Iranian Evolutionary Educational Psychology Journal
IEEPJ
University of Hormozgan
Online ISSN: 2588 - 4395
Homepage: https://ieepj.hormozgan.ac.ir

Psychometric Properties of Intellectual Humility Questionnaire: The Classical Test Theory and Graded Item Response Theory

Elham Ranjbari[1], Azita Amirfakhraei[2]✉, Noshin Taghinejad[3]

1. Ph.D. Student, Department of Psychology, BA.C., Islamic Azad University, Bandar Abbas, Iran
2. Associate Professor, Department of Psychology, BA.C., Islamic Azad University, Bandar Abbas, Iran, afakhraei2002@iau.ac.ir
3. Assistance Professor, Department of Psychology , BA.C., Islamic Azad University , Bandar Abbas, Iran

Article Info

Article type:
Research Article

Article history:
Received 03 Jun. 2025
Received in revised form 15 Jul. 2025
Accepted 27 Sep. 2025
Published online 01 Jun. 2026

Keywords:
Intellectual Humility,
Psychometrics,
Classical theory of test scores,
Graded item-response theory

ABSTRACT

Objective: This study aimed to evaluate the psychometric properties of the Intellectual Humility Questionnaire among students at Islamic Azad University, Bandar Abbas Branch, using a developmental psychometric approach.

Methods: A total of 500 students were selected through cluster sampling, and data from 492 participants were analyzed. The instrument was the Intellectual Humility Questionnaire developed by Cromery and Rose (2016). Content validity was assessed using the Waltz and Bausell method, calculating the Content Validity Ratio (CVR) and Content Validity Index (CVI). Item analysis was conducted using Classical Test Theory (CTT) and Graded Item Response Theory (IRT). Factor structure, item information functions, and reliability indices were examined.

Results: Content validity indices indicated satisfactory CVR and CVI values. Factor analysis supported a four-factor structure: intellectual independence, respect for others' opinions, openness to reconsideration, and overconfidence, with factor loadings ranging from 0.60 to 0.92. Item 1 showed the highest loading (0.85) for intellectual independence, and item 19 showed the highest loading (0.92) for overconfidence. Item information function analysis revealed that most items provided maximum information around the average theta level. Item 1 was most informative within $\theta = 0$ to $+1.6$, and item 19 within $\theta = -2$ to $+1.6$. IRT marginal reliability coefficients ranged from 0.80 to 0.90. Internal consistency analysis showed an overall Cronbach's alpha of 0.899, with subscale reliabilities ranging from 0.778 to 0.846.

Conclusions: The findings indicate that the Intellectual Humility Questionnaire demonstrates strong validity, reliability, and measurement precision, making it a suitable tool for assessing intellectual humility among university students.

Cite this article: Ranjbari, E., Amirfakhraei, A. & Taghinejad, N. (2026). Psychometric properties of intellectual humility questionnaire: the classical test theory and graded item response theory. *Iranian Evolutionary Educational Psychology Journal*, 8 (2), 1-29.
DOI: https//doi.org/ 10.22034/8.2.1

Publisher: University of Hormozgan.
DOI: https// doi.org/ 10.22034/8.2.1

Introduction

Intellectual humility is a mindset, disposition, or personality characteristic that guides our responses to evidence and information, as individuals seek truth and strive to avoid error. Over the years, intellectual humility has been examined from various perspectives and has been regarded as a virtue. In contrast to arrogant individuals, intellectually humble individuals tend to evaluate their abilities and achievements more accurately. They acknowledge their flaws and limitations, remain open to conflicting information, place less focus on themselves and more on others, and appreciate the value of all perspectives (Alfano et al., 2017).

With the rise of positive psychology, intellectual humility has become one of the most compelling areas of psychological research, as it has been shown to exert a broad influence on human functioning. Nevertheless, research on intellectual humility remains in its early stages, and progress in understanding the construct has been challenged by diverse, multidimensional definitions and measurement approaches. Moreover, inconsistent findings have been reported in this field (Cromery & Rose, 2016).

Humility may be considered a virtue that promotes moral character, whereas intellectual humility can be classified as an epistemic virtue that promotes being a good thinker. Diverse theoretical foundations have led to varying definitions of intellectual humility. Some scholars have defined intellectual virtues as the absence of their counterparts, such as intellectual arrogance, suggesting that intellectual humility becomes a virtue when individuals are attentive to their own epistemic condition. Another conceptualization frames intellectual humility as an Aristotelian mean between two vices, such as a balanced and accurate understanding of one's cognitive capacities, positioned between intellectual grandiosity and intellectual diffidence. Accordingly, although intellectual humility involves openness to new ideas (McElroy et al., 2014) and willingness to revise one's views when necessary (Hopkin, Hoyle, & Toner, 2014), it does not imply indiscriminate deference to others. An intellectually humble individual is able to strike an appropriate balance between dogmatically rejecting opposing views and conceding too quickly in the face of intellectual disagreement.

Thus, intellectual humility entails holding one's beliefs with confidence while remaining open to alternative evidence. This capacity stems from awareness of one's epistemic limitations and the fallibility of one's knowledge, as well as the ability to distinguish what one knows from what one

does not. Other theories emphasize that knowledge and strength of belief should derive from one's epistemic position rather than external sources. For instance, intellectual humility involves reluctance to treat one's beliefs as unquestionably true and avoiding excessive self-investment in one's viewpoints. Consequently, intellectually humble individuals can generate or exchange differing perspectives without hostility (McElroy et al., 2014) and show respect toward those who hold opposing views.

Based on existing theories, intellectual humility is defined as a non-threatening awareness of one's intellectual fallibility. According to the literature, such awareness should lead to openness to revising one's views, avoidance of overconfidence in one's knowledge, respect for others' opinions, and a non-defensive stance toward intellectual disagreement. Accordingly, intellectual humility is conceptualized as both an intrapersonal and interpersonal construct associated with outcomes such as open-mindedness and tolerance toward others (Cromery & Rose, 2016).

In recent years, research on humility has expanded into more specific domains, such as relational humility (Davis et al., 2011), cultural humility (Hook et al., 2015), and spiritual humility (Davis et al., 2016). These forms of humility can be viewed as reflecting the moral dimension of virtues. However, virtues are not solely moral; they can also possess an intellectual dimension (Baehr, 2011). Intellectual humility represents a pioneering field of study that integrates these two dimensions and requires acknowledging that one's knowledge and cognitive abilities are limited and imperfect (Alfano et al., 2017).

Researchers in psychology have defined intellectual humility as "insight into the limits of one's knowledge" (McElroy et al., 2014), "a non-threatening awareness of one's intellectual fallibility" (Cromery & Rose, 2016), and "the recognition that a particular personal belief may be fallible, accompanied by appropriate attention to the limitations of the evidential basis for that belief and one's own limitations in acquiring and evaluating relevant information" (Leary et al., 2017). In summary, the core feature of intellectual humility is the recognition that one's beliefs and knowledge may be incorrect or limited.

Whitcomb et al. (2017) further proposed that intellectual humility consists of four fundamental components: cognitive, behavioral, motivational, and emotional responses. First, individuals must recognize and accept their intellectual limitations and the potential negative consequences arising from them. Second, as a desirable behavioral outcome, individuals should acknowledge these

limitations to others in order to correct or compensate for them; for example, by adjusting their viewpoints and responding appropriately to new information. Third, individuals should be motivated to identify their intellectual strengths and weaknesses and be willing to revise their views when necessary. Fourth, emotional responses should be appropriate; for instance, regret or disappointment may be acceptable, whereas hostility is not.

From another perspective, Samuelson et al. (2015) suggested that intellectual humility comprises epistemic and social dimensions. The epistemic dimension refers to truth-seeking and the pursuit of knowledge, including curiosity and a desire to learn. The social dimension can be divided into two aspects: an intrapersonal aspect, which concerns how individuals view themselves (e.g., avoiding self-aggrandizement and self-promotion), and an interpersonal aspect, which relates to how individuals interact with others. These two dimensions represent fundamentally distinct processes (Danovitch et al., 2019).

Intellectual humility can be conceptualized as a trait that varies from one individual to another; that is, individuals generally differ in their overall level of intellectual humility across situations and over time. It can also be understood as a state that fluctuates within individuals, reflecting the extent to which a person is intellectually humble in a specific situation or at a particular time. Accordingly, intellectual humility can be considered relatively stable, yet susceptible to change at a given moment due to internal or external factors. Furthermore, intellectual humility may be either general or domain-specific. General intellectual humility varies across individuals, while a single individual may exhibit intellectual humility regarding certain beliefs but intellectual arrogance regarding others (Hill et al., 2016).

With respect to the measurement of intellectual humility, studies such as those by Grossmann et al. (2010) and Kross and Grossmann (2012) employed a coding scheme based on structured interviews addressing social situations, including recognition of uncertainty and limitations of knowledge. Trained coders rated interview transcripts for intellectual humility on a scale ranging from 1 (not at all) to 3 (very much). The initial study did not report inter-rater reliability for intellectual humility separately from overall wisdom ratings. However, a subsequent study reported inter-rater reliability coefficients exceeding 0.90 for intellectual humility. According to Cromery and Rose (2016), and in light of critiques of the findings reported by Grossmann et al. (2010) and Kross and Grossmann (2012), which aimed to develop a more comprehensive and

versatile measure of intellectual humility, two intellectual humility scales were used as foundations: a domain-specific self-report measure of religious intellectual humility developed by Hopkin et al. (2014), and a general intellectual humility scale developed by McElroy et al. (2014). Hopkin et al. (2014) identified items related to religious beliefs that validly reflected intellectual humility, yielding four factors: awareness of belief fallibility (e.g., "When it comes to religious or spiritual beliefs, my beliefs are more accurate than others'"), caution in expressing beliefs (e.g., "Even when I hold strong religious or spiritual beliefs, not everyone needs to know about them"), comfort with holding private beliefs (e.g., "It is important to share my religious or spiritual views with others regardless of whether they agree"), and respect for others' beliefs (e.g., "I listen to others' religious or spiritual beliefs even when I think I am right"). Internal consistency coefficients ranged from 0.71 to 0.89. Thus, this measure provides a relatively broad conceptualization of religious intellectual humility, encompassing both belief awareness and respect for others' beliefs. However, because the scale was designed specifically to assess religious intellectual humility and includes items explicitly referring to religious and spiritual beliefs, it was not intended for use as a general measure of intellectual humility.

The only general intellectual humility scale was developed by McElroy et al. (2014). This scale assesses perceptions of a target's intellectual humility based on the theoretical assumption that intellectual humility is fundamentally relational in nature and involves regulating interactions with others concerning belief- and worldview-related issues. McElroy et al. (2014) suggested that intellectual humility is particularly relevant in situations involving competition or negotiation of ideas within relationships or groups. Accordingly, the items primarily describe observable behaviors presumed to reflect intellectual humility or intellectual arrogance. The scale comprises two factors: intellectual openness (7 items; e.g., "Is good at mediating controversial topics") and intellectual arrogance (9 reverse-scored items; e.g., "Always has to have the last word in an argument"). Confirmatory factor analysis demonstrated good model fit for this scale (Cromery & Rose, 2016).

Cromery and Rose (2016) subsequently developed a 22-item scale measured on a five-point Likert scale (from strongly disagree to strongly agree) to assess individuals' levels of humility and intellectual humility (e.g., "I am willing to reconsider my important beliefs in the face of new information"). This scale consists of four factors: intellectual independence (5 items), openness to

revising one's viewpoint (5 items), respect for others' opinions (6 items), and lack of overconfidence (6 items). The scale demonstrated good internal consistency, test–retest reliability, and construct validity (Alfano et al., 2017).

According to the findings reported by Cromery and Rose (2016), participants were recruited across the United States via Amazon Mechanical Turk (MTurk). The dataset was randomly split into two halves (n = 380 each) for use in Study 1 and Study 2. Participants in Study 1 were 56% male and 44% female, aged between 18 and 71 years. The sample identified as 77% Caucasian, 9% Asian, 8% Hispanic, 4% Black or African American, and 2% other. Regarding highest completed education level, 36% had completed high school, 56% college, and 8% graduate education.

Using exploratory factor analysis to obtain a relatively concise measure, a maximum of six items per factor with the highest factor pattern and structural coefficients (minimum of 0.40) were selected, and items with cross-loadings exceeding 0.30 were removed. Although some items exceeded the 0.30 cross-loading threshold, those with the greatest separation between the two structural coefficients were retained. Principal axis factoring of the selected 22 items yielded a KMO value of 0.89 and a significant Bartlett's test of sphericity. All items loaded on their respective factors. The four factors accounted for 57.24% of the total variance and 48.25% of the common variance. As expected, correlations among factors ranged from 0.31 to 0.67. Cronbach's alpha was 0.88 for the total scale and ranged from 0.73 to 0.89 for the subscales.

Confirmatory factor analysis was conducted in Study 2 with participants (N = 380), of whom 59% were male and 41% female, aged 18 to 72 years. The sample consisted of 77% Caucasian, 8% Asian, 5% Hispanic, 5% Black or African American, 3% multiracial, and 2% other participants. In terms of education, 33% had completed high school, 55% undergraduate education, and 12% graduate education. The 22 items selected in Study 1 were entered into the CFA in Study 2. The results indicated that the four-factor higher-order model demonstrated strong fit indices. In Study 2, Cronbach's alpha was 0.87 for the total scale and ranged from 0.74 to 0.89 for the subscales.

With the expansion of public awareness regarding the importance, limitations, and consequences of measurement and assessment, criticisms of testing practices have also increased. Although numerous tests are developed and widely used, one of the most critical issues is ensuring that measurement instruments possess sound psychometric properties (Kiani, 2022). Psychometrics provides effective methodological approaches to address such challenges, with the ultimate goal

of achieving the greatest possible accuracy in measuring the target construct. The development and validation of tests are primarily grounded in Classical Test Theory (CTT) and Item Response Theory (IRT).

Classical Test Theory was originally introduced by Spearman in the early twentieth century and was subsequently refined over time. However, despite its development, certain limitations have persisted, motivating measurement specialists to propose alternative frameworks to overcome these shortcomings. This effort led to the emergence of Item Response Theory. Although IRT is not without limitations, it has addressed many of the constraints inherent in Classical Test Theory. The advancement of IRT reflects ongoing efforts by measurement and assessment experts to resolve the inadequacies of Classical Test Theory (Bertolotti et al., 2013).

Classical Test Theory is a relatively simple model that examines measurement error based on observed scores and is therefore often referred to as a weak true-score theory. In contrast, Item Response Theory offers several advantages over CTT. Within IRT, individual items contribute differently to the estimation of item parameters, and item selection is tailored to an individual's level of the latent trait. Consequently, IRT is considered a more appropriate and valid framework than Classical Test Theory for test development and validation (Gierl, 2013).

According to Edwards, Houts, and Cai (2018), Item Response Theory encompasses a family of latent variable measurement models that attempt to explain respondents' answers to a set of test items as a function of both item parameters and person parameters. In other words, IRT examines the relationship between varying levels of a latent trait and item responses through item characteristic functions (Hambleton, Swaminathan, & Rogers, 2010). Item Response Theory offers additional advantages, including assigning greater weight to items with higher discrimination power, resulting in scores that may be more valid than total scores derived from Classical Test Theory. Moreover, IRT-based scoring places scores from different test forms on a common scale, enabling the selection of the most appropriate test form for each examinee (DeMars, 2010).

Despite these advances, the lack of application of modern psychometric approaches—particularly Graded Item Response Theory—in estimating item parameters and overall scale parameters, along with the absence of standardized screening tools for assessing a broad range of intellectual humility characteristics, and significant shortcomings in the validation of existing instruments, have been

identified. Therefore, based on the reviewed evidence, the development and validation of a robust instrument for measuring intellectual humility is essential.

Accordingly, in the present study, to reduce measurement error and maximize the informational value of both individual items and the overall questionnaire, the Graded Item Response Theory approach was employed. This modern framework for examining psychometric properties and interpreting measurement results provides valid and reliable criteria for evaluating scale development and offers practitioners a precise assessment tool to facilitate accurate identification and effective intervention. Taking into account the limitations of previous instruments, the aim of this study was to examine the psychometric properties of the Intellectual Humility Questionnaire developed by Cromery and Rose (2016) among students at Islamic Azad University, Bandar Abbas Branch. The central research question was whether the Intellectual Humility Questionnaire demonstrates adequate psychometric properties based on Classical Test Theory and Graded Item Response Theory.

Material and Methods

This study was conducted using a developmental psychometric approach and a survey research method to examine the psychometric properties of the Intellectual Humility Questionnaire based on both Classical Test Theory (CTT) and Graded Item Response Theory (G-IRT). Graded Item Response Theory is specifically designed to estimate individuals' true scores in questionnaires whose items include ordered response categories. This model allows researchers to examine the probability that an examinee with a given level of a latent trait will select a particular response option (Ostini & Nering, 2011).

Translation and Cultural Adaptation

To implement the study, the Intellectual Humility Questionnaire developed by Cromery and Rose (2016) was first translated from English into Persian by a professional English language expert. Subsequently, a back-translation was conducted and compared with the original version. Any translation-related issues at the item level were reviewed and resolved through consultation with translation specialists and psychology faculty members.

Face validity, Content Validity Ratio (CVR), and Content Validity Index (CVI) were assessed by psychology experts using the method proposed by Waltz and Bausell. In addition, the translated

questionnaire was administered to a pilot sample of 30 students from Islamic Azad University, Bandar Abbas Branch, to evaluate clarity, comprehensibility, and simplicity. Based on participants' feedback, the necessary revisions were made.

Sample Size and Sampling Method

In psychometric studies, Comrey and Lee (2013) recommend a sample size of at least 10 participants per item. Klein (2010) suggests 10 to 20 participants per variable for exploratory factor analysis; however, for confirmatory factor analysis, the minimum sample size is determined based on the number of factors rather than variables (i.e., 10 to 20 participants per factor). According to Hooman (2011), a sample size of 100 is considered weak, 200 relatively adequate, 300 good, 500 very good, and 1,000 excellent for factor analysis.

Regarding sample size requirements for Item Response Theory models, Jiang, Wang, and Weiss (2016) demonstrated that a sample size of 500 participants is sufficient for accurate parameter estimation. Similarly, Blanchin et al. (2013) reported that a sample of 500 examinees provides adequate precision in estimating model parameters.

Concerning sampling strategy in IRT-based research, Embretson and Reise (2000) emphasized that model parameters should be estimated using a heterogeneous sample, particularly with respect to the latent trait. Such heterogeneity is crucial in IRT models for obtaining accurate estimates of item parameters, especially item discrimination.

Accordingly, considering potential sample attrition, the statistical population of this study consisted of students at Islamic Azad University, Bandar Abbas Branch. Using cluster sampling, 500 students were selected as the initial sample. After data collection and the removal of incomplete questionnaires, data from 491 participants were included in the statistical analyses.

Instrument

Intellectual Humility Questionnaire: The Intellectual Humility Scale developed by Cromery and Rose (2016) consists of 22 items rated on a five-point Likert scale ranging from strongly disagree to strongly agree. The scale measures individuals' levels of humility and intellectual humility (e.g., "I am willing to reconsider my important beliefs in the face of new information"). The scale comprises four factors: Intellectual Independence (5 items), Openness to Revising One's Viewpoint (5 items), Respect for Others' Opinions (6 items) and Lack of Overconfidence (6

items). The instrument has demonstrated good internal consistency, test–retest reliability, and construct validity (Alfano et al., 2017).

According to Cromery and Rose (2016), participants were recruited across the United States via Amazon Mechanical Turk (MTurk). The dataset was randomly divided into two equal subsamples (n = 380 each) for Study 1 and Study 2. Participants in Study 1 were 56% male and 44% female, aged between 18 and 71 years. The sample consisted of 77% Caucasian, 9% Asian, 8% Hispanic, 4% Black or African American, and 2% other ethnicities. In terms of education, 36% had completed high school, 56% college, and 8% graduate education.

Using exploratory factor analysis (EFA) to obtain a relatively concise measure, a maximum of six items per factor with the highest factor pattern and structural coefficients (minimum of 0.40) were retained, and items with cross-loadings exceeding 0.30 were removed. Although some items exceeded the 0.30 cross-loading threshold, those with the greatest separation between structural coefficients were retained.

Principal axis factoring of the selected 22 items yielded a KMO value of 0.89 and a significant Bartlett's test of sphericity, $\chi^2(231) = 3458.65$, $p < .001$. All items loaded on their respective factors. The four factors accounted for 57.24% of the total variance and 48.25% of the common variance. As expected, inter-factor correlations ranged from 0.31 to 0.67. Cronbach's alpha was 0.88 for the total scale and ranged from 0.73 to 0.89 for the subscales.

In Study 2, confirmatory factor analysis (CFA) was conducted with 380 participants (59% male and 41% female) aged 18 to 72 years. The sample consisted of 77% Caucasian, 8% Asian, 5% Hispanic, 5% Black or African American, 3% multiracial, and 2% other participants. Regarding education, 33% had completed high school, 55% undergraduate education, and 12% graduate education.

The 22 items retained from Study 1 were entered into the CFA. The results indicated that the higher-order four-factor model demonstrated strong fit indices. In Study 2, the Cronbach's alpha for the total scale was 0.87 and ranged from 0.74 to 0.89 for the subscales. In the present study, the Intellectual Humility Questionnaire developed by Cromery and Rose (2016) was used for the first time in Iran. Using the Waltz and Bausell method, the Content Validity Ratio (CVR) for intellectual humility was calculated as 0.97, and the Content Validity Index (CVI) values were obtained as follows: clarity = 0.95, simplicity = 0.96, and relevance = 0.97 for all items.

The assumption of local independence was supported based on Pearson's chi-square (χ^2) index under the application of Samejima's Graded Item Response Theory model, and the assumption of one-dimensionality was confirmed using multidimensional item response theory analysis as well as exploratory factor analysis (principal axis factoring). Based on both Classical Test Theory and Graded Item Response Theory, the estimated item parameters were satisfactory.

Item factor loadings ranged from +0.60 to +0.92, with Item 7 showing the lowest loading and Item 19 the highest. Specifically, for the Intellectual Independence component, Item 4 had the lowest loading (+0.65) and Item 1 the highest (+0.85). For Respect for Others' Opinions, Item 7 had the lowest loading (+0.60) and Item 11 the highest (+0.86). For Openness to Revising One's Viewpoint, Item 16 showed the lowest loading (+0.69) and Item 13 the highest (+0.75). Finally, for Lack of Overconfidence, Item 20 had the lowest loading (+0.61) and Item 19 the highest (+0.92).

The item information function indicated that the questionnaire provided the maximum information around the average theta (θ) levels. The greatest item information for Intellectual Independence was obtained from Item 1, with peak information across the theta range of 0 to +1.6. For Respect for Others' Opinions, Item 11 provided the highest information, with optimal information across –0.8 to +1.6. For Openness to Revising One's Viewpoint, Item 13 yielded the greatest information, with optimal information in the range of –2.8 to +0.8. For Lack of Overconfidence, Item 19 showed the highest information, with peak information across –2.0 to +1.6.

To estimate reliability, empirical (marginal) reliability was employed. Given that, within the Item Response Theory framework, the standard error of measurement varies across individuals, the results indicated that marginal reliability coefficients were 0.80 for Intellectual Independence, 0.86 for Respect for Others' Opinions, 0.81 for Openness to Revising One's Viewpoint, and 0.90 for Lack of Overconfidence.

Additionally, internal consistency was examined using Classical Test Theory. The internal consistency reliability of all 22 items was analyzed simultaneously. The results indicated a total scale reliability of 0.899. Reliability coefficients for the subscales were 0.827 for Intellectual Independence (5 items), 0.846 for Respect for Others' Opinions (6 items), 0.778 for Openness to Revising One's Viewpoint (5 items), and 0.843 for Lack of Overconfidence (6 items).

Ethical Considerations

This study was conducted in accordance with the ethical principles of human research and complied with the Declaration of Helsinki and the guidelines of the University Research Ethics Committee. Participation in the study was voluntary, and all participants were informed about the objectives and procedures of the research prior to data collection. Informed consent was obtained from all participants. They were assured of the confidentiality and anonymity of their responses and informed of their right to withdraw from the study at any stage without any consequences. All collected data were used solely for research purposes and were analyzed and reported in aggregate form to protect participants' privacy.

Results

To examine the face validity of the Intellectual Humility Questionnaire, the opinions of eight psychology experts were solicited regarding item wording, clarity, appropriateness, comprehensibility, and avoidance of technical terminology. Based on their feedback, the necessary revisions were made. Subsequently, the questionnaire was administered to the same eight experts to evaluate content validity. For this purpose, each item was rated on a three-point scale (not necessary = 0, useful = 1, essential = 2). The Content Validity Ratio (CVR) for the Intellectual Humility Questionnaire was calculated as 0.97. In addition, the Content Validity Index (CVI) values were obtained as 0.95 for clarity, 0.96 for simplicity, and 0.97 for relevance, indicating excellent content validity.

Prior to addressing the research questions, the overall quality of the data was examined. First, out-of-range values were identified and corrected. Next, missing data were evaluated, and questionnaires with more than 5% missing responses were excluded from further analyses. After verifying the statistical assumptions, the final analyses were conducted on a sample of 492 participants.

Table 1. Descriptive Statistics, Discrimination Indices, and Reliability Coefficients of the Intellectual Humility Items

Component	Item	Mean	SD	Discrimination	α if Item Deleted
Intellectual Independence	1	1.80	1.275	0.470	0.896
	2	2.19	1.358	0.548	0.893
	3	2.13	1.324	0.472	0.896
	4	2.67	1.360	0.528	0.894
	5	2.70	1.449	0.495	0.895
Respect for Others' Opinions	6	2.36	1.216	0.509	0.894
	7	1.76	1.070	0.390	0.899
	8	1.95	1.039	0.407	0.899
	9	2.07	1.179	0.492	0.895
	10	2.31	1.276	0.423	0.897
	11	2.36	1.160	0.429	0.896
Openness to Revising One's Viewpoint	12	3.91	1.007	0.668	0.891
	13	3.78	1.039	0.660	0.891
	14	3.47	1.034	0.657	0.891
	15	3.61	1.013	0.695	0.890
	16	3.76	0.960	0.624	0.892
Lack of Overconfidence	17	2.53	1.045	0.326	0.899
	18	3.01	1.043	0.586	0.893
	19	3.05	1.012	0.571	0.893
	20	3.11	0.988	0.439	0.896
	21	3.13	1.031	0.625	0.892
	22	3.06	0.889	0.537	0.894

Based on the results presented in Table 1, the item discrimination parameters of the Intellectual Humility Questionnaire, estimated using Classical Test Theory, ranged from +0.326 to +0.695, with Item 17 showing the lowest discrimination and Item 15 the highest.

Within the Intellectual Independence component, Item 2 demonstrated the highest discrimination (0.548) and Item 1 the lowest (0.470). For Respect for Others' Opinions, Item 6 showed the highest discrimination (0.509) and Item 7 the lowest (0.390). In the Openness to Revising One's Viewpoint component, Item 15 exhibited the highest discrimination (0.695) and Item 16 the lowest (0.624). Finally, within the Lack of Overconfidence component, Item 21 showed the highest discrimination (0.625) and Item 17 the lowest (0.326). Overall, the discrimination parameters indicated that the items demonstrated acceptable to strong discriminative power under Classical Test Theory.

Internal consistency was also examined using Classical Test Theory. Reliability was estimated using two approaches. First, all 22 items were analyzed simultaneously, yielding an overall Cronbach's alpha of 0.899, indicating excellent internal consistency. Second, reliability was calculated separately for each component. The results showed reliability coefficients of 0.827 for

Intellectual Independence (5 items), 0.846 for Respect for Others' Opinions (6 items), 0.778 for Openness to Revising One's Viewpoint (5 items), and 0.843 for Lack of Overconfidence (6 items). In addition, the contribution of each item to internal consistency was evaluated using the alpha-if-item-deleted method. The results indicated that removing any item did not increase the internal consistency of any component. Therefore, all items were retained, as they demonstrated strong homogeneity and substantial correlations with their respective components. Overall, the findings confirm that the Intellectual Humility Questionnaire exhibits satisfactory reliability based on Classical Test Theory.

Item Analysis Based on the Graded Item Response Theory

In this section, the psychometric properties of the Intellectual Humility Questionnaire were examined using the Graded Item Response Theory (G-IRT). The graded response model is capable of analyzing polytomous categorical responses along an ordinal continuum, such as Likert-type scales, in which respondents' attitudes or tendencies toward a statement are expressed through multiple ordered categories. In this framework, respondents' attitudes are represented along a latent continuum using discrete response categories.

Graded Item Response Theory assumes that item responses are ordinal categorical variables for which score intervals are not equally spaced (Lee, 2013). Accordingly, given that the Intellectual Humility Questionnaire consists of polytomous Likert-type items, the graded response model was employed to evaluate its psychometric indices. The assumptions of local independence, one-dimensionality, and item information functions were examined using the IRT-PRO software.

Before conducting item analyses based on Item Response Theory, it is essential to ensure that two key assumptions are satisfied: local independence and one-dimensionality.

Local Independence

The assumption of local independence states that if examinees' ability level is considered the sole underlying factor, their responses to test items are statistically independent of one another. When this assumption holds, an individual's response to one item should not be influenced by their responses to other items (Sijtsma & Molenaar, 2002). Hambleton (2010) noted that when the assumption of one-dimensionality is satisfied, local independence is also likely to be met. Specifically, when a dominant factor is extracted, the covariance among variables becomes zero or near zero, and the relationships among items disappear once the dominant factor is removed.

According to Cai, du Toit, and Thissen (2011), items with Pearson's χ^2 values greater than 10 suggest the potential presence of local dependence. Similarly, Thissen, Cai, and Bock (2010) indicated that Marginal χ^2 values exceeding 10 reflect violations of local independence. A lack of local independence implies that items do not load on a single dimension; rather, items with χ^2 values greater than 10 may form an additional dimension. Therefore, the presence of local dependence indicates that a single dimension is insufficient to model the structure of those items.

To evaluate the assumption of local independence for the Intellectual Humility Questionnaire, Pearson's χ^2 and Marginal χ^2 indices were computed under Samejima's graded response model, assuming a single latent dimension for all items of each component, using the IRT-PRO software. The results showed that the χ^2 values for all items were below 10, confirming that the assumption of local independence was satisfied. Thus, responses to each item of the Intellectual Humility Questionnaire are independent, and no item influences responses to other items.

One-dimensionality

Another fundamental assumption of Item Response Theory is one-dimensionality, which implies the existence of a single dominant underlying factor that loads on most test items and governs respondents' choice of response categories. In other words, all test items measure only one ability or latent trait. Violations of this assumption lead to biased parameter estimates (Reckase, 2009).

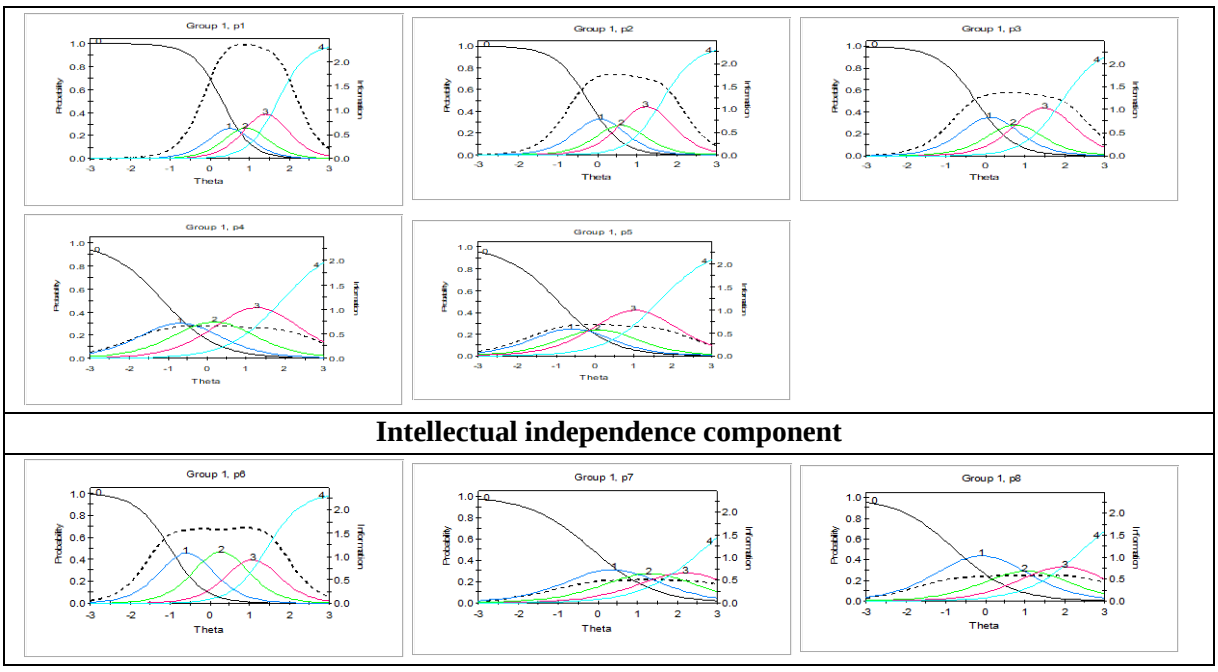
To assess the one-dimensionality of the Intellectual Humility Questionnaire, multidimensional item response theory analysis was conducted. In this procedure, one-dimensional, two-dimensional, and three-dimensional models were estimated and compared to infer the dimensionality of the scale. Based on the obtained results, the comparison of log-likelihood values indicated that the one-dimensional model yielded a smaller value than the two- and three-dimensional models, suggesting a superior fit. Additionally, the one-dimensional model reached convergence with fewer iterations, further supporting its better fit.

Subsequent comparisons among the three models revealed that the chi-square difference tests between the one-dimensional and two-dimensional models, as well as between the one-dimensional and three-dimensional models, were statistically significant. Moreover, the Akaike Information Criterion (AIC) and the corrected Akaike Information Criterion (AICc) values for the one-dimensional model were more favorable compared to the two- and three-dimensional models. These findings indicate that the one-dimensional model provides a better fit, thereby

confirming the one-dimensionality of the Intellectual Humility Questionnaire. After confirming the assumptions of local independence and one-dimensionality, the graded response model was estimated using IRT-PRO software, and the results are presented in Table 2.

Table 2. Item Parameters of the Intellectual Humility Questionnaire Based on the Graded Response Model

Component	Item	A	b1	b2	b3	b4	λ
Intellectual Independence	1	2.73	0.31	0.71	1.11	1.70	0.85
	2	2.37	−0.23	0.35	0.83	1.63	0.81
	3	2.09	−0.23	0.47	1.03	1.92	0.78
	4	1.45	−1.13	−0.27	0.63	1.92	0.65
	5	1.47	−0.94	−0.25	0.42	1.63	0.67
Respect for Others' Opinions	6	2.33	−1.00	−0.16	0.71	1.43	0.81
	7	1.28	−0.14	0.87	1.74	2.64	0.60
	8	1.38	−0.77	0.60	1.47	2.47	0.63
	9	1.51	−0.76	0.29	1.26	1.90	0.66
	10	2.45	−0.89	−0.10	0.83	1.29	0.82
	11	2.85	−1.03	−0.34	0.63	1.52	0.86
Openness to Revising One's Viewpoint	12	1.87	−3.19	−2.23	−0.94	0.60	0.74
	13	1.95	−2.96	−1.96	−0.69	0.81	0.75
	14	1.63	−3.18	−1.56	−0.22	1.40	0.71
	15	1.86	−3.11	−1.84	−0.47	1.08	0.74
	16	1.61	−4.07	−2.19	−0.60	0.99	0.69
Lack of Overconfidence	17	2.10	−1.47	0.13	1.17	2.12	0.78
	18	3.64	−1.79	−0.76	0.42	1.61	0.91
	19	4.11	−1.88	−0.78	0.38	1.57	0.92
	20	1.37	−2.75	−1.08	0.46	2.65	0.61
	21	1.64	−2.25	−1.15	0.21	2.23	0.70
	22	1.38	−2.80	−1.47	0.63	3.32	0.63



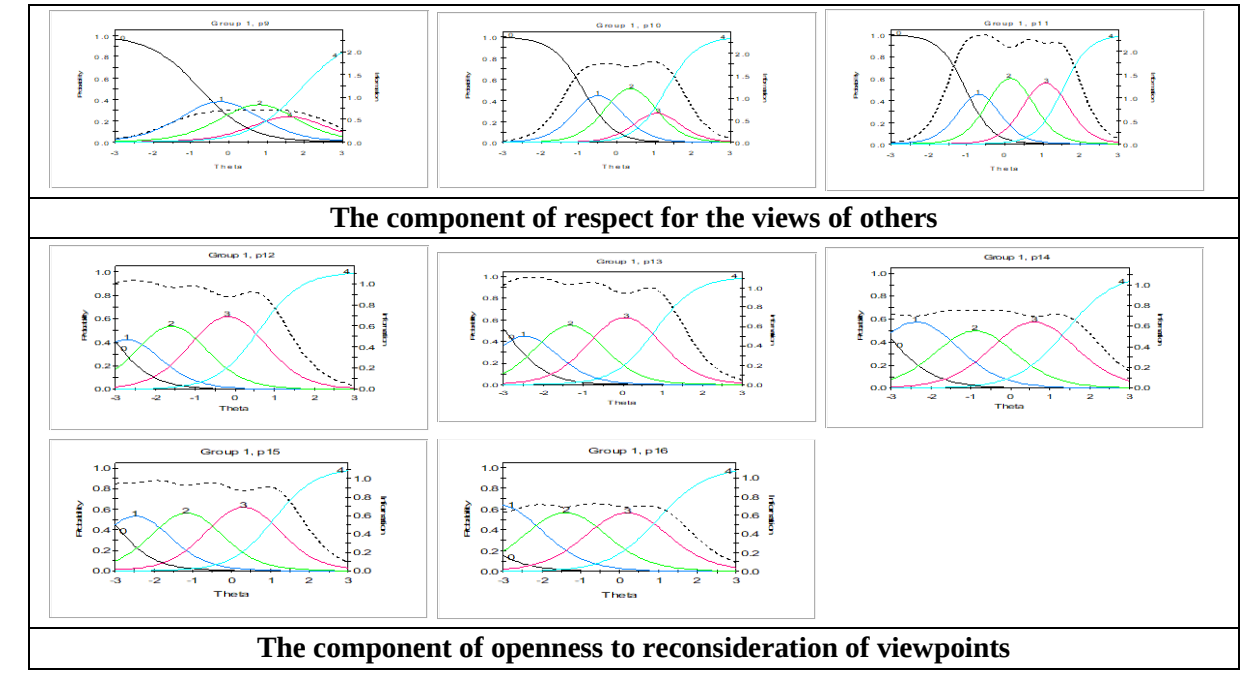


Figure 1. Awareness functions and response curves of the items of the intellectual acceptance components

Goodness-of-Fit Indices for Items Based on the Graded Item Response Theory

The examination of the item fit indices based on the Graded Item Response Theory (G-IRT) indicated that the chi-square significance level calculated for all items was greater than 0.05. Therefore, all items demonstrated acceptable model fit and were well represented by the graded response model. The item discrimination parameters (a) across the entire scale ranged from +1.28 to +4.11, reflecting good to very high discrimination power according to Hambleton and

Swaminathan's (1985) interpretation guidelines. Specifically, the discrimination parameters for each component were as follows:

Intellectual Independence: between +1.45 and +2.73

Respect for Others' Opinions: between +1.28 and +2.85

Openness to Revising One's Viewpoint: between +1.61 and +1.95

Lack of Overconfidence: between +1.37 and +4.11

These values demonstrate that all items possess satisfactory sensitivity in differentiating respondents with varying levels of the latent trait (intellectual humility).

Regarding the threshold parameters (b_1–b_4), which indicate the difficulty levels between response categories, the obtained ranges were as follows:

First threshold (b_1): from −4.07 to +0.31

Second threshold (b_2): from −2.23 to +0.87

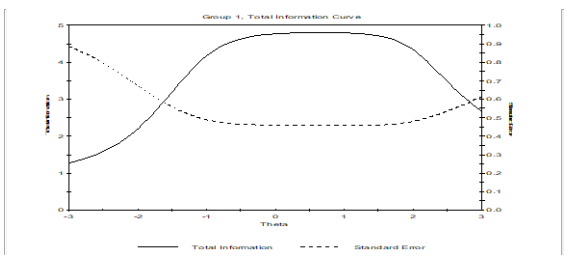
Third threshold (b_3): from −0.94 to +1.74

Fourth threshold (b_4): from +0.60 to +3.32

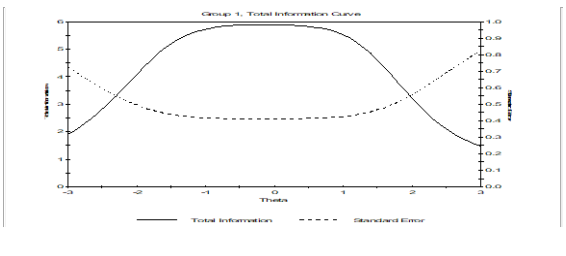
These ordered threshold values confirm that the response categories function properly along the latent continuum of intellectual humility. As shown in Table 2 and Figure 1, the category response curves exhibit a clear and well-ordered pattern. The curves are distinct and non-overlapping, with no flattening or entanglement observed among adjacent categories. Moreover, the thresholds for all items are sufficiently spaced, ensuring that each response category operates independently and dominates its own region of the latent trait (θ). This pattern indicates that the Likert-type rating scale performs efficiently and provides distinct response zones for participants at different ability levels. The factor loadings (λ) of items ranged from +0.60 to +0.92, reflecting appropriate item representation. Among the items, Item 7 showed the lowest loading (λ = +0.60) and Item 19 exhibited the highest loading (λ = +0.92).

At the component level: For Intellectual Independence, Item 4 had the lowest (λ = +0.65) and Item 1 the highest (λ = +0.85); For Respect for Others' Opinions, Item 7 had the lowest (λ = +0.60) and Item 11 the highest (λ = +0.86); For Openness to Revising One's Viewpoint, Item 16 had the lowest (λ = +0.69) and Item 13 the highest (λ = +0.75); For Lack of Overconfidence, Item 20 had the lowest (λ = +0.61) and Item 19 the highest (λ = +0.92).

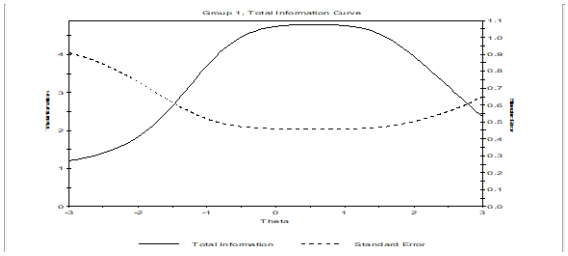
Overall, the results confirm that all items demonstrated adequate to strong factor loadings ($\lambda \geq 0.60$), consistent with the accepted psychometric standard ($\lambda \geq 0.70$ desirable), and thus support the construct validity and item efficiency of the Iranian version of the Intellectual Humility Questionnaire.



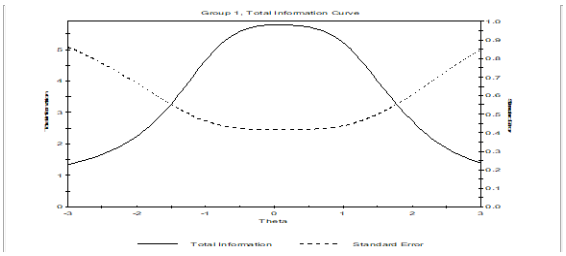
Individual agency component



Self-efficacy component



Self-sufficiency component



Self-management component

Figure 2. Awareness function of the intellectual acceptance component test

Item and Test Information Functions and Marginal Reliability Based on G-IRT

Based on the results reported in Table 4-11 and Figure 4-2, the test information function demonstrated its highest level of information around the average levels of the latent trait (θ). This indicates that the Intellectual Humility Questionnaire provides the greatest measurement precision for individuals with moderate levels of intellectual humility.

At the component level, the findings revealed that: For Intellectual Independence, Item 1 provided the highest amount of information, with the component showing its maximum information within the range of $\theta = 0$ to $+1.6$. For Respect for Others' Opinions, Item 11 yielded the greatest information, and the highest information for this component was observed within the range of $\theta = -0.8$ to $+1.6$. For Openness to Revising One's Viewpoint, Item 13 demonstrated the highest

information, with the component's peak information occurring within the range of θ = −2.8 to +0.8. For Lack of Overconfidence, Item 19 provided the greatest amount of information, and the highest information for this component was observed within the range of θ = −2.0 to +1.6.

These results indicate that the most informative items within each component effectively discriminate respondents across meaningful regions of the latent continuum. Furthermore, the remaining items also demonstrated adequate levels of information, confirming that all items contribute satisfactorily to the overall measurement precision of the scale.

To estimate reliability, the marginal (empirical) reliability coefficient was employed. Given that, within the Item Response Theory framework, the standard error of measurement varies across individuals, marginal reliability provides a more appropriate index of reliability than classical coefficients. The results indicated that the IRT-based marginal reliability coefficients were 0.80 for Intellectual Independence, 0.86 for Respect for Others' Opinions, 0.81 for Openness to Revising One's Viewpoint, and 0.90 for Lack of Overconfidence.

Overall, these findings demonstrate that the Intellectual Humility Questionnaire exhibits high measurement precision across a wide range of the latent trait, particularly at moderate levels of intellectual humility, and that all components possess satisfactory to excellent reliability based on the graded response model.

Discussion

There is no doubt that tests are an essential part of everyday life, and therefore, the validation of measurement instruments is of critical importance. The application of modern measurement theories can substantially contribute to achieving this objective. Given the significance of this construct in individuals' life processes and overall satisfaction, the present study aimed to examine the psychometric properties of the Intellectual Humility Questionnaire developed by Cromery and Rose (2016) based on both Classical Test Theory (CTT) and Item Response Theory (IRT).

Considering the response format of the items, the Graded Item Response Theory (GR-IRT) model was employed to evaluate the psychometric indices of the scale. The graded response model is capable of analyzing ordered categorical responses along a latent continuum representing individuals' attitudes or tendencies toward a given phenomenon. To estimate item parameters and to examine the assumptions of local independence, one-dimensionality, and item information

functions for each component of the Intellectual Humility Questionnaire, the IRT-PRO software was utilized.

In the Intellectual Humility Questionnaire, responses to each item are assumed to be independent, such that no item influences responding to or refraining from responding to other items. To evaluate the assumption of local independence among the items, marginal chi-square (χ^2) values were computed based on Samejima's graded response model using the IRT-PRO software. According to Cai et al. (2011), items with marginal χ^2 values greater than 10 indicate the possible presence of local dependence. Similarly, Thissen (2010) suggested that marginal χ^2 values exceeding 10 reflect violations of local independence. A lack of local independence implies that items do not load on a single dimension; rather, items with χ^2 values greater than 10 may form an additional latent dimension. Consequently, the presence of local dependence indicates that a unidimensional model is insufficient to adequately represent the structure of the items.

The results showed that the marginal χ^2 values for all items of the Intellectual Humility Questionnaire were below 10, indicating that the assumption of local independence was satisfied across all items.

Another fundamental assumption of Item Response Theory is one-dimensionality, which refers to the existence of a single dominant underlying factor that loads on most test items and governs respondents' selection of response categories. In other words, all items in a test should measure only one ability or latent trait. Failure to meet this assumption results in biased parameter estimates (Reckase, 2009). If one-dimensionality is not supported, not only does score interpretation become ambiguous, but comparisons among individuals' scores also become problematic, and item parameter estimates may be severely biased (Champlain, 2010).

Accordingly, prior to applying the graded response model, the assumption of one-dimensionality for the Intellectual Humility Questionnaire was examined using the NOHARM program. Based on the procedure proposed by Tate (2003), if the root means square residual (RMSR) for the one-dimensional solution decreases by less than 10% compared to the two-dimensional solution, the one-dimensional model is considered more appropriate. The results indicated that the RMSR for the one-dimensional solution decreased by less than 10% relative to the two-dimensional solution. In addition, Tanaka's fit index was calculated as 0.9201381 for the one-dimensional

model and 0.9126221 for the two-dimensional model. Therefore, based on these findings, the Intellectual Humility Questionnaire was confirmed to be unidimensional.

To further assess one-dimensionality, exploratory factor analysis (EFA) using the principal axis factoring method was conducted. In this approach, the first factor should explain at least 20% of the total variance, and the eigenvalue of the first factor should be compared with that of the second factor (Reckase, 2009). For evaluating data suitability for factor analysis, two statistical indices are essential: the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity. The KMO index determines whether the sample size is adequate for factor analysis; values closer to 1 indicate that the collected data are suitable for factor extraction. Bartlett's test of sphericity examines whether the correlation matrix significantly differs from an identity matrix; rejection of the null hypothesis supports the appropriateness of factor analysis.

The analysis of the Intellectual Humility Questionnaire data revealed a KMO value of 0.910, and Bartlett's test of sphericity yielded a chi-square value of 3781.83 with 231 degrees of freedom, which was statistically significant at the 99% confidence level. These results indicate that the sample size was sufficient for factor analysis and that the correlation matrix was not an identity matrix. Based on both criteria, the application of factor analysis was justified for the studied sample, and the assumption of one-dimensionality was supported. Taken together, the results of the RMSR analysis and exploratory factor analysis demonstrate that the data obtained from the Intellectual Humility Questionnaire conform to a one-dimensional measurement model.

An effective questionnaire must be capable of discriminating between individuals with high and low levels of the latent trait. To address this issue, the item discrimination (slope) parameters were examined. In the graded response model, the slope parameter reflects the extent to which an item is related to the latent trait and indicates how rapidly the probability of endorsing a particular response category increases with a one-unit increase in the latent trait. In other words, it represents the rate at which the expected item score changes as the latent trait level changes.

According to Baker and Kim (2004), discrimination values below 0.65 indicate low discrimination, values between 0.65 and 1.34 indicate moderate discrimination, and values of 1.35 or higher reflect high discrimination. Baker (2001) further suggested that a good item should have a discrimination parameter of at least 0.65 under normally distributed latent traits, and items with discrimination values below this threshold should be excluded from further analysis.

The examination of item fit indices based on the graded response theory showed that the chi-square significance levels for all items were greater than 0.05, indicating that all items demonstrated good model fit and were appropriately represented by the graded response model.

The findings related to the item discrimination parameters indicated that, for the total scale, discrimination values ranged from +1.28 to +4.11. At the component level, discrimination parameters ranged from +1.45 to +2.73 for Intellectual Independence, +1.28 to +2.85 for Respect for Others' Opinions, +1.61 to +1.95 for Openness to Revising One's Viewpoint, and +1.37 to +4.11 for Lack of Overconfidence. According to the criteria proposed by Baker and Kim (2004), all items demonstrated moderate to high discrimination power, indicating that the items effectively differentiate individuals with varying levels of the latent trait of intellectual humility. These findings confirm that none of the items fall within the low-discrimination range and that all items contribute meaningfully to measurement precision.

Regarding the threshold parameters, the results showed that the first category threshold ranged from −4.07 to +0.31, the second threshold ranged from −2.23 to +0.87, the third threshold ranged from −0.94 to +1.74, and the fourth threshold ranged from +0.60 to +3.32. The orderly progression of these threshold parameters indicates that the response categories functioned as intended across increasing levels of the latent trait.

Examination of the category response curves further demonstrated that the response options exhibited a well-ordered structure, with clearly distinguishable distributions. The curves showed no flattening or overlap, and the thresholds for the response categories of all items were sufficiently spaced such that no category was subsumed by another. Each response option independently dominated a specific region of the latent trait continuum, indicating the effective functioning of the rating scale and supporting the adequacy of the scoring system.

The factor loadings (λ) of the items ranged from +0.60 to +0.92, with Item 7 showing the lowest loading and Item 19 the highest. At the component level, for Intellectual Independence, factor loadings ranged from +0.65 (Item 4) to +0.85 (Item 1); for Respect for Others' Opinions, from +0.60 (Item 7) to +0.86 (Item 11); for Openness to Revising One's Viewpoint, from +0.69 (Item 16) to +0.75 (Item 13); and for Lack of Overconfidence, from +0.61 (Item 20) to +0.92 (Item 19). These results indicate that all items had acceptable to excellent factor loadings, confirming their strong association with their respective latent components.

Analysis of the test and item information functions revealed that the questionnaire provided its highest level of information around moderate values of theta (θ), indicating optimal measurement precision for individuals with average levels of intellectual humility. The greatest information for Intellectual Independence was provided by Item 1, with peak information occurring in the range of θ = 0 to +1.6. For Respect for Others' Opinions, Item 11 was the most informative, with peak information in the range of θ = −0.8 to +1.6. For Openness to Revising One's Viewpoint, Item 13 yielded the greatest information, with peak information observed in the range of θ = −2.8 to +0.8. Finally, for Lack of Overconfidence, Item 19 provided the highest level of information, with peak information in the range of θ = −2.0 to +1.6. These findings demonstrate that the scale effectively measures the latent trait across a broad and meaningful range of intellectual humility.

To estimate reliability, empirical (marginal) reliability coefficients were employed. Because, within the Item Response Theory framework, the standard error of measurement varies across individuals, marginal reliability provides a more appropriate index of reliability than classical coefficients. The results indicated marginal reliability coefficients of 0.80 for Intellectual Independence, 0.86 for Respect for Others' Opinions, 0.81 for Openness to Revising One's Viewpoint, and 0.90 for Lack of Overconfidence, reflecting good to excellent reliability across all components.

In addition, the internal consistency reliability of the items was examined based on Classical Test Theory, with all 22 items analyzed simultaneously. The results yielded a total scale reliability coefficient of 0.899. The reliability coefficients for the subscales were 0.827 for Intellectual Independence (5 items), 0.846 for Respect for Others' Opinions (6 items), 0.778 for Openness to Revising One's Viewpoint (5 items), and 0.843 for Lack of Overconfidence (6 items). These findings further support the high internal consistency of the questionnaire.

Overall, the results of the present study indicate that the Intellectual Humility Questionnaire is a valid and reliable instrument for use by Iranian researchers. The scale provides a sound framework for assessing intellectual humility among students of the Islamic Azad University, Bandar Abbas Branch, and can be effectively used to design and implement interventions aimed at improving intellectual humility, as well as to examine its relationships with other psychological variables.

Nevertheless, the present study is subject to certain limitations, including its implementation within a specific institutional and cultural context and the use of a relatively limited sample size,

which may restrict the generalizability of the findings to other populations. Therefore, future research is recommended to examine additional psychometric indices, such as convergent and predictive validity, in relation to other relevant constructs, and to administer the questionnaire across diverse populations and regions of the country.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by ethics committee of Islamic Azad University.

Author contributions

All authors contributed to the study conception and design, material preparation, data collection and analysis. All authors contributed to the article and approved the submitted version.

Funding

The authors did (not) receive support from any organization for the submitted work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Alfano, M., Iurino, K., Stey, P., Robinson, B., Christen, M., Yu, F., & Lapsley, D. (2017). Development and validation of a multi-dimensional measure of intellectual humility. *PLOS ONE*, *12*(8), e0182950. https://doi.org/10.1371/journal.pone. 0182950

Baehr, J. (2011). The inquiring mind: On intellectual virtues and virtue epistemology. New York, NY: Oxford University Press.

Baker F.B., & Kim S. H. (2004). Item response theory: Parameter estimation techniques (2nd ed). Boca Raton, FL: CRC Press.

Baker, F. B. (2001). The basics of item response theory. ERIC clearinghouse on assessment and evaluation.

Blanchin, M., Hardouin, J. B., Guillemin, F., Falissard, B., & Se´ bille, V. (2013). Power and sample size determination for the group comparison of patient-reported outcomes with rasch family models. PLoS One. 8(2), 1-13.

Bortolotti, S. L.V., Tezza, R., de Andrade, D. F., Bornia, A. C., & de Sousa Júnior, A. F. (2013). Relevance and advantages of using the item response theory. Quality & Quantity: International Journal of Methodology, 47(4), 2341–2360.

Cai, L., Du Toit, S. H. C., & Thissen, D. (2011). IRTPRO: User guide. Lincolnwood, IL: Scientific Software International.

Champlain, (2010). A primer on classical test theory and item response theory for assessments in medical education. Published in Medical Education 1 January 2010 Medicine, Education. DOI:10.1111/j.1365-2923.2009.03425.xCorpus ID: 205358152.

Comrey AL, Lee HB. A first course in factor analysis: Psychology Press; 2013.

Danovitch, J. H., Fisher, M., Schroder, H., Hambrick, D. Z., & Moser, J. (2019). Intelligence and neurophysiological markers of error monitoring relate to children's intellectual humility. *Child Development*, *90*(3), 924–939. https://doi. org/10.1111/cdev.12960

Davis, D. E., Hook, J. N., Worthington, E. L., Van Tongeren, D. R., Gartner, A. L., Jennings II, D. J., and Emmons, R. A. 2011. "Relational Humility: Conceptualizing and Measuring Humility as a Personality Judgment." *Journal of Personality Assessment*,*93*(3), 225–234. https://doi.org/10.1080/ 00223891.2011.558871

Davis, D. E., McElroy, S. E., Rice, K. G., Choe, E., Westbrook, C., Hook, J. N., DeBlaere, C., Hill, P., Placares, V., Worthington, E. L., & Van Tongeren, D. R. (2016). Is modesty a subdomain of humility? *The Journal of Positive Psychology*, *11*(4), 439–446. https://doi.org/10.1080/17439760.2015.1117130

DeMars, C. (2010). Item response theory. Understanding statistics measurement. City: Oxford University Press.

Edwards, M. C., Houts, C. R., & Cai, L. (2018). A diagnostic procedure to detect departures from local independence in item response theory models. Psychological Methods, 23(1), 138-149.

Embretson SE, Reise SP. Item response theory for psychologists. 2000. Lawrence Earlbaum Associates, Mahwah, NJ.

Grossmann, I., Na, J., Varnuma, M. W., Park, D. C., Kitayama, S., & Nisbett, R. E. (2010). Reasoning about social conflicts improves into old age. Proceedings of the National Academy of Sciences of the United States of America, 107, 7246-7250. doi:10.1073/pnas.1001715107.

Hambleton, R. K., Van der Linden, W. J., & Wells, C. S. (2010). IRT models for the analysis of polytomously scored data: Brief and selected history of model building advances. In M. L. Nering & R. Ostini (Eds.), Handbook of polytomous item response theory models. Published in the Taylor & Francis e-Library.

Hook, J. N., Davis, D. E., Van Tongeren, D. R., Hill, P. C., Worthington, E. L., Jr., Farrell, J. E., & Dieke, P. (2015). Intellectual humility and forgiveness of religious leaders. *The Journal of Positive Psychology*, *10*(6), 499–506. https:// doi.org/10.1080/17439760.2015.1004554

Hooman, H. A. (2011). *Multivariate data analysis in behavioral research* (2nd ed.). Tehran: Peyk-e Farhang Publications.

Hopkin, C. R., Hoyle, R. H., & Toner, K. (2014). Intellectual humility and reactions to opinions about religious beliefs. *Journal of Psychology and Theology*, *42*(1), 50–61. https://doi. org/10.1177/009164711404200106 .

Hoyle, R. H., Davisson, E. K.,Diebels, K. J., & Leary, M. R. (2016). Holding specific views with humility: Conceptualization and measurement of specific intellectual humility. *Personality and Individual Differences*, *97(July)*, 165–172. https://doi. org/10.1016/j.paid.2016.03.043

Jiang, S., Wang, C., &Weiss, D. (2016). The sample size requirements for estimation of item parameters in the multidimensional graded response model. Frontiers in Psychology. 7(1), 1-10.

Kiani, S. (2022). *Determining psychometric indices of the human agency traits scale based on graded item response theory*. *Educational Measurement Quarterly*, 12(47), 1–20.

Kline, R.B. Principles and practice of structural equation modeling (3rded.). New York: Guilford Press. (2010).

Kross, E., & Grossmann, I. (2012). Boosting wisdom: Distance from the self-enhances wise reasoning, attitudes, and behavior. Journal of Experimental Psychology: General, 141, 43-48. Doi: 10.1037/a0024158.

Krumrei-Mancuso, E. J., and Rouse, S. V. (2016). The development and validation of the comprehensive intellectual humility scale. J. Pers. Assess. 98, 209–221. doi: 10.1080/00223891.2015.1068174.

Le, D. T. (2013). Applying item response theory modeling in educational research. Maggio, K. M. (2015). Why do you see it that way? The preferences of confrontational behavior as seen through attachment style and facework behavior. Doctoral dissertation, University of Delaware.

Leary, M. R., Diebels, K. J., Davisson, E. K., Jongman-Sereno, K. P., Isherwood, J. C., Raimi, Raimi, K. T., Deffler, S. A., & Hoyle, R. H. (2017). Cognitive and interpersonal features of intellectual humility. *Personality and Social Psychology Bulletin*, *43*(6), 793–813. https://doi.org/10.1177/0146167217697695

Mark J Gierl. (2013). Evaluating the quality of medical multiple-choice items created with automated processes. Medical Education 2013: 47: 726–733doi: 10.1111/medu.12202.Centre for Research in Applied Measurement and Evaluation, University of Alberta, 6–110 Education North, Edmonton, Alberta T6G 2G5, Canada. Tel: 00 1 780 492 2396; E-mail: mark.gierl@ualberta.ca.

McElroy, S. E., Rice, K. G., Davis, D. E., Hook, J. N., Hill, P. C., Worthington, E. L., Jr., & Van Tongeren, D. R. (2014). Intellectual humility: Scale development and theoretical elaborations in the context of religious leadership. *The Journal of Psychology and Theology*, *42*(1), 19–30. https://doi.org/10. 1177/009164711404200103 .

Michael Nering, Remo Ostini. (2010). Handbook of Polytomous Item Response Theory Models. Imprint Routledge New York 19 January 2011,1st Edition. eBook ISBN9780203861264. https://doi.org/10.4324/9780203861264.

Reckase, M. D. (2009). Multidimensional item response theory. New York: Springer Verlag.

Samuelson, P. L., & Church, I. M. (2015). When cognition turns vicious: Heuristics and biases in light of virtue epistemology. *Philosophical Psychology*, *28*(8), 1095–1113. https://doi.org/10.1080/09515089.2014.904197

Sijtsma, K., Molenaar, W. (2002). Introduction to Nonparametric Item Response Theory. Sage, Thousand Oaks, CA.

Tate R. A (2003). comparison of selected empirical methods for assessing the structure of response to test items. Applied Psychological Measurement. 2003; 27(3): 159-203.

Thissen, D., Cai, L., & Bock, R. D. (2010). The nominal item response model. In M. Nering & R. Ostini (Eds.), Handbook of polytomous item response theory models (pp. 43-75). New York, NY: Routledge.

Whitcomb, D., Battaly, H., Baehr, J., & Howard-Snyder, D. (2017). Intellectual humility: Owning our limitations. *Philosophy and Phenomenological Research*, *94*(3), 509–539. https://doi.org/10.1111/phpr.12228.