketiga kelas ini ditentukan oleh jenis kelamin atau usia. Berdasarkan Tabel 7 dan 8, distribusi keanggotaan kelas tidak mengikuti pola berdasarkan jenis kelamin atau usia. Oleh karena itu, dapat dipastikan bahwa ketiga kelompok yang ada tidak dibagi berdasarkan usia atau jenis kelamin. Analisis lebih lanjut menggunakan model Rasch multidimensi direkomendasikan untuk mengeksplorasi interaksi potensial antara dimensi laten dan memperdalam pemahaman tentang struktur yang mendasari.

## SIMPULAN

Beberapa kesimpulan dapat ditarik dari studi ini. Uji unidimensi menunjukkan bahwa setiap subskala atau dimensi BREQ-3 memenuhi asumsi unidimensi. Penggunaan lima kategori respons dalam setiap subskala berfungsi dengan baik, dengan ambang batas kategori meningkat dari nilai negatif ke positif—menunjukkan bahwa pilihan respons berfungsi sebagaimana mestinya, kecuali dalam subskala Amotivasi. Pengecualian ini menunjukkan potensi ambiguitas dalam cara partisipan menafsirkan dan menilai item dalam dimensi Amotivasi. Di antara keenam subskala, hanya item C15, E11, dan E17 yang diidentifikasi sebagai tidak sesuai, karena korelasi titik ukur yang tinggi menunjukkan bahwa item-item ini mungkin tidak berfungsi optimal dalam kerangka model Rasch. Setelah analisis DIF, kami menemukan bahwa item F6, C3, dan C9 diduga menyebabkan bias dalam penilaian, sehingga mengganggu validitas alat ukur.

Temuan ini menunjukkan bahwa, selain item C3, C9, C15, E11, E17, dan F6, item-item lainnya secara efektif mengukur konstruk regulasi latihan. Item-item yang bermasalah perlu direvisi untuk meningkatkan akurasi pengukuran. Dalam hal kesulitan, semua item dalam subskala A (*Identified Regulation*) umumnya mudah hingga sedang, sementara item B20 dalam subskala B merupakan yang paling sulit. Distribusi jawaban dalam subskala C menunjukkan bahwa subskala ini relatif mudah bagi partisipan, karena hanya sedikit yang memilih jawaban ekstrem. Temuan penting lainnya dari analisis DIF adalah bahwa item F6 menunjukkan bias terkait gender. Pria cenderung memilih peringkat yang lebih tinggi daripada wanita pada item ini, yang sejalan dengan pengamatan bahwa pria sering berolahraga karena motivasi eksternal.

Keterbatasan studi ini terletak pada proses pengumpulan data, yang mungkin tidak sepenuhnya mewakili dewasa muda di Indonesia karena penggunaan metode sampling nonprobabilitas. Namun, karena analisis Rasch tidak terlalu bergantung pada karakteristik sampel, temuan mengenai sifat pengukuran di enam subskala BREQ-3 masih dapat digeneralisasikan dengan hati-hati. Selain itu, analisis menggunakan MRM menunjukkan bahwa tiga kelas yang diidentifikasi tidak dibedakan berdasarkan jenis kelamin atau usia. Penelitian selanjutnya dapat mengeksplorasi karakteristik dan

klasifikasi yang bermakna dari kelompok-kelompok ini. Selain itu, penelitian selanjutnya didorong untuk meningkatkan reliabilitas individu dengan merevisi item yang bermasalah dan mengembangkan item baru—khususnya untuk subskala C—untuk meningkatkan indeks *person* dan *item separation*.

Penelitian lebih lanjut diperlukan untuk menilai validitas dan reliabilitas instrumen ini pada populasi atau kelompok demografis tertentu. Penelitian semacam itu dapat memberikan pemahaman yang lebih komprehensif tentang pengukuran regulasi olahraga dalam konteks yang berbeda. Secara keseluruhan, BREQ-3 versi Indonesia dapat diterapkan dalam penelitian mendatang dan aplikasi praktis dalam organisasi, dengan mempertimbangkan hal-hal yang telah disebutkan.

#### **ACKNOWLEDGEMENTS / UCAPAN TERIMA KASIH**

We would like to express our gratitude to God Almighty, our families, teachers, participants in this research, and all other parties who have assisted in the successful implementation of this research. / *Puji Syukur kami ucapkan kepada Tuhan yang Maha Esa, keluarga, pengajar, dan seluruh pihak yang terlibat untuk membantu kelancaran pelaksanaan penelitian ini.*

#### **DECLARATION OF POTENTIAL CONFLICTS OF INTEREST / DEKLARASI POTENSI TERJADINYA KONFLIK KEPENTINGAN**

Bryan Andika, Ingrid Tunggawan, and Ananta Yudiarso do not work for, consult, own shares in, or receive funding from any company or organization that might profit from the publication of this manuscript. / *Bryan Andika, Ingrid Tunggawan, and Ananta Yudiarso tidak bekerja, menjadi konsultan, memiliki saham, atau menerima dana dari perusahaan atau organisasi manapun yang mungkin akan mengambil untung dari diterbitkannya naskah ini.*

#### **REFERENCES / REFERENSI**

- Association, A. E. R., Association, A. P., on Measurement in Education, N. C., on Standards for Educational, J. C., & (U.S.), P. T. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological Methods & Research*, 10(2), 141-163. <https://doi.org/10.1177/004912418101000205>
- Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.
- Cavicchiolo, E., Sibilio, M., Lucidi, F., Cozzolino, M., Chirico, A., Girelli, L., ... & Alivernini, F. (2022). The psychometric properties of the behavioural regulation in exercise questionnaire (BREQ-3): factorial structure, invariance and validity in the Italian context. *International Journal of Environmental Research and Public Health*, 19(4), 19-37. <https://doi.org/10.3390/ijerph19041937>
- Chai, S., Kueh, Y. C., Majdi Yaacob, N., & Kuan, G. (2022). Psychometric properties of the Malay version of the Behavioural Regulation in Exercise Questionnaire (BREQ-3). *Plos one*, 17(6). <https://doi.org/10.1371/journal.pone.0269099>

- Chemolli, E., & Gagné, M. (2014). Evidence against the continuum structure underlying motivation measures derived from self-determination theory. *Psychological Assessment*, 26(2), 575–586. <https://doi.org/10.1037/a0035979>
- Chung, P. K., & Liu, J. D. (2012). Examination of the psychometric properties of the Chinese translated Behavioral Regulation in Exercise Questionnaire-2. *Measurement in Physical Education and Exercise Science*, 16(4), 300–315. <https://doi.org/10.1080/1091367X.2012.693364>
- Cid, L., Monteiro, D., Teixeira, D., Teques, P., Alves, S., Moutão, J., ... & Palmeira, A. (2018). The behavioral regulation in exercise questionnaire (BREQ-3) Portuguese-version: Evidence of reliability, validity and invariance across gender. *Frontiers in Psychology*, 9, 19-40. <https://doi.org/10.3389/fpsyg.2018.01940>
- Clancy, R. B., Herring, M. P., & Campbell, M. J. (2017). Motivation measures in sport: A critical review and bibliometric analysis. *Frontiers in Psychology*, 8(348), 1-12. <https://doi.org/10.3389/fpsyg.2017.00348>
- Creswell, J. W. (2014). *Research design: qualitative, quantitative, and mixed methods approaches*. 4th ed. SAGE Publications.
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1-4. <https://doi.org/10.11648/j.ajtas.20160501.11>
- Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, 21(1), 1095.
- Gregoire, J. (2018). ITC guidelines for translating and adapting tests. *International Journal of Testing*, 18(2), 101-134.
- Hagger, M., & Chatzisarantis, N. (2008). Self-determination theory and the psychology of exercise. *International Review of Sport and Exercise Psychology*, 1(1), 79–103. <https://doi.org/10.1080/17509840701827437>
- Holster, T. A., & Lake, J. (2016). Guessing and the rasch model. *Language Assessment Quarterly*, 13(2), 124–141. <https://doi.org/10.1080/15434303.2016.1160096>
- Khine, M. S. (2020). *Rasch measurement*. Springer Singapore.
- Linacre, J. M., Heinemann, A. W., Wright, B. D., Granger, C. V., & Hamilton, B. B. (1994). The structure and stability of the Functional Independence Measure. *Archives of Physical Medicine and Rehabilitation*, 75(2), 127-132. [https://doi.org/10.1016/0003-9993\(94\)90384-0](https://doi.org/10.1016/0003-9993(94)90384-0)
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement transactions*, 16(2), 878-882.
- Linacre, J. (2007). RUMM2020 item-trait chi-square and winsteps dif size. *Rasch Mea Trans*, 21(1), 1096.
- Linacre, J. M. (2012). A user's guide to WINSTEPS MINISTEP. *Rasch model computer programs*. Beaverton, Oregon: Winsteps. com.
- Markland, D., & Tobin, V. (2004). A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. *Journal of Sport and Exercise Psychology*, 26(2), 191-196. <https://doi.org/10.1123/jsep.26.2.191>

- McCreary, L. L., Conrad, K. M., Conrad, K. J., Scott, C. K., Funk, R. R., & Dennis, M. L. (2013). Using the Rasch measurement model in psychometric analysis of the family effectiveness measure. *Nursing Research*, 62(3), 149-159. [10.1097/NNR.0b013e31828eafe6](https://doi.org/10.1097/NNR.0b013e31828eafe6)
- Nam, T. T., Le, C. N., Phu, D. H., Stanikzai, M. H., Shohaimi, S., Dadras, O., ... & Suwanbamrung, C. (2023). Assessment of self-determined motivation in exercise: A systematic review and meta-analysis. *Journal of Human, Earth, and Future*, 4(2), 241-256. <https://doi.org/10.1002/jeh2.12111>
- Pietrabissa, G., Rossi, A., Borrello, M., Manzoni, G. M., Mannarini, S., Castelnuovo, G., & Molinari, E. (2020). Development and validation of a self-determination theory-based measure of motivation to exercise and diet in children. *Frontiers in Psychology*, 11, 1299-?. <https://doi.org/10.3389/fpsyg.2020.01299>
- Rogowska, A. M., & Morouço, P. (2024). Editorial: Motivations for physical activity. *Frontiers in Psychology*, 15, 1437220. <https://doi.org/10.3389/fpsyg.2024.1437220>
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14(3), 271-282. <https://doi.org/10.1177/014662169001400305>
- Teixeira, P. J., Carraça, E. V., Markland, D., Silva, M. N., & Ryan, R. M. (2012). Exercise, physical activity, and self-determination theory: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 9, 1-30. <https://doi.org/10.1186/1479-5868-9-78>
- Yu, C. H. (2020). Objective measurement: How Rasch modeling can simplify and enhance your assessment. *Rasch measurement: Applications in quantitative educational research*, 47-73.
- Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. *Applied Latent Class Analysis*, 11(89-106), 60.
- Wilson, P. M., Rodgers, W. M., & Fraser, S. N. (2002). Examining the psychometric properties of the Behavioral Regulation in Exercise Questionnaire. *Measurement in Physical Education and Exercise Science*, 6(1), 1-21. [https://doi.org/10.1207/S15327841MPEE0601\\_1](https://doi.org/10.1207/S15327841MPEE0601_1)
- Wilson, P.M., Rodgers, W.M., Loitz, C.C., & Scime, G. (2006). "It's who I am...really!" The importance of integrated regulation in exercise contexts. *Journal of Biobehavioral Research*, 11, 79-104. <https://doi.org/10.1111/j.1751-9861.2006.tb00021.x>
- Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement*, 36(1), 1-28.