JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia)
DOI: http://dx.doi.org/10.15408/jp3i.v14i1.38854
http://journal.uinjkt.ac.id/index.php/jp3i

Psychometric Properties of the Resistance to Framing Scale: Evidence from Indonesia

Rahmat Hidayat[1], Ajeng Putri Pertiwi[1]

Faculty of Psychology, Gadjah Mada University, Indonesia[1]

E-mail: r.hidayat@ugm.ac.id

Abstract

This paper reports the validation of the Indonesian version of the resistance to framing effects scale, which is part of the Adult Decision-Making Competence Index (A-DMC). Framing refers to an individual's tendency to be influenced by how information is structured. The resistance to framing effects scale is a measurement tool designed to assess an individual's susceptibility to framing. This scale comprises two dimensions: attribute framing and risky-choice framing. A total of 217 participants (60 men and 157 women) completed the 14-item scale. Item Response Theory and the Multidimensional Graded Response Model (MGRM) were used to evaluate the psychometric properties of the scale. The MGRM analysis results indicated that the data fit the model, as evidenced by global fit statistics. Additionally, all items showed a good fit with the MGRM model. The reliability of this scale was 0.697 for the attribute dimension and 0.722 for the risky-choice dimension. However, we found that one item had low discrimination (ATT5 with a = 0.638), while the remining 13 items had optimal discrimination. Based on these results, we conclude that the 14-item Indonesian version of the resistance to framing effects scale is a reliable measurement tool that can be used for future research in behavioral economics or economic psychology in Indonesia. There are limitations of this study: the relatively small sample size and the lack of convergent validity testing with other instruments. However, the strong psychometric properties observed in this study suggest that this instrument is suitable for use in future research and may also be applied for practical measurement purposes.
Keywords: construct validity, framing effect, item response theory, resistance to framing.

Abstrak

Paper ini melaporkan validasi dari skala resistensi terhadap efek pembingkaian versi bahasa Indonesia, yang merupakan salah satu bagian dari rangkaian skala pengukuran Adult-Decision Making Competence Index (A-DMC). Pembingkaian adalah tendensi dari seseorang untuk terpengaruh oleh bagaimana informasi disusun. Skala resistensi terhadap efek pembingkaian adalah sebuah alat pengukuran untuk menguji kerentanan seseorang terhadap pembingkaian. Terdapat dua dimensi dalam skala ini, yakni attribute dan risky-choice. Secara keseluruhan, 217 partisipan (60 pria dan 157 wanita) menyelesaikan 14 aitem skala. Analisis Item Response Theory dan Multidimensional Graded Response Model (MGRM) digunakan untuk mengevaluasi properti psikometrik dari skala resistensi terhadap efek pembingkaian. Hasil analisis MGRM menunjukkan bahwa data yang diperoleh fit dengan model dilihat dari statistik global-fit-nya. Selain itu, seluruh aitem juga menunjukkan kesesuaiannya dengan MGRM. Reliabilitas dari skala ini adalah 0.697 untuk dimensi attribute dan 0.722 untuk dimensi risky-choice. Namun demikian, kami menemukan bahwa satu aitem memiliki daya beda yang rendah (ATT5 dengan a = 0.638) dan 13 aitem lainnya memiliki daya beda yang optimal. Berdasarkan hasil ini, kami berkesimpulan bahwa 14 aitem dari skala resistensi terhadap efek pembingkaian versi bahasa Indonesia itu merupakan alat ukur yang dapat dipertanggungjawabkan dan bisa digunakan untuk penelitian ekonomi perilaku ataupun psikologi ekonomi di Indonesia di masa depan. Terdapat dua keterbatasan dalam penelitian ini: ukuran sampel yang relatif kecil dan tidak dilakukannya pengujian validitas konvergen dengan instrumen lain. Namun, properti psikometrik yang kuat yang diamati dalam penelitian ini menunjukkan bahwa instrumen ini layak digunakan dalam penelitian berikutnya dan mungkin juga dapat diterapkan untuk tujuan pengukuran praktis.

***Kata kunci**: efek pembingkaian, resistensi terhadap pembingkaian, teori respons butir, validitas konstruk.*

Introduction

The framing effect refers to a phenomenon in behavioral research where different presentations of the same problem can result in varying decisions, even when the underlying information remains identical. This phenomenon was first identified as preference reversal (Kahneman & Tversky, 1979) and later formally introduced as the framing effect (Tversky & Kahneman, 1981). Among the many cognitive biases that shape human judgment (Hidayat, 2016), the framing effect is one of the most widely recognized and extensively studied (e.g., Beratsova, Krchova, Gazova, & Jirasek, 2016; (Kühberger, 2023). While other biases, such as anchoring, confirmation bias, and the availability heuristic, have also received significant attention, the framing effect is notable for its broad public awareness and widespread application. It occurs when individuals' decisions are influenced by how information is presented—whether framed as potential gains or losses—despite the underlying facts remaining the same. Its prominence in public discourse is due to its relevance across various fields, including marketing, political campaigns, and health messaging, where the framing of choices can dramatically impact outcomes. Consequently, the framing effect is not only a key area of study in cognitive psychology but also a practical concern in real-world decision-making.

The framing effect violates the invariance principle of the subjective expected utility model (Tversky & Kahneman, 1986). In the expected utility theory, the invariance principle states that the preference between two choices should remain consistent, regardless of how the options are presented (i.e., the framing). This principle assumes that rational decision-makers will evaluate choices based solely on their expected outcomes, without being influenced by how the options are framed (e.g., in terms of gains or losses). However, the framing effect shows that people often make different decisions when the same choice is framed differently, even when the expected utility of the outcomes remains the same (e.g., Levin & Gaeth, 1988). This inconsistency violates the assumption of invariance in the expected utility model, revealing that human decision-making is not always rational as the model assumes.

Evidence and theorical explanations

Most research that has revealed framing effects has been conducted using experimental methods. In these studies, participants were presented with various decision-making scenarios in which two identical choices, offering equivalent outcomes, were framed differently—usually in terms of gains and losses. A classic example of risky-choice framing is the "Asian Disease Problem" (Tversky & Kahneman, 1981), presented as follows:

> *Problem 1:*
>
> *Imagine that the US is preparing for the out-break of an unusual disease, which is ex-pected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific es-timate of the consequences of the programs are as follows:*
>
> *If Program A is adopted, 200 people will be saved.*
>
> *If Program B is adopted, there is ⅓probabil-ity that 600 people will be saved, and ⅔ probability that no people will be saved.Which program would you choose?*
>
> *Which of the two programs would you favor?*

The two choices presented to participants are essentially equivalent. In Choice B, there is a 1/3 chance that 600 people will be saved and a 2/3 chance that no one will be saved, yielding an expected value of [(1/3 x 600) + (2/3 x 0)] = 200. Rationally, participants should view Choice A and Choice B as equal. Given that participants were only presented with these two options, the number selecting Choice A should be similar to those selecting Choice B. However, in this experiment, 75% of participants opted for Choice A, while the remaining 25% chose Choice B.

Although the tendency to choose A may seem irrational, this alone does not illustrate the framing effect. The framing effect emerges when participants are presented with an alternative version of the Asian Disease Problem, as shown below:

Problem 2:

If Program C is adopted, 400 people will die.

If Program D is adopted, there is 1/3 probability that nobody will die and 2/3

probability that 600 people will die.

Which of the two programs would you favor?

Choices C and D are actually equivalent, as both result in 400 people dying. Specifically, for Choice D, since life and death are mutually exclusive, death is represented as 0, leading to the calculation: [(1/3 x 600) + (2/3 x 0)] = 200. Therefore, the proportion of participants choosing Program C should match those choosing Program D. Moreover, Program C is essentially identical to Program A, a certain choice, because 400 people dying (Program C) is equivalent to 200 people being saved (Program A). Similarly, Program D corresponds to Program B, as both involve risky choices. However, the pattern of choices reported by Tversky and Kahneman (1981) deviates significantly from the predictions of the Subjective Expected Utility Theory. In contrast to Problem 1, only 25% of participants chose Program C (compared to 75% for Program A), while 75% opted for Program D (compared to 25% for Program B).

Labeling choices as "saved," which implies gains, and "dead," which implies losses, clearly influences participants' decisions. This pattern has been observed in various forms of experimental manipulation. Research on the framing effect has investigated factors that might explain behavior deviating from rationality predictions, including formal characteristics of the choice problem, such as varying value and weighting functions. In line with this, studies have used different disease scenarios, such as unusual infections, leukemia, and AIDS, with varying levels of risk (Diederich et al., 2018), as well as different numbers of affected individuals and probabilities of survival/death (Wyszynski & Diederich, 2023). Research has also explored different problem types, including the coin flip problem, savings problem, tennis problem, and movie problem (Stanovich & West, 1998), consumer choice of meat (Levin & Gaeth, 1988), and economic crises (Giuliani et al., 2023). The conditions under which participants make decisions have also been manipulated, for example, by introducing time pressure (Roberts et al., 2022). The framing effect has been examined in different task presentations, whether through vignette-based or reward-based gambling tasks (Zhen & Yu, 2016). Additionally, studies have been conducted across different age groups, cohorts, and testing times(Mayhorn et al., 2022).

Additionally, research has examined the influence of individual differences on vulnerability to the framing effect. Cognitive style factors such as cognitive reflection, subjective numeracy, actively open-minded thinking, need for cognition, and hemispheric dominance have been shown to affect susceptibility to the framing effect (LeBoeuf & Shafir, 2003; Mandel & Kapler, 2018; Rachev et al., 2022). While most studies have used between-subject designs, research utilizing within-subject designs has also been conducted (Aczel, Szollosi, & Bago, 2018; LeBoeuf & Shafir, 2003).

The robust phenomenon of the framing effect, observed across numerous studies, can be elegantly explained by Prospect Theory, developed by Daniel Kahneman and Amos Tversky (1979). According to Prospect Theory, people evaluate outcomes relative to a reference point (often the status quo) rather than in absolute terms. Whether an option is framed as a "gain" or a "loss" relative to this reference point significantly influences decision-making. The theory posits that individuals are more sensitive to losses than to gains of the same size—in other words, the pain of losing something outweighs the pleasure of gaining something of equal value. Consequently, when choices are framed as avoiding losses, people tend to become risk-seeking, whereas choices framed as achieving gains tend to encourage risk-averse behavior.

http://journal.uinjkt.ac.id/index.php/jp3i

In the context of the framing effect, Prospect Theory predicts that when a problem is framed in terms of gains (e.g., "200 people will be saved"), individuals are likely to prefer the sure option, demonstrating risk aversion. This is evident in the 75% of participants who favored Program A. Conversely, when the same problem is framed in terms of losses (e.g., "400 people will die"), individuals are more inclined to choose the riskier option in an effort to avoid the loss, exhibiting risk-seeking behavior. Thus, Prospect Theory explains how different framings of the same problem—whether in terms of gains or losses—lead to different decisions, despite the actual outcomes being identical. This deviation from expected utility theory forms the basis of the framing effect.

Measurement of individual differences

Research has identified various factors that influence susceptibility to the framing effect. Beratsova et al. (2016) identified four broader contributing factors, namely decision situation setup (amount of information, additional presentation of options), experience (knowledge, engagement), effort (attention, complexity, amount of information to process), and demographics (gender, nationality). Demographic factors clearly represent individual differences (Kühberger, 2023), as do cognitive style factors such as the need for cognition (LeBoeuf & Shafir, 2003; Mandel & Kapler, 2018; Rachev et al., 2022).

The presence of individual factors that lead to different responses to decision situations suggests variability in the degree of vulnerability to the framing effect across individuals. This has led to the development of various instruments designed to measure this vulnerability. One of the most widely used instruments is the Adult Decision-Making Competence (ADMC), developed by de Bruin, Parker, and Fischhoff (2007). The ADMC is a psychometric tool designed to assess an individual's ability to make effective and rational decisions across various contexts. It evaluates multiple aspects of decision-making by measuring how well individuals avoid common cognitive biases and apply rational decision-making strategies. The ADMC consists of several subtests, one of which is resistance to the framing effect. This subtest includes 14 items, equally divided to measure both risky-choice framing and attribute framing. Responses are measured on a 6-point strength-of-preference rating scale, with no midpoint to encourage participants to express a relative preference between options, even if only weakly. The primary aim of the resistance to framing subtest is to assess whether individuals are influenced by how choices are framed (e.g., as gains vs. losses). Individuals with high decision-making competence are expected to make consistent choices regardless of the framing of the options.

To the best of our knowledge, the A-DMC subtest is the only measurement tool specifically designed to assess individual susceptibility to the framing effect. Despite its widespread application, recent studies suggest that its psychometric properties are rarely thoroughly evaluated (Geiger et al., 2022a). Furthermore, previous research has found that the internal consistency of the instrument often falls below optimal criteria (< 0.70). For example, a study reported that the Bulgarian version had reliabilities of 0.31 and 0.35 using classical reliability coefficients (ω), and 0.52 for both aspects based on IRT reliability (Geiger et al., 2022). Similarly, the Slovak version demonstrated a reliability of 0.72 (Bavolar, 2013), the Italian version reported 0.67 (Weller et al., 2015), and the French version showed a reliability of 0.76 (Berthet, 2021). In its original version, the reliability of the resistance to framing effects subtest was 0.62 (Bruine de Bruin et al., 2007). The Chinese A-DMC demonstrated relatively good reliability, with Cronbach's alpha above 0.6 and test–retest reliability coefficients ranging from 0.44 to 0.78 across all subscales, comparable to the original version (Peng et al., 2019). These findings highlight the challenges and importance of evaluating the psychometric properties of this instrument to better understand its validity and reliability across different cultures, and to determine whether consistently low reliability is present in other samples.

Studies reporting the validation and reliability testing of scales measuring resistance to framing have been conducted across various countries and cultures, including Slovakia (Bavolar, 2013), Italy ((Weller et al., 2015), Croatia ((Gabor & Knezović, 2016), Russia (Katerina, 2018), China (Liang & Zou, 2018), Bulgaria, and in comparisons with North American samples (Geiger et al., 2022; Rachev et al., 2022). In Indonesia, several studies have explored the interrelationships between framing effects and other

variables, such as sunk costs, from both economic (e.g., (Permatasari et al., 2020) and psychological perspectives (e.g., (Hidayat & Putra, 2022). Other studies have discussed the basic concept of framing effects and their relationship with different aspects within the A-DMC (e.g., (Hidayat, 2016). However, research specifically focused on examining the validity and reliability of the resistance to framing effects instrument using Indonesian samples remains very limited.

Focusing on validation studies, this instrument has been tested using various modern psychometric analysis methods such as item response theory, network analysis, and confirmatory factor analysis (e.g., Bruine de Bruin et al., 2007; Geiger et al., 2022). These studies have found that the measurement model of resistance to framing effects follows a two-factor structure, with one factor representing risky-choice framing and the other representing attribute framing (Geiger et al., 2022). Therefore, it is crucial for researchers in Indonesia to determine whether this factor structure can be replicated in Indonesian samples.

This study aims to evaluate the psychometric properties of the instrument measuring resistance to framing effects at both the construct level (to confirm the factor structure) and the item level. The research is expected to provide further insights into the validity and reliability of the Indonesian version of the instrument for measuring resistance to framing effects.

Methods

Participants

Data were obtained from 217 students (60 men [27.6%], 157 women [72.4%]) from various faculties and departments at Universitas Gadjah Mada. The sample's mean age was 21.92 years old (SD = 3). All participants were active students. Data were collected online. Students were provided with information about the purpose of this research and were informed that their data would be protected for privacy. Data collection was carried out after obtaining research ethics approval from the Research Ethics Committee of the Faculty of Psychology, Universitas Gadjah Mada number 14560/UN1/FPsi.1.3/SD/PT.01.04/2023

Instruments

The Resistance to Framing Scale (Bruine de Bruin et al., 2007) is one of the subscales of the Adult Decision-Making Competence Index (A-DMC; (Bruine de Bruin et al., 2007). This scale consists of 14 items, comprising 7 items measuring risky-choice and 7 items measuring attributes framing. Each item is presented in both positive and negative forms. Participants are asked to rate each item on a 6-point scale ranging from 1 (Definitely Choose A) to 6 (Definitely Choose B). Resistance to framing effects is then assessed based on the absolute difference between the rankings of positive and negative frames for each item (Bruine de Bruin et al., 2007), where lower scores indicate lower vulnerability to framing effects. Absolute difference scoring is the procedure used in this study, while another scoring method, directed scoring, is not used because it has been criticized in previous studies (i.e., Geiger et al., 2022). In the Indonesian version, modifications were made to three items whose wording was changed to better align with the Indonesian cultural context and the context of being a student. For example, original item of ATT2 "Imagine the following situation. You are entertaining a special friend by inviting them for dinner. You are making your favorite lasagna dish with ground beef. Your roommate goes to the grocery store and purchases a package of ground beef for you. The label says 80% lean ground beef. What's your evaluation of the quality of this ground beef?" were adapted and modified into "*Bayangkan situasi berikut ini. Anda sedang membantu panitia di Fakultas untuk menjamu makan malam seorang dosen tamu dari luar negeri. Dosen ini dikenal hati-hati dalam memilih makanan. Anda ditugaskan membeli daging fillet di sebuah supermarket. Di label kemasan tertulis 80% daging sapi tanpa lemak."* These modifications were also made in previous studies conducted by Geiger et al. (2022).

All measures were translated into Bahasa Indonesia following the International Test Commission Guidelines for Test Adaptation (International Test Commission, 2010). The original English versions of

http://journal.uinjkt.ac.id/index.php/jp3i

the questionnaires were translated into Bahasa Indonesia by two qualified Indonesian translators who are English lectures with postgraduate degrees from universities in the US and UK. Subsequently, the Bahasa versions were translated back into English by native English speakers with diplomas in Bahasa and over 12 years of experience living in Indonesia. These translators were unaware of the purpose of the instruments.

A review committee, consisting of the first author and two clinical psychology lecturers fluent in English with over 15 years of clinical experience, examined the original English versions, the English back translations, and the Bahasa versions of all questionnaires. During the review, it was noted that the Bahasa versions were longer than the original English and back-translated versions due to the limited number of emotional terms in Bahasa. The committee decided to prioritize the clarity of the items over the text length.

Multidimensional Graded Response Model (MGRM)

Given the multidimensional structure of the Resistance to Framing Scale, we use the extension of the Graded Response Model (GRM; Samejima, 1969) called the Multidimensional Graded Response Model (MGRM; (de Ayala, 1994) has been developed. Geiger et al. (2022) previously used this model to assess the psychometric properties of the same instruments. The Multidimensional Graded Response Model (hereinafter referred to as MGRM) requires a set of multidimensional parameters. In MGRM, an individual's response to item i is categorized into $m_i + 1$ ordered categories, where the highest category indicates a higher θ level, and m_i is the number of category thresholds. Category scores for item i, x_i take values of 0,…, x_i. MGRM is expressed as follows:

$$P_{x_i}(\Theta) = \frac{exp[D\Sigma a_{ih}(\theta_h + d_{xi})]}{1 + exp[D\Sigma a_{ih}(\theta_h + d_{xi})]}$$

θ_h is the latent trait of dimension h (h = 1, ..., r dimensions), a_{ih} is the discrimination parameter for item i in dimension h, d_{xi} is the difficulty level parameter for category x of item i, and their sum across dimensions (de Ayala, 1994). According to Standard 3.9 of the Standards for Educational and Psychological Testing (American Educational Research Association et al., 2014), both overall model fit indices and item-level fit should be reported when an IRT model is used.

In this study, the model fit indices used for testing the overall model fit (overall goodness-of-fit) are M_2 and $RMSEA_2$. If it is found that M_2 is not significant (p-value > 0.05), it means that the model fits the data (Maydeu-Olivares & Joe, 2006); whereas if $RMSEA_2$ shows a value < 0.05, it also means that the MGRM model fits the data (Maydeu-Olivares, 2013). The M2 and RMSEA2 also indicates the construct validity given the hypothesized model (i.e., two-factor model). In addition, the SRMR were also reported with the value < 0.08 indicates the fulfillment of the local independence assumptions (Rahayu et al., 2023).

Furthermore, reliability in IRT is estimated by the marginal reliability coefficient (Green et al., 1984), which is equivalent to the alpha coefficient in classical test theory (Reise, 1999); if its value is higher than 0.80, then the instrument has good internal consistency (e.g., (Petscher et al., 2015). However, it should be noted that the reliability is the properties of the scores, not the test itself (Rahayu et al., 2023). After obtaining the model fit for the overall model. The fit of the model is tested at the item level using the S-χ^2 (Kang & Chen, 2008); this method has been tested and proven to work well in MGRM (Su et al., 2021). An item is considered to fit the model if S-χ^2 is not significant at $p < 0.05$ (Kang & Chen, 2008).

To estimate person parameter and items, MGRM was used in this study through the 'mirt' package (Chalmers, 2012) in R. In using the mirt package, person parameter and item are estimated using maximum marginal likelihood.

Data Availability Statement

The anonymous datasets presented in this study can be found in online repositories through Open Science Framework **(**https://osf.io/xs75e/?view_only=9f7fd99d8d1b4dd997953a8b29e95979**)**.

Results and Discussion

Correlation Matrix

In the first stage, correlation analysis was conducted on the 14 items of Indonesian version of the resistance to framing scale. This analysis was performed to obtain a simple overview of the direction and magnitude of correlations among the items, as this procedure was done in previous studies (i.e., (Bavolar, 2013; Bruine de Bruin et al., 2007; Geiger et al., 2022). The polychoric correlation matrix can be seen in Table 1.

As can be seen in the Table 1, it was found that the correlation between items on the Indonesian version of the resistance to framing effects scale was relatively low with a range of 0.003 to 0.351. Pretty much items have negative correlations with other items. Bivariate correlation between items can further be presented in the form of network analysis to provide a more comprehensive view of the relationship between items.

Table 1. Inter-item Correlation Matrix

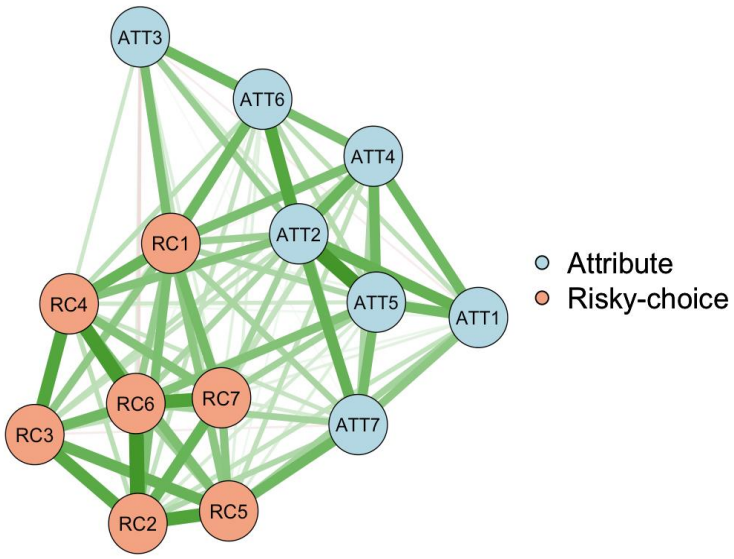
Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14
ATT1	-													
ATT2	0.256	-												
ATT3	-0.026	0.158	-											
ATT4	0.234	0.24	0.123	-										
ATT5	0.23	0.348	0.066	0.237	-									
ATT6	0.1	0.292	0.232	0.211	0.116	-								
ATT7	0.18	0.253	0.014	0.129	0.224	0.188	-							
RC1	0.035	0.155	0.212	0.211	0.127	0.232	0.135	-						
RC2	0.093	0.003	-0.053	0.085	0.061	0.093	0.122	0.211	-					
RC3	0.007	0.092	0.078	0.1	0.029	0.068	-0.027	0.177	0.282	-				
RC4	0.024	0.195	-0.004	-0.014	0.073	0.105	0.085	0.245	0.103	0.298	-			
RC5	0.143	0.025	0.096	0.1	0.051	0.047	0.232	0.182	0.305	0.253	0.209	-		
RC6	0.031	0.131	0.048	0.062	0.199	0.028	0.143	0.166	0.351	0.232	0.335	0.155	-	
RC7	0.044	0.143	0.053	0.091	0.146	0.046	0.065	0.223	0.249	0.1	0.2	0.186	0.307	-

Source: Personal Data (2025).

Network Plot

In the second stage, the findings on the correlation between items, which were previously presented in matrix form, can be presented in the following network form in Figure 1.

As can be seen in Figure 1, it is found that in the risky-choice (RC) dimension, the items have correlations that are classified as significant in value as seen from the thicker green lines such as the relationship between RC1 to RC7. In addition, RC items also have correlations with ATT dimension items. Unexpected correlations, namely those with negative values, are shown on the pink line, the number of which is relatively small. Like RC items, ATT items tend to cluster and be close to other ATT items where the farthest item is ATT3. However, item ATT3 bridges the relationship between ATT and RC as seen from the large correlation between ATT3 and RC1. Figure 1 contains the initial foundation of testing the two-factor model where ATT and RC are expected to be significantly correlated with each other.



Source: Personal Data (2025).

Figure 1. Network Plot of the RtFE

IRT Analysis

In the third stage, the Indonesian version of the resistance to framing scale was calibrated. This analysis is carried out determine whether there is a model specification error or not. The MGRM goodness-of-fit statistics for both instruments can be seen in Table 2.

Table 2. Goodness-of-fit Statistics of the Indonesian Version of the Resistance to Framing Effects Scale

Fit Indices	Value
M_2	18.740
df	21
p-value of M_2	0.602
RMSEA	0.000
90% C.I. of RMSEA	0.000, 0.051
SRMR	0.070
CFI	1.000

Source: Personal Data (2025).

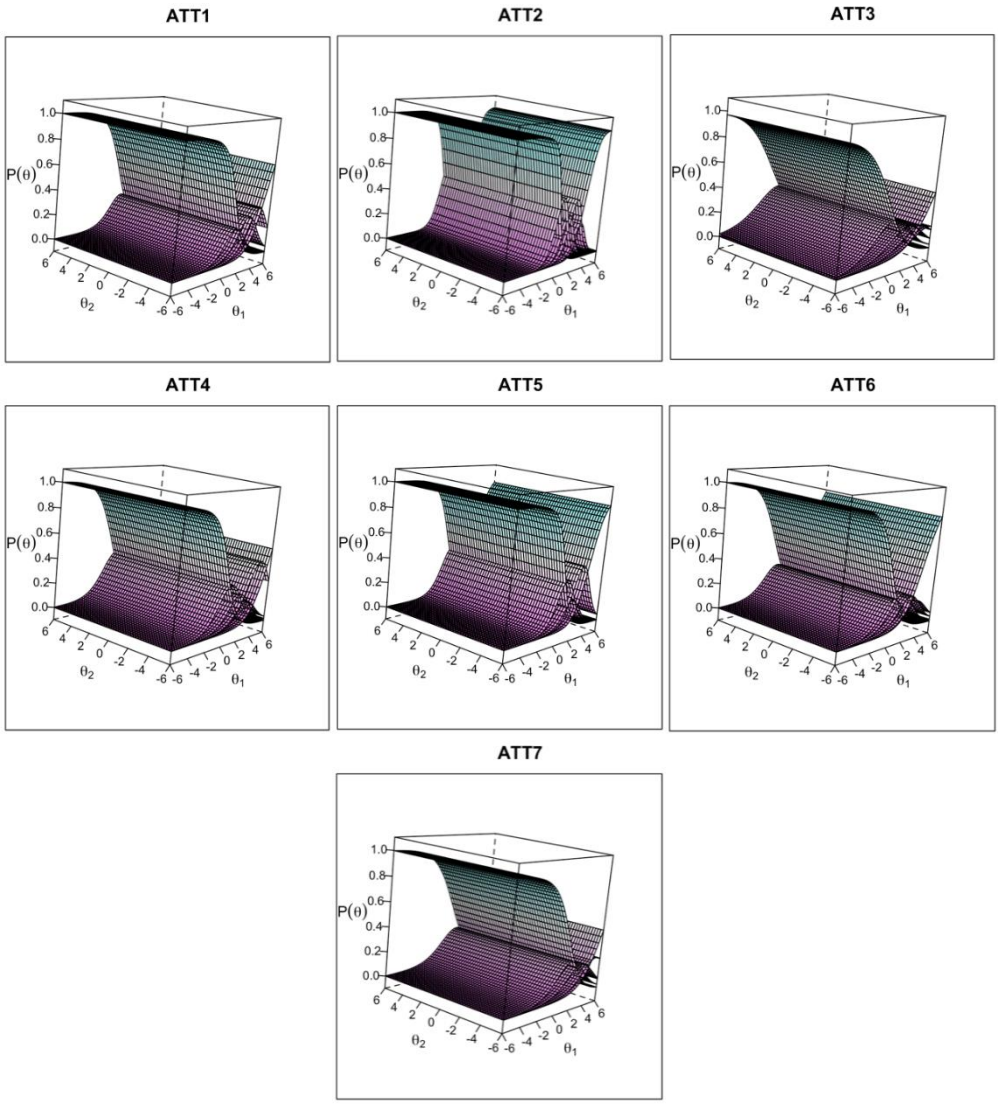
Based on the information in Table 2, it is known that the Indonesian version of the resistance to framing scale fits the MGRM with a two-factor solution. This is proven by the non-significant M_2 value as well as RMSEA, SRMR, and CFI values that are acceptable. After obtaining statistical evidence regarding the overall model fit, an examination of item fit to the model was conducted (see Table 3). Based on this examination, it is known that all items on the Indonesian version of the resistance to framing scale fit the model with a non-significant S-χ^2 ($p > 0.05$).

However, it can be seen that item ATT5 has four threshold levels, unlike other items that have five threshold levels. This is because one response option was not chosen by any participants. Nevertheless, this item still fits the model (S-χ^2 = 20.922, df = 30, p = 0.890). Therefore, it can be concluded that the psychometric properties of resistance to effect scale in Indonesian version is proven to fit the MGRM model and confirm that the two-factor model is replicated in the Indonesian sample. The item characteristic curves for all aspects of attributes can be seen in Figure 2.

Table 3. Item Fit Analysis

Item	Item discrimination	Threshold					S-χ^2	*df*	*p*
		B1	B2	B3	B4	B5			
Attribute (IRT reliability = 0.697)									
1	1.152	0.907	1.887	2.470	3.712	5.155	24.506	26	0.547
2	2.043	-0.006	1.112	1.872	2.907	3.472	29.971	31	0.519
3	0.638	-0.511	1.704	3.359	4.853	6.135	48.176	41	0.205
4	1.045	-0.014	1.505	2.616	3.434	5.610	50.648	36	0.053
5	1.399	0.521	1.745	2.423	3.955	-	20.922	30	0.890
6	1.001	-0.135	0.991	2.244	3.328	4.171	46.888	40	0.211
7	0.827	1.197	2.693	3.969	4.645	6.023	25.552	23	0.322
Risky-choice (IRT reliability = 0.722)									
8	1.038	-0.496	1.224	1.821	3.435	4.951	37.074	35	0.374
9	1.255	-0.394	0.947	1.762	2.624	4.246	31.932	36	0.663
10	1.110	-0.683	0.651	1.447	2.467	3.841	37.105	47	0.849
11	1.139	-0.281	0.846	1.617	2.513	3.759	38.771	44	0.695
12	0.903	-0.482	1.306	2.430	3.390	4.799	42.289	39	0.331
13	1.573	-0.145	1.107	1.649	2.682	3.649	38.998	32	0.184
14	1.160	0.248	1.293	1.747	2.790	2.911	36.422	32	0.270

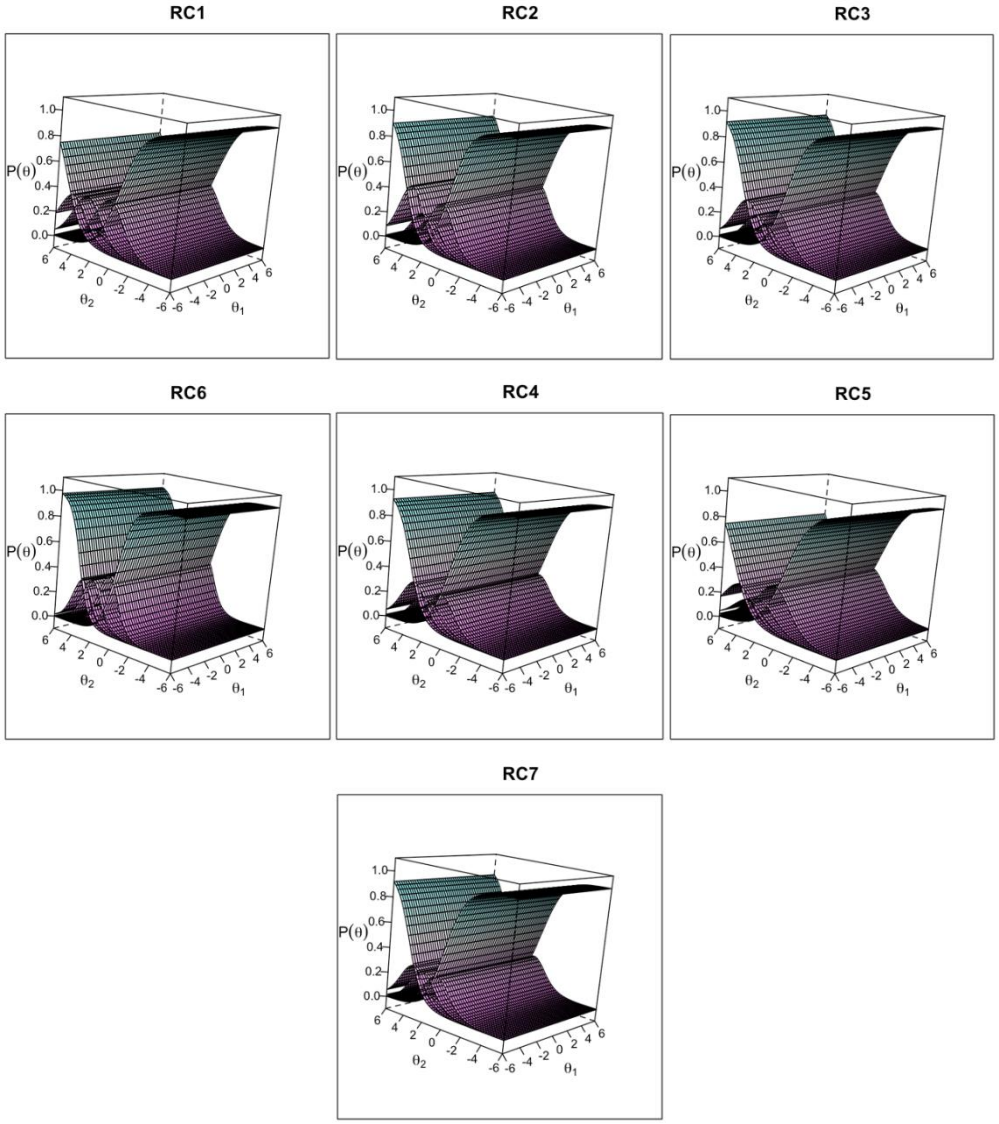
Source: Personal Data (2025).



Source: Personal Data (2025).

Figure 2. Item Characteristic Curve of Attribute Dimension (ATT)

As seen in Figure 2, it can be observed that item ATT3 which has the lowest discrimination power when compared to other items has a sloping curve. Meanwhile, item ATT5 which has optimal discriminating power (a = 1.399) but has a different number of threshold compared to other items. This means that there are response options from ATT5 that were not chosen by any of the participants. With the finding that the item discriminating power of almost all items is sufficiently optimal, this finding indicates that the items theoritically measuring the attribute dimension have sufficient strength to differentiate between individuals with high resistance to framing effect and those with low resistance to framing effect in attribute aspect. Meanwhile, Figure 3 contains the item characteristics curves for the RC dimension.



Source: Personal Data (2025).

Figure 3. Item Characteristic Curve of Risky-choice Dimension (RC)

As can be seen in Figure 3, all RC items have optimal discrimination (range of a = 0.903 to 1.573). Furthermore, all items have the same number of thresholds. This means that there are no response options from any RC item that were not chosen by any participants. With the finding that the discrimination of items from almost all items is sufficiently optimal, this finding indicates that the items theoretically measuring risky-choice dimension have sufficient strength to differentiate between individuals with high resistance to framing effects and those with low resistance to framing effects in the risky-choice dimension.

Discussion

This is the first study to provide an empirical validity evidence based on internal structure of the measurement of the resistance to framing effect in Indonesian version. At the test or instrument level, based on the analysis results using modern test theory (IRT), it was found that the two-factor model of resistance to framing effect fits the data. This finding is consistent with recent studies conducted using the same method on two different samples, Bulgaria and North America (Geiger et al., 2022). However, the two-factor model containing two dimension (attribute and risky-choice) differs from the findings of the first validation study of this instrument conducted by Bruine de Bruin et al. (2007), where the model that fit the data was a one-factor model.

Nevertheless, it is important to note that Bruine de Bruin et al. (2007) also tested the two-factor model in their study, similar to what was found in our research, although in their study, the model did not fit. The two-factor model (attribute and risky-choice framing) aligns with the typology introduced in earlier studies by Levin et al. (1998), stating that they are different based on what framing is (choices with varying levels of risk as opposed to object evaluations), and how the effect is tested (comparing choices against object attractiveness) – all of which are expected to influence the framing effect. The same model was also found in measuring resistance to framing effects in the Chinese population (Peng et al., 2019). Furthermore, recent studies have compiled systematic literature reviews on risky-choice framing effects, supporting the two-factor model of the resistance to framing scale (Kühberger, 2023), complementing earlier meta-analysis studies (Piñon & Gambara, 2005). Therefore, we suggest users of this instrument in Indonesian samples to use the two-factor model as found in this study and some previous studies that are in line with it.

Furthermore, at the item level, we found that all items fit the multidimensional GRM model. This finding is consistent with recent studies that also used MGRM in different samples (i.e., Geiger et al., 2022). In the Indonesian version, modifications were made to the items aimed at making them more suitable for Indonesian culture have been successful. However, one item, ATT3, had a relatively low item discrimination if 0.638.

Regarding the reliability or internal consistency of the resistance to framing scale, we found that the reliability of the resistance to framing scale in the Indonesian version is 0.672 for the attribute dimension and 0.722 for the risky-choice dimension. This finding is not much different from previous studies that found that this instrument has internal consistency that is below optimal criteria (< 0.70), including studies that found that the Bulgarian version of the instrument has reliabilities of 0.31 and 0.35 with classical reliability coefficients (ω) and 0.52 for both aspects based on IRT reliability (Geiger et al., 2022). Meanwhile, various other studies found reliabilities of 0.72 for the Slovak version (Bavolar, 2013), 0.67 for the Italian version (Weller et al., 2015), 0.67 and 0.62 for the Chinese version (Liang & Zou, 2018; Peng et al., 2019), and 0.76 for the French version (Berthet et al., 2022). In its original version, the resistance to framing scale had a reliability of 0.62 (Bruine de Bruin et al., 2007). These findings undoubtedly present challenges and underscore the importance of evaluating the psychometric properties of this instrument for its users.

Mathematically, we suspect that the low reliability is related to the small number of items (7 for each aspect). Previous studies have found that a small number of items is a cause of low reliability (Bell & Lumsden, 1980; Fitzpatrick & Yen, 2001; Tavakol & Dennick, 2011; Wright, 1996). That is why we suspect that the low reliability of the resistance to framing scale is consistently found across different cultures due to the factor of numbers item. Another factor suspected to cause low reliability is the familiarity of Indonesian respondents with Likert scale test formats and response options result in low test score reliability (Lozano et al., 2008; Solomon & Kopelman, 1984).

These technical issues are discussed in depth in Geiger et al. (2022) regarding response format and scoring format for the resistance to framing scale. We follow the recommendations of that study to use absolute difference scoring format and not use directed scoring format because this format results in very low reliability. It is highly recommended for users of the resistance to framing scale in the Indonesian version to use the same scoring format (see Bruine de Bruin et al., 2007). Other studies specifically summarize various possible forms of scoring and options in measuring resistance to framing effects that can serve as references for researchers to understand this issue more deeply (Kühberger & Gradl, 2013).

Lastly, it cannot be denied that this study is not without limitations. The first limitation is sample size. Although a sample size of 200 (met in this study) is considered sufficient for the application of modern test theories like MGRM in this research (e.g., Rahayu et al., 2023; Wright, 1977), other studies suggest a minimum sample size of 500 to produce better power and confidence estimations (Fitzpatrick & Yen, 2001). Future research is expected to conduct psychometric evaluation processes on the resistance to framing scale with a sample size larger than 500. The second limitation is that validity evidence based on

relations to other variables (e.g., convergent validity) has not been conducted in this study since no other instruments were administered simultaneously with Indonesian version of the resistance to framing scale, although in other versions, convergent validity has been tested (e.g., Zhen & Yu, 2016). These two limitations provide significant ideas to extend the findings of the present study and contribute further to studies specifically focusing on evaluating the psychometric properties of the resistance to framing effect scale, not only in Indonesia but globally.

Conclusion

The conclusion of this study is that the resistance to framing scale in Indonesian version has good psychometric properties and has a multidimensional factor structure that can be used as a reference for future research. Additionally, the reliability of this instrument is sufficient for the risky-choice dimension but slightly below the optimal threshold. It is important for future research to conduct psychometric property evaluation processes in other samples in Indonesia to provide a more comprehensive picture of the stability of the psychometric properties of the resistance to framing scale in the Indonesian version. Finally, it is hoped that this instrument will contribute to the development of economic psychology research in Indonesia.

Acknowledgment

The authors would like to thank all participants for the voluntary participations in this research. We also would like to thank all parties in the Faculty of Psychology Universitas Gadjah Mada for the support throughout this research. We specially thank to Muhammad Dwirifqi Kharisma Putra for his suggestions on the data analysis.

Conflict of Interest

The authors declare no potential conflict of interest in this study.

Authors Contribution

RH conducted instrument adaptation, designed the study, lead in the data collection and writing the manuscript. APP translated and manuscript formatting, designed and conducted the data analysis. Both authors wrote and approved the final version of the manuscript.

References

Aczel, B., Szollosi, A., & Bago, B. (2018). The effect of transparency on framing effects in within-subject designs. *Journal of Behavioral Decision Making*, *31*(1), 25–39. https://doi.org/10.1002/bdm.2036

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Bavolar, J. (2013). Validation of the adult decision-making competence in Slovak students. *Judgment and Decision Making*, *8*(3), 386–392. https://doi.org/10.1017/S1930297500006057

Bell, R., & Lumsden, J. (1980). Test length and validity. *Applied Psychological Measurement*, *4*(2), 165–170. https://doi.org/10.1177/014662168000400203

Berthet, V. (2021). The measurement of individual differences in cognitive biases: A review and improvement. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.630177

Berthet, V., Autissier, D., & de Gardelle, V. (2022). Individual differences in decision-making: A test of

a one-factor model of rationality. *Personality and Individual Differences*, *189*, 111485. https://doi.org/10.1016/j.paid.2021.111485

Bock, R. D., & Gibbons, R. D. (2021). *Item response theory*. John Wiley & Sons.

Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, *92*(5), 938–956. https://doi.org/10.1037/0022-3514.92.5.938

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6). https://doi.org/10.18637/jss.v048.i06

de Ayala, R. J. (1994). The influence of multidimensionality on the graded response model. *Applied Psychological Measurement*, *18*(2), 155–170. https://doi.org/10.1177/014662169401800205

De Martino, B. (2012). The effect of context on choice and value. In R. Dolan & T. Sharot (Eds.), *Neuroscience of Preference and Choice* (pp. 93–119). Elsevier. https://doi.org/10.1016/B978-0-12-381431-9.00005-X

Diederich, A., Wyszynski, M., & Ritov, I. (2018). Moderators of framing effects in variations of the Asian disease problem: Time constraint, need, and disease type. *Judgment and Decision Making*, *13*(6), 529–546. https://doi.org/10.1017/s1930297500006574

Fitzpatrick, A. R., & Yen, W. M. (2001). The effects of test length and sample size on the reliability and equating of tests composed of constructed-response items. *Applied Measurement in Education*, *14*(1), 31–57. https://doi.org/10.1207/S15324818AME1401_04

Gabor, A. M., & Knezović, A. (2016). Decision making styles of investors in capital market. *Proceedings of the Faculty of Economics & Business in Zagreb/Zbornik Ekonomskog Fakulteta u Zagrebu*, *14*(1).

Geiger, S. J., Vintr, J., & Rachev, N. R. (2022a). A reassessment of the resistance to framing scale. *Behavior Research Methods*, *55*(5), 2320–2332. https://doi.org/10.3758/s13428-022-01876-7

Geiger, S. J., Vintr, J., & Rachev, N. R. (2022b). A reassessment of the resistance to framing scale. *Behavior Research Methods*, *55*(5), 2320–2332. https://doi.org/10.3758/s13428-022-01876-7

Giuliani, F., Cannito, L., Gigliotti, G., Rosa, A., Pietroni, D., & Palumbo, R. (2023). The joint effect of framing and defaults on choice behavior. *Psychological Research*, *87*(4), 1114–1128. https://doi.org/10.1007/s00426-022-01726-3

Geiger, S. J., Vintr, J., & Rachev, N. R. (2022). A reassessment of the resistance to framing scale. *Behavior Research Methods*, *55*(5), 2320–2332. https://doi.org/10.3758/s13428-022-01876-7

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assesing computerized adaptive tests. *Journal of Educational Measurement*, *21*(4), 347–360. https://doi.org/10.1111/j.1745-3984.1984.tb01039.x

Hidayat, R. (2016). Rasionalitas: Overview terhadap pemikiran dalam 50 tahun terakhir. *Buletin Psikologi*, *24*(2), 101–122. https://doi.org/10.22146/buletinpsikologi.26772

Hidayat, R., & Putra, M. D. K. (2022). Pilot testing and preliminary psychometric validation of situational judgment test of resistance to sunk cost. *Jurnal Psikologi*, *49*(3), 211–228. https://doi.org/10.22146/jpsi.73737

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292. https://www.jstor.org/stable/1914185

Kang, T., & Chen, T. T. (2008). Performance of the generalized S-χ2 item fit index for polytomous IRT models. *Journal of Educational Measurement*, *45*(4), 391–406. https://doi.org/10.1111/j.1745-3984.2008.00071.x

Katerina, G. (2018). *Development of a local version of the scale resistance to the framing effect on the basis of the comprehensive assesment of rational thinking* [Thesis, HSE University]. https://www.hse.ru/en/edu/vkr/219398134

Kühberger, A. (2023). A systematic review of risky-choice framing effects. *EXCLI Journal*, *22*, 1012–1031. https://doi.org/10.17179/excli2023-6169

Kühberger, A., & Gradl, P. (2013). Choice, rating, and ranking: Framing effects with different response modes. *Journal of Behavioral Decision Making*, *26*(2), 109–117. https://doi.org/10.1002/bdm.764

LeBoeuf, R. A., & Shafir, E. (2003). Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making*, *16*(2), 77–92. https://doi.org/10.1002/bdm.433

Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, *15*(3), 374–378. https://www.jstor.org/stable/2489471

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, *76*(2), 149–188. https://doi.org/10.1006/obhd.1998.2804

Liang, S., & Zou, Y. (2018). Validation of adult decision-making competence in Chinese college students. *Judgment and Decision Making*, *13*(4), 393–400. https://doi.org/10.1017/S1930297500009268

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, *4*(2), 73–79. https://doi.org/10.1027/1614-2241.4.2.73

Mandel, D. R., & Kapler, I. V. (2018). Cognitive style and frame susceptibility in decision-making. *Frontiers in Psychology*, *9*(AUG). https://doi.org/10.3389/fpsyg.2018.01461

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/BF02296272

Maydeu-Olivares, A. (2013). Why should we assess the goodness-of-fit of IRT models? *Measurement: Interdisciplinary Research & Perspective*, *11*(3), 127–137. https://doi.org/10.1080/15366367.2013.841511

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, *71*(4), 713–732. https://doi.org/10.1007/s11336-005-1295-9

Miyapuram, K. P., & Pammi, V. S. C. (2013). Understanding decision neuroscience: A multidisciplinary perspective and neural substrates. *Progress in Brain Research*, *202*, 239–266. https://doi.org/10.1016/B978-0-444-62604-2.00014-9

Pabst, S., Brand, M., & Wolf, O. T. (2013). Stress and decision making: A few minutes make all the difference. *Behavioural Brain Research*, *250*, 39–45. https://doi.org/10.1016/j.bbr.2013.04.046

Peng, J., Feng, T., Zhang, J., Zhao, L., Zhang, Y., Chang, Y., Zhang, Y., & Xiao, W. (2019). Measuring decision-making competence in Chinese adults. *Journal of Behavioral Decision Making*, *32*(3), 266–279. https://doi.org/10.1002/bdm.2114

Peng, J., Li, H., Miao, D., Feng, X., & Xiao, W. (2013). Five different types of framing effects in medical situation: A preliminary exploration. *Iranian Red Crescent Medical Journal*, *15*(3), 161–165. https://doi.org/10.5812/ircmj.8469

Permatasari, A. W., Budiarti, L., & Srirejeki, K. (2020). The effect of sunk cost, framing effect, and educational background on the escalation of commitment. *Journal of Contemporary Accounting*, *2*(1), 49–61. https://doi.org/10.20885/jca.vol2.iss1.art5

Petscher, Y., Mitchell, A. M., & Foorman, B. R. (2015). Improving the reliability of student scores from speeded assessments: An illustration of conditional item response theory using a computer-administered measure of vocabulary. *Reading and Writing*, *28*(1), 31–56. https://doi.org/10.1007/s11145-014-9518-z

Piñon, A., & Gambara, H. (2005). A meta-analytic review of framing effect: Risky, attibute and goal framing. *Psicothema*, *17*(2), 325–331. http://www.psicothema.com/pi?pii=3107

Rachev, N. R., Geiger, S. J., Vintr, J., Kirilova, D., Nabutovsky, A., & Nelsson, J. (2022). Actively open-

minded thinking, bullshit receptivity, and susceptibility to framing: Evaluating the dual-process account in North America and Bulgaria. *European Journal of Psychological Assessment*, *38*(6), 440–451. https://doi.org/10.1027/1015-5759/a000685

Rahayu, W., Hayat, B., & Putra, M. D. K. (2023). *Analisis Rasch: Aplikasi dan interpretasi*. UNJ Press.

Reise, S. P. (1999). Personality measurement issues viewed through the eyes of IRT. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 219–241). Erlbaum. https://doi.org/10.4324/9781410603593

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, *34*(Suppl 1), 1–97. https://doi.org/10.1007/BF03372160

Solomon, E., & Kopelman, R. E. (1984). Questionnaire format and scale reliability: An examination of three modes of item presentation. *Psychological Reports*, *54*(2), 447–452. https://doi.org/10.2466/pr0.1984.54.2.447

Su, S., Wang, C., & Weiss, D. J. (2021). Performance of the S−χ2 statistic for the multidimensional graded response model. *Educational and Psychological Measurement*, *81*(3), 491–522. https://doi.org/10.1177/0013164420958060

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Toplak, M. E. (2022). *Cognitive sophistication and the development of judgment and decision-making*. Academic Press Inc.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, *59*(4), S251–S278. https://www.jstor.org/stable/2352759

Weller, J. A., Ceschi, A., & Randolph, C. (2015). Decision-making competence predicts domain-specific risk attitudes. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00540

Wright, B. D. (1977). Misunderstanding the rasch model. *Journal of Educational Measurement*, *14*(3), 219–225. https://doi.org/10.1111/j.1745-3984.1977.tb00039.x

Wright, B. D. (1996). Reliability and separation. *Rasch Measurement Transactions*, *9*(4), 472. https://www.rasch.org/rmt/rmt94n.htm

Wyszynski, M., & Diederich, A. (2023). Individual differences moderate effects in an unusual disease paradigm: A psychophysical data collection lab approach and an online experiment. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1086699

Zhen, S., & Yu, R. (2016). All framing effects are not created equal: Low convergent validity between two classic measurements of framing. *Scientific Reports*, *6*(1), 30071. https://doi.org/10.1038/srep30071