Measuring Affinity for Technology Interaction with an Image-Based Testing Approach

Master's Thesis

Nico O. Reisch

s2329352

Faculty of Behavioural, Management and Social Sciences

Human Factors & Engineering Psychology

University of Twente

1st Supervisor: Dr. Martin Schmettow

2nd Supervisor: Dr. Simone Borsci

October 18th, 2023

Measuring Affinity for Technology Interaction with an Image-Based Testing Approach

Abstract

Personality characteristics are frequently assessed with traditional text-based self-report scales despite concerns about limited engagement, subjective interpretation of scale levels and faking. The study explored whether image-based testing approaches could serve as a viable alternative to text-based methods. Thus, the affinity for technology interaction (ATI) scale, a validated text-based rating scale, was transformed into the image-based affinity for technology interaction (IBATI) scale. Generative artificial intelligence was utilised to visualise the response options. Responses of 178 participants were analysed to assess the rating scales' validity and reliability on item and scale levels. Bayesian Confirmatory Factor Analysis was conducted to evaluate the scales' factor structure and psychometric properties. While the data collection yielded an outstanding completion rate, the IBATI scale displayed slightly lower reliability than the ATI scale. Still, outcomes on the text-based and image-based rating scales were mostly congruent on scale and item levels. The text-based Animal Attitude Scale (AAS) was introduced to examine if the IBATI scale specifically measures affinity for technology interaction and displays discriminant validity to the AAS, which aims at a separate construct. The image-based testing approach revealed notable potential, reflected by robust construct validity and internal consistency. The research highlights that extensive pretesting and iterative refinements of image-based response options are vital. Additional research is required to explore the potential of image-based personality testing further. Future studies could introduce external measures for concurrent and predictive validity, examine test-retest reliability and combine image-based and text-based rating scales to potentially obtain higher validity and reliability and facilitate a better testing experience.

Keywords: Image-Based Testing, Affinity for Technology Interaction, Bayesian Confirmatory Factor Analysis, Generative Artificial Intelligence

Table of Contents

Measuring Affinity for Technology Interaction with an Image-Based Testing Approach

The field of psychometrics focuses on quantifying and measuring psychological constructs, such as mental attributes, behaviour and performance (VandenBos, 2015). Various measurement instruments are utilised to assess these psychological constructs and observe interindividual differences. While most self-reporting tests in personality testing rely on text-based approaches, image-based approaches have recently gained increased attention (Hilliard et al., 2022a; Krainikovsky et al., 2019; Leutner et al., 2017; Sang et al., 2016; Zhang et al., 2017). Despite growing interest in image-based testing, few validation studies have been conducted, leaving a gap between available peer-reviewed literature and the interest in leveraging the innovative assessment method (Leutner et al., 2017). This study aims to establish an image-based rating scale to determine whether the image-driven approach yields comparable results to traditional text-based testing. The rating scale targets the construct of affinity for technology interaction, which addresses people's subjective experience when interacting with technical systems (Attig et al., 2017). The psychometric qualities of the newly created image-based rating scale will be explored on scale and item level to evaluate convergent and discriminant validity and reliability.

Personality Traits and Their Importance for Work and Technology Interaction

Individuals' personality comprises several characteristics, such as traits, interests, values, and emotional patterns that shape their approach to life (Kankaraš, 2017; VandenBos, 2015). Personality traits are 'stable, inner, personal dispositions that determine relatively consistent patterns of behaviour' (Chamorro-Premuzic & Furnham, 2010, p. 129). Consequently, trait psychology assumes individual differences persist across situations, and personality traits largely remain the same throughout individuals' lifespan (Chamorro-Premuzic, 2015; McCrae & Costa Jr, 1995). As an individual's personality consistently influences their behaviour, information about personality traits can be leveraged to anticipate future behaviour (Barrick & Mount, 2005; Chamorro-Premuzic, 2015; Kankaraš, 2017; McCrae & Costa Jr, 1995). However, it is crucial to emphasise that behaviour is intricate and influenced by various elements, such as situational factors, for example, motivation or mood (Chamorro-Premuzic, 2015). Despite moderating and mediating variables that impact behaviour, personality is a central element in understanding how people tend to act (Barrick & Mount, 2005). Hence, valid assessments of individuals' personality traits promise value in many fields, for example, in the context of work.

Many studies were conducted to assess the influence of personality on occupational success and identified that personality is a valid predictor of individuals' engagement and performance at work (Barrick & Mount, 2005; Sackett & Walmsley, 2014; Young et al., 2018). Personality tests are used primarily in personnel selection, talent development and team building (Lundgren et al., 2017). Especially in personnel selection, personality testing has steadily gained more popularity (Chamorro-Premuzic & Furnham, 2010; Rothstein & Goffin, 2006). Bad hires are detrimental to the organisational culture and lead to substantial financial costs, which can be limited by using valid personality tests as helping tools to support decision-making (Baez, 2013). Despite the potential advantages, leveraging personality assessments at work has been discussed and criticised for decades (Morgeson et al., 2007). One commonly criticised aspect is that characteristics to succeed at work diverge between job fields (Chamorro-Premuzic & Furnham, 2010). A clear trend in many job fields is the increased prevalence of technology at work, which requires individuals to interact with it successfully (Wessel et al., 2020). While technology is integral for many tasks in professional and private settings, individuals' attitudes and associations towards it vary. Many people primarily use technology as a tool to complete their tasks; however, some users are very enthusiastic about using new technology (Schmettow & Drees, 2014; Schmettow et al., 2013; Wessel et al., 2020). Multiple studies in human factors and engineering have emphasised the relevance of users' interindividual differences in human-technology interaction (Attig et al., 2017; Schmettow & Drees, 2014; Wessel et al., 2020).

Affinity for Technology Interaction

Measuring users' attitudes towards technology can be beneficial for several purposes. The results could be leveraged to identify individuals likely to engage and succeed with new technology (Wessel et al., 2020), for example, in the context of job selection. Another fundamental purpose is that user diversity in system usability tests can be monitored to obtain meaningful sampling data in research and development (Franke et al., 2019; Wessel et al., 2020). Thus, several psychometric measures have been introduced to examine users' attitudes towards technology interaction. Schmettow and Drees (2014) introduced geekism as a continuous trait and constructed the Gex scale to quantify users' enthusiasm towards computers. In a more recent study, Franke et al. (2019) developed and validated the unidimensional affinity for technology interaction scale, which is grounded in the psychological construct of the need for cognition.

At the individual level, the need for cognition refers to being open to exploring relevant information sources and dealing with cognitively challenging tasks (Bauer & Stiner, 2020; Cacioppo & Petty, 1982; Schmettow et al., 2013). Previous research suggests the link between the need for cognition, problem-solving and interaction with technology. Schmettow and Drees (2014) demonstrated that an individual's need for cognition directly influences how people interact with technology. In line with these findings, empirical studies indicate that the need for cognition is related to innovativeness (Hoffmann & Soyez, 2010, as cited in Franke et al., 2019) and efficient problem-solving when interacting with computers (Ebelhäuser, 2015, as cited in Franke et al., 2019). Therefore, in line with previous research, the need for cognition seems to be a suitable psychological construct to draw on when creating a psychometric instrument that measures individuals' attitudes towards technology interaction (Franke et al., 2019; Schmettow & Drees, 2014).

Text-Based Testing Approaches in Psychometrics: Limitations and Critique

Many commonly used psychometric scales rely on self-report and are text-based. This applies to widespread personality tests, for example, the Big Five Inventory (BFI), but also to measures focusing on users' attitudes towards technology, such as the Gex or the ATI scales (Attig et al., 2017; Franke et al., 2019; Schmettow & Drees, 2014; Zhang et al., 2017). Despite the popularity of these scales, researchers have pointed out several limitations of text-based testing, including limited engagement, comparability, and easiness of faking.

Limited Engagement

All test takers need to read, process, and understand the item content, which makes the user experience effortful and time-consuming (Sang et al., 2016; Zhang et al., 2017). Thus, text-based assessments are often perceived as not engaging, yielding low response rates and poor response quality due to test-taker fatigue caused by the extensive length of scales (Hilliard et al., 2022a; Leutner et al., 2022).

Comparability

The results of personality assessments have to be comparable across all users. As the psychometric instruments are often used internationally, precise translations are required when translating the items into other languages to ensure test fairness (Hilliard et al., 2022a; Sang et al., 2016; Zhang et al., 2017). Differences in the meaning of translations can affect how test takers interpret the item content and skew the test results. Inconsiderate translations harm the accuracy and comparability of results. Another criticism of traditional personality assessment related to comparability is the ambiguity of items and response options that lead to interpretative subjectivity. Text-based items usually entail some degree of ambiguity because of words that leave room for interpretation, for instance, often or regularly (Lilienfeld et al., 2000). When using Likert scales, test takers might have different associations with what each response option on the scale represents (Mischel, 1968). Thus, users deviating in their content interpretations might negatively impact the comparability of the measure.

Easiness to Fake

Another major criticism of text-based self-report testing is the risk that users can fake their responses (Chamorro-Premuzic & Furnham, 2010; Leutner et al., 2017; Sang et al., 2016; Ziegler et al., 2012). Faking can be observed when test takers' answers do not depict reality to achieve personal goals (Ziegler et al., 2012). Text-based self-reporting has drawn criticism, especially in high-stakes situations, such as job selection, as the test taker can easily discern the measured construct (Ihsan & Furnham, 2018; Leutner et al., 2017). Thus, the test taker's choices are likely impacted by what they consider socially desirable (Sang et al., 2016). Many studies have reported score inflation when comparing the results of high-stakes and low-stakes testing situations, which suggests participants can present themselves favourably on text-based assessments (Furnham et al., 2013; Hilliard et al., 2022a). Despite the widespread use of text-based personality assessments, several obstacles to obtaining valid and reliable test results have been identified. Consequently, alternative testing approaches to assess personality should be explored.

Projective Techniques in Personality Testing

An alternative procedure to traditional text-based testing is the use of projective techniques. Lilienfeld et al. (2000) highlighted several available projective techniques such as Association, Construction and Selection, which require test takers to respond to ambiguous questions. Depending on the projective technique, items are often open-ended, which allows an infinite number of responses. Thus, projective instruments, such as the Rorschach Inkblot Test and the Thematic Apperception Test, have received criticism because of restricted validity and standardisation (Lilienfeld et al., 2000; Piotrowski, 2015; Sartori, 2010). However, some projective tests offer limited response options, for example, when participants are prompted to select a specific element (Lilienfeld et al., 2000). Despite controversies around projective tests, it has been argued that the limited validity can be attributed to suboptimal design and test construction rather than projective testing. It is therefore recommended to utilise an iterative process when creating instruments that use projective techniques to obtain meaningful outcomes (Lilienfeld et al., 2000). Various elements, such as incomplete statements or images, can be displayed as ambiguous stimuli in projective testing

(Sartori, 2010). Projective tests are then evaluated based on individuals' responses to the displayed stimuli.

Image-Based Testing Approaches in Psychometrics: Potential Benefits

A promising development related to projective testing is using image-based testing approaches. Instead of relying solely on text, image-based testing approaches present visual stimuli as interpretable reference points for test takers. Several studies that introduced images instead of text indicated that image-based tests could help overcome numerous problems associated with traditional testing methods and lead to several benefits (Hilliard et al., 2022a; Hilliard et al., 2022b; Leutner et al., 2017; Sang et al., 2016; Zhang et al., 2017). These benefits contribute to theoretical and practical advantages related to engagement, objectivity and resistance to faking.

Higher Engagement

Previous studies suggest that using an image-based testing approach is likely to be perceived as more engaging (Leutner et al., 2017). According to Sang et al. (2016), using images instead of text allows a more natural interaction, making users more relaxed when engaging with the assessment. Hence, users perceive the testing procedure as less stressful. Furthermore, introducing images in personality tests has been reported to evoke stronger user preferences than text-based testing approaches (Leutner et al., 2017). Meissner and Rothermund (2015), who compared the effect of displaying text and images, found that images led to equal or stronger attitudes towards the presented stimuli. Stronger attitudes towards the presented stimuli are associated with shorter testing times, as users take less time to contemplate the questions (Hilliard et al., 2022b; Leutner et al., 2017; Leutner et al., 2022). Thus, image-based measures might allow for a more efficient testing procedure and could potentially yield higher completion rates than text-based approaches.

Increased Objectivity

Personality assessments are commonly used to compare test takers from diverse backgrounds. When creating any personality assessment, linguistic differences and interpretative subjectivity are two threats to objectivity that must be considered. Previous research suggests that image-based testing is usable for broader audiences, including users with different cultural and linguistical backgrounds (Hilliard et al., 2022a; Leutner et al., 2017; Paunonen et al., 2001; Sang et al., 2016; Zhang et al., 2017). While text-based approaches often require the creation of language-specific test items, image-based testing approaches uphold objectivity with language-neutral stimuli (Hilliard et al., 2022b). Relying solely on text might lead to misunderstandings and disparities in how the content is interpreted (Leutner et al., 2017). Using images as response options could guide the interpretation of different trait levels. Consequently, introducing image-based stimuli might contribute to less interpretative subjectivity in the measure if the images are understood as intended in the test design. Image-based testing approaches might also reduce the language translation interpretation bias impacting text-based assessments. Thus, using image-based approaches might improve objectivity and reduce the effect of response style biases.

Resistance to Faking

One of the most prominent criticisms of traditional psychometric testing is the easiness of faking (Chamorro-Premuzic & Furnham, 2010; Ziegler et al., 2012). Leutner et al. (2017) claim that assessments could be more resistant to faking by introducing image-based elements instead of text. The rationale behind this is that the measured constructs and their associated trait levels cannot be identified as easily. However, meaningful research on faking image-based measures has yet to be conducted. An eye-tracking study identified nuances of behaviour when participants fake to achieve more desirable outcomes on text-based assessments, as respondents tend to focus on extreme response options (Van Hooft & Born, 2012). This finding suggests that identifying the desirable trait levels of image-based response options might be more challenging than on traditional text-based Likert scales, as the order of image-based response options can be randomised to restrict faking.

Existing Research on Image-Based Personality Testing

Several studies have introduced visual elements in the testing procedure; however, there are evident disparities between the methods of integrating images. The Nonverbal Personality Questionnaire presents illustrations displaying behaviour as the item. It requires test takers to rate, on a seven-point Likert scale, how likely they would be to engage in the behaviour. The test takers' personality traits are mapped with the Big Five based on their responses. The convergent validity between the nonverbal questionnaire and a text-based questionnaire was examined in a normative sample and across cultures. Correlations for the normative sample were moderate, ranging from 0.45 to 0.59, and were weaker for the cross-cultural sample, with values between 0.35 and 0.54 (Paunonen et al., 2001).

Leutner et al. (2017) displayed a text-based question and introduced image-based response options to measure creativity on three scales. The items were designed with response options representing various trait levels or response options assessing distinct traits. The convergent validity between the image-based and text-based assessments resulted in correlations of 0.35, 0.50 and 0.52.

Another approach measured the Big Five by introducing forced-choice image pairs and asking the test taker to indicate which of the presented images was more like them (Hilliard et al., 2022a; Hilliard et al., 2022b). Thus, images were introduced as the response option, and the question pool consisted of unidimensional and multidimensional image pairs. Various models were tested to explore which scoring approach would yield the highest convergent validity. The traditional summative ipsative scoring approach resulted in convergent validities between 0.41 and 0.73. Machine learning-based Lasso models were used to train a predictive scoring algorithm, which led to the highest convergent validities ranging from 0.60 to 0.78 (Hilliard et al., 2022b).

The approaches introduced above required the test taker to indicate which response option represents them most or to what extent the image represents their behaviour. Meanwhile, some other studies have explored image-based testing approaches (Krainikovsky et al., 2019; Sang et al., 2016; Zhang et al., 2017); however, in these studies, test takers were instructed to choose the image they liked most rather than the one that represented themselves or their behaviour. In one study, additional, potentially irrelevant questions regarding colour and music were introduced in addition to image-based test questions, leading to unsatisfactory convergent validities between 0.06 and 0.28 when measuring the Big Five (Krainikovsky et al., 2019). Other studies, such as Sang et al. (2016) and Zhang et al. (2017), did not report on the convergent validities of their studies.

Implications for Image-Based Test Design

After reviewing the existing research on image-based personality testing, it is vital to consider the practical implications of these findings for the design of future tests and to leverage these studies to inform decision-making in image-based test design. Leutner et al. (2017) criticised some modern assessment tools for focusing on entertainment and neglecting the most critical elements: test validity and reliability. No matter how entertaining and well users perceive a tool to be, if an assessment is invalid or unreliable, it does not fulfil its primary purpose. Therefore, establishing validity and reliability must be the top priority when designing any psychometric test, including image-based personality tests. As mentioned above, only a limited number of validation studies have been conducted so far. Several researchers have indicated that further research is required to examine the implications of using images instead of text and connect research to practical application (Hilliard et al., 2022a; Leutner et al., 2017).

Testing approaches that require the test taker to indicate which image represents them or their behaviour have performed better than those that instruct them to select their preferred image (Hilliard et al., 2022a; Hilliard et al., 2022b; Leutner et al., 2017; Paunonen et al., 2001). Furthermore, Leutner et al. (2017) pointed out that replacing the question with images puts constraints on the content of the assessment because only visually representable questions can be introduced. Alternatively, introducing images as response options to text-based questions provides more flexibility when designing the test (Leutner et al., 2017).

As the responses of individuals depend on their interpretations of the images, it is crucial to align the image-based measure with a text-based measure (Hilliard et al., 2022a). Thus, the test takers' interpretations of the newly designed items are more likely to match the intended constructs and do not leave too much room for interpretation. Substantial disparities between the test takers' interpretations can harm the measure's validity. Especially when using the assessment across cultures, differences in the interpretation of images could negatively affect the comparability of the tool (Hilliard et al., 2022a; Hilliard et al., 2022b).

An important consideration for any psychometric test design concerns the number of items that comprise the scales. The number of required items depends on the complexity of the measured construct; however, at least three meaningful, well-performing items are required per scale to obtain valid and reliable results. Designing additional items per scale is highly recommended if any item does not perform as expected (Robinson, 2018). Especially for image-based items, additional items per scale should be introduced in case some of the content is misinterpreted and fails to measure the intended constructs.

A frequent point of discussion when designing a new rating scale is the number of response options. Robinson (2018) points out that no consensus on the optimal number of response options exists but scales with five or seven response options often yield high-quality data. An even number of response options can help gain new insights, as participants cannot select a neutral option. However, image-based tests with an odd number of response options might not be as susceptible to participants selecting the middle option, as the corresponding scores are not discernible as quickly as on a regular, numeric Likert scale. In line with Robinson (2018), Sang et al. (2016) suggested using five response options in image-based studies.

Artificial Intelligence and Text-to-Image Generative Models

While image-based testing has already gained increased attention in previous years, recent technological innovation makes the use of images in assessments even more appealing. With the rapid advancements regarding generative artificial intelligence, online text-to-image diffusion models, such as DALL-E, Stable Diffusion and Midjourney, offer an enormous breadth of applications (Weisz et al., 2023; Zhang et al., 2023). Many existing image-based personality tests rely on stock images (Hilliard et al., 2022a; Hilliard et al., 2022b; Krainikovsky et al., 2019; Leutner et al., 2017). Contrarily, text-to-image generative models provide an accessible, flexible and low-cost alternative for sourcing images. However, successful image creation might require extensive prompt engineering and refinements until high-quality images are produced. In natural language processing, prompt engineering refers to the process of crafting textual definitions to dictate which output the generative model should create to achieve the desired outcome (Liu & Chilton, 2022). Creating images with a text-to-image generative model allows displaying of test elements visually instead of textually.

Aim of the Study

The present study explores whether image-based assessments can serve as a valid, alternative measure to assess personality traits, specifically focusing on individuals' affinity for technology interaction. An image-based rating scale is established with a text-to-image generative model. The instrument is validated with an existing, parallel text-based rating scale as a benchmark to test for convergent validity. Additionally, a separate text-based rating scale is introduced to examine discriminant validity and identify potential response style biases. Hence, it can be evaluated to what extent the new image-based rating scale succeeds in assessing the construct affinity for technology interaction it intends to measure.

RQ1: How effectively can an image-based testing approach replace a text-based approach in measuring affinity for technology interaction without compromising psychometric quality?

To receive valid results, it is essential that test takers' interpretation of the test items and the corresponding image-based response options matches the intended trait level. The initial test validation outcomes are used to examine potential problems in the item design that could lead to disparities. Since each image-based item is aligned with a text-based item, response options deviating from the expected outcomes can be detected. Thus, it can be explored if the chosen questions and corresponding images are suitable to obtain valid outcomes in line with the text-based items.

RQ2: To what extent do image-based and text-based testing approaches yield results that are congruent at the item level?

Methods

Participants

Ethical approval was obtained by the Ethical Committee of the Behavioural, Management and Social Sciences at the University of Twente (UT). Participants were primarily recruited from the UT's test subject pool, notice board postings and convenience sampling. UT students who signed up through the test subject pool received 0.5 test subject credits for completing the study. A minimum age of 18 and sufficient English skills were set as the criteria to participate. Subjects were informed about the procedures and purpose of the study (see Appendix A) and had to indicate their consent to participate (Appendix B). Participants were told they could withdraw from the study at any point. Of the 189 subjects who participated in the study, 11 were excluded from the analysis due to missing consent or incomplete data. Ultimately, the data of 178 participants were processed for the analysis (Table 1). The age of the participants ranged from 18 to 61 (M = 25.62, SD = 9.50).

Table 1

Gender and Nationality Distribution of Participants

Gender	*n*
Male	73
Female	103
Prefer not to say	2
Nationality	
German	109
Dutch	36
Other European	22
Other Non-European	10
Prefer not to say	1
Total	178

Measures

Two rating scales were introduced to assess participants' affinity for technology interaction. The participants were assessed with the existing text-based Affinity for Technology Interaction scale and the newly designed image-based test version. Furthermore, the short version of the Animal Attitude Scale was introduced to observe potential response style biases and evaluate if the image-based scale exclusively measures affinity for technology interaction, thereby assessing its discriminant validity.

Affinity for Technology Interaction (ATI)

The ATI scale, created by Thomas Franke and colleagues, is a text-based rating scale designed to assess individuals' perspectives toward technology on a unidimensional scale (Franke et al., 2019). The normative rating scale is grounded in the Need for Cognition theory and comprises nine items to measure the construct of affinity for technology interaction. The scale requires participants to self-report and rank themselves on a six-point Likert scale. The mean score of all items is calculated to obtain a total score that allows categorising individuals' affinity for technology interaction, making the ATI instrument a valuable tool for several purposes. The ATI scale is primarily used to monitor user diversity in usability tests. This study introduced the ATI scale as the text-based benchmark test (Appendix C).

Image-Based Affinity for Technology Interaction (IBATI)

The IBATI scale, a normative image-based rating scale, was created considering the findings and recommendations from previous studies about image-based testing. Leutner et al. (2017) pointed out that relying on images as items strongly restricts the content. Therefore, text was used to describe the item, while images were introduced as response options. In line with recommendations by Robinson (2018) and Sang et al. (2016), the image-based items were introduced with five response options each. A corresponding image-based item was introduced for each text-based item in the ATI scale. Consequently, the IBATI scale comprises nine items with 45 total response options. Every newly designed item was designed to be unidimensional and to focus solely on people's affinity for technology interaction. All images were created with the generative artificial intelligence tool DALL-E 2 by OpenAI. The prompts used to create the images were written manually (Appendix D). Depending on the image quality, additional image variations were generated. In total, 552 images were reviewed to identify suitable response options measuring different levels of the measured trait. The images were then assigned to a Likert scale score representing the trait level. An iterative approach was taken to the item development process to optimise the response options' accuracy. Three psychology graduates who had expertise in psychometric testing were consulted to re-rank the image-based response options. Ranked images that deviated by more than one position from the intended scale level were replaced. The process was repeated until no substantial disparities were reported. Ultimately, an image-based item pool was created to measure affinity for technology interaction (Appendix E).

Animal Attitude Scale (AAS)

The AAS is a normative rating scale to assess individuals' perspectives towards animals, which is distinct from affinity for technology interaction. For this study, the AAS was included to assess discriminant validity and investigate response style biases. While the scales' original version comprises 20 items, Herzog et al. (2015) created a brief, unidimensional version with five items, the AAS-5, which was used in this study (Appendix F). Participants have to indicate their answer on a five-point Likert scale.

Design

The study employed a correlational design to examine the relationships between scores on the ATI scale, the new IBATI scale and the AAS. As the primary purpose was the validation of the IBATI scale against the ATI scale, the relationship between these scales was most crucial. The relationship between the IBATI scale and the AAS was assessed to explore if the IBATI scale specifically measures affinity for technology interaction.

Procedures

The study was conducted online, and participants accessed the study through Qualtrics with their devices. Participants were informed that the study would take about 15 minutes to complete. After providing consent, participants were asked to answer demographic questions regarding age, nationality and gender. Next, participants had to answer all items of three rating scales. The tests were administered separately, and the order of the instruments was randomised to mitigate test fatigue and priming effects. The AAS and the ATI scale were displayed as Likert scales, with each scale's items being displayed on a single web page. The nine IBATI items were shown individually. The order of the response options on the IBATI items was randomised. After completing the three personality rating scales, subjects were thanked for their participation, and it was confirmed that their response was recorded. Every participant had to complete the IBATI and ATI scales and the AAS to be considered in the analysis. Participants had to answer 27 questions, excluding the consent form, and click through 13 web pages to complete the survey.

Data Analysis

Data Preparation and Initial Analysis

After completion of the data collection, the responses were exported to R 4.3.1 (R Core Team, 2023). The statistical analysis was conducted with the packages ‘blavaan’ and ‘loo’ (Merkle & Rosseel, 2015). The data underwent both descriptive and inferential analysis to obtain an extensive overview of the rating scale’s psychometric properties. Participants with incomplete data were removed from the dataset. Next, negatively worded items in the ATI scale and the AAS were reverse-coded. Moreover, all scores were rescaled to values between zero and one to mitigate the effect of disparities in the number of response options.

Descriptive Statistics

Afterwards, descriptive statistics were calculated for the rescaled items to evaluate distributions, outliers and correlations in the item pool. In line with the study by Othman et al. (2011), the frequency at which each response option was selected was considered to identify items with limited ability to discriminate between different trait levels. Reliability was evaluated with Cronbach’s alpha to assess the scales’ internal consistency, which indicates the extent to which all items within a scale measure the same construct. Following Norman (2010) and Revelle (2023), Pearson's r was utilised to determine the items’ bootstrapped correlations. The corrected item-total correlation was computed for each scale by measuring the correlation between the selected item and the sum of all item scores of the scale minus the selected item. As all scales were designed to be unidimensional, all items should correlate at least 0.40 when considering the corrected item-total correlation, and items below the threshold should be reconsidered and potentially removed (Vaske et al., 2017). The item-total correlation supports evaluating the direction of scoring and how individual items contribute to the total scale score (Di Lillo et al., 2009). Mean scores for each scale were determined, as Franke et al. (2019) suggested, which provides a comparable rescaled total score regardless of the number of items and response options. Correlations and scoring ranges for the mean scores were considered to evaluate their relationships and discriminative ability.

Inferential Analysis

A parallel analysis was performed to explore the eigenvalues and define the number of factors for the factor analysis (Appendix G). Next, Bayesian Confirmatory Factor Analysis (BCFA) was conducted to evaluate the hypothesised factor structure. The Bayesian approach was introduced as it allows a probabilistic interpretation of results with consideration of uncertainty and enables the calculation of credibility intervals rather than point estimates (Schmettow, 2021). Based on the outcomes of the BCFA, the model's performance was evaluated and compared to an alternative model based on their information criteria. As suggested by Schmettow (2021), the Watanabe-Akaike Information Criterion (WAIC) and the Leave-One-Out Information Criteria (LOOIC) were taken into consideration. For both information criteria, a lower outcome indicates better forecasting accuracy. The WAIC served as the primary criterion, with the LOOIC being employed as an alternative if needed. Schmettow (2021) outlines that WAIC performs an internal check of the estimates' integrity to prevent wrong approximations. The internal check can lead to warnings in the WAIC because of the estimated effective number of parameters (p_{WAIC}). According to Vehtari et al. (2017), the outcomes are at risk of being unreliable if any estimates exceed the threshold of 0.4. Schmettow (2021) suggests replacing the WAIC with another criterion when in doubt about the estimates' reliability, such as the LOOIC. However, LOOIC can trigger warnings when Pareto k diagnostic values exceed the desired threshold of 0.5. Surpassing this threshold indicates that the reliability of the estimates might be in danger (Vehtari et al., 2017). The Bayesian Root Mean Square of Approximation (BRMSEA) is another useful metric to evaluate the models' performance when conducting Bayesian Confirmatory Factor Analysis (Hoofs et al., 2018). The BRMSEA contains a penalty term for model complexity, so it can be used to evaluate forecasting accuracy. Hoofs et al. (2018) point out that BRMSEA values below .05 indicate good model fit, and outcomes between .05 and .08 suggest adequate fit. The individual items were evaluated based on factor loadings and 95% credibility interval, which indicate the strength of the relationship to the latent variables. The factor loadings were standardised to enhance their interpretability by fixing the factors' variances to 1 (Zhang & Wang, 2017). The standardisation simplifies the loadings comparison and highlights the items' contribution to the factors. Factor loadings of 0.6 were considered acceptable, while loadings above 0.7 were classified as optimal (Hair et al., 2010). When an item's credibility interval does not surpass the minimally acceptable threshold, it indicates a weak association between the item and the latent variable. Fixing factors' variances to 1 also allows transforming covariances to obtain direct estimates for the credibility intervals of Pearson correlations between the latent variables (Koen & Yonelinas, 2016; Zhang & Wang, 2017).

Results

Descriptive Statistics

The rescaled results depict patterns of responses per item (Table 2). The mean scores for almost all IBATI items and all AAS items were higher than 0.50. In contrast, the mean scores for the ATI items were distributed around 0.50. Notably, the mean scores for the five reverse-coded items on the AAS and the ATI scale were lower than those of their respective scales' other items. In addition, the dispersion of scores, as indicated by the standard deviations, revealed a standard deviation ranging from 0.23 to 0.35 for image-based items. Thus, the image-based items' standard deviation was larger than for most text-based items. For the ATI scale, standard deviations ranged from 0.22 to 0.27, while for the AAS, they fluctuated between 0.21 and 0.27.

Table 2

Mean and Standard Deviation of all Items

Item	IBATI		ATI		AAS	
	M	*SD*	*M*	*SD*	*M*	*SD*
1	0.70	0.23	0.56	0.25	0.84	0.22
2	0.60	0.32	0.63	0.22	0.64	0.27
3	0.55	0.30	0.48	0.25	0.54	0.27
4	0.60	0.24	0.59	0.23	0.83	0.21
5	0.68	0.29	0.58	0.24	0.68	0.27
6	0.44	0.35	0.44	0.27		
7	0.50	0.31	0.48	0.25		
8	0.63	0.29	0.45	0.24		
9	0.58	0.24	0.59	0.23		

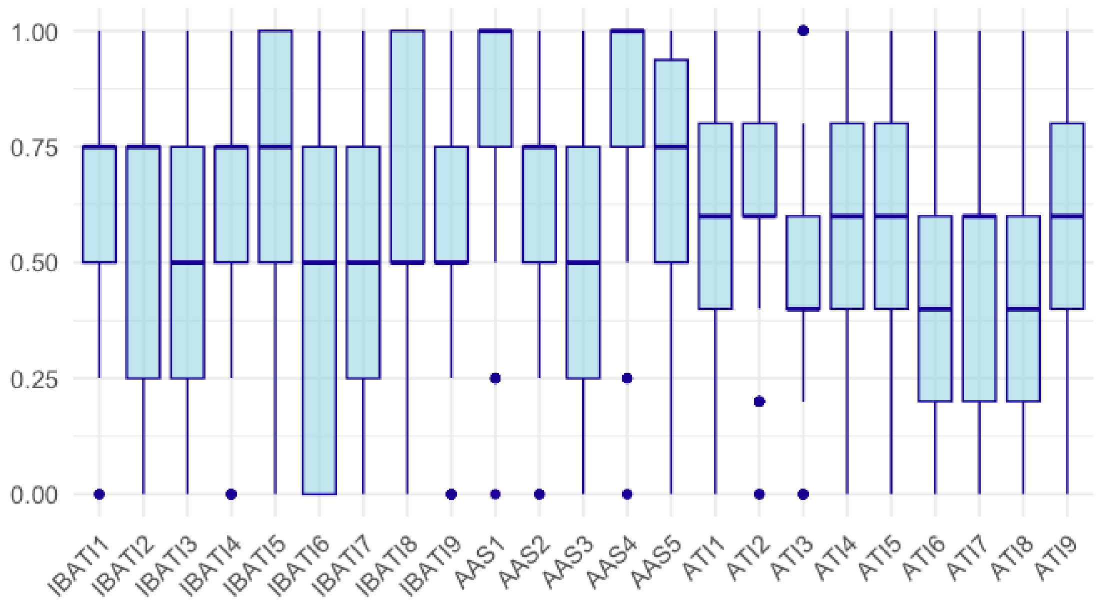
Response Tendencies

All response options to the ATI and IBATI items were chosen at least four times, and the response options to the AAS were selected at least twice (Appendix H). When comparing the selection frequency of each response option, the response options of most AAS items were more polarised than on the IBATI and ATI items. IBATI1, IBATI4 and IBATI9 stood out as many participants selected the same response option. However, the corresponding answers on ATI4 and ATI9 are also the most frequently chosen response options. As items on the ATI scale offer more response options than on the IBATI scale and the AAS, the predicted frequency of any response option being chosen is lower by default. When accounting for the disparity in the number of response options, little differences between frequencies in the ATI and the IBATI scales remain.

The box plot in Figure 1 displays patterns in the response distribution. Due to the discrepancies in the number of response options, the data was rescaled, but differences in the relative numbers of the rescaled data will remain. Answers to the IBATI items were mainly spread around 0.5 and 0.75, corresponding to the third and fourth out of five response options. Responses to the ATI items were centred around 0.4 and 0.6, equivalent to the third and fourth out of six response options. The median of the reverse-coded items ATI3, ATI6 and ATI8 was 0.4, while the median of all other ATI items was 0.6. The pattern was also observable for the reverse-coded items AAS2 and AAS3, as other AAS items elicited even more extreme responses.

Figure 1

Box Plot of Item Answer Distributions

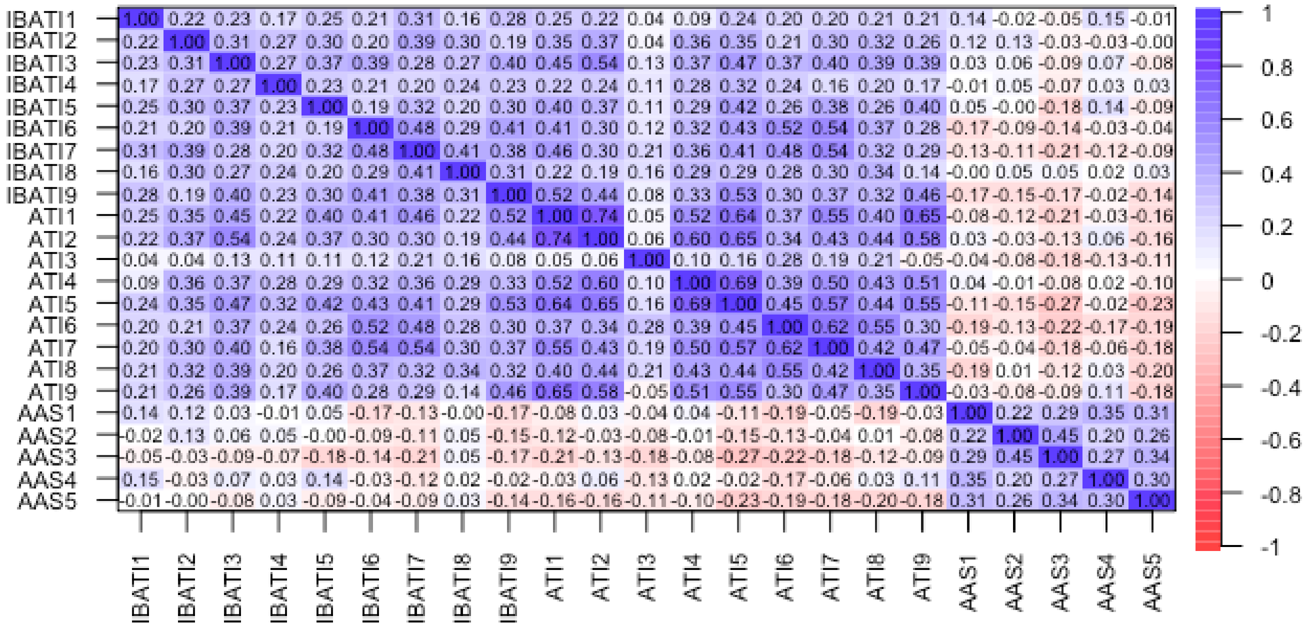


Internal Consistency

The image-based test reached an alpha of 0.78, which is considered good and indicates that the items within the scale are reasonably correlated. The ATI scale achieved an excellent alpha of 0.87, which indicated high internal consistency. Cronbach's alpha for the AAS resulted in a minimally acceptable alpha of 0.68.

Figure 2

Bootstrapped Correlation Heat Map on Item Level



Item-Level Correlations Within Rating Scales

Participants' answers were examined to evaluate the bootstrapped item correlations within and between the three rating scales (Figure 2). All correlations between the IBATI items were positive, with values spanning from 0.16 to 0.48. In contrast, the correlations between the ATI items ranged from -0.05 to 0.74, with ATI3 being distinctive and displaying weaker correlations with several items of the ATI scale. All AAS items exhibited positive interrelations ranging from 0.20 to 0.45.

Item-Level Correlation Between Rating Scales

The correlations between items on IBATI and ATI ranged from 0.04 to 0.54, which resulted in strong relationships for 72 of the 81 item pairs. Again, ATI3 stood out, as it was the only item with weak correlations with more than one IBATI item. Items on the AAS and the IBATI scale had weak correlations between -0.21 and 0.15. Moreover, correlations for items on the AAS and the ATI scale ranged from -0.27 to 0.06. When considering item pairs from different scales, five of the nine IBATI items had the strongest correlations with the ATI item they were derived from. These correlations ranged from 0.13 to 0.54 and were particularly high for eight of the nine parallel item pairs.

Item-Total Correlation

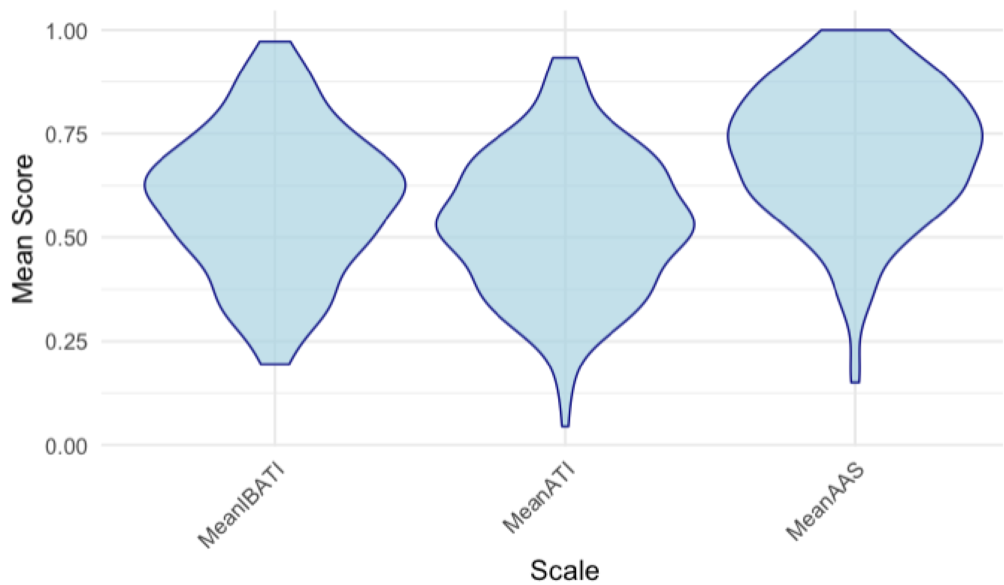
The correlation between the individual items and the total scale scores was considered, and the corrected item-total correlation was calculated to adjust for items contributing to the total scale score (Appendix I). All image-based items were positively associated with the IBATI total score, with corrected item-total correlations ranging from 0.37 to 0.59. The items IBATI1 and IBATI4 reached corrected correlations of 0.37 and did not surpass the threshold of 0.40. Furthermore, six of the nine IBATI items correlated stronger than 0.40 regarding the ATI total score, with correlations ranging from 0.27 to 0.56. All ATI items besides ATI3 had high corrected correlations ranging from 0.58 to 0.75 when related to the ATI total score and between 0.48 and 0.64 when considering the IBATI total score. No IBATI or ATI item correlated stronger than 0.27 with the AAS total score. The AAS items had corrected item-total correlations between 0.39 and 0.51. Furthermore, items of the AAS had a maximum correlation of -0.24 to the ATI mean score and of -0.16 to the IBATI mean score.

Mean Score Distributions

The individuals' mean scores ranged from 0.19 to 0.97 (*M* = 0.59, *SD* = 0.17) for the IBATI scale, 0.04 to 0.93 (*M* = 0.53, *SD* = 0.17) for the ATI scale and 0.15 to 1 (*M* = 0.71, *SD* = 0.16) for the AAS (Figure 3). These outcomes provide insights into the variability and central tendency of the mean scores on each scale.

Figure 3

Mean Score Distribution per Scale

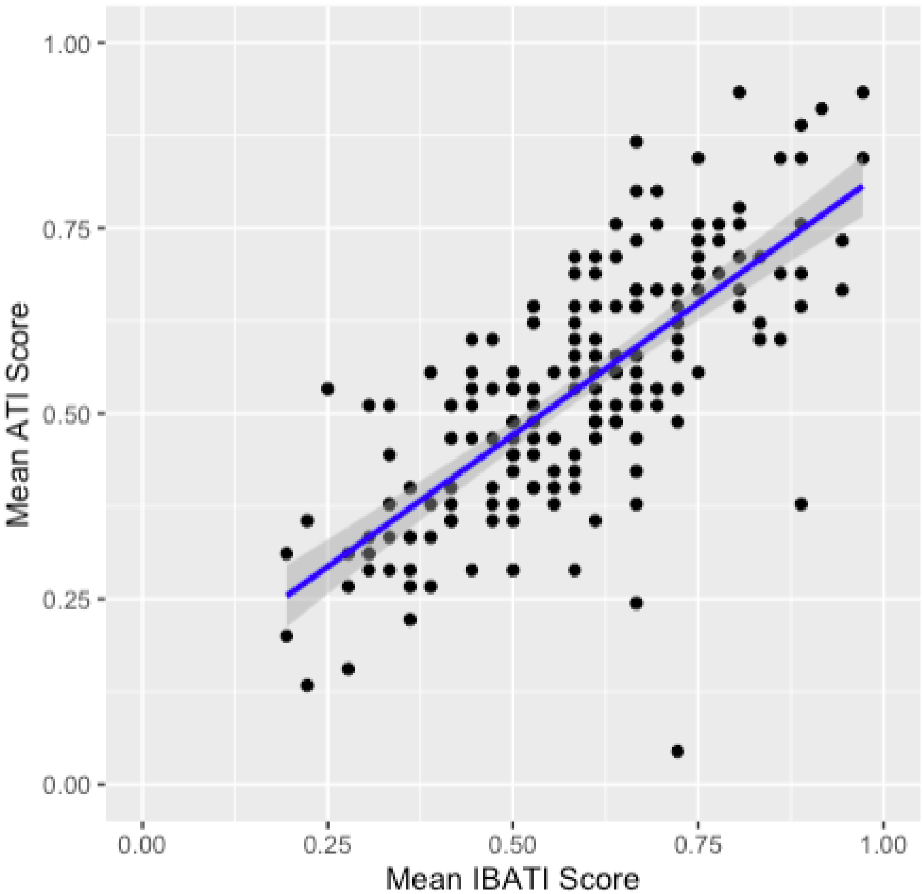


Mean Score Correlations

The correlation between the mean score on the IBATI and ATI scales was very robust, indicated by a correlation coefficient of 0.74 (Figure 4). The high correlation symbolises that higher scores on one rating scale are associated with higher scores on the second rating scale. While the correlation between the mean scores suggests high consistency, three subjects deviated from the pattern and reported much higher scores on the image-based scale than on the text-based scale. These outliers align with the mean score distribution and underline that many subjects reported slightly higher scores on the IBATI scale than on the ATI scale.

Figure 4

Correlation Mean Scores on IBATI and ATI Scales



The correlation between mean scores on the AAS and the ATI scale was negative (r = -0.23), suggesting that those with a higher affinity for technology interaction, as measured by the ATI scale, tend to have lower scores on the AAS. The correlation between mean scores on the AAS and IBATI scale was weak (r = -0.08), indicating that the scales are barely related.

Bayesian Confirmatory Factor Analysis

Bayesian Confirmatory Factor Analysis was conducted to evaluate the factor structure of the latent variables. Markov Chain Monte Carlo sampling was used to estimate the parameter distribution with three chains of 15000 iterations each. The models' forecasting accuracy was evaluated using the LOOIC, WAIC and BRMSEA (Table 3). Standardised factor loadings were considered to estimate the relationship between the latent variables and the data.

Model 1

Each of the rating scales was introduced as a separate factor. Thus, the model is composed of three different factors. Factor 1 comprises the IBATI items, Factor 2 consists of the ATI items, and Factor 3 entails the AAS items. Both WAIC and LOOIC resulted in a warning during the computation. Therefore, BRMSEA was introduced to evaluate the model's forecasting accuracy. BRMSEA indicates an acceptable model fit, as the centre estimate is below the threshold of 0.08. Thus, most of the credibility interval of BRMSEA is positioned in this range, which suggests that an acceptable model fit is likely.

Model 2

As an alternative to the three factors of the first model, the second model solely consists of two factors. Factor 1 consists of all IBATI and ATI items. Factor 2 comprises the AAS items. As in Model 1, WAIC and LOOIC led to errors indicating restricted. BRMSEA was computed to assess the model's forecasting accuracy and indicates a centre estimate above the acceptable threshold of 0.08.

Table 3

Information Criteria to Estimate Predictive Accuracy

Model	WAIC	LOOIC	BRMSEA
Model 1 (3-Factor Model)	11477.01*	11477.43*	0.079 [90 % CrI 0.077 – 0.082]
Model 2 (2-Factor Model)	11497.47*	11497.38*	0.083 [90 % CrI 0.080 – 0.085]

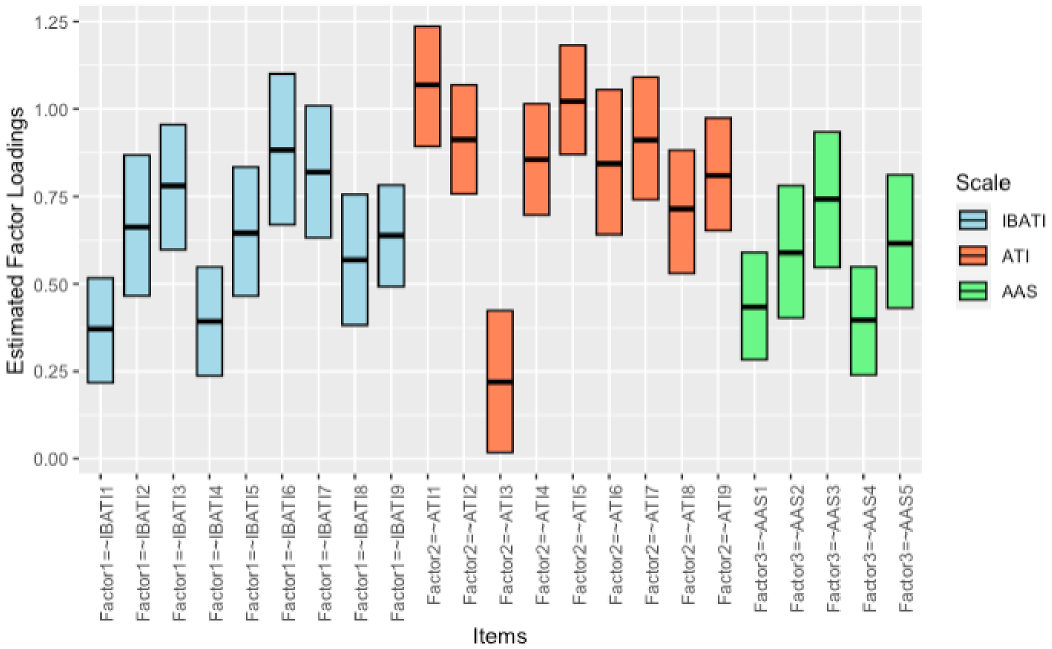
Note. * indicates warnings in the statistical modelling restricting the estimates' reliability.

Factor Loadings

The 90 % credibility interval for the BRMSEA of Model 2 exceeds the estimated threshold of Model 1, which indicates that Model 1 has a superior fit. As a result, Model 1 was chosen for further evaluation on item and factor levels. The standardised factor loadings and their credibility intervals indicate the strength of the relationships between the latent variables and the items (Figure 5). Items from the IBATI scale, the ATI scale and the AAS were assessed to determine the quality of the items for each scale.

Figure 5

Factor Loadings of 3-Factor Model per Item with 95% Credibility Interval



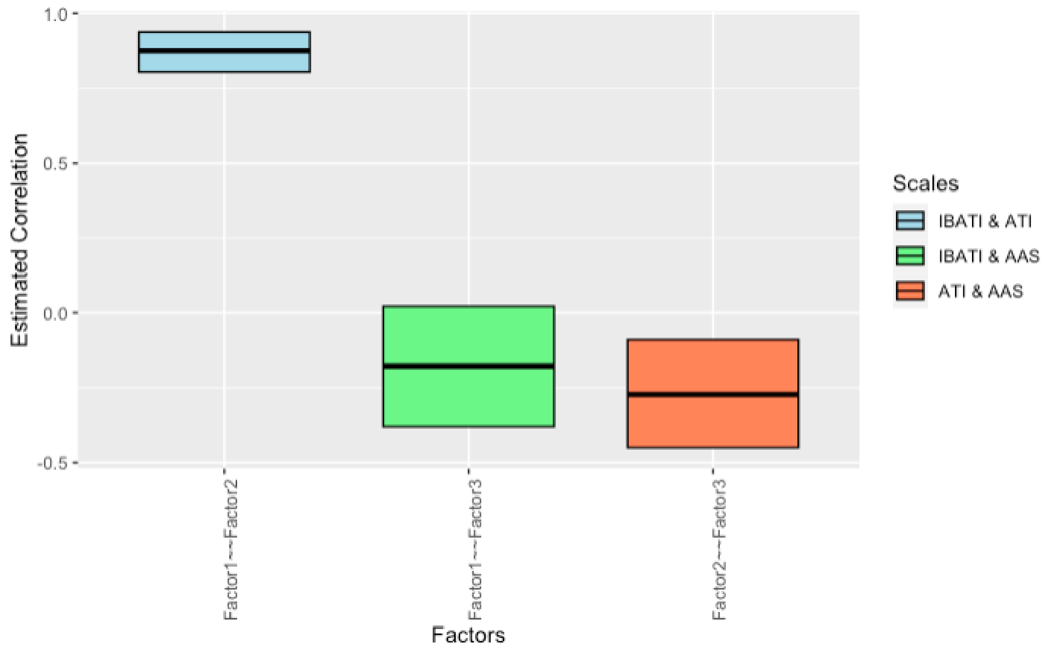
The estimated factor loading of three of nine items in Factor 1 (IBATI) was below the desired minimum threshold of 0.6. The credibility interval for two of these items, IBATI1 and IBATI4, displayed no overlap to the minimally acceptable threshold, suggesting the items might not adequately represent the construct they are supposed to measure. Three IBATI items had acceptable factor loadings, and three items' factor loadings were classified as optimal. For Factor 2 (ATI), most items exhibit strong relationships to their latent factors, with estimated loadings ranging from 0.71 to 1.07. However, item ATI3 stands out with a much lower factor loading of 0.22, suggesting a weak relationship with its latent factor. The estimated factor loadings for Factor 3 (AAS) were lower than expected, with three of five centre estimates below the acceptable threshold. The complete overview of all factor loadings and their credibility intervals can be found in Appendix J. For a more comprehensive understanding of the convergence and stability of the model, trace plots visualising the distribution of loadings throughout the iterations were included in Appendix K. These factor loadings and their credibility intervals provide insights into the reliability of the BCFA model.

Latent Variable Correlation

As the factors' variances were fixed to 1, the covariances between the factors were leveraged to estimate 95 % credibility intervals for the correlation (Figure 6). Strong positive associations between Factor 1 (IBATI) and Factor 2 (ATI) can be observed with an estimated correlation of 0.875 [95 % CrI 0.799, 0.934]. Factor 1 (IBATI) and Factor 3 (AAS) had a weak negative correlation with an estimated correlation of -0.178 [95 % CrI -0.374, 0.029]. Lastly, Factor 2 (ATI) and Factor 3 (AAS) had an estimated correlation of -0.273 [95 % CrI -0.448, -0.087].

Figure 6

Factor Loadings of 3-Factor Model per Item with 95% Credibility Interval



In conclusion, the Bayesian Confirmatory Factor Analysis revealed a robust 3-Factor model that aligns with the IBATI, ATI, and AAS. The high correlations between ATI and IBATI on scale and item level provide valuable insights regarding convergent validity, while the AAS helps to establish discriminant validity. The standardised factor loadings are vital to assess reliability, item quality and compare the image-based rating scale to the text-based rating scales. These findings establish a solid foundation for further investigation, interpretation and application of the newly created IBATI scale.

Discussion

The current study explored whether image-based testing approaches can be used as an alternative to traditional text-based approaches. The newly created image-based scale IBATI and its psychometric qualities were compared to the text-based scales ATI and AAS. The study provided insights into the quality of both image-based tests on the scale and item levels. The robust reliability and validity of the IBATI scale open up new avenues for research and practical applications. In the context of existing literature, the outcomes add to the growing body of evidence supporting the use of image-based scales in psychometric evaluations and provide a blueprint for transforming text-based scales to image-based formats.

Reliability of the IBATI Scale

Any psychometric scale needs to be reliable to be useful. Reliability refers to the scales' ability to yield consistent results and was assessed with Cronbach's Alpha on the scale level and the factor loadings on the item level (Borsci et al., 2023). While the internal consistency of the ATI scale was as high as in previous studies (Franke et al., 2019), the AAS yielded lower internal consistency than expected (Herzog et al., 2015). Cronbach's Alpha for the ATI scale was slightly higher than for the IBATI scale; however, both scales display strong reliability in terms of internal consistency. The factor loadings are pivotal in evaluating to what extent individual items contribute to the measurement of their underlying factor (Hair et al., 2010). While most ATI items exhibited slightly higher loadings than IBATI items, almost all items from both scales generally displayed strong loadings. However, some inconsistencies for items on either scale indicate minor disparities in the alignment with the underlying factors and will be addressed in the section about suggestions for improvement. Still, the findings suggest robust reliability for the IBATI scale on both scale and item levels.

Validity of the IBATI Scale

Validity is required to ensure that the scale successfully measures the intended construct. The correlations between the ATI and IBATI scales were so pronounced that by model selection, it was assessed whether the scales were entirely exchangeable. This assumption was not confirmed as the 3-Factor model was the better fit. However, the factor structure might be influenced by the presence of reverse-coded items in the ATI scale. Individuals' affinity for technology interaction draws on the need for cognition theory (Franke et al., 2019), which was initially introduced as a unidimensional construct but commonly displays more than one factor because of the directionality of the wording of the items (Bauer & Stiner, 2020). This study might reflect the phenomenon, which means that the reverse-coded items likely affected the factor structure, leading to a better fit of the 3-Factor model. Nevertheless, the outstanding correlation between the ATI and IBATI scales establishes convergent validity and underlines that the IBATI scale successfully assesses the intended construct. In contrast, the AAS taps into a different dimension, establishing discriminant validity between the AAS and the IBATI scale. The convergent validity between the image-based and text-based response scales in this study is notably stronger than in other studies exploring image-based testing approaches (Hilliard et al., 2022a; Hilliard et al., 2022b; Krainikovsky et al., 2019; Leutner et al., 2017; Paunonen et al., 2001). This was expected, as the IBATI scale was derived from the validated ATI scale, while other studies aimed at implementing standalone image-based rating scales. Based on the assumption that the ATI scale is valid in assessing affinity for technology interaction, either IBATI or ATI can be used and will succeed in measuring the intended construct.

Suggested Improvements to the IBATI and ATI Scales

While the study established robust reliability and construct validity, it also showed that some aspects on item-level in the IBATI and ATI scales can be improved further. Regarding the IBATI scale, some image-based response options were possibly considered more appealing and relatable than others (e.g. IBATI1 and IBATI4) and were not as effective in differentiating individuals as the remaining image-based items. This phenomenon could be controlled by conducting extensive pre-tests, where participants assign a score to indicate the response options' attractiveness and limit discrepancies between the likability of the images. By refining the response options of the distinctive image-based items, their discriminative abilities and, consequently, the scales' psychometric properties could be improved further.

Concerning the ATI scale, reverse-coded text-based items (e.g. ATI3 and ATI8) stood out and were less reliable in assessing the underlying factor than the other text-based items. Generally, the reverse-coded items did not match the psychometric quality of the remaining text-based items and should be reconsidered. Merging items from text-based and image-based scales could be a valuable exploration to maximise engagement and increase test validity (Leutner et al., 2017). Combining valid items of both types of rating scales potentially allows for an improved testing experience and a more precise measurement of the underlying factor. Valid image-based items from the IBATI scale could replace the reverse-coded text-based items and might yield more consistent and meaningful test results.

Recommended Use of the IBATI Scale

The pronounced correlations between the underlying factors on the response scales show that the IBATI scale is a viable alternative to the ATI scale. Although the ATI scale displayed superior factor loadings compared to the IBATI scale, the outcomes underline that the image-based response scale can differentiate individuals. The ATI scale has demonstrated convergent validity to numerous related rating scales (Franke et al., 2019). Due to the strong convergence between ATI and IBATI scales, it can be expected that both response scales succeed in measuring affinity for technology interaction. Consequently, either scale can be selected to assess affinity for technology interaction based on the available data. Decisions about which scale to choose depend on personal preferences, target group and purpose of use.

Limitations and Recommendations for Further Research

Despite strong evidence for the validity and reliability of the IBATI scale, two limitations must be considered and can inform future research regarding the IBATI scale. Firstly, and most importantly, while IBATI and ATI scales were congruent on scale and item level, validity is only based on the outcomes of the rating scales. To establish ecological validity, an external measure of affinity for technology interaction is required. By comparing the relationship between the rating scales and an external measure, it can be examined which rating scale succeeds in forecasting actual behaviour concurrently and predictively. Previous experiments have shown that individual differences in attitudes about technology interaction are connected to observable behaviour (Schmettow & Drees, 2014). When applying psychometric assessments in high-stakes situations, for example, to predict success at work, ecological validity is vital (Cizek & Rosenberg, 2011). Secondly, as participants answered all items only once, test-retest reliability was not assessed yet. Exploring which rating scale yields more consistent results over time and across situations would be valuable. The scales' stability and situational robustness can be evaluated by completing the scales at multiple points in time and in different emotional states. Repeated measures allow observation of whether the assessed trait level is similar each time a person completes the measure (Cizek & Rosenberg, 2011). Additional considerations for future research are listed below.

Number of Response Options: The decisions about the number of response options in the IBATI and ATI scales were aligned with existing literature (Franke et al., 2019; Robinson, 2018; Sang et al., 2016), which led to a disparity between the scales. Using the same number of response options across the scales could likely improve the rating scales' comparability.

Design Guidelines: The limited number of items does not allow the establishment of design guidelines for image-based response options, which could be the focus of future studies.
Construct Validity: One image-based scale was introduced, so convergent and discriminant validity between image-based scales was not explored and could be examined in the future.
Faking: Image-based testing approaches might restrict faking (Hilliard et al., 2022a; Leutner et al., 2017). Future studies could instruct subjects to fake their answers to obtain a high score. By monitoring score inflation, it can be evaluated which testing approach is harder to fake.

Reflection on Converting a Test to an Image-Based Format

Using text-to-image generative models to create image-based response options is a nuanced approach. It promises a flexible procedure to decrease manual labour and cost while obtaining robust validity and reliability. While traditional approaches to establishing psychometric scales can be expensive and time-consuming (Robinson, 2018), the item design of the IBATI scale was completed in less than 30 hours and cost less than 50 €. Thus, it seems that text-based rating scales with a strong theoretical foundation and robust psychometric properties can be efficiently transformed into a congruent image-based rating scale. The initial motivation for establishing an image-based response scale was three-fold and related to engagement, objectivity and resistance to faking.

High Engagement

Several researchers have pointed out that using image-based testing might increase the completion rate (Hilliard et al., 2022a; Hilliard et al., 2022b; Leutner et al., 2017); however, no clear evidence about the completion rates of image-based tests was available. As the study comprised text-based and image-based questions, the completion rate must be considered cautiously, but most web pages displayed an image-based item (9 of 13). 178 of the 189 subjects who started the study also finished it, which represents a completion rate of 94.1%. Liu and Wronski (2018) have analysed over 25000 web surveys to predict the completion rate based on the number of questions and web pages. The estimated completion rate was about 89 % based on the number of questions and 86 % based on the number of web pages. The completion rate of this study surpassed both estimations, which is remarkable as subjects were not paid to complete the survey and suggest that participants perceived the inclusion of image-based testing as more engaging than traditional text-based testing. Despite high engagement, it must be considered that image-based rating scales might not be accessible to everyone, for example, visually impaired people. Therefore, image-based and text-based testing approaches should not be considered direct competitors, but complementing tools to provide an ideal testing experience for different population groups.

Objectivity

Text-based items might require precise translations when tested internationally; however, these text-based translations are time-insensitive. If used like in this study, image-based items provide guidance to distinguish trait levels, which might positively affect the rating scales' objectivity. In the context of affinity for technology interaction, constant innovation might influence what is perceived as regular (Wessel et al., 2020), which requires regular refinement of response options. As technological advancements deviate between cultures, cross-cultural comparisons could be challenging for the IBATI scale. Moreover, research about image-based tests with children showed that image-based tests are strongly tied to the cultural background and ideas of how an object looks might differ (Carter et al., 2005). While the study shows that image-based response options can effectively distinguish between individuals, additional refinements must be expected when using image-based scales across longer periods of time and various cultures.

Resistance to Faking

While both response scales rely on Likert scales to distinguish various trait levels, the ATIs' response options are displayed in order, and the IBATIs' response options order is randomised. Van Hooft and Born (2012) demonstrated that participants mostly fixate on the most extreme response options when faking in text-based approaches. By shuffling the order of the image-based response option, the most extreme trait level is not discernible as easily, which might make the IBATI scale harder to fake. The congruence between items on the ATI and IBATI scales indicates that most response options were perceived to depict the designated trait levels accurately and were chosen because of the content rather than the order. Ultimately, it can be concluded that image-based personality testing is a promising method to explore further. Negative consequences caused by using image-based testing approaches are unlikely, as it seems to be an engaging and fair alternative to traditional text-based testing.

Conclusion

In conclusion, this study explored to what extent image-based testing approaches are congruent to traditional text-based testing approaches on scale and item level. Generative artificial intelligence was leveraged to create the IBATI scale, an image-based version of the ATI scale by Franke et al. (2019). The results suggest that image-based testing is an engaging, innovative approach and can yield robust construct validity and internal consistency. The IBATI scale was congruent to the ATI scale on item and scale levels and displayed discriminant validity to the AAS. Further research is recommended to establish ecological validity, test-retest reliability and investigate the potential benefits of combining text-based and image-based items. While this study emphasised the potential of image-based testing, it also highlights the need for refinements and research to explore its full potential.

References

Attig, C., Wessel, D., & Franke, T. (2017). Assessing personality differences in human-technology interaction: an overview of key self-report scales to predict successful interaction. In *HCI International 2017–Posters' Extended Abstracts: 19th International Conference, HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings, Part I 19* (pp. 19-29). Springer International Publishing.

Baez, H. B. (2013). Personality tests in employment selection: Use with caution.

Barrick, M. R., & Mount, M. K. (2005). Yes, Personality Matters: Moving on to More Important Matters. *Human Performance*, *18*(4), 359-372.

Bauer, B., & Stiner, E. (2020). Need for cognition. *Encyclopedia of personality and individual differences*, 3122-3125.

Borsci, S., Schmettow, M., Malizia, A., Chamberlain, A., & Van Der Velde, F. (2023). A confirmatory factorial analysis of the Chatbot Usability Scale: a multilanguage validation. *Personal and Ubiquitous Computing*, *27*(2), 317-330.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of personality and social psychology*, *42*(1), 116-131.

Carter, J. A., Lees, J. A., Murira, G. M., Gona, J., Neville, B. G., & Newton, C. R. (2005). Issues in the development of cross-cultural assessments of speech and language for children. *International Journal of Language & Communication Disorders*, *40*(4), 385-401.

Chamorro-Premuzic, T. (2015). *Personality and individual differences*. John Wiley & Sons.

Chamorro-Premuzic, T., & Furnham, A. (2010). *The psychology of personnel selection*. Cambridge University Press.

Cizek, G. J., & Rosenberg, S. L. (2011). Psychometric methods and high-stakes assessment: Contexts and methods for ethical testing practice. *Handbook of ethics in quantitative methodology*.

Di Lillo, M., Cicchetti, A., Scalzo, A. L., Taroni, F., & Hojat, M. (2009). The Jefferson Scale of Physician Empathy: preliminary psychometrics and group comparisons in Italian physicians. Academic Medicine, 84(9), 1198-1202.

Franke, T., Attig, C., & Wessel, D. (2019). A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale. *International Journal of Human–Computer Interaction*, *35*(6), 456-467.

Furnham, A., Guenole, N., Levine, S. Z., & Chamorro-Premuzic, T. (2013). The NEO Personality Inventory–Revised: Factor structure and gender invariance from exploratory structural equation modeling analyses in a high-stakes setting. *Assessment*, *20*(1), 14-23.

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010). Multivariate Data Analysis. 7th Edition, Pearson, New York.

Herzog, H., Grayson, S., & McCord, D. (2015). Brief measures of the animal attitude scale. *Anthrozoös*, *28*(1), 145-152.

Hilliard, A., Kazim, E., Bitsakis, T., & Leutner, F. (2022a). Measuring personality through images: Validating a forced-choice image-based assessment of the big five personality traits. *Journal of Intelligence*, *10*(1), 12. https://doi.org/10.3390/jintelligence10010012

Hilliard, A., Kazim, E., Bitsakis, T., & Leutner, F. (2022b). Scoring a forced-choice image-based assessment of personality: A comparison of machine learning, regression, and summative approaches. *Acta Psychologica*, *228*, 103659.

Hoofs, H., van de Schoot, R., Jansen, N. W., & Kant, I. (2018). Evaluating model fit in Bayesian confirmatory factor analysis with large samples: Simulation study introducing the BRMSEA. Educational and psychological measurement, 78(4), 537-568.

Ihsan, Z., & Furnham, A. (2018). The new technologies in personality assessment: A review. *Consulting Psychology Journal: Practice and Research*, *70*(2), 147-166.

Kankaraš, M. (2017). Personality matters: Relevance and assessment of personality characteristics.

Koen, J. D., & Yonelinas, A. P. (2016). Recollection, not familiarity, decreases in healthy ageing: Converging evidence from four estimation methods. *Memory*, *24*(1), 75-88.

Krainikovsky, S., Melnikov, M., & Samarev, R. (2019). Estimation of psychometric data based on image preferences. In *Conference Proceedings for Education and Humanities, WestEastInstitute* (Vol. 2019, pp. 75-82).

Leutner, F., Akhtar, R., & Chamorro-Premuzic, T. (2022). Gamification and Game-Based Assessments. In *The Future of Recruitment*. Emerald Publishing Limited.

Leutner, F., Yearsley, A., Codreanu, S.-C., Borenstein, Y., & Ahmetoglu, G. (2017). From Likert Scales to images: Validating a novel creativity measure with image-based response scales. *Personality and Individual Differences*, *106*, 36–40. https://doi.org/10.1016/j.paid.2016.10.007.

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological science in the public interest*, *1*(2), 27-66.

Liu, V., & Chilton, L. B. (2022, April). Design guidelines for prompt engineering text-to-image generative models. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-23).

Liu, M., & Wronski, L. (2018). Examining completion rates in web surveys via over 25,000 real-world surveys. Social Science Computer Review, 36(1), 116-124.

Lundgren, H., Kroon, B., & Poell, R. F. (2017). Personality testing and workplace training: Exploring stakeholders, products and purpose in Western Europe. *European Journal of Training and Development*.

McCrae, R. R., & Costa Jr, P. T. (1995). Trait explanations in personality psychology. *European Journal of Personality*, *9*(4), 231-252.

Meissner, F., & Rothermund, K. (2015). A thousand words are worth more than a picture? The effects of stimulus modality on the implicit association test. *Social Psychological and Personality Science*, *6*(7), 740-748.

Merkle, E. C., & Rosseel, Y. (2015). blavaan: Bayesian structural equation models via parameter expansion. *arXiv preprint arXiv:1511.05604*.

Mischel, W. (1968). *Personality and assessment*. Psychology Press.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel psychology*, *60*(3), 683-729.

Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. Advances in health sciences education, 15, 625-632.

Othman, A. R., Yin, T. S., Sulaiman, S., Ibrahim, M. I. M., & Razha-Rashid, M. (2011). Application of mean and standard deviation in questionnaire surveys. *Menemui Matematik (Discovering Mathematics)*, *33*(1), 11-22.

Paunonen, S. V., Ashton, M. C., & Jackson, D. N. (2001). Nonverbal assessment of the Big Five personality factors. *European Journal of Personality*, *15*(1), 3-18.

Piotrowski, C. (2015). Projective techniques usage worldwide: A review of applied settings 1995-2015. *Journal of the Indian Academy of Applied Psychology*, *41*(3), 9-19.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Revelle, W. (2023, July 9). Cor.ci: Bootstrapped and normal confidence intervals for raw and composite correlations in psych: Procedures for psychological, psychometric, and Personality Research. https://rdrr.io/cran/psych/man/cor.ci.html

Revelle, W. (n.d.). *fa.parallel: Scree plots of data or correlation matrix compared to random ``parallel" matrices*. R Documentation. https://www.personality-project.org/r/html/fa.parallel.html

Robinson, M. A. (2018). Using multi-item psychometric scales for research and practice in human resource management. *Human resource management*, *57*(3), 739-750.

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support?. *Human resource management review*, *16*(2), 155-180.

Sackett, P. R., & Walmsley, P. T. (2014). Which personality attributes are most important in the workplace?. *Perspectives on Psychological Science*, *9*(5), 538-551.

Sang, J., Zhang, H., & Xu, C. (2016). Visual BFI: An exploratory study for image-based personality test. In *Advances in Multimedia Information Processing-PCM 2016: 17th Pacific-Rim Conference on Multimedia, Xi′ an, China, September 15-16, 2016, Proceedings, Part I* (pp. 95-106). Springer International Publishing.

Sartori, R. (2010). Face validity in personality tests: psychometric instruments and projective techniques in comparison. *Quality & quantity*, *44*, 749-759.

Schmettow, M. (2021). *New statistics for design researchers*. Cham: Springer International Publishing.

Schmettow, M., & Drees, M. (2014, September). What drives the geeks? Linking computer enthusiasm to achievement goals. In *Proceedings of the 28th International BCS Human Computer Interaction Conference (HCI 2014) 28* (pp. 234-239).

Schmettow, M., Noordzij, M. L., & Mundt, M. (2013). An implicit test of UX: individuals differ in what they associate with computers. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2039-2048).

VandenBos, G. R. (2015). *Apa Dictionary of Psychology* (2nd ed.). American Psychological Association.

Van Hooft, E. A., & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, *97*(2), 301-316.

Vaske, J. J., Beaman, J., & Sponarski, C. C. (2017). Rethinking internal consistency in Cronbach's alpha. *Leisure sciences*, *39*(2), 163-173.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing*, *27*, 1413-1432.

Weisz, J. D., Muller, M., He, J., & Houde, S. (2023). Toward General Design Principles for Generative AI Applications. *arXiv preprint arXiv:2301.05578*.

Wessel, D., Heine, M., Attig, C., & Franke, T. (2020). Affinity for technology interaction and fields of study: implications for human-centered design of applications for public administration. In *Proceedings of Mensch und Computer 2020* (pp. 383-386).

Young, H. R., Glerum, D. R., Wang, W., & Joseph, D. L. (2018). Who are the most engaged at work? A meta-analysis of personality and employee engagement. *Journal of Organizational Behavior*, *39*(10), 1330–1346. https://doi.org/10.1002/job.2303

Zhang, Z. & Wang, L. (2017). Advanced statistics using R. Granger, IN: ISDSA Press. https://doi.org/10.35566/advstats. ISBN: 978-1-946728-01-2.

Zhang, H., Zhang, J., Sang, J., & Xu, C. (2017). A demo for image-based personality test. *MultiMedia Modeling*, 433–437. https://doi.org/10.1007/978-3-319-51814-5_36

Zhang, C., Zhang, C., Zhang, M., & Kweon, I. S. (2023). Text-to-image diffusion model in generative ai: A survey. *arXiv preprint arXiv:2303.07909*.

Ziegler, M., MacCann, C., & Roberts, R. (Eds.). (2012). *New perspectives on faking in personality assessment*. Oxford University Press.

Appendix A:

Information Sheet 'Image-Based Personality Testing'

Purpose of the research:

This study is conducted to observe the validity of a newly created image-based personality test assessing interindividual differences. While most existing personality tests are text-based, image-based personality tests provide advantages as increased engagement, improved objectivity and are considered to be harder to fake. By examining how participants answer on the text-based and image-based personality test that relate to the same construct, convergent validity can be established. Thus, the study seeks if image-based tests can be used as an alternative to traditional text-based tests.

Risks of participating:

No risk for the participants is involved when completing the study. The study has been approved by the BMS Ethics Committee/Domain Humanities & Social Sciences).

Withdrawal from the study:

Participants can decide that they want to withdraw from the study at any point in time. It is not required to state a reason for withdrawal from the study.

Personal information:

Before the start of the study, personal information about the participants' age, gender and nationality will be asked for validation purposes. Participants can choose 'Prefer not to say', if they do not want their personal information to be recorded. The personal information will be treated confidentially and will not be shared. If participants want to erase their personal data, they can contact the researcher via mail to be removed from the study.

Data storage:

The test data and personal data will be collected through Qualtrics and stored in the system for data archiving purposes. Individual data will be not be shared outside of the research team and only considered for validation purposes. After completion the data will be stored securely.

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher, please contact the Secretary of the Ethics Committee/domain Humanities & Social Sciences of the Faculty of Behavioural, Management and Social Sciences at the University of Twente by ethicscommittee-hss@utwente.nl

Appendix B

Consent Form for 'Image-Based Personality Testing'

Please tick the appropriate boxes (integrated in Qualtrics)	***Yes***
Taking part in the study:	
I have read and understood the study information dated [13/06/2023], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	□
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	□
I understand that taking part in the study involves completing image and text-based items through Qualtrics.	□
Use of the information in the study:	
I understand that information I provide will be used for validation purposes of the image-based questionnaire and the results will be processed and used for a Master's Thesis.	□
I understand that personal information collected about me that can identify me, such as [e.g. my name or age], will not be shared beyond the study team.	□
Future use and reuse of the information by others:	
I give permission for the questionnaire answers that I provide to be archived in Qualtrics so it can be used for future research and learning.	□

Study contact details for further information:

Nico Reisch (n.o.reisch@student.utwente.nl)

Appendix C

Items in the Text-Based Affinity for Technology Interaction Scale

The response options for the text-based ATI scale range from 1 (Completely Disagree) to 6 (Completely Agree). Items that are indicated with (*) are reverse scored.

1. I like to occupy myself in greater detail with technical systems.
2. I like testing the functions of new technical systems.
3. I predominantly deal with technical systems because I have to. (*)
4. When I have a new technical system in front of me, I try it out intensively.
5. I enjoy spending time becoming acquainted with a new technical system.
6. It is enough for me that a technical system works; I don't care how or why. (*)
7. I try to understand how a technical system exactly works.
8. It is enough for me to know the basic functions of a technical system. (*)
9. I try to make full use of the capabilities of a technical system

Appendix D

Prompts for Generative Artificial Intelligence Image Generation

1. Multiple computer screens with code, geek
2. One computer screen with a video editing software
3. Casual smartphone user
4. Writing on a piece of paper
5. Technology novice using phone for the first time
6. Person being excited and happy
7. Person watching a video on a computer
8. Person flying a drone
9. Person watching a movie on TV
10. Watching a movie on TV
11. Beginner using a computer
12. Person immersed in phone use
13. Technology expert using phone
14. Young adult interacting with phone
15. Young adult interacting with multiple technological devices
16. Young adult wearing headphones and smiling at phone
17. Disengaged person with crossed arms
18. Person that shrugs
19. Indifferent person shrugging
20. Moderately engaged person
21. Smiling person looking forward to something fun
22. Person that is neutral towards their next task
23. Representation of task that needs to be done
24. Person playing video games with a good set-up
25. Person playing video games on a large screen
26. Person calling a friend with their phone
27. Person watching a movie on a computer
28. Person talking on the phone
29. Content person
30. Focused person
31. Picture of person that is excited and looking forward to what is next

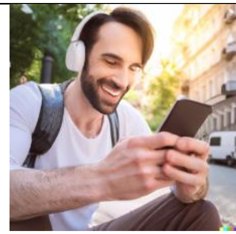
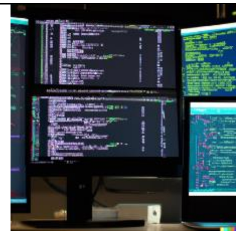
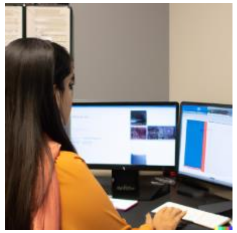
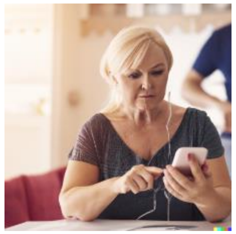
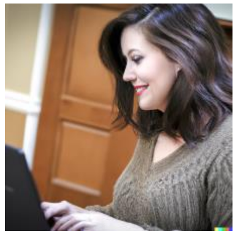
32. Person waving off because of annoyment
33. Neutral person
34. Closed laptop with frustrated person
35. Satisfied person using a computer
36. Person in trance using a computer
37. Person throwing computer away
38. Person throwing computer on the floor
39. Person purposely throwing computer on the floor
40. Person purposely stepping on a laptop
41. Geek being super excited to use computer
42. IT expert being curious about coding
43. Person that does not care about technology at all
44. Person putting computer in the trash
45. Bored and tired person
46. Highly curious person using computer
47. Highly curious person looking at computer code
48. Highly curious person
49. Highly curious person looking at the inside of computer
50. Inside of computer
51. Person that does not care at all
52. Person walking away from computer
53. Person using VR
54. Fascinated person
55. Fascinated person using computer
56. Person without interest in computer
57. Yawning person
58. Impressive technical set-up with high-tech tools
59. Maximise the potential of computer
60. Large computer server room
61. Large computer server room with green and red lights
62. Computer with multiple screens, mobile phone and VR glasses
63. Person with large interest in new technology
64. Person being angry that they cannot get technology to work
65. Person living a life without any new technology

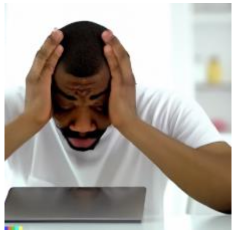
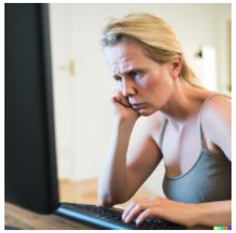
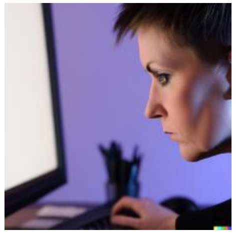
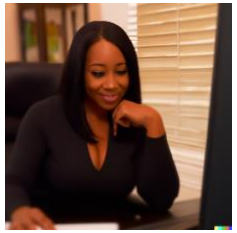
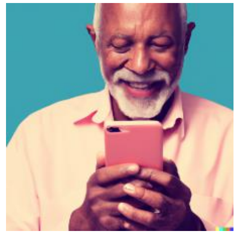
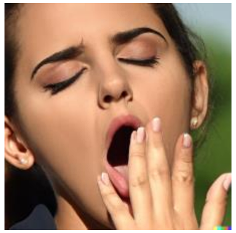
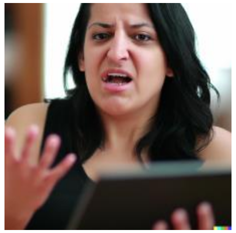
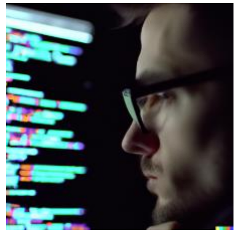
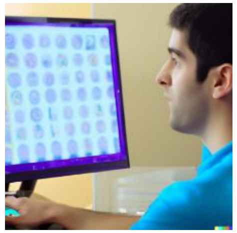
66. Highly curious person using technology
67. Moderately happy person using technology
68. Hacker with multiple screens, shot from behind
69. Person giving up on computer use
70. Neutral person using computer
71. Person smiling at phone and being fully immersed in use
72. Person happy that there were able to solve a task on their computer
73. Hand working on the inside of computer
74. Phone with a broken screen
75. Maximising potential of technical tool
76. Grandma using phone and having problems
77. Advanced computer use
78. Advanced computer user
79. Data visualisation on computer
80. Person working on data visualisation on computer
81. Person without much interest in their computer
82. Happy IT user in front of multiple screens
83. Person in flow working on technology
84. Happy person celebrating success
85. Person calling friend on old phone
86. Old man using phone to call a friend
87. Person working with artificial intelligence
88. Person working with two computer screens
89. Person working with two computer screens, from side
90. Advanced computer user interacting with software
91. Women using touchscreen
92. Touchscreen user
93. Advanced computer user with touchscreen
94. Casual computer user
95. Person on computer losing focus
96. Unhappy person using computer
97. Person smiling at phone screen
98. Person smirking in front of computer

Appendix E

Items in the Image-Based Affinity for Technology Interaction Scale

The response options for the scale range from 1 (Low) to 5 (High).

	Item	Response Option 1	Response Option 2	Response Option 3	Response Option 4	Response Option 5
Q1	Please select the image that best represents **your interest** in new technical systems:					
Q2	Please select the image that best represents **your engagement** with testing the functions of new technical systems:					
Q3	Please select the image that best represents **your motivation** for dealing with technical systems:					
Q4	Please select the image that best represents **your approach** to trying out a new technical system:					

Q5	Please select the image that best represents **your enjoyment** in becoming acquainted with a new technical system:					
Q6	Please select the image that best represents **your attitude** towards the inner workings of a technical system:					
Q7	Please select the image that best represents **your curiosity** about understanding how a technical system works:					
Q8	Please select the image that best represents **your level of interest** in fully exploring and utilising the capabilities of a technical system:					
Q9	Please select the image that best represents **your intention to maximise the potential** of a technical system:					

Appendix F

Items in the Animal Attitudes Scale-5

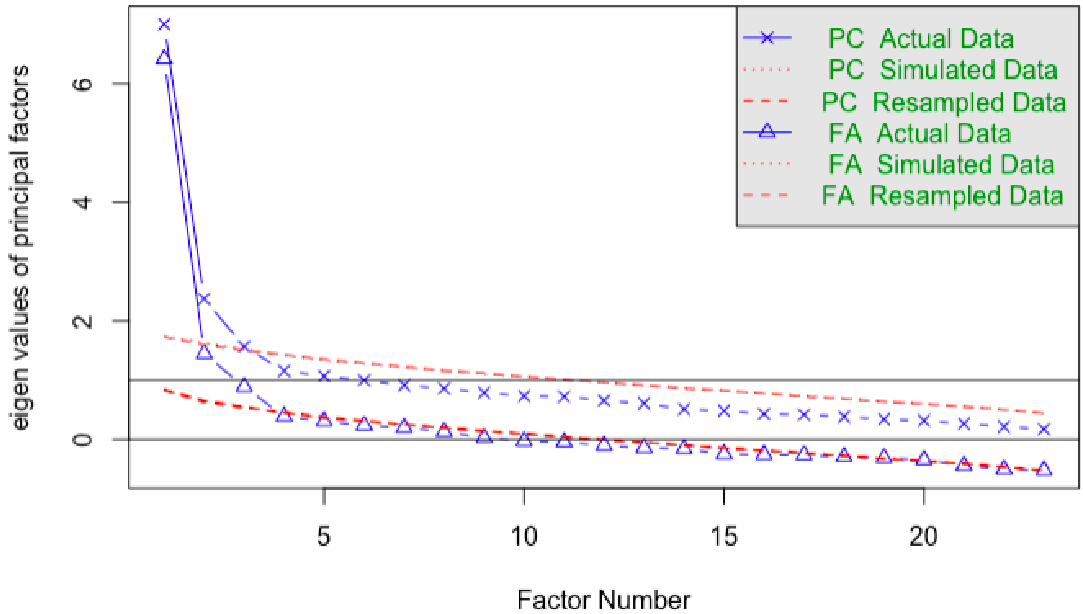
The response options for the AAS-5 range from 1 (Strongly Disagree) to 5 (Strongly Agree). Items that are indicated with (*) are reverse scored.

1. It is morally wrong to hunt wild animals just for sport.
2. I do not think that there is anything wrong with using animals in medical research. (*)
3. I think it is perfectly acceptable for cattle and hogs to be raised for human consumption. (*)
4. The slaughter of whales and dolphins should be immediately stopped even if it means some people will be put out of work.
5. I sometimes get upset when I see wild animals in cages at zoos.

Appendix G

Parallel Analysis

A parallel analysis was conducted to confirm the number of relevant factors to include into the final model. Thus, the eigenvalues in the actual data were compared to the eigenvalues of 20 simulated parallel matrices and mapped in a scree plot as suggested by Revelle (n.d.). The quantile for the eigenvalues in the simulated matrices was set to 0.95. The number of extracted factors ultimately depends on where the eigenvalues of the actual data diverge from the eigenvalues of the simulated parallel data. Consequently, the parallel analysis indicated to include three factors into the model. The eigenvalues of these factors were 6.47, 1.45 and 0.87. The eigenvalues for all remaining factors combined were less than 1.25, which suggests that three factors are sufficient to explain the variance in the results.



Appendix H

Frequency of Selected Response Options

Item	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6
Image-Based Affinity for Technology Interaction (IBATI)						
IBATI1	4	14	30	95	35	
IBATI2	15	38	25	61	39	
IBATI3	9	49	47	40	33	
IBATI4	9	20	52	84	13	
IBATI5	8	23	38	51	58	
IBATI6	48	31	44	26	29	
IBATI7	29	22	70	32	25	
IBATI8	10	19	67	31	51	
IBATI9	9	16	79	54	20	
Affinity for Technology Interaction (ATI)						
ATI1	5	28	34	53	44	14
ATI2	4	12	26	60	62	14
ATI3	8	28	43	55	34	10
ATI4	5	15	35	65	48	10
ATI5	5	16	42	52	52	11
ATI6	8	24	47	36	42	21
ATI7	9	42	34	58	30	5
ATI8	7	21	40	59	45	6
ATI9	6	15	27	70	52	8
Animal Attitude Scale (AAS)						
AAS1	2	8	9	61	98	
AAS2	35	70	37	32	4	
AAS3	23	44	56	47	8	
AAS4	3	4	15	65	91	
AAS5	4	29	23	77	45	

Note. IBATI and AAS only consist of five response options per item.

Appendix I

Item-Total Correlation Matrices

	IBATI1	IBATI2	IBATI3	IBATI4	IBATI5	IBATI6	IBATI7	IBATI8	IBATI9
IBATI	0.37	0.44	0.52	0.37	0.44	0.49	0.59	0.45	0.52
ATI	0.27	0.41	0.56	0.31	0.46	0.53	0.54	0.36	0.53
AAS	0.05	0.05	-0.01	0.01	-0.04	-0.14	-0.20	0.04	-0.20

Note. Values in the IBATI row are corrected item-total correlations.

	ATI1	ATI2	ATI3	ATI4	ATI5	ATI6	ATI7	ATI8	ATI9
IBATI	0.61	0.55	0.19	0.50	0.64	0.54	0.60	0.51	0.48
ATI	0.70	0.69	0.17	0.67	0.75	0.59	0.68	0.58	0.59
AAS	-0.19	-0.08	-0.16	-0.05	-0.25	-0.27	-0.16	-0.14	-0.09

Note. Values in the ATI row are corrected item-total correlations.

	AAS1	AAS2	AAS3	AAS4	AAS5
IBATI	-0.03	-0.01	-0.16	0.03	-0.07
ATI	-0.10	-0.10	-0.24	-0.04	-0.24
AAS	0.42	0.42	0.51	0.39	0.44

Note. Values in the AAS row are corrected item-total correlations.

Appendix J

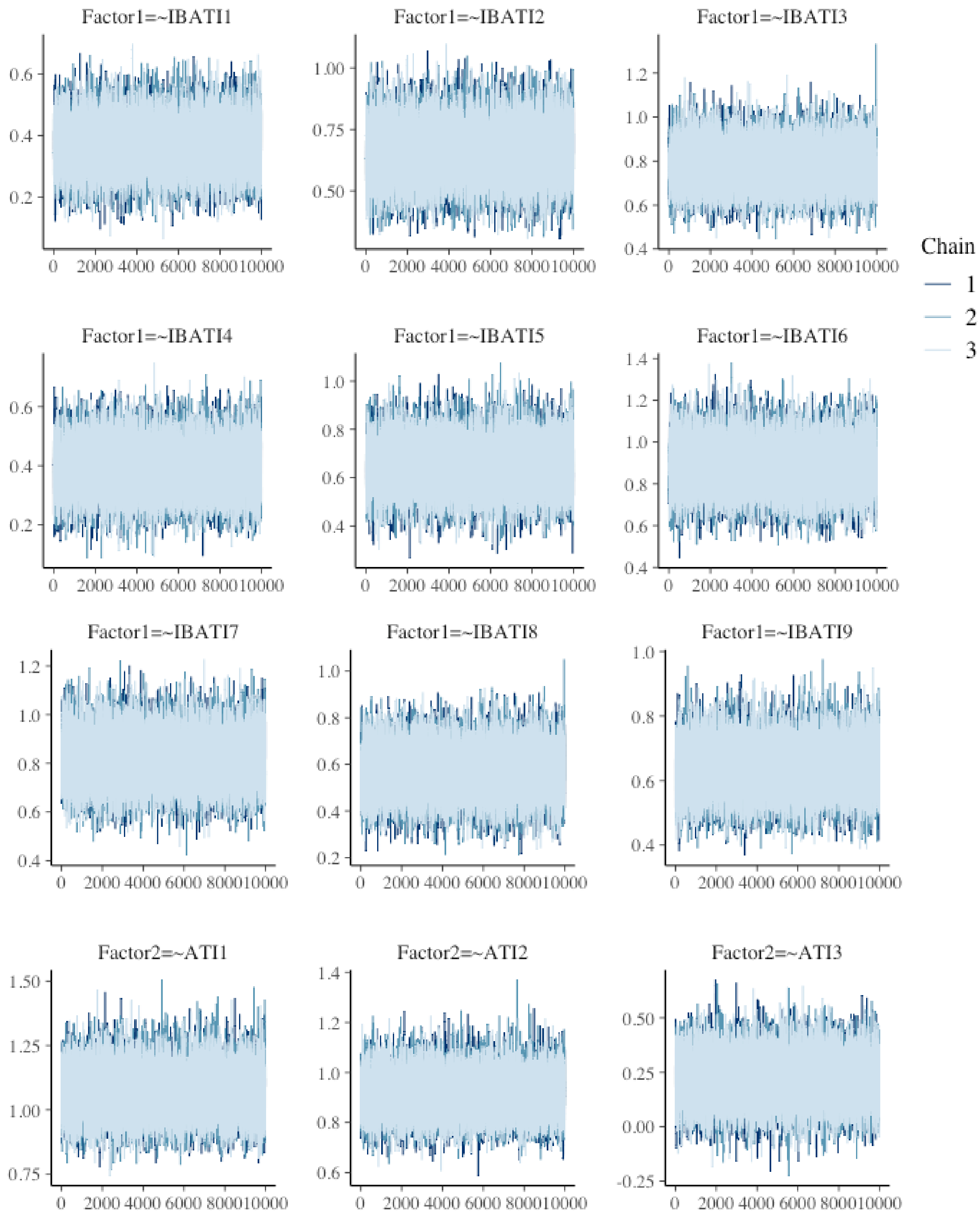
3-Factor Model Factor Loadings Table

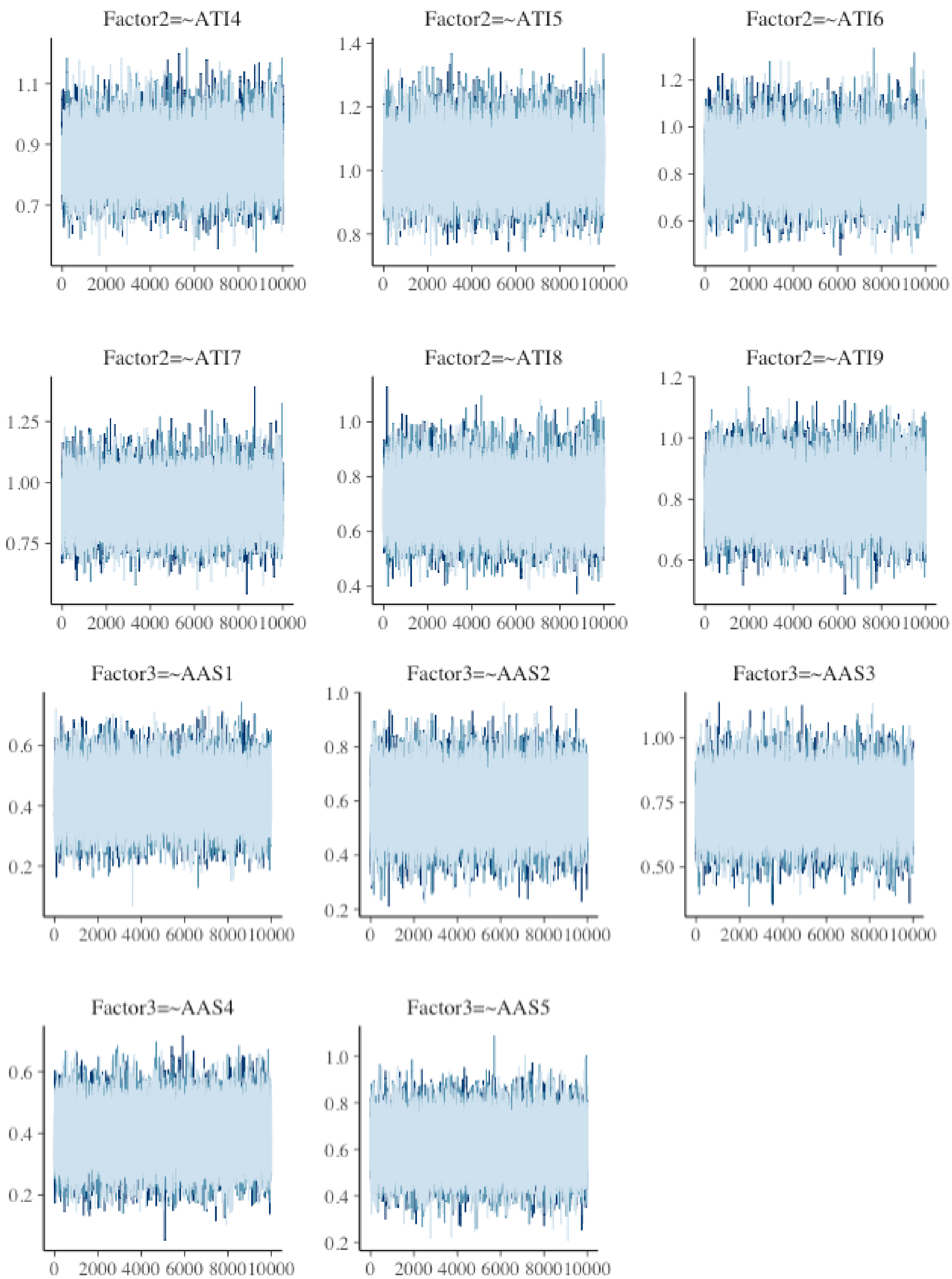
Factor	Item	Centre	Lower	Upper
Factor 1	IBATI1	0.37	0.22	0.52
	IBATI2	0.66	0.46	0.87
	IBATI3	0.78	0.61	0.96
	IBATI4	0.39	0.24	0.55
	IBATI5	0.65	0.46	0.83
	IBATI6	0.88	0.67	1.11
	IBATI7	0.82	0.64	1.01
	IBATI8	0.57	0.39	0.76
	IBATI9	0.64	0.50	0.79
Factor 2	ATI1	1.07	0.90	1.25
	ATI2	0.91	0.76	1.08
	ATI3	0.22	0.02	0.42
	ATI4	0.86	0.70	1.02
	ATI5	1.02	0.87	1.19
	ATI6	0.84	0.64	1.06
	ATI7	0.91	0.74	1.09
	ATI8	0.71	0.54	0.90
	ATI9	0.81	0.65	0.98
Factor 3	AAS1	0.43	0.28	0.59
	AAS2	0.59	0.40	0.78
	AAS3	0.74	0.55	0.94
	AAS4	0.40	0.24	0.55
	AAS5	0.62	0.43	0.81

Note. Centre indicates the estimated factor loading for each item. Lower and Upper indicate the 95% credibility interval for each item.

Appendix K

Trace Plots of Markov Chain Monte Carlo Sampling per Item





Note. The y-axis indicates the standardised factor loadings and the x-axis the number of iterations per chain.

Appendix L

R Syntax

```
library(readxl)
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────
── tidyverse 1.3.2 ──
## ✓ ggplot2 3.4.0     ✓ purrr   1.0.1
## ✓ tibble  3.2.1     ✓ dplyr   1.1.2
## ✓ tidyr   1.3.0     ✓ stringr 1.5.0
## ✓ readr   2.1.3     ✓ forcats 0.5.2
## ── Conflicts ──────────────────────────────────────────── ti
dyverse_conflicts() ──
## ✕ dplyr::filter() masks stats::filter()
## ✕ dplyr::lag()    masks stats::lag()

library(lavaan)

## This is lavaan 0.6-15
## lavaan is FREE software! Please report any bugs.

library(ggplot2)
library(bayr)

##
## Attaching package: 'bayr'
##
## The following object is masked from 'package:tidyr':
##
##     expand_grid

library(dplyr)
library(stats)
library(psych)

##
## Attaching package: 'psych'
##
## The following object is masked from 'package:lavaan':
##
##     cor2cov
##
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha

library(blavaan)

## Loading required package: Rcpp
## This is blavaan 0.4-8
```

```
## On multicore systems, we suggest use of future::plan("multicore") or
##   future::plan("multisession") for faster post-MCMC computations.
```

```r
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
library(loo)
```

```
## This is loo version 2.6.0
## - Online documentation and vignettes at mc-stan.org/loo
## - As of v2.0.0 loo defaults to 1 core but we recommend using as many as possible. Use the 'cores' argument or set options(mc.cores = NUM_CORES) for an entire session.
```

```r
library(Hmisc)
```

```
##
## Attaching package: 'Hmisc'
##
## The following object is masked from 'package:psych':
##
##     describe
##
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
##
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```r
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```r
IBT <- read_excel("Image-Based Affinity for Technology Measure_July 18, 2023_04.52.xlsx
")
```

```
## New names:
## • `Q22` -> `Q22...20`
## • `Q22` -> `Q22...24`
```

```r
names(IBT) <- gsub("_", "", names(IBT))
names(IBT) <- gsub("Q", "IBATI", names(IBT))

# Focus on main variables to exclude metadata
relevant_columns <- c("Finished" ,"Gender", "Nationality", "Age", "IBATI1", "IBATI2", "IB
ATI3","IBATI4", "IBATI5","IBATI6","IBATI7", "IBATI8","IBATI9","ATI1","ATI2","ATI3
","ATI4","ATI5","ATI6","ATI7","ATI8","ATI9","AAS1","AAS2","AAS3","AAS4","AAS5"
)

# Create IBT_CLEAN by selecting the relevant columns from IBT without metadata
IBT_CLEAN <- subset(IBT, select = relevant_columns)
```

```
# Remove rows with "Finished" equal to 0 and add variable for participant number
IBT_CLEAN <- IBT_CLEAN[complete.cases(IBT_CLEAN) & IBT_CLEAN$Finished != 0,
]
IBT_CLEAN <- IBT_CLEAN[-1,]
IBT_CLEAN$Participant_Number <- 1:nrow(IBT_CLEAN)

# Mutate all test items as numeric for analysis
numeric_cols <- c("IBATI1", "IBATI2", "IBATI3", "IBATI4", "IBATI5", "IBATI6", "IBATI
7", "IBATI8", "IBATI9", "ATI1", "ATI2", "ATI3", "ATI4", "ATI5", "ATI6", "ATI7", "ATI8"
, "ATI9", "AAS1", "AAS2", "AAS3", "AAS4", "AAS5")
IBT_CLEAN <- IBT_CLEAN %>%
  mutate_at(vars(all_of(numeric_cols)), as.numeric)%>%
  mutate(Participant_Number = as.character(Participant_Number))
head(IBT_CLEAN)

## # A tibble: 6 × 28
##   Finished Gender Nationality Age   IBATI1 IBATI2 IBATI3 IBATI4 IBATI5 IBATI6
##   <chr>    <chr>  <chr>       <chr>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 1.0      1.0    2.0         23         5      5      4      3      5      1
## 2 1.0      2.0    2.0         24         4      4      2      4      5      3
## 3 1.0      1.0    2.0         24         5      5      4      5      5      4
## 4 1.0      1.0    2.0         24         3      4      2      3      3      1
## 5 1.0      1.0    2.0         61         4      4      3      4      5      3
## 6 1.0      1.0    2.0         29         3      4      4      4      5      3
## # ℹ 18 more variables: IBATI7 <dbl>, IBATI8 <dbl>, IBATI9 <dbl>, ATI1 <dbl>,
## #   ATI2 <dbl>, ATI3 <dbl>, ATI4 <dbl>, ATI5 <dbl>, ATI6 <dbl>, ATI7 <dbl>,
## #   ATI8 <dbl>, ATI9 <dbl>, AAS1 <dbl>, AAS2 <dbl>, AAS3 <dbl>, AAS4 <dbl>,
## #   AAS5 <dbl>, Participant_Number <chr>

# Five items were reverse coded, so their scoring had to be changed to match the other items.
Because of disparities in the number of response options, the answer on the ATI was calculate
d with a different number that the AAS.
IBT_CLEAN$ATI3 <- 7 - IBT_CLEAN$ATI3
IBT_CLEAN$ATI6 <- 7 - IBT_CLEAN$ATI6
IBT_CLEAN$ATI8 <- 7 - IBT_CLEAN$ATI8
IBT_CLEAN$AAS2 <- 6 - IBT_CLEAN$AAS2
IBT_CLEAN$AAS3 <- 6 - IBT_CLEAN$AAS3

# Because of the varying number of response options, scores were rescaled to fit between 0 an
d 1.
IBT_CLEAN_normalised <- IBT_CLEAN %>%
  mutate(across(all_of(numeric_cols), scales::rescale))
head(IBT_CLEAN_normalised)

## # A tibble: 6 × 28
##   Finished Gender Nationality Age   IBATI1 IBATI2 IBATI3 IBATI4 IBATI5 IBATI6
##   <chr>    <chr>  <chr>       <chr>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 1.0      1.0    2.0         23      1      1      0.75   0.5    1      0
## 2 1.0      2.0    2.0         24      0.75   0.75   0.25   0.75   1      0.5
## 3 1.0      1.0    2.0         24      1      1      0.75   1      1      0.75
## 4 1.0      1.0    2.0         24      0.5    0.75   0.25   0.5    0.5    0
```

```
## 5 1.0     1.0   2.0        61      0.75   0.75   0.5    0.75    1     0.5
## 6 1.0     1.0   2.0        29      0.5    0.75   0.75   0.75    1     0.5
## # ℹ 18 more variables: IBATI7 <dbl>, IBATI8 <dbl>, IBATI9 <dbl>, ATI1 <dbl>,
## #   ATI2 <dbl>, ATI3 <dbl>, ATI4 <dbl>, ATI5 <dbl>, ATI6 <dbl>, ATI7 <dbl>,
## #   ATI8 <dbl>, ATI9 <dbl>, AAS1 <dbl>, AAS2 <dbl>, AAS3 <dbl>, AAS4 <dbl>,
## #   AAS5 <dbl>, Participant_Number <chr>
```

```
#Obtain Frequencies, Max, Min, Mean and SD for all variables. (Table 1 & 2, Appendix H)
descriptive_stats <- describe(IBT_CLEAN_normalised)
print(descriptive_stats)
```

```
## IBT_CLEAN_normalised
##
##  28  Variables      178  Observations
## --------------------------------------------------------------------------------
## Finished
##        n  missing distinct    value
##      178        0        1      1.0
##
## Value        1
## Frequency  178
## Proportion   1
## --------------------------------------------------------------------------------
## Gender
##        n  missing distinct
##      178        0        3
##
## Value          1     2     3
## Frequency     73   103     2
## Proportion 0.410 0.579 0.011
## --------------------------------------------------------------------------------
## Nationality
##        n  missing distinct
##      178        0        5
##
## Value          1     2     3     4     5
## Frequency     36   109    22    10     1
## Proportion 0.202 0.612 0.124 0.056 0.006
## --------------------------------------------------------------------------------
## Age
##        n  missing distinct
##      178        0       23
##
## lowest : 18 19 20 21 22, highest: 56 57 59 60 61
## --------------------------------------------------------------------------------
## IBATI1
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.835   0.7008   0.2352
##
## Value       0.00  0.25  0.50  0.75  1.00
## Frequency      4    14    30    95    35
## Proportion 0.022 0.079 0.169 0.534 0.197
```

```
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## IBATI2
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.936   0.5997   0.3538
##
## Value       0.00  0.25  0.50  0.75  1.00
## Frequency     15    38    25    61    39
## Proportion 0.084 0.213 0.140 0.343 0.219
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## IBATI3
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.943   0.5548   0.3322
##
## Value       0.00  0.25  0.50  0.75  1.00
## Frequency      9    49    47    40    33
## Proportion 0.051 0.275 0.264 0.225 0.185
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## IBATI4
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.868   0.6011   0.2514
##
## Value       0.00  0.25  0.50  0.75  1.00
## Frequency      9    20    52    84    13
## Proportion 0.051 0.112 0.292 0.472 0.073
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## IBATI5
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5     0.93   0.6798   0.3237
##
## Value       0.00  0.25  0.50  0.75  1.00
## Frequency      8    23    38    51    58
## Proportion 0.045 0.129 0.213 0.287 0.326
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## IBATI6
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.953   0.4396   0.3991
##
## Value       0.00  0.25  0.50  0.75  1.00
## Frequency     48    31    44    26    29
## Proportion 0.270 0.174 0.247 0.146 0.163
##
```

```
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## IBATI7
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.924   0.5028   0.3415
##
## Value       0.00  0.25  0.50  0.75  1.00
## Frequency     29    22    70    32    25
## Proportion 0.163 0.124 0.393 0.180 0.140
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## IBATI8
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.917    0.632    0.323
##
## Value       0.00  0.25  0.50  0.75  1.00
## Frequency     10    19    67    31    51
## Proportion 0.056 0.107 0.376 0.174 0.287
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## IBATI9
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.882   0.5843   0.2571
##
## Value       0.00  0.25  0.50  0.75  1.00
## Frequency      9    16    79    54    20
## Proportion 0.051 0.090 0.444 0.303 0.112
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## ATI1
##        n  missing distinct     Info     Mean      Gmd
##      178        0        6    0.947   0.5629   0.2837
##
## Value        0.0   0.2   0.4   0.6   0.8   1.0
## Frequency      5    28    34    53    44    14
## Proportion 0.028 0.157 0.191 0.298 0.247 0.079
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## ATI2
##        n  missing distinct     Info     Mean      Gmd
##      178        0        6    0.916   0.6315   0.2418
##
## Value        0.0   0.2   0.4   0.6   0.8   1.0
## Frequency      4    12    26    60    62    14
## Proportion 0.022 0.067 0.146 0.337 0.348 0.079
##
## For the frequency table, variable is rounded to the nearest 0.01
```

```
## ---------------------------------------------------------------------------
## ATI3
##        n  missing distinct     Info     Mean      Gmd
##      178        0        6    0.945   0.4775   0.2776
##
## Value       0.0   0.2   0.4   0.6   0.8   1.0
## Frequency     10    34    55    43    28     8
## Proportion 0.056 0.191 0.309 0.242 0.157 0.045
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## ATI4
##        n  missing distinct     Info     Mean      Gmd
##      178        0        6    0.923   0.5865   0.2467
##
## Value       0.0   0.2   0.4   0.6   0.8   1.0
## Frequency      5    15    35    65    48    10
## Proportion 0.028 0.084 0.197 0.365 0.270 0.056
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## ATI5
##        n  missing distinct     Info     Mean      Gmd
##      178        0        6    0.936   0.5831   0.2601
##
## Value       0.0   0.2   0.4   0.6   0.8   1.0
## Frequency      5    16    42    52    52    11
## Proportion 0.028 0.090 0.236 0.292 0.292 0.062
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## ATI6
##        n  missing distinct     Info     Mean      Gmd
##      178        0        6    0.956   0.4393   0.3097
##
## Value       0.0   0.2   0.4   0.6   0.8   1.0
## Frequency     21    42    36    47    24     8
## Proportion 0.118 0.236 0.202 0.264 0.135 0.045
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## ATI7
##        n  missing distinct     Info     Mean      Gmd
##      178        0        6     0.94    0.482   0.2764
##
## Value       0.0   0.2   0.4   0.6   0.8   1.0
## Frequency      9    42    34    58    30     5
## Proportion 0.051 0.236 0.191 0.326 0.169 0.028
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
```

```
## ATI8
##        n  missing distinct     Info     Mean      Gmd
##      178        0        6    0.934   0.4517   0.2588
##
## Value       0.0   0.2   0.4   0.6   0.8   1.0
## Frequency     6    45    59    40    21     7
## Proportion 0.034 0.253 0.331 0.225 0.118 0.039
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## ATI9
##        n  missing distinct     Info     Mean      Gmd
##      178        0        6     0.91   0.5921   0.2413
##
## Value       0.0   0.2   0.4   0.6   0.8   1.0
## Frequency     6    15    27    70    52     8
## Proportion 0.034 0.084 0.152 0.393 0.292 0.045
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## AAS1
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.793   0.8441   0.2046
##
## Value      0.00  0.25  0.50  0.75  1.00
## Frequency     2     8     9    61    98
## Proportion 0.011 0.045 0.051 0.343 0.551
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## AAS2
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.917   0.6404   0.2932
##
## Value      0.00  0.25  0.50  0.75  1.00
## Frequency     4    32    37    70    35
## Proportion 0.022 0.180 0.208 0.393 0.197
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## AAS3
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.933   0.5379   0.3035
##
## Value      0.00  0.25  0.50  0.75  1.00
## Frequency     8    47    56    44    23
## Proportion 0.045 0.264 0.315 0.247 0.129
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## AAS4
```

```
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.817   0.8329   0.2074
##
## Value       0.00  0.25  0.50  0.75  1.00
## Frequency      3     4    15    65    91
## Proportion 0.017 0.022 0.084 0.365 0.511
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## AAS5
##        n  missing distinct     Info     Mean      Gmd
##      178        0        5    0.896   0.6826   0.2904
##
## Value       0.00  0.25  0.50  0.75  1.00
## Frequency      4    29    23    77    45
## Proportion 0.022 0.163 0.129 0.433 0.253
##
## For the frequency table, variable is rounded to the nearest 0.01
## ---------------------------------------------------------------------------
## Participant_Number
##        n  missing distinct
##      178        0      178
##
## lowest : 1   10  100 101 102, highest: 95  96  97  98  99
## ---------------------------------------------------------------------------
```

```
statistics <- sapply(IBT_CLEAN_normalised, function(x) c(Mean = mean(x), SD = sd(x)))
```

```
## Warning in mean.default(x): argument is not numeric or logical: returning NA

## Warning in mean.default(x): argument is not numeric or logical: returning NA

## Warning in mean.default(x): argument is not numeric or logical: returning NA

## Warning in mean.default(x): argument is not numeric or logical: returning NA

## Warning in mean.default(x): argument is not numeric or logical: returning NA
```

```
print(statistics)
```

```
##      Finished    Gender Nationality      Age    IBATI1    IBATI2    IBATI3
## Mean       NA        NA          NA       NA 0.7008427 0.5997191 0.5547753
## SD          0 0.5135441   0.7758536 9.442104 0.2302341 0.3181646 0.2961947
##         IBATI4    IBATI5    IBATI6    IBATI7    IBATI8    IBATI9      ATI1
## Mean 0.6011236 0.6797753 0.4396067 0.5028090 0.6320225 0.5842697 0.5629213
## SD   0.2398184 0.2947008 0.3538591 0.3087573 0.2937433 0.2419431 0.2537324
##           ATI2      ATI3      ATI4      ATI5      ATI6      ATI7      ATI8
## Mean 0.6314607 0.4775281 0.5865169 0.5831461 0.4393258 0.4820225 0.4516854
## SD   0.2247889 0.2489539 0.2265891 0.2356556 0.2747852 0.2475013 0.2348758
##           ATI9      AAS1      AAS2      AAS3      AAS4      AAS5
## Mean 0.5921348 0.8441011 0.6404494 0.5379213 0.8328652 0.6825843
## SD   0.2253558 0.2157146 0.2669536 0.2728954 0.2137933 0.2703943
```

```
##      Participant_Number
## Mean                 NA
## SD             51.52831

#A correlation matrix with p-values was printed for all items (Appendix I) and the correlation heat map is plotted. (Figure 2)
numeric_cols <- sapply(IBT_CLEAN_normalised, is.numeric)
numeric_data <- IBT_CLEAN_normalised[, numeric_cols]

corCi(numeric_data, keys = NULL, n.iter = 1000,  p = 0.05,overlap = FALSE,
 poly = FALSE, method = "pearson", plot=TRUE,minlength=5,n=NULL, cex.axis=0.7, las=2,
cex = 0.38)

#%>%select(- MeanIBATI,- MeanATI,- MeanAAS)
#cor_matrix <- cor(numeric_data, method = "pearson")
#corrplot(cor_matrix, method = "color", tl.col = 'black', tl.cex = 0.8)
#rcorr(as.matrix(numeric_data),type="pearson")

# Reshape data to long format and plot box plots for answer distribution (Figure 1)
data_long <- reshape2::melt(numeric_data)

## No id variables; using all as measure variables

ggplot(data_long, aes(x = variable, y = value)) +
  geom_boxplot(fill = "lightblue", color = "darkblue", alpha = 0.6) +
  xlab("Item") +
  ylab("Value") +
  ggtitle("Box Plots") +
  theme_minimal(base_size = 14) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1),
        plot.title = element_text(hjust = 0.5))

# The corrected item-total correlation was calculated to observe if all items were in line with t
he total score and contributed to the score. The loop was used to correct the correlation by re
moving the item itself from the total score when calculating the correlation to the total score.

calculate_correlations <- function(data, start_col, end_col) {
  total_col_name <- paste0("total_", start_col, "_", end_col)
  data[[total_col_name]] <- rowSums(data[, start_col:end_col])
  corrected_totals <- data[[total_col_name]] - data[, start_col:end_col]

# Prepare matrices to store correlations and p-values
  n_items <- end_col - start_col + 1
  correlations <- numeric(n_items)
  p_values <- numeric(n_items)

for (i in start_col:end_col) {
    cor_test <- cor.test(data[[i]], corrected_totals[, i - start_col + 1])
    correlations[i - start_col + 1] <- cor_test$estimate
    p_values[i - start_col + 1] <- cor_test$p.value}
  return(list(correlations = correlations, p_values = p_values))
  }
```

```
# Calculate for each set of items
resultsIBATI <- calculate_correlations(numeric_data, 1, 9)
resultsATI <- calculate_correlations(numeric_data, 10, 18)
resultsAAS <- calculate_correlations(numeric_data, 19, 23)
resultsIBATI$correlations

## [1] 0.3674679 0.4435161 0.5229146 0.3659504 0.4355822 0.4907037 0.5924469
## [8] 0.4478190 0.5228350

resultsIBATI$p_values

## [1] 4.517372e-07 5.638149e-10 7.011051e-14 5.078080e-07 1.227331e-09
## [6] 3.566227e-12 3.016063e-18 3.666470e-10 7.083154e-14

resultsATI$correlations

## [1] 0.7019992 0.6888021 0.1721392 0.6659778 0.7543830 0.5917182 0.6810842
## [8] 0.5796394 0.5938612

resultsATI$p_values

## [1] 9.667503e-28 2.248882e-26 2.158336e-02 3.558101e-24 5.306894e-34
## [6] 3.394221e-18 1.312057e-25 2.304368e-17 2.396115e-18

resultsAAS$correlations

## [1] 0.4186504 0.4199174 0.5095849 0.3924121 0.4386213

resultsAAS$p_values

## [1] 6.045385e-09 5.381705e-09 3.747409e-13 6.046217e-08 9.132397e-10

ImageBased <- c("IBATI1","IBATI2","IBATI3","IBATI4","IBATI5", "IBATI6", "IBATI7","
IBATI8", "IBATI9")
ATI <- c("ATI1","ATI2","ATI3","ATI4","ATI5","ATI6","ATI7","ATI8","ATI9")
AAS <- c("AAS1","AAS2","AAS3","AAS4","AAS5")

# Calculate total score & number of items per scale
IBT_CLEAN_normalised$IBATITotal <- rowSums(IBT_CLEAN_normalised[, ImageBased]
, na.rm = TRUE)
IBT_CLEAN_normalised$ATITotal <- rowSums(IBT_CLEAN_normalised[, ATI], na.rm =
TRUE)
IBT_CLEAN_normalised$AASTotal <- rowSums(IBT_CLEAN_normalised[, AAS], na.rm =
TRUE)
num_IBATI <- length(ImageBased)
num_ATI <-length(ATI)
num_AAS <-length(AAS)

# Calculate mean score by dividing total score by number of items as not all scales had the sa
me number of items.
IBT_CLEAN_normalised$MeanIBATI <- IBT_CLEAN_normalised$IBATITotal / num_IBA
TI
IBT_CLEAN_normalised$MeanATI <- IBT_CLEAN_normalised$ATITotal / num_ATI
IBT_CLEAN_normalised$MeanAAS <- IBT_CLEAN_normalised$AASTotal / num_AAS
```

```
# Calculate Cronbach's alpha for each scale
alpha1 <- psych::alpha(IBT_CLEAN_normalised[, ImageBased])$total$raw_alpha
alpha2 <- psych::alpha(IBT_CLEAN_normalised[, ATI])$total$raw_alpha
alpha3 <- psych::alpha(IBT_CLEAN_normalised[, AAS])$total$raw_alpha
print(paste("Alpha for Image Based Test:", alpha1))

## [1] "Alpha for Image Based Test: 0.779725380107107"

print(paste("Alpha for ATI:", alpha2))

## [1] "Alpha for ATI: 0.865244205961746"

print(paste("Alpha for AAS:", alpha3))

## [1] "Alpha for AAS: 0.679667792645351"

#Reshape total scores to long format
FinalScores <- IBT_CLEAN_normalised %>%
  select(MeanIBATI,MeanATI,MeanAAS)
Final_long <- reshape2::melt(FinalScores)

## No id variables; using all as measure variables

# Violin plot displays the distribution of scores for all scales. (Figure 3)
ggplot(Final_long, aes(x = variable, y = value)) +
  geom_violin(fill = "lightblue", color = "darkblue", alpha = 0.7) +
  xlab("Scale") +
  ylab("Mean Score") +
  ggtitle("Violin Plots") +
  theme_minimal(base_size = 14) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1),
        plot.title = element_text(hjust = 0.5))

# Scatterplot of Mean Scores on ATI and IBATI (Figure 4)
IBT_CLEAN_normalised %>%
  ggplot(aes(x = MeanIBATI, y = MeanATI)) +
  geom_point(color="black" ,size = 1.5) +
  geom_smooth(method = "lm", color = "blue") +
  labs(x = "Mean IBATI Score", y = "Mean ATI Score") +
  coord_fixed(ratio = 1) +
  xlim(0, 1)+
  ylim(0,1)

## `geom_smooth()` using formula = 'y ~ x'

# Conduct parallel analysis and obtain eigenvalues in line with Revelle (n.d.)
parallel_result <- fa.parallel(IBT_CLEAN[, numeric_cols], nfactors = 1,
                               main = "Parallel Analysis Scree Plots",
                               n.iter = 20, error.bars = FALSE, se.bars = FALSE,
                               SMC = FALSE, ylabel = NULL, show.legend = TRUE,
                               sim = TRUE, quant = 0.95, cor = "cor",
                               use = "pairwise", plot = FALSE, correct = 0.5, sqrt = FALSE)
```

```
## Parallel analysis suggests that the number of factors =  3  and the number of components =
2

# Plot scree plot with parallel analysis results (Appendix G)
plot(parallel_result, show.legend = TRUE, fa = "both")

# Define the BCFA model syntax in line with Merkle & Rosseel (2015)
bcfa_model3 <- "
  # Specify the latent factors
  Factor1 =~  IBATI1 + IBATI2 + IBATI3 + IBATI4 + IBATI5 + IBATI6 + IBATI7 + IBAT
I8 + IBATI9
  Factor2 =~ ATI1 + ATI2 + ATI3 + ATI4 + ATI5 + ATI6 + ATI7 + ATI8 + ATI9
  Factor3 =~ AAS1 + AAS2 + AAS3 + AAS4 + AAS5

  # Specify covariances between latent factors
  Factor1 ~~ Factor2
  Factor2 ~~ Factor3
  Factor1 ~~ Factor3
"
# Fit the BCFA model with 3 chains of MCMC Sampling and Standardise factor loadings
bcfa_result3 <- bcfa(bcfa_model3, data=IBT_CLEAN, mcmcfile = T, std.lv=TRUE, cp = "srs
",
    dp = NULL, n.chains = 3, burnin = 5000, adapt = 1000, sample = 10000, inits = "simple",
    convergence = "manual", target = "stan", save.lvs = TRUE,
    wiggle = NULL, wiggle.sd = 0.1, prisamp = FALSE, jags.ic = FALSE,
    seed = 15, bcontrol = list())

##
## SAMPLING FOR MODEL 'stanmarg' NOW (CHAIN 1).
## Chain 1:
## Chain 1: Gradient evaluation took 0.002918 seconds
## Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 29.18 seconds.
## Chain 1: Adjust your expectations accordingly!
## Chain 1:
## Chain 1:
## Chain 1: Iteration:     1 / 15000 [  0%]  (Warmup)
## Chain 1: Iteration:  1500 / 15000 [ 10%]  (Warmup)
## Chain 1: Iteration:  3000 / 15000 [ 20%]  (Warmup)
## Chain 1: Iteration:  4500 / 15000 [ 30%]  (Warmup)
## Chain 1: Iteration:  5001 / 15000 [ 33%]  (Sampling)
## Chain 1: Iteration:  6500 / 15000 [ 43%]  (Sampling)
## Chain 1: Iteration:  8000 / 15000 [ 53%]  (Sampling)
## Chain 1: Iteration:  9500 / 15000 [ 63%]  (Sampling)
## Chain 1: Iteration: 11000 / 15000 [ 73%]  (Sampling)
## Chain 1: Iteration: 12500 / 15000 [ 83%]  (Sampling)
## Chain 1: Iteration: 14000 / 15000 [ 93%]  (Sampling)
## Chain 1: Iteration: 15000 / 15000 [100%]  (Sampling)
## Chain 1:
## Chain 1:  Elapsed Time: 91.619 seconds (Warm-up)
## Chain 1:                182.265 seconds (Sampling)
## Chain 1:                273.884 seconds (Total)
## Chain 1:
```

```
##
## SAMPLING FOR MODEL 'stanmarg' NOW (CHAIN 2).
## Chain 2:
## Chain 2: Gradient evaluation took 0.003104 seconds
## Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 31.04 seconds.
## Chain 2: Adjust your expectations accordingly!
## Chain 2:
## Chain 2:
## Chain 2: Iteration:     1 / 15000 [  0%]  (Warmup)
## Chain 2: Iteration:  1500 / 15000 [ 10%]  (Warmup)
## Chain 2: Iteration:  3000 / 15000 [ 20%]  (Warmup)
## Chain 2: Iteration:  4500 / 15000 [ 30%]  (Warmup)
## Chain 2: Iteration:  5001 / 15000 [ 33%]  (Sampling)
## Chain 2: Iteration:  6500 / 15000 [ 43%]  (Sampling)
## Chain 2: Iteration:  8000 / 15000 [ 53%]  (Sampling)
## Chain 2: Iteration:  9500 / 15000 [ 63%]  (Sampling)
## Chain 2: Iteration: 11000 / 15000 [ 73%]  (Sampling)
## Chain 2: Iteration: 12500 / 15000 [ 83%]  (Sampling)
## Chain 2: Iteration: 14000 / 15000 [ 93%]  (Sampling)
## Chain 2: Iteration: 15000 / 15000 [100%]  (Sampling)
## Chain 2:
## Chain 2:  Elapsed Time: 88.596 seconds (Warm-up)
## Chain 2:                190.334 seconds (Sampling)
## Chain 2:                278.93 seconds (Total)
## Chain 2:
##
## SAMPLING FOR MODEL 'stanmarg' NOW (CHAIN 3).
## Chain 3:
## Chain 3: Gradient evaluation took 0.001622 seconds
## Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 16.22 seconds.
## Chain 3: Adjust your expectations accordingly!
## Chain 3:
## Chain 3:
## Chain 3: Iteration:     1 / 15000 [  0%]  (Warmup)
## Chain 3: Iteration:  1500 / 15000 [ 10%]  (Warmup)
## Chain 3: Iteration:  3000 / 15000 [ 20%]  (Warmup)
## Chain 3: Iteration:  4500 / 15000 [ 30%]  (Warmup)
## Chain 3: Iteration:  5001 / 15000 [ 33%]  (Sampling)
## Chain 3: Iteration:  6500 / 15000 [ 43%]  (Sampling)
## Chain 3: Iteration:  8000 / 15000 [ 53%]  (Sampling)
## Chain 3: Iteration:  9500 / 15000 [ 63%]  (Sampling)
## Chain 3: Iteration: 11000 / 15000 [ 73%]  (Sampling)
## Chain 3: Iteration: 12500 / 15000 [ 83%]  (Sampling)
## Chain 3: Iteration: 14000 / 15000 [ 93%]  (Sampling)
## Chain 3: Iteration: 15000 / 15000 [100%]  (Sampling)
## Chain 3:
## Chain 3:  Elapsed Time: 82.207 seconds (Warm-up)
## Chain 3:                148.494 seconds (Sampling)
## Chain 3:                230.701 seconds (Total)
## Chain 3:
## Computing posterior predictives...
```

```
# Calculate Information Criteria and BRMSEA and its CI (Table 3)
fit_indices3 <- fitMeasures(bcfa_result3)

## Warning:
## 35 (19.7%) p_waic estimates greater than 0.4. We recommend trying loo instead.

## Warning: Some Pareto k diagnostic values are too high. See help('pareto-k-diagnostic') for
details.

print(fit_indices3)

##      npar      logl       ppp       bic       dic     p_dic      waic
##    49.000  -5683.231     0.000  11620.092  11463.912    48.725  11477.005
##    p_waic    se_waic     looic     p_loo    se_loo margloglik
##    60.391    158.570  11477.430    60.604   158.663  -5857.788

BRMSEA <- blavFitIndices(bcfa_result3)

## Warning:
## 35 (19.7%) p_waic estimates greater than 0.4. We recommend trying loo instead.

## Warning: Some Pareto k diagnostic values are too high. See help('pareto-k-diagnostic') for
details.

summary(BRMSEA, central.tendency = c("mean"), prob = .90, hpd = T)

##
## Posterior summary statistics and highest posterior density (HPD) 90% credible intervals fo
r devm-based fit indices:
##
##              EAP    SD lower upper
## BRMSEA       0.079 0.002 0.077 0.082
## BGammaHat    0.895 0.004 0.888 0.901
## adjBGammaHat 0.865 0.005 0.857 0.873
## BMc          0.508 0.014 0.485 0.532

#Display Parameters with Credibility Intervals (Appendix J)
summary(bcfa_result3)

## blavaan (0.4-8) results of 10000 samples after 5000 adapt/burnin iterations
##
##   Number of observations                           178
##
##   Statistic                           MargLogLik        PPP
##   Value                                -5857.788      0.000
##
## Latent Variables:
##                  Estimate  Post.SD pi.lower pi.upper     Rhat    Prior
##   Factor1 =~
##     IBATI1          0.371    0.076    0.224    0.524    1.000    normal(0,10)
##     IBATI2          0.662    0.103    0.462    0.866    1.000    normal(0,10)
##     IBATI3          0.781    0.092    0.605    0.964    1.000    normal(0,10)
##     IBATI4          0.392    0.079    0.239    0.550    1.000    normal(0,10)
##     IBATI5          0.645    0.095    0.464    0.834    1.000    normal(0,10)
```

```
##    IBATI6          0.883   0.111   0.673   1.105   1.000   normal(0,10)
##    IBATI7          0.820   0.096   0.635   1.013   1.000   normal(0,10)
##    IBATI8          0.569   0.096   0.385   0.760   1.000   normal(0,10)
##    IBATI9          0.638   0.074   0.497   0.787   1.000   normal(0,10)
##   Factor2 =~
##    ATI1          1.069   0.087   0.904   1.248   1.000   normal(0,10)
##    ATI2          0.912   0.079   0.763   1.075   1.000   normal(0,10)
##    ATI3          0.219   0.103   0.017   0.423   1.000   normal(0,10)
##    ATI4          0.855   0.081   0.703   1.022   1.000   normal(0,10)
##    ATI5          1.022   0.080   0.874   1.186   1.000   normal(0,10)
##    ATI6          0.844   0.105   0.641   1.055   1.000   normal(0,10)
##    ATI7          0.911   0.090   0.742   1.093   1.000   normal(0,10)
##    ATI8          0.714   0.090   0.541   0.895   1.000   normal(0,10)
##    ATI9          0.810   0.082   0.654   0.978   1.000   normal(0,10)
##   Factor3 =~
##    AAS1          0.434   0.078   0.281   0.588   1.000   normal(0,10)
##    AAS2          0.588   0.096   0.401   0.779   1.000   normal(0,10)
##    AAS3          0.742   0.099   0.549   0.937   1.000   normal(0,10)
##    AAS4          0.396   0.079   0.242   0.552   1.000   normal(0,10)
##    AAS5          0.615   0.098   0.427   0.809   1.000   normal(0,10)
##
## Covariances:
##                  Estimate  Post.SD pi.lower pi.upper   Rhat   Prior
##   Factor1 ~~
##    Factor2         0.875   0.035   0.799   0.934   1.000   lkj_corr(1)
##   Factor2 ~~
##    Factor3        -0.273   0.092  -0.448  -0.087   1.000   lkj_corr(1)
##   Factor1 ~~
##    Factor3        -0.178   0.103  -0.374   0.029   1.000   lkj_corr(1)
##
## Variances:
##                  Estimate  Post.SD pi.lower pi.upper   Rhat   Prior
##   .IBATI1          0.733   0.082   0.588   0.910   1.000 gamma(1,.5)[sd]
##   .IBATI2          1.235   0.141   0.987   1.540   1.000 gamma(1,.5)[sd]
##   .IBATI3          0.859   0.104   0.674   1.085   1.000 gamma(1,.5)[sd]
##   .IBATI4          0.791   0.089   0.635   0.980   1.000 gamma(1,.5)[sd]
##   .IBATI5          1.023   0.118   0.816   1.274   1.000 gamma(1,.5)[sd]
##   .IBATI6          1.308   0.155   1.034   1.639   1.000 gamma(1,.5)[sd]
##   .IBATI7          0.924   0.115   0.718   1.169   1.000 gamma(1,.5)[sd]
##   .IBATI8          1.099   0.125   0.879   1.369   1.000 gamma(1,.5)[sd]
##   .IBATI9          0.573   0.070   0.450   0.724   1.000 gamma(1,.5)[sd]
##   .ATI1          0.575   0.076   0.440   0.737   1.000 gamma(1,.5)[sd]
##   .ATI2          0.511   0.065   0.395   0.651   1.000 gamma(1,.5)[sd]
##   .ATI3          1.525   0.165   1.238   1.884   1.000 gamma(1,.5)[sd]
##   .ATI4          0.626   0.077   0.492   0.790   1.000 gamma(1,.5)[sd]
##   .ATI5          0.440   0.060   0.335   0.568   1.000 gamma(1,.5)[sd]
##   .ATI6          1.254   0.144   1.004   1.563   1.000 gamma(1,.5)[sd]
##   .ATI7          0.784   0.094   0.618   0.988   1.000 gamma(1,.5)[sd]
##   .ATI8          0.926   0.105   0.740   1.154   1.000 gamma(1,.5)[sd]
##   .ATI9          0.680   0.080   0.539   0.854   1.000 gamma(1,.5)[sd]
##   .AAS1          0.571   0.072   0.442   0.726   1.000 gamma(1,.5)[sd]
```

```
##    .AAS2           0.819   0.109   0.622   1.048   1.000 gamma(1,.5)[sd]
##    .AAS3           0.671   0.119   0.446   0.917   1.000 gamma(1,.5)[sd]
##    .AAS4           0.588   0.073   0.457   0.745   1.000 gamma(1,.5)[sd]
##    .AAS5           0.816   0.113   0.610   1.053   1.000 gamma(1,.5)[sd]
##     Factor1        1.000
##     Factor2        1.000
##     Factor3        1.000

#Plot Traceplot (Appendix K)
plot(bcfa_result3, pars = 1:6, plot.type = "trace")

# Calculate Factor Loadings, Covariances and CIs in dataframe
postmean <- blavInspect(bcfa_result3, "postmean")
hpd_intervals <- blavInspect(bcfa_result3, "hpd", level = .95)
factor_names <- names(postmean)[grep("Factor", names(postmean))]
factor_loadings_df <- data.frame(
  Element = factor_names,
  Estimate = postmean[factor_names],
  Lower = hpd_intervals[factor_names, "lower"],
  Upper = hpd_intervals[factor_names, "upper"])

# Visualising Factor Loadings and CIs (Figure 5)
factor_loadingsonly_df <- factor_loadings_df %>%
  filter(!str_detect(Element, "~~"))
factor_loadingsonly_df$FactorColours <- ifelse(grepl("Factor1", factor_loadingsonly_df$Ele
ment), "lightblue",ifelse(grepl("Factor2", factor_loadingsonly_df$Element), "salmon1", "light
green"))
ggplot(factor_loadingsonly_df, aes(x = Element, y = Estimate, fill =FactorColours)) +
  geom_crossbar(aes(ymin = Lower, ymax = Upper), width = 0.7) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Items", y = "Estimated Factor Loadings")+
  scale_fill_identity(name = "Scale", labels = c("IBATI", "ATI", "AAS"),
  guide = guide_legend(override.aes = list(fill = c("lightblue", "salmon1", "lightgreen"))))

# Visualising Covariances and CIs (Figure 6)
covariancesonly_df <- factor_loadings_df %>%
  filter(!str_detect(Element, "=~"))
covariancesonly_df$FactorColours <- c("lightblue", "salmon1", "lightgreen", rep("lightgreen"
, nrow(covariancesonly_df) - 3))
ggplot(covariancesonly_df, aes(x = Element, y = Estimate, fill=FactorColours)) +
  geom_crossbar(aes(ymin = Lower, ymax = Upper),width = 0.7) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(x = "Factors", y = "Estimated Correlation")+
  scale_fill_identity(name = "Scales", labels = c("IBATI & ATI", "IBATI & AAS", "ATI & A
AS"),
  guide = guide_legend(override.aes = list(fill = c("lightblue", "lightgreen", "salmon1"))))

bcfa_model2 <- "
   Factor1 =~  ATI1 + ATI2 + ATI4 + ATI5 + ATI7 + ATI8 + ATI9 + IBATI1 + IBATI2 + IB
ATI3 + IBATI4 + IBATI5 + IBATI6 + IBATI7 + IBATI8 + IBATI9
   Factor2 =~ AAS1 + AAS2 + AAS3 + AAS4 + AAS5
```

```
  Factor1 ~~ Factor2
"
# Fit the BCFA model as above in the 3-Factor Model
bcfa_result2 <- bcfa(bcfa_model2, data=IBT_CLEAN, mcmcfile = T, std.lv=TRUE, cp = "srs
",
    dp = NULL, n.chains = 3, burnin = 5000, adapt = 1000, sample = 10000, inits = "simple",
    convergence = "manual", target = "stan", save.lvs = TRUE,
    wiggle = NULL, wiggle.sd = 0.1, prisamp = FALSE, jags.ic = FALSE,
    seed = 15, bcontrol = list())

##
## SAMPLING FOR MODEL 'stanmarg' NOW (CHAIN 1).
## Chain 1:
## Chain 1: Gradient evaluation took 0.005942 seconds
## Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 59.42 seconds.
## Chain 1: Adjust your expectations accordingly!
## Chain 1:
## Chain 1:
## Chain 1: Iteration:     1 / 15000 [  0%]  (Warmup)
## Chain 1: Iteration:  1500 / 15000 [ 10%]  (Warmup)
## Chain 1: Iteration:  3000 / 15000 [ 20%]  (Warmup)
## Chain 1: Iteration:  4500 / 15000 [ 30%]  (Warmup)
## Chain 1: Iteration:  5001 / 15000 [ 33%]  (Sampling)
## Chain 1: Iteration:  6500 / 15000 [ 43%]  (Sampling)
## Chain 1: Iteration:  8000 / 15000 [ 53%]  (Sampling)
## Chain 1: Iteration:  9500 / 15000 [ 63%]  (Sampling)
## Chain 1: Iteration: 11000 / 15000 [ 73%]  (Sampling)
## Chain 1: Iteration: 12500 / 15000 [ 83%]  (Sampling)
## Chain 1: Iteration: 14000 / 15000 [ 93%]  (Sampling)
## Chain 1: Iteration: 15000 / 15000 [100%]  (Sampling)
## Chain 1:
## Chain 1:  Elapsed Time: 70.031 seconds (Warm-up)
## Chain 1:                182.121 seconds (Sampling)
## Chain 1:                252.152 seconds (Total)
## Chain 1:
##
## SAMPLING FOR MODEL 'stanmarg' NOW (CHAIN 2).
## Chain 2:
## Chain 2: Gradient evaluation took 0.001082 seconds
## Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 10.82 seconds.
## Chain 2: Adjust your expectations accordingly!
## Chain 2:
## Chain 2:
## Chain 2: Iteration:     1 / 15000 [  0%]  (Warmup)
## Chain 2: Iteration:  1500 / 15000 [ 10%]  (Warmup)
## Chain 2: Iteration:  3000 / 15000 [ 20%]  (Warmup)
## Chain 2: Iteration:  4500 / 15000 [ 30%]  (Warmup)
## Chain 2: Iteration:  5001 / 15000 [ 33%]  (Sampling)
## Chain 2: Iteration:  6500 / 15000 [ 43%]  (Sampling)
## Chain 2: Iteration:  8000 / 15000 [ 53%]  (Sampling)
## Chain 2: Iteration:  9500 / 15000 [ 63%]  (Sampling)
## Chain 2: Iteration: 11000 / 15000 [ 73%]  (Sampling)
```

```
## Chain 2: Iteration: 12500 / 15000 [ 83%]  (Sampling)
## Chain 2: Iteration: 14000 / 15000 [ 93%]  (Sampling)
## Chain 2: Iteration: 15000 / 15000 [100%]  (Sampling)
## Chain 2:
## Chain 2:  Elapsed Time: 66.961 seconds (Warm-up)
## Chain 2:                132.497 seconds (Sampling)
## Chain 2:                199.458 seconds (Total)
## Chain 2:
##
## SAMPLING FOR MODEL 'stanmarg' NOW (CHAIN 3).
## Chain 3:
## Chain 3: Gradient evaluation took 0.00108 seconds
## Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 10.8 seconds.
## Chain 3: Adjust your expectations accordingly!
## Chain 3:
## Chain 3:
## Chain 3: Iteration:     1 / 15000 [  0%]  (Warmup)
## Chain 3: Iteration:  1500 / 15000 [ 10%]  (Warmup)
## Chain 3: Iteration:  3000 / 15000 [ 20%]  (Warmup)
## Chain 3: Iteration:  4500 / 15000 [ 30%]  (Warmup)
## Chain 3: Iteration:  5001 / 15000 [ 33%]  (Sampling)
## Chain 3: Iteration:  6500 / 15000 [ 43%]  (Sampling)
## Chain 3: Iteration:  8000 / 15000 [ 53%]  (Sampling)
## Chain 3: Iteration:  9500 / 15000 [ 63%]  (Sampling)
## Chain 3: Iteration: 11000 / 15000 [ 73%]  (Sampling)
## Chain 3: Iteration: 12500 / 15000 [ 83%]  (Sampling)
## Chain 3: Iteration: 14000 / 15000 [ 93%]  (Sampling)
## Chain 3: Iteration: 15000 / 15000 [100%]  (Sampling)
## Chain 3:
## Chain 3:  Elapsed Time: 62.423 seconds (Warm-up)
## Chain 3:                128.656 seconds (Sampling)
## Chain 3:                191.079 seconds (Total)
## Chain 3:
## Computing posterior predictives...
```

```
# Print Information Criteria, BRMSEA and its CI
fit_indices2 <- fitMeasures(bcfa_result2)
```

```
## Warning:
## 29 (16.3%) p_waic estimates greater than 0.4. We recommend trying loo instead.

## Warning: Some Pareto k diagnostic values are too high. See help('pareto-k-diagnostic') for
details.
```

```
print(fit_indices2)
```

```
##      npar      logl       ppp       bic       dic     p_dic      waic
##    43.000  -5132.908     0.000  10488.390  10351.433    42.809  10361.924
##    p_waic   se_waic     looic     p_loo    se_loo margloglik
##    52.103   142.086  10362.461    52.372   142.269  -5289.478
```

```
BRMSEA <- blavFitIndices(bcfa_result2)
```

```
## Warning:
## 29 (16.3%) p_waic estimates greater than 0.4. We recommend trying loo instead.

## Warning: Some Pareto k diagnostic values are too high. See help('pareto-k-diagnostic') for
details.
```

```
summary(BRMSEA, central.tendency = c("mean"), prob = .90, hpd = T)
```

```
##
## Posterior summary statistics and highest posterior density (HPD) 90% credible intervals fo
r devm-based fit indices:
##
##                EAP    SD lower upper
## BRMSEA       0.078 0.002 0.074 0.081
## BGammaHat    0.907 0.004 0.900 0.914
## adjBGammaHat 0.880 0.005 0.871 0.889
## BMc          0.584 0.016 0.560 0.610
```