Analyzing The Construct Validity of the Indonesian Basic Empathy Scale using the Rasch model

Noviana Diswantika  
Indonesia University of Education

Uman Suherman  
Indonesia University of Education

Nandang Rusmana  
Indonesia University of Education

Dodi Suryana (dodisuryana@upi.edu)  
Indonesia University of Education

Abdullah Qurbi  
Senior High School 3 Bandar Lampung

Amelia Mohd Noor  
Sultan Idris Education University

Research Article

Keywords: Basic empathy scale, empathy, post-traumatic stress disorder, Rasch model

Posted Date: October 13th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3364544/v1

License: ☕️ ☀️ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License

Additional Declarations: No competing interests reported.
Abstract

**Background** The COVID-19 pandemic has brought many changes to human social life. The ability for empathy plays an essential role in these changes because empathy enables the process of sharing experiences, needs, and desires between individuals. In this study, the author aims to evaluate, develop and modify the Basic Empathy Scale in Adults (BES-A) to adapt it to Indonesian culture, which will be transformed into Indonesian Basic Empathy Scale (I-BES).

**Methods** The methodology used in this research is a survey design. Research using a survey design, the participants, Indonesia (n= 101, male 17, female 84 regions 5 region), evaluation was done by analyzing the Indonesian Empathy Scale using the RASCH model.

**Results** Show that the instrument does not have unidimensionality. Based on the results of fitting the item model, several items on the I-BES were omitted because they were deemed inappropriate, including at least five items. The instrument has good dispersion reliability among items but low dispersion reliability among individuals. The Indonesian Basic Empathy Scale needs further improvement to obtain response categories based on its functions.

**Conclusion** This study can be used as teaching material for counseling and guidance groups, and the personality theory of sharpening the material regarding the development of empathy, in addition to the measurement instruments for the development of empathy, which can be used as material for counseling and guidance assessment courses.

**Introduction**

The COVID-19 pandemic has significantly affected human existence. The social lives of people have been altered by the COVID-19 pandemic, causing various mental health issues for individuals that may exacerbate pre-existing problems [1]. Many events during the COVID-19 pandemic are traumatic and can have emotional effects, such as anxiety, depression, and trauma [2], [3]. Individuals who have difficulties adapting to a traumatic event have serious consequences. One is that individuals can experience post-traumatic stress disorder [4], [5].

COVID-19 has become an impetus for individuals to share, understand, and help others through the affective side [6]–[8]. Although the COVID-19 pandemic had a tremendous negative impact, it also had a positive impact that promotes the psychological side of individuals, which is to increase the awareness of empathy [9]. Empathy promotes the individual's motivation to follow all existing regulations to help others, especially those regulations that are in place during the COVID-19 pandemic [10]. Behavior generated by empathy - which is not based on any interests - can positively affect the environment to behave similarly. In other words, behavior that arises from empathy can lead other individuals to empathize and to prosocial behavior [10]–[12].
Empathy enables individuals to relate to the feelings of others. Empathy is a factor that encourages someone to engage in social interactions that are attentive to the well-being of others [13]. In other words, empathy plays a vital role in interpersonal and social relationships. Empathy enables individuals to share experiences, needs, and desires. This process leads people to experience satisfaction in social relationships and become an emotional bridge that can promote positive behavior [14]–[16]. For people who have experienced a traumatic event, emotional encouragement through empathy can help them escape the symptoms of Post Traumatic Stress Disorder (PTSD) [17].

The World Health Organization (WHO), which continuously urges every country to continue vaccinating, and many countries have now reached the specified vaccination limit, suggests that the COVID-19 pandemic may slowly be overcome. However, it is impossible to predict when the endemcity will begin [18]. This possibility must be considered when considering that traumatic events such as the COVID-19 pandemic significantly increase empathy. When the COVID-19 pandemic begins to become endemic, the things that trigger increased empathy also decrease. This information does not mean that it takes other traumatic events to elicit empathy in the individual capacity but that there must be efforts that can improve or even increase the empathy ability of individuals. One of the efforts that can be made is to measure empathy instruments to know what individuals need to get information about their needs regarding empathy abilities. In this way, people can be offered the right services to maintain or improve their empathy abilities.

Its development was done as a revision and reconstruction of the shortcomings of the Basic Empathy Scale (BES) to make it more comprehensive. This instrument is based on [20] definition of empathy as "sharing and understanding other people's emotional states or circumstances as a result of experiencing emotional states and comprehending individual emotions."

The process of developing an empathy instrument showed that the results of the instrument were declared valid using the Classic Test Theory method [21]. However, the instrument being analyzed has limitations in the Classic Test Theory validation process. When two different types of tests are given to two groups, the results cannot be compared. In addition, the difficulty level and weighting of the questions greatly depend on the measurements taken [22]. In addition, the instrument has weaknesses from using classic theory to measure the validity and reliability of an instrument construct. The measurement results will depend on the test characteristics used, and item parameters depend on the test takers' ability. Measurement errors can only be known for non-individual groups [23]. The Rasch model is an alternative way to investigate instrument validity. The Rasch model presents the scale structure of an instrument so that the disclosure of empathy instruments is valid or invalid [24], [25].

The Rasch model can be used to analyze the results of the instrument. The advantage of the Rasch model is that it provides accurate results regarding participants and instrument quality by creating a measurement scale with equal intervals. The advantages of the Rasch model are: (1) it can predict missing data; (2) it provides a linear scale with the same interval; (3) it detects inaccuracies; (4) it provides more accurate estimates/estimates; and (5) it provides replicable measurements [26].
research was conducted to fill the gaps in other instrument analysis models by conducting instrument tests that measure empathy in adolescents using the Rasch model. The BES-A instrument developed by [19] was further adapted and modified to fit the respondents’ circumstances in Indonesia. Adjustments were made regarding the language and context of the statements used, which did not eliminate the concepts that Carré had developed. The items resulting from the adaptation and modification process were fifteen items. Therefore, the instrument had to be analyzed and evaluated to maintain its measurement accuracy. In this context, the development and modification of BES-A are called the Indonesia Basic Empathy Scale (I-BES).

The RASCH model measures I-BES to determine its validity and reliability. Several studies evaluating the instruments have mentioned that the RASCH model is a powerful tool for determining validity and reliability, even with a small sample size [27], [28]. The RASCH model provides a probabilistic view of the object of measurement. Therefore, a RASCH model analysis is not deterministic and can more accurately identify the observed item. The RASCH model can also address metric disparities between items and difficulties in data intervals [29].

In this article, we aimed to evaluate the I-BES modified and adapted based on the Indonesian context through comprehensive techniques by analyzing the RASCH model, Based on the research problem presented. The results of this analysis will contribute to the evaluation of the questionnaire and significantly influence the literature in different contexts. Specifically, In the study, we identified appropriate respondents with I-BES, identified invalid items, confirmed their construction, and indicated the reliability between items and respondents.

**Methodology**

**Research Design**

The methodology used in this research is a survey design. Research using a survey design describes quantitative (numbers) related to empathy [30]. This research was conducted by collecting, classifying, and analyzing or processing data, drawing conclusions, and writing reports to objectively show the quality of the (I-BES) instrument in a description.

The study design was a survey study. A survey is a quantitative research technique in which the researcher interviews a sample or an entire population to determine the population's attitudes, beliefs, behaviors, or characteristics. For this study, a random sample was used, focusing on the population of Lampung, Indonesia. The basis for consideration is that the philosophical values of the Lampung cultural subculture demonstrate an openness to diversity that supports the development of the value of empathy. The social philosophy of Nengah Nyappur in Lampung culture is rooted in the cultural attitudes of the Lampung (ethnic) people. Thus, it can help avoid discriminatory social attitudes that can potentially cause unproductive social rifts in building community unity (integration) [31].

**Rasch Model**
In the 1960s, Georg Rasch constructed an analytical method of item response theory or item response theory (IRT) known as 1PL (one parameter logistic) [32], [33]. Ben Wright later popularized this mathematical model [34]. Rasch developed a model that connects students and things using raw data in dichotomous data (true and false) representing students' abilities [33].

In the field of social sciences, the data were obtained in the form of source numbers, which can usually be obtained in the form of attitudes and opinions about statement items or questions in a given instrument [35]. The instrument was designed based on satisfactorily defined variables. Subsequently, the relevant constructs were identified, and the items to measure the variables in question were developed. At the same time, the response options offered generally follow the scoring pattern of classical test theory (CTT). In the context of the Rasch model, this 'settled' scoring pattern is nothing more than a measurement whose results depend on who is being measured (test-dependent scoring), whereas quantitative research in the social sciences is concerned with objective measurement [36], [37].

According to Mok and Wright [32], objective measurement in the social sciences must meet five criteria: 1. providing a linear measure with equal intervals; 2. conducting an appropriate estimation procedure; 3. finding inappropriate items (misfits) or unusual ones (outliers); 4. overcoming data loss; 5. generating replicable measurements (independent of the parameters studied). Only the Rasch model can satisfy the five conditions of the five conditions. In other words, the quality of measurements in the social sciences made with the Rasch model will be the same as measurements in the field of physics [32], [33], [38].

Sample and Data Collection

The participants were guidance and counseling students of STKIP PGRI Bandar Lampung, a total of hundred and two students. Forty-eight students were in 2018-2019, and as many as fifty-four started their studies in 2020-2021. The distribution was based on the COVID-19 pandemic, which emerged at the end of 2019. The social interaction restrictions that impacted the world of education required students to take lectures online after the policy was enacted in mid-2020, around June-July. Namely, students in the 2018-2019 cohort were still delivering lectures offline and interacting a lot with their peers in lectures.

Meanwhile, students in the class of 2020-2021 did not hold intensive offline lectures because no policy allows lectures to be held with a 100% capacity in a class. In addition, participants were also categorized by gender - a total of eighty-five female students and seventeen male students. Specifically, the distribution of participants is described in Table 1.

Table 1. Demographic Characteristics of the Respondents
The Basic Empathy Scale in Adult (BES-A)

The Basic Empathy Scale-A (BES-A) measures items derived from the Basic Empathy Scale (BES) [39]. BES-A was developed by Carré et al. (2013) as a revision and reconstruction of the shortcomings of BES so that it is more comprehensive than the previous item. BES has forty items, whereas BES-A consists of twenty items aimed at adults. In the context of development, the average student is already in the adult stage, so this item should be able to measure empathy comprehensively.

BES is based on Cohen and Strayer's (1996) definition of empathy as "sharing and understanding the emotional state or context of others as a result of experiencing the emotional state (affective) and understanding the emotions (cognitive) of others" The BES assesses five basic emotions (fear, sorrow, anger, and joy) as well as broader cognitive and affective empathy measures than nonspecific emotional states (e.g., anxiety). For a forty-item scale, items were provided with reversed wording, with twenty items requiring a positive answer and another twenty items requiring a negative response. A shortened twenty-item version is also available for use with adults in France. However, several studies have found that the BES, a two-factor scale, does not consider newer ideas of empathy, which have found that empathy depends on three components [40], [41].

Modification of BES-A and Transformation to Indonesia Basic Empathy Scale (I-BES)

To develop a comprehensive empathy instrument and refer to the recommendations of several previous studies [40], [42], [43], the Indonesia Basic Empathy Scale (I-BES) refers to three main components of empathy. The empathy instrument construct measures empathy contagion (affect empathy), cognitive empathy, and emotional disconnection. Empathy contagion is a situation in which a person
automatically adjusts his or her emotions to the emotions or states of others. Cognitive empathy refers to a person's ability to understand and articulate emotions based on the effect of other people's emotions or situations. In addition, emotional disconnection is a regulatory component that involves self-protection from anguish, suffering, and emotional impact caused by other people's emotions or situations.

I-BES is an instrument analyzed and evaluated for validity and reliability. Since the research was conducted in Indonesia, the items were translated into Indonesian and adapted to the context of the participants. The language change was made to avoid irrelevant construct variables caused by the participants' language proficiency. Translators included a psychologist, a career counselor, and an English language expert. For the instrument to be comprehensively adapted to Indonesian culture regarding empathy, language translation equality, concepts, and metrics, considering the linguistic expressions and language culture of the local community under this study. Language translation involved forward translation, expert panel, back translation, pre-testing and cognitive interviewing, and final revision.

The instrument initially contained twenty items and was adjusted to fifteen based on empirical test results. The primary consideration of the instrument used had fifteen items; that is, the instrument was conducted through the stage of unidimensionality, and the item analysis consisted of item fit and item difficulty, using procedures and assumptions from the Rasch model; the results are fifteen items; by eliminating items not included in the criteria, procedures, and assumptions of the Rasch model—a 4-point Likert scale in the form of statements with sixty statements. Participants were asked to select from four statements for each item the one that applied to them. A summary of the correspondence between I-BES and item taxonomy is provided in Table 2.

Table 2. Summary of the I-BES

<table>
<thead>
<tr>
<th>Construct</th>
<th>Control Taxonomy</th>
<th>Number of Items</th>
<th>Item Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Empathy</td>
<td>Emotional Contagion</td>
<td>5</td>
<td>1-5</td>
</tr>
<tr>
<td></td>
<td>Cognitive Empathy</td>
<td>5</td>
<td>6-10</td>
</tr>
<tr>
<td></td>
<td>Emotional Disconnection</td>
<td>5</td>
<td>11-15</td>
</tr>
</tbody>
</table>

Data Coding

Participants' responses to the items in I-BES were coded using a predetermined Likert scale. The statement describing the most positive behavior received a score of 4, and the statement describing the most negative behavior received a score of 1.

Data Analysis Procedure
The RASCH model was used to analyze the instrument I-BES. An application that can analyze the RASCH model, WINSTEP [44], was used for the analysis. The RASCH model can be used to determine whether the responses given by participants correspond to their abilities [45]. The RASCH model estimates an item's ability and difficulty by setting the parameters of the person and the item to the same logit measure. The association between the parameters must be linear to satisfy the RASCH criterion of relative invariance [22], [46]–[49].

The term "item model fit" refers to how well an item fits the model. The statistical value of the item means fit square (MNSQ) was determined to evaluate the extent to which the items reflect the underlying concept [22], [35], [37], [49].

The MNSQ values infit and outfit are used to measure the item's fit to the model. The infit value can predict abnormal patterns in respondents' observation of items based on their ability. The infit values can also predict abnormal patterns in respondents' observation of items regardless of their ability level [50]. Ideally, the MNSQ value, which indicates that the item is consistent with the RASCH model, is 1.0, and researchers believe that the criteria for an acceptable MNSQ value range from 0.6 to 1.4 [34], [50], [51]. The item is considered a misfit if the MNSQ score does not meet these criteria. Misfit items indicate that the item does not measure what it is supposed to measure [34].

Rating scale analysis was used to analyze the degree of accuracy of the scale. I-BES is suitable for rating scale analysis because it uses a 4-point Likert scale that assumes a constant threshold at each item, a common characteristic in scales that measure personality or attitudes [52]. Analysis was conducted to determine whether the relative difficulties of steps within the items were constant, the psychological distances between statement choices, and whether or not they had the same position [53].

The difficulty of the items can be determined using the RASCH model analysis. The RASCH analysis creates a distribution map between respondents and items that graphically represents the items' difficulty level and the respondent's ability level. The RASCH model estimates the item difficulty and respondent ability parameters using logit to form the same linear interval scale [35]. Thus, both parameters can be compared simultaneously to determine whether the items available match the respondent's ability.

Reliability measures were analyzed using the RASCH model, which was conducted to determine the reliability and separation index level for individuals and items [54], [55]. The reliability value was obtained to determine how much the scale can discriminate between persons and items. The reliability value must be between 0 and 1 [36]. The separation index value indicates how much of the distribution of persons or items can be identified in the measured variable. The separation value is at least 2 to provide adequate separation between persons, items, or [56][57].

Unidimensionality is an analysis used to determine the construct validity of the instrument. This analysis aims to identify the ability of the instrument to measure the range of respondents' abilities [24], [49]. Instruments that have a valid construct measure respondents from highest to lowest ability. This analysis determines whether the respondent's measured ability is correct. In addition, the one-dimensionality
analysis measures the variables comprehensively. The analysis is conducted to determine if the items used in the instrument measured what should be measured [37], so that the instrument can provide accurate information. The value to be considered is the natural variance explained by the measures. The instrument measures the diversity of respondents' abilities if the value is at least 20%. In addition, the value of unexplained variance in the first contrast is accepted if the value does not exceed 15%, so the instrument does not have significant noise. The eigenvalues are also considered to determine if the items are not measuring what should be measured. Initially, if the eigenvalue of unexplained variance is above 2, other items are not measuring what should be measured. This result is evidenced in the standardized residual contrast plot.

**Findings**

**Item Model Fit**

Item-model fit for the I-BES instrument was measured by looking at the MNSQ value on infit and outfit. Misfit items with MNSQ values greater than 1.4 will be analyzed further. This result is because items that do not fit will give a picture of information that should not be measured. Four items are identified as misfits: item 2, item 1, item 5, and item 14. Detailed descriptions are presented in Table 3.

<table>
<thead>
<tr>
<th>Entry Number</th>
<th>Item Difficulty</th>
<th>Infit Mean-Square</th>
<th>Outfit Mean-Square</th>
<th>Control taxonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.84</td>
<td>1.74</td>
<td>1.82</td>
<td>Emotional Contagion</td>
</tr>
<tr>
<td>1</td>
<td>-0.35</td>
<td>1.77</td>
<td>1.43</td>
<td>Emotional Contagion</td>
</tr>
<tr>
<td>5</td>
<td>-1.37</td>
<td>1.60</td>
<td>1.06</td>
<td>Emotional Contagion</td>
</tr>
<tr>
<td>14</td>
<td>0.22</td>
<td>0.55</td>
<td>0.57</td>
<td>Emotional Diconnection</td>
</tr>
</tbody>
</table>

Note. MNSQ = mean square

The four items declared to be misfits were removed, leaving eleven items. Therefore, ten items are declared fit with the model after going through the item-model fit process twice in the analysis process. Table 4 describes the twelve items that were declared misfits in the second analysis after the items that were declared misfits in the first analysis were removed. The following is also Table 5, which displays the items that were declared fit with the model.

**Table 4. Descriptions and Item fit Statistics of items 12**

<table>
<thead>
<tr>
<th>Item No</th>
<th>Item Difficulty</th>
<th>Infit Mean-Square</th>
<th>Outfit Mean-Square</th>
<th>Control taxonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>-1.58</td>
<td>1.48</td>
<td>0.72</td>
<td>Cognitive Empathy</td>
</tr>
</tbody>
</table>
Table 5. Item Fit Statistics of the Final Refit Model

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Item Difficulty</th>
<th>Infit Mean-Square</th>
<th>Outfit Mean-Square</th>
<th>Control taxonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.55</td>
<td>1.08</td>
<td>1.08</td>
<td>Emotional Contagion</td>
</tr>
<tr>
<td>4</td>
<td>0.70</td>
<td>1.01</td>
<td>0.90</td>
<td>Emotional Contagion</td>
</tr>
<tr>
<td>6</td>
<td>-0.04</td>
<td>0.85</td>
<td>0.85</td>
<td>Cognitive Empathy</td>
</tr>
<tr>
<td>7</td>
<td>-0.80</td>
<td>1.15</td>
<td>0.83</td>
<td>Cognitive Empathy</td>
</tr>
<tr>
<td>8</td>
<td>0.98</td>
<td>1.25</td>
<td>1.16</td>
<td>Cognitive Empathy</td>
</tr>
<tr>
<td>9</td>
<td>0.20</td>
<td>0.65</td>
<td>0.71</td>
<td>Cognitive Empathy</td>
</tr>
<tr>
<td>10</td>
<td>-0.59</td>
<td>1.10</td>
<td>0.98</td>
<td>Emotional Disconnection</td>
</tr>
<tr>
<td>11</td>
<td>-0.19</td>
<td>1.25</td>
<td>1.20</td>
<td>Emotional Disconnection</td>
</tr>
<tr>
<td>13</td>
<td>0.42</td>
<td>0.85</td>
<td>0.83</td>
<td>Emotional Disconnection</td>
</tr>
<tr>
<td>15</td>
<td>-1.23</td>
<td>1.16</td>
<td>0.85</td>
<td>Emotional Disconnection</td>
</tr>
</tbody>
</table>

After making adjustments by eliminating items that do not fit or have infit values and MNSQ outfits that are less or more than 0.6 and 1.4, 10 items were found to have been declared fit. The average MNSQ infit and outfit scores were 1.03 (SD = 0.18) and 0.94 (SD = 0.15), indicating that the items in the instrument fit the model.

Rating Scale

Rating scale analysis was conducted to determine whether each statement on each item could be understood and used by the respondent as intended. The rating scale values are presented in Table 6 as follows.

Table 6. Rating Scale Structures for the I-BES
<table>
<thead>
<tr>
<th>Category</th>
<th>Observed Average</th>
<th>Expected Average</th>
<th>Infit Mean-Square</th>
<th>Outfit Mean-Square</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.28</td>
<td>-0.20</td>
<td>0.91</td>
<td>0.84</td>
<td>None</td>
</tr>
<tr>
<td>2</td>
<td>0.32</td>
<td>0.25</td>
<td>1.08</td>
<td>0.98</td>
<td>0.26</td>
</tr>
<tr>
<td>3</td>
<td>0.86</td>
<td>0.82</td>
<td>0.91</td>
<td>0.86</td>
<td>-1.14</td>
</tr>
<tr>
<td>4</td>
<td>1.59</td>
<td>1.62</td>
<td>1.18</td>
<td>1.08</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Note. Obsv. Avrg = Observed Average, Expt. Avrg = Expect Average, MNSQ = Mean Square

Table 6 shows that respondents did not understand the difference in answers if the observer’s average and Andrich’s threshold values decreased in the alternative third, so it was suspected that the third choice confused the respondent, the assumptions of the Rasch analysis model of the confusing choice had to be removed, to examine more sharply can be analyzed through figure 1 shows the probability curve of the answer choice categories to describe the probability of the responses in each category of answer choices which are estimated based on the score and difficulty level of the item.

Category probability curves show that category one curves do not intersect through categories 2, 3, and 4, respectively. The threshold value in Table 5, which has decreased on a scale of 2 to 3, is proven to be a discrepancy, resulting in an assumption, not by the intended scale. In other words, scales 2 and 3 are often considered similar, causing bias. The omitted scale is a scale of 3 with a decreasing threshold value.

**Compatibility of difficulty level with sample**

The item’s difficulty level and the respondent’s ability level will be illustrated in Figure 2, the map variable. The item difficulty level and the respondent’s skill level are ordered from most difficult to least complicated. Respondents with more excellent empathy abilities and the most difficult things in sequence will be at the top, while those with lesser empathy abilities and the most accessible items will be at the bottom.

**Reliability Measures**

Table 6 shows the summary of the reliability of the person and item. The summary is the result after eliminating items that are declared misfits. Table 7 summarizes the reliability of persons and items.

**Table 7. Person and Item Reliability Summary Statistics**
Table 7 shows that the person has a separation value of 1.17, and the item has a value of 4.15. This indicates that the results obtained will likely change when repeated measurements are made.

**Dimensionality**

Dimensionality is a Rasch-residual-based principal component analysis (PCA) that identifies whether I-BES can be stated as dimensional. Previously, I-BES had been adjusted based on item-model fit analysis by eliminating items declared to be a misfit. The results of dimensionality after adjustment with before adjustment has changed. The following is Table 8, which displays a summary of PCA results.

**Table 8. Summary of PCA Results**

<table>
<thead>
<tr>
<th>Original Scale</th>
<th>First factor Eigenvalue units</th>
<th>First contrast Eigenvalue units</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.4 (29.8%)</td>
<td>2.0 (9.2%)</td>
</tr>
<tr>
<td>First Revision</td>
<td>6.4 (36.6%)</td>
<td>1.7 (9.6%)</td>
</tr>
<tr>
<td>Second Revision</td>
<td>5.4 (35.0%)</td>
<td>1.6 (10.2%)</td>
</tr>
</tbody>
</table>

Note. First Revision = items 2, 1, 5, and 14 were removed from the original scale; Second Revision = item 12 was removed from the first revision scale.

Table 8 shows a change in the value of first-factor and first-contrast eigenvalue units. The first-factor eigenvalue unit shows that the original scale up to the second revision has natural variance explained by quite reasonable measures because it is above 20%. Even the increase occurred after the first revision, although it decreased in the second revision but was insignificant. Thus, the instrument can quite measure what should be measured, especially using an instrument revised second.

**Discussion**
The MNSQ value for infit and outfit was used to determine the fit of the item model for the instrument I-BES. To obtain better results or "Rasch-compliant" [34], results, the four items that were declared misfits were removed, leaving eleven items. The remaining items were reanalyzed using the same analysis. It is known that one item, namely item twelve, was declared a misfit. Therefore, ten items are declared as matching the model after going through the item-model matching process twice in the analysis process. Table 4 describes the twelve items declared as misfits in the second analysis after removing the items declared as misfits in the first analysis.

The observed and expected averages increase, but the threshold does not. A scale of 2 to 3 shows that the threshold has decreased. This result indicates that the scale used is still not functioning correctly [44], [58]. However, each scale's MNSQ infit and outfit values range from 0.8 to 1.2. This result shows that the scale used can be well understood and does not generate noise that leads to misinterpretation because the value is below 2.0 [58].

The threshold in Table 6, which fell on a scale of 2 to 3, is a discrepancy that results in an assumption, not by the intended scale. In other words, scales 2 and 3 are often considered similar, leading to bias. To overcome this bias, one of the categories is simplified or omitted [58] to become a 3-point Likert scale. The omitted scale is a 3-point scale with a decreasing threshold. Theory of Empathy [59] Students who understand relationships with others are categorized as being in an empathy learning process if they are categorized at a high level in the personality structure. Conversely, students who only do or act are weak in interacting with others so they may be categorized as weak in empathy. Students with weak relationships with others also affect their social life in the school environment and the community [60]. The difficulty of the items and the respondent's ability level are arranged from highest to lowest difficulty. Respondents with more excellent empathy abilities and the most difficult items in the sequence are at the top, while respondents with lower empathy abilities and the easiest are at the bottom. Figure 2 shows that the difficulty of items varies with different difficulty levels. From the comparison between the difficulty level of the item and the ability level of the respondents, it appears that most respondents were able to answer each item. This result shows that the diversity of respondents' abilities is less visible. Thus, the solution is to increase the number of items with different difficulty levels and functions. In this way, the diversity of respondents in terms of difficulty and function of each item becomes greater.

Nevertheless, no items have the same function, which can be appeared from the fact that no items are in the same position [56]. The items obtained from the analysis results will comprehensively describe individual empathic behavior, starting with identification; they will be considered for implementing preventive or curative efforts that can be carried out by educators or a team of experts, which include guidance and counseling (BK). BK, the efforts of educators are a form of help, understanding, and facilitating the optimal development of individual potential; if left unchecked, individuals will commit destructive acts. The 2020 United Nations Children's Fund (UNICEF) research states that out of 39,8675 students, 41% of adolescents at age fifteen had low empathy development and engaged in bullying behaviors at least several times a month, which, if left unchecked, would increase the risk of
psychological disorders later in life and poor social functioning [61], [62]. Strengthened by the research findings [63] which stated that 367 students out of a total of 1167 students had the weakest social interactions, it was predicted that students would experience acts of violence, bullying, and lower quality of learning in tertiary institutions. If students' weak development of empathy is left unchecked, they are at risk for frustration, acts of violence, beatings, sexual harassment, and murder [64].

It is known that the most challenging item is item 8 (Q8), with a value of 0.98, and the most accessible item is item fifteen (Q15), with a value of -1.23. The respondent with the highest ability has a value of 3.29, and the lowest ability has a value of -0.57. The difference between the maximum and minimum scores between the respondent's ability and the item's difficulty indicates that I-BES may not provide sufficient information, especially for respondents who scored much higher than that of the most challenging item. This result is because the gap between the scores of the most difficult items and high-ability respondents is greater than 5%, so the information provided by the instrument for respondents with scores greater than 5% of the most challenging items provides weak or less accurate information [58]. This result is caused by items with lower difficulty levels that are higher or more different than those measured. The number of items needs to be increased to maintain the variety of items that provide accurate information, especially for respondents with high empathy ability.

Table 7 shows that the person has a separation score of 1.17, and the item has a score of 4.15, indicating that the person has weak diversity while the item has excellent diversity [58]. This result shows that the instrument could not provide diverse information about the respondent's empathy ability. This result may be caused by the low difficulty level of the item, or in other words, it is too easy for the respondent. This result has already been explained in the section discussing the suitability of the difficulty level for the sample. Moreover, the reliability value of the person is 0.58, and that of the item is 0.95. The reliability value of the items can be considered excellent, but the reliability value of the person is relatively weak [58]. This finding indicates that item difficulty separation is declared reliable, but person separation reliability is not. This result suggests that the results obtained will likely change with repeated measurements.

Table 8 shows that the values of the eigenvalue units of the first factor and the first contrast have changed. The first factor, the eigenvalue unit, shows that until the second revision, the original scale had a natural variance that can be considered quite good since it is above 20%. After the first revision, there was an increase, while in the second revision, it decreased but was not significant. Thus, it can be concluded that the instrument is quite capable of measuring what should be measured, especially when using an instrument revised in the second revision.

The original scaling appears to have decreased in the first contrast eigenvalues, and the percentage increased in the second revision. This result shows that after the second revision, there are no items left in the instrument that does not measure what should be measured. The percentage of eigenvalue items in the first contrast increased, although not significantly. However, for the second improvement, the
percentage did not exceed 15 %, so it can be noted that the unexplained variance in the contrasts was quite good.

On the original scale, the value of the first contrast eigenvalue unit is 2.0, indicating the possibility that an item does not measure what it should. This value touches the tolerance value of the contrast eigenvalue units, 2.0. So it is necessary to eliminate or improve items that do not measure what should be measured. However, after the first and second improvements, the value of the first contrast eigenvalue unit has decreased so that it has not touched the value of 2.0. So, the adapted instrument has an item that measures what it should.

**Conclusion**

After some analysis and evaluation, the I- BES instrument is exceptionally well constructed. Initially, I-BES had 15 items, and now ten items after analysis. In the first item model fit analysis, four items (items 1, 2, 5, 14) were found to be inappropriate or did not measure what was intended to be measured, so they had to be removed. In addition, in the second item-model fit analysis, by eliminating four items declared as misfits, another item was found to be declared as a misfit, item 12. As a result, I-BES contains ten items that are indicated to fit the model or can measure what should be assessed.

In the response category analysis section, I-BES has a problem in one of the response categories. The third or third response category is not functioning correctly, and its function is still the same as that of the second scale or response category. In other words, respondents still consider answer choices 2 and 3 to be the same. Regarding understanding the sentences, the answer choices are still easy to understand and do not create noise that leads to misinterpretation.

After analysis, the difficulty with the sample is known to be the most challenging item in the instrument and, after revision, the easiest. For the items measured based on the logit parameter, it is known that the items have their respective functions, and there are no equal values, so the items are considered uniform. In addition, the ability of the highest and lowest respondents was known. As measured by the logit parameter, it is known that the highest level of response-ability exceeds the difficulty level of the items, well above 5 %. This result indicates that the instrument provides less accurate information to respondents when the logit value of the respondent's ability is 5 % higher than the logit item value. Thus, although the items have different functions, they cannot correctly identify the diversity of the respondents. This result indicates that increasing the number of items with different difficulty levels and functions is necessary.

In analyzing the reliability of the diversity of persons and objects, it is also known that the diversity of persons is only 1.17. This result illustrates the previous discussion that the instrument can identify the diversity of groups only up to 1 group (rounded 1.17). This result means that the instrument will only provide weak information about the diversity of respondents' abilities. Moreover, the item's reliability value can be considered excellent, but the reliability value of the person is weak. This finding shows that the item's difficulty level is the same throughout. However, the person's reliability to be weak indicates that the
value is unreliable. If the respondent is allowed to complete the instrument again, there will likely be a significant change in ability. This result may be caused by the inconsistency of the answers given by the respondent with the respondent's abilities.

In the dimensionality test, it is known that from the original scale to the second revision, nothing is below the minimum limit of the eigenvalue units of the first factor. In addition, for the eigenvalue units of the first contrast, it is known that there is evidence in the original scale that some items do not measure what should be measured. After the first and second revisions were made based on the results of the model-item fit and a dimensionality test was conducted, it was found that the items that did not measure what should be measured no longer existed. Therefore, it can be concluded that the instrument has a reasonably good construct and has items that measure what should be measured. The results of this study can be used as teaching material for counseling and guidance groups, and the personality theory of sharpening the material regarding the development of empathy, in addition to the measurement instruments for the development of empathy, which can be used as material for counseling and guidance assessment courses.

Declarations

Acknowledgments

The authors would like to thank all the participants involved in this research.

Authors’ contributions

All authors reviewed the manuscript. ND and DS drafted the text of main manuscript text. US, NR and AMN supervision article, ND, AQ and DS collected the data and conducted analyses. AMN edited the manuscript for content. All Authors read and approved the final manuscript.

Funding

This research is not funded by the university.

Data Availability

Data is available from the corresponding author upon reasonable request

Competing interest

The authors declare no competing interests.

Ethics approval and consent to participate

The research conducted hereby states that Analyzing The Construct Validity of the Indonesian Basic Empathy Scale using the Rasch model, Compiled and created exclusively by me (and co-authors, if any)
and is what I claim, and will claim, to be correspondence author, fully approved by the co-author (if any). certify, on behalf of myself and co-authors (if any), that I/we have no conflict of interest relation to this work. certifies that this article is not being reviewed by any other journal. certify that this article contains original data, the integrity and accuracy of which I (and the co-authors, ifwhatever) assume responsibility for. This research used a sample of students, and informed consent was given to collect data to ensure willingness to provide data voluntarily and accurately.

Consent for publication

Not applicable.

References


64. J. G. Culhane, *Sticks and Stones: Defeating the Culture of Bullying and Rediscovering the Power of Character and Empathy, by Emily Bazelon*. HeinOnline, 2012.

**Figures**
Figure 1

Category Probability Curves for the I-BES
Figure 2

Map of Persons and Items for I-BES