Original Article

# Psychometric Evaluation of the Swedish Version of the Empathic Experience Scale (EES)

Andreas Sarling[1,2], Örjan Sundin[3], and Billy Jansson[1]

[1]Department of Psychology and Social Work, Mid Sweden University, Östersund, Sweden
[2]Department of Psychology, Stockholm University, Sweden
[3]Catolica Research Centre for Psychological – Family and Social Wellbeing, Universidade Portuguesa, Lisbon, Portugal

**Abstract:** *Background*: This study assessed the factor structure and psychometric properties of the Swedish version of the Empathic Experience Scale (EES), a recently developed scale for empathic traits (Innamorati et al., 2019). According to previous research, EES has two dimensions: Vicarious Experience and Intuitive Understanding. *Methods*: We used a split-sample method with a combination of exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) for an adult population ($N$ = 805) from both Sweden and Finland. *Results*: Our findings support previous research, favoring a two-factor model over a unidimensional model. The final model provided support for measurement invariance across different grouping variables. Consistent with previous research on empathic traits, women obtained higher scores on both dimensions of the scale compared to men. Furthermore, the EES showed high internal consistency, good discriminant, and convergent validity.

**Keywords:** empathy, vicarious experience, intuitive understanding, factor structure, psychometric properties

OPEN DATA OPEN MATERIALS OPEN CODE

Empathy is widely recognized as encompassing both state and trait dimensions, with state empathy involving immediate responses that mirrors others' emotions, and the latter represented by individual differences in latent capacities and tendencies to enact these empathic responses (Cuff et al., 2016; Hall & Schwartz, 2019). In this study, we will focus on measuring empathy as a trait. Hall and Schwartz (2019) highlighted the often unclear and inconsistent definitions of empathy in their comprehensive review of 393 studies. They also discussed that the term "empathy" has, in many cases, played out its role, making it more meaningful to assess the lower-lever constructs utilized in empathy measures (Hall & Schwartz, 2019). In addition, they found that most studies used multidimensional concepts of empathy, with cognitive empathy and affective empathy being the most common components (Hall & Schwartz, 2019). Cognitive empathy has been defined as involving the understanding or identifying another person's emotions, while affective empathy concerns sharing others' emotions (Coll et al., 2017; Cuff et al., 2016; Gerdes et al., 2010). Recognizing the diverse definitions of empathy in existing literature, we acknowledge the value of various conceptualizations in both research and practical contexts. In our study, we propose a model that primarily focuses on the two dimensions of empathy: emotional and cognitive. This approach does not claim to be definitive but highlights the importance of exploring and validating this two-dimensional conceptualization in understanding empathic processes.

## Theoretical Background

Empathy is a multifaceted construct that has been interpreted variously within psychological research. Cuff et al. (2016) described it as an emotional sharing and understanding of others' emotions, distinct from sympathy (feeling for others), compassion (a blend of sympathy and a desire to help), and tenderness (a nurturing attitude). Innamorati et al. (2019) proposed making empathy measures more distinct from similar constructs. Our study examined empathy as defined by Innamorati et al. (2019) for the Empathic Experience Scale (EES) by validating the scale with a Swedish-speaking population, contributing to the broader discourse on its multifaceted nature.

Both Cuff et al. (2016) and Innamorati et al. (2019) explored the contrasting perspectives on cognitive empathy, categorizing it as either automatic (effortless and intuitive) or


© 2024 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2698-1866/a000079 - Friday, September 20, 2024 6:23:20 AM - IP Address:158.162.233.152

https://econtent.hogrefe.com/doi/pdf/10.1027/2698-1866/a000079 - Friday, September 20, 2024 6:23:20 AM - IP Address:158.162.233.152

controlled (effortful). Cuff et al. (2016) described support for both views in prior research, highlighting their possible co-occurrence in empathic responses. However, Innamorati et al. (2019) observed that preexisting measures lacked the intuitive aspect of cognitive empathy. As Innamorati et al. (2019) also pointed out psychometric weaknesses in previous measures of empathy, they advocated a need for a new scale for empathic traits, leading them to develop the Empathic Experience Scale (EES), to capture both the affective and cognitive components of empathy. These components were labeled Vicarious Experience and Intuitive Understanding, respectively, with a specific focus on the effortless nature of the cognitive aspect. The affective items of the EES reflect spontaneous emotional congruence with others' experiences, such as feeling teary-eyed when witnessing someone crying or feeling pain when seeing someone get hurt. This highlights the automatic and effortless nature of the affective empathy component, similar to the intuitive understanding seen in the cognitive empathy items. Unlike previous instruments that focus on reflective empathy, the EES uniquely measures the intuitive dimensions of empathetic engagement, making it a novel contribution to the field.

Vicarious experience has been defined as "participating in someone else's emotional state by experiencing similar emotions," and intuitive understanding as "the (effortless) cognitive awareness of the emotional or sensorimotor state of someone else" (Innamorati et al., 2019, p. 5). When Innamorati et al. (2019) compared a correlated two-factor model with a bifactor model, the results supported the two-factor model measuring the two components separately. Their final model of 30 items (15 for each factor) showed high reliability, and good convergent and discriminant validity with respect to other measures of empathy. The two factors of the EES have been examined twice with the Italian version, showing moderate correlations between the factors ($r$ = .32 and $r$ = .37), suggesting some overlapping variance but measuring two distinct concepts (Ebisch et al., 2022; Innamorati et al., 2019). Moreover, there is evidence that Vicarious Experience and Intuitive Understanding are related to different functional brain networks. In their fMRI study, Ebisch et al. (2022) identified associations between Vicarious Experience and the frontoparietal network, but also between Intuitive Understanding and both the subcortical network and the somatomotor network, strengthening the evidence of discriminant validity. They also found that two EES factors overlapped on correlates with salience network regions, which lends support to convergent validity. The EES has also shown to have higher internal consistency than previous frequently used measures, including the Interpersonal Reactivity Index (Innamorati et al., 2019). To the best of our knowledge, no studies have tested the replicability of the factor structure of the EES, neither in the original language nor in any translated version. Translating the scale is crucial for cross-cultural research, enabling a broader understanding of empathic traits in diverse populations and facilitating comparisons between different cultural contexts.

Gerdes et al. (2010) have proposed that empathy can be nurtured and cultivated over time. Different components of empathy evolve in gradual stages from early childhood onward, driven by a complex interplay of biological predispositions, social interactions, and emotional bonds with others (Decety, 2010; Decety & Jackson, 2004). Notably, this developmental process has demonstrated a gender advantage for women in empathic development compared to men (Christov-Moore et al., 2014). Building upon these insights, when investigating gender differences using the EES, Ebisch et al. (2022) found that women scored higher than men on the Vicarious Experience dimension (but not for Intuitive Understanding), highlighting their propensity for feeling emotions indirectly through the experiences of others and underscoring the importance of understanding the nuances within empathy and its gender-related dynamics.

The aim of this study was to examine the factor structure and psychometric properties of the EES in a Swedish translation using a combination of exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) with a split-sample method for a general adult population. The purpose of testing the scale is to establish the utility of the EES for assessing empathic traits across diverse populations in research settings, which could enhance our understanding of empathy and informing further research in various domains. In line with Innamorati et al. (2019), we expected that a correlated two-factor model containing Vicarious Experience (affective empathy) and Intuitive Understanding (cognitive empathy) would be favored over a bifactor model with the overarching empathy factor. As shown before (Ebisch et al., 2022; Innamorati et al., 2019), we expected that the two dimensions would show a moderate positive correlational with each other. Also, drawing on previous research (Ebisch et al., 2022), it was expected that women would score higher on the Vicarious Experience dimension of the EES compared to men. We also investigated potential gender differences on the Intuitive Understanding dimension.

# Methods

## Participants

To reach a diverse range of ages within different geographical areas in Sweden, as well as areas with a high density of Swedish-speaking populations in Finland, participants for this study were recruited both via webpages from two Swedish-speaking universities (one in Sweden and

© 2024 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2698-1866/a000079 - Friday, September 20, 2024 6:23:20 AM - IP Address:158.162.233.152

one in Finland) and through marketed posts on Facebook and Instagram. The study targeted Swedish-speaking adults aged 18 years or older. In total, 1,180 participants opened the survey, of whom 363 terminated the survey prior to the completion of all stages and 12 completed the survey but did not provide answers to all questions. Due to the complete case approach design, cases with any missing data were excluded, leaving a final sample of 805 participants between the ages of 18 and 83 years ($M_{\text{age}}$ = 45.13, $SD$ = 14.04). Most participants were Swedish (70.4%), 27.5% were Finnish, and the remaining participants had other nationalities (2.1%). Of the total sample, 21.7% were students, while 78.3% were not. The sample was predominantly women (87.0%). The survey was administered online, and no incentives were provided for participation.

Participants were given information about the anonymity and confidentiality of the data that participation was voluntary and could be terminated at any point. After being informed about the study's purpose (*to develop measures to better understand social interactions*), participants were required to give their informed consent. The study received ethical approval from the local committee as the subjects were anonymous, and similar surveys (with similar subjects) had been previously approved, making it unnecessary to seek approval from the regional ethics committee.

## Procedure

The survey consisted of four questionnaires presented in a random order: the EES questionnaire and three other questionnaires measuring attachment security, compassion, and self-compassion. This study is dedicated to the validation of the EES in a Swedish-speaking sample. While the collected data included other measures, those have been analyzed and reported elsewhere. Our focus here is solely on the EES's psychometric properties, ensuring our analysis remains distinct and nonredundant, without overlap to previous studies. Prior to completing the questionnaires, participants were asked to answer four demographic questions (age, gender, nationality, and student status). The survey was administered using Qualtrics (http://www.qualtrics.com) between March 14 and May 5, 2020. To prevent potential biases from multiple responses, Qualtrics placed a cookie in participants' web browsers to exclude them from taking the survey more than once.

## Measures

### The Empathic Experience Scale

Empathic traits were measured with the EES (Innamorati et al., 2019) in a Swedish translation. The scales contain 30 items – 15 each for the subscales Intuitive Understanding and Vicarious Experience (see Table S1 for the items in Swedish, Table S2 for the items in English, and the Swedish instruction text for answering the items in Appendix A in the supplementary materials available at https://osf.io/g5mvt). Items are answered on a 5-point scale (1 = *not at all true* to 5 = *completely true*). Two examples of Intuitive Understanding items are "I can quickly intuit the state of mind of a person I know even if he/she tries to hide his/her real emotions" and "I am good at intuitively understanding the feelings of others," and two examples of Vicarious Experience items are "Seeing an adult cry because of pain makes me suddenly get teary-eyed" and "When I see someone get hurt, I feel his/her pain as if I were hurt, without being able to distance myself from the pain." Cronbach's α is .944 for Intuitive Understanding and .910 for Vicarious Experience.

### Translations

Permission to translate the scales was given by the second author of the original study of the EES (Innamorati et al., 2019). The first draft of a translation was carried out by the first and second authors of this study, both native speakers of Swedish. The translations were then sent to a translation agency, where a professional back-translator with extensive experience in Swedish-to-English translation performed the back-translations.

To validate the translation, a quantitative process recommended by Sperber (2004) was employed. All items from the original English version were presented in pairs with their back-translated English versions, and an independent panel of nine individuals with proficiency in English (all but one with at least a bachelor's degree in psychology) judged the language comparability and interpretation similarity. Then, some items were revised before sending the original items side-by-side with the Swedish translations and back-translations to the third author of this study for written qualitative feedback, resulting in the final versions of the translations.

### Data Analytic Procedures

To validate the factor structure of the EES, two separate analyses (EFA and CFA) were used to be able to compare the best fit between a possible exploratory model find from the EFA with other models in the CFA. The EFA was computed using JASP (JASP Team, 2023), while the CFA was computed using the R (R Core Team, 2021) package lavaan (Rosseel, 2012).

A total of 805 participants were split into two subsamples in order of participation by putting odd number to the exploratory sample ($N$ = 403) and even number to the confirmatory sample ($N$ = 402). From rule of thumb, this gave a good sample power for both the EFA and the CFA as


© 2024 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

there were over 300 participants in each subsample (Kyriazos, 2018). The exploratory sample consisted of 53 men ($M_{age}$ = 42.87, $SD$ = 15.44) and 345 women ($M_{age}$ = 45.66, $SD$ = 13.66), whereas the confirmatory sample consisted of 42 men ($M_{age}$ = 39.50, $SD$ = 13.78) and 355 women ($M_{age}$ = 45.67, $SD$ = 14.04).

We used parallel analysis to test whether the EFA could yield an alternative factor structure. The estimation method used for factor extraction was Maximum Likelihood with oblique rotation (promax). If the EFA suggests an alternative factor model (other than the hypothesized two-factor model), this model will be compared with the other models in the CFA.

In the confirmatory sample, we used CFA to compare a one-factor model (where all items loaded on a single empathy factor) with the hypothesized two-factor model, and a bifactor model in which items are assumed to load on both a general factor and two specific factors. Bifactor models with psychometric support of both unidimensionality and specified group factors allow researchers to examine both the overall construct and the specific factors independently (Rodriguez et al., 2015). As the data in this study both showed some deviations to normality and were of ordinal character, we used a mean- and variance-adjusted weighted least squares (WLSMV) estimator with a polychoric correlation matrix for the CFAs (see, e.g., Brauer et al., 2023).

We used Dueber's (2017) calculator to retrieve several psychometric indices specific for bifactor models. For internal reliability of the multidimensional unit-weighted composite for both the general factor and specific factors, we used omega (ω) and omega subscale ($\omega_S$), respectively. Two other omega measures, omega hierarchical ($\omega_H$) and omega hierarchical subscale ($\omega_{HS}$), were used. While omega hierarchical tested the proportion of reliable systematic variance attributed to the general factor, omega hierarchical subscale tested this for the specific factors (Vicarious Experience and Intuitive Understanding) after portioning out variability stemming from the general factor. The omega hierarchical should be at least .80 or above to consider the general factor unidimensional (Rodriguez et al., 2015). Two other indices that were used to ensure quality of the measurement model were factor determinacy (FD) and construct replicability (H). FD represents the correlation between factor scores and the factors, and H represents the correlation between a factor and an optimally weighted item composite. FD should be over .90 and H over .80 to be considered good, but values of $FD^2$ (FD squared) and H of at least .70 are acceptable to specify group factors (Rodriguez et al., 2015). Besides the omega hierarchical index, we used two other indices to test unidimensionality of the bifactor model: explained common variance (ECV) and percentage of uncontaminated correlations (PUC). ECV was used to capture the proportion of variance explained by the general factor in the bifactor model, but also ECV for the subscales ($ECV_S$) to capture the strength of a specific factor explaining the proportion of variance for all items within the bifactor model. PUC captures the percentage of covariance terms solely explained by the general factor. Both ECV and PUC should be .70 or above (Rodriguez et al., 2015).

Four model fit indices were analyzed according to a priori cutoffs assessing measurement model validity. For this purpose, we used the root-mean-square error of approximation (RMSEA), the standardized root mean residual (SRMR), the comparative fit index (CFI), and the relative chi-square statistics ($\chi^2/df$). We mainly used cutoffs recommended by Schermelleh-Engel et al. (2003) for assessing model fit. With respect to the RMSEA, values equal to or below .05 are considered a good-fitting model, values between .05 and .08 indicate an acceptable fit, and values between .08 and .10 indicate a mediocre fit. SRMR values equal to or under .05 indicate a good fit to the data, while the value equal to or under .10 is considered acceptable. Based on commonly accepted thresholds for model fit, CFI values equal to or greater than .97 are indicative of a good fit, while values equal to or above .95 indicate an acceptable fit (Schermelleh-Engel et al., 2003). However, some studies have used a more lenient threshold of .90 for acceptable fit (e.g., Williams et al., 2014). In the current study, we consider CFI values of at least .90 to indicate mediocre fit. For $\chi^2/df$, an acceptable fit is reached when $\chi^2/df$ is below 3 and a good fit below 2. After the assessment of fit of the factor structures, the modification indices were examined to detect improvement of the models. We chose to not cross-correlate between different factors as these correlations complicate interpretation of the factor model.

The final model of the EES was also tested for measurement invariance for the factor structure using several multigroup models. Due to uneven distribution within certain groups, rather than using the confirmatory sample for measurement invariance testing, we opted for using the entire study sample for these analyses, comparing invariance for genders (men and women, $N$ = 95 and 700), nationality (Swedish and Finnish, $N$ = 567 and 221), and age (< 45 years; ≥ 45 years, $N$ = 375 and 430). Invariance was assessed in four sequential steps, assessing configural, metric, scalar, and residual invariance (Putnick & Bornstein, 2016). Configural invariance was assessed according to a baseline model in which parameters were freely estimated across groups, with the model evaluated based on cutoffs used for CFI, SRMR, and RMSEA mentioned earlier. The other levels of invariance were assessed according to decrements or increments in fit between the other steps of testing. Although the sample size was 805,

© 2024 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2698-1866/a000079 - Friday, September 20, 2024 6:23:20 AM - IP Address:158.162.233.152

https://econtent.hogrefe.com/doi/pdf/10.1027/2698-1866/a000079 - Friday, September 20, 2024 6:23:20 AM - IP Address:158.162.233.152

the more conservative cutoff values described by Chen (2007) for sample sizes under 300 were used for gender due to substantial difference in gender distribution. This decision was made to ensure statistical validity despite uneven group sizes, thereby addressing potential biases in the measurement invariance analysis across genders. Therefore, for metric invariance, decrements in CFI ≥ .005 together with increments in SRMR of ≥ .025 and RMSEA of ≥ .01 indicate noninvariance between groups for gender. For scalar and residual invariance, the same thresholds apply for CFI and RMSEA, but for SRMR, increments ≥ .005 indicate noninvariance. As the nationality and age groups had considerably larger group sizes, the less conservative cutoff values were used, i.e., decrements in CFI ≥ .01 together with increments in SRMR ≥ .03 and RMSEA ≥ .015 indicate noninvariance between groups for metric invariance (as well as for scalar and residual invariance for CFI and RMSEA), while increments in SRMR ≥ .01 indicate noninvariance for scalar and residual invariance (Chen, 2007). Adequate steps of measurement invariance testing would mean equivalence in factor structure (configural invariance), an equal contribution from items to latent factors (metric invariance), an equally captured shared variance of items by the latent factors (scalar invariance), and a comparable item and error variance (residual variance; Putnick & Bornstein, 2016) across groups.

Finally, the confirmatory sample was used to assess the final model of the EES for internal consistency, and convergent and discriminant validity. This assessment was done with guidelines for cutoffs (Hair et al., 2014). Good internal consistency, shown by composite reliability (CR), should be ≥ .7. For convergent validity, all standardized factor loadings within a factor, as well as average variance extracted (AVE), should be ≥ .5. An AVE of .5 or higher indicates that, on average, more variance is explained by the items' factor loadings than the measurement error. Discriminant validity was tested by comparing the AVE of each factor with the maximum shared variance (MSV) with any other factor in the model. AVE should be larger than MSV as it indicates that more variance is explained within the factor than through intercorrelations with another factor.

## Results

The parallel analysis of EFA in the exploratory sample suggested that two factors should be extracted. All items correlated with their respective factor without any cross-loadings over .32, indicating that the two factors are unidimensional (see Table S1 in the supplementary materials for all factor loadings).

**Table 1.** CFA of the EES: model-fit indices for a one-factor, two-factor, bifactor, and modified two-factor model

| Model | $\chi^2$ (*df*) | $\chi^2/df$ | CFI | SRMR | RMSEA |
|---|---|---|---|---|---|
| One-factor | 4,324.194 (405) | 10.68 | .79 | .15 | .16 |
| Two-factor | 1,392.149 (404) | 3.45 | .95 | .07 | .08 |
| Bifactor | 1,177.065 (375) | 3.14 | .96 | .05 | .07 |
| Modified two-factor | 1,058.697 (399) | 2.65 | .97 | .06 | .06 |

*Note*. CFI = comparative fit index, SRMR = standardized root-mean-square residual, RMSEA = root mean-square-error of approximation.

The results of the CFA are presented in Table 1, showing a one-factor, two-factor, bifactor, and modified two-factor model. The one-factor model had unacceptable fit on three fit indices ($\chi^2/df$, SRMR, and RMSEA) while CFI was mediocre. The two-factor model showed unacceptable fit according to $\chi^2/df$, but acceptable CFI, SRMR, and RMSEA. All standardized loadings between items and corresponding two factors were above .5 (see Table S3 in the supplementary materials for specific loadings), i.e., no need to remove items. The standardized loading between the factors Vicarious Experience and Intuitive Understanding in the two-factor model was .52.

The bifactor model showed slightly better fit than the two-factor model, however, still with unacceptable fit according to $\chi^2/df$, and acceptable CFI and RMSEA, but good fit according to SRMR. The psychometric indices for the bifactor model showed that internal reliability was high for both the general factor and specific factors with $\omega = .97$, and $\omega_S$ (Vicarious Experience) = .95 and $\omega_S$ (Intuitive Understanding) = .96. Considering the reliable systematic variance, the general factor did not show support for unidimensionality as $\omega_H = .69$, i.e., below the threshold .80. Variability left to be explained by specific factors showed $\omega_{HS}$ (Vicarious Experience) = .10 and $\omega_{HS}$ (Intuitive Understanding) = .68. Both factor determinacy and construct replicability were good for the general factor (FD = .97; *H* = .96) and Intuitive Understanding (FD = .97; *H* = .93). Factor determinacy for Vicarious Experience was on the threshold for good, and construct replicability was acceptable (FD = .90; *H* = .72). When the final two indices assessing unidimensionality of the general factor (ECV and PUC) was considered, we did not find support for unidimensionality (ECV = .53 and PUC = .52). Specific factors showed $ECV_S$ (Vicarious Experience) = .10 and $ECV_S$ (Intuitive Understanding) = .37. As none of these psychometric indices supported unidimensionality of the general factor, we suggest that a multidimensional two-factor model is more suitable than a bifactor model for the EES, leading us to consider the two-factor model for possible modifications.

Modification indices for the two-factor model showed several suggestions that would decrease $\chi^2$ values substantially by covarying residuals of items within the same latent


© 2024 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

**Table 2.** Results of the multigroup tests of invariance between gender, nationality, and age for the modified two-factor model of the EES

| Model | $\chi^2$ (*df*) | CFI | ΔCFI | SRMR | ΔSRMR | RMSEA | ΔRMSEA |
|---|---|---|---|---|---|---|---|
| Gender invariance | | | | | | | |
| Configural | 1,320.507 (798) | .931 | — | .054 | — | .041 | — |
| Metric | 1,116.243 (826) | .962 | .031 | .055 | .001 | .030 | −.011 |
| Scalar | 1,155.561 (854) | .960 | −.002 | .056 | .001 | .030 | .000 |
| Residual | 1,182.939 (884) | .961 | .001 | .057 | .001 | .029 | −.001 |
| Nationality invariance | | | | | | | |
| Configural | 1,450.458 (798) | .913 | — | .052 | — | .046 | |
| Metric | 1,255.429 (826) | .943 | .030 | .056 | .004 | .036 | −.010 |
| Scalar | 1,295.931 (854) | .941 | −.002 | .057 | .001 | .036 | .000 |
| Residual | 1,316.517 (884) | .942 | .001 | .058 | .001 | .035 | −.001 |
| Age invariance | | | | | | | |
| Configural | 1,492.537 (798) | .911 | — | .052 | — | .047 | |
| Metric | 1,268.484 (826) | .943 | .032 | .056 | .004 | .037 | −.010 |
| Scalar | 1,335.841 (854) | .938 | −.005 | .057 | .001 | .037 | .000 |
| Residual | 1,376.823 (884) | .937 | −.001 | .059 | .002 | .037 | .000 |

*Note*. Δ values are with respect to the previous level of measurement invariance.

factors. Two of these were between IU10 and IU13 (modification index = 182.678), and VE8 and VE9 (modification index = 56.648), which both also were covaried in the study for the original Italian scale (Innamorati, et al., 2019). As the Swedish items were also judged as very similar, they were allowed to covary. Three other pairs that had high modification indices that were allowed to covary were VE3 and VE5 (modification index = 125.231), VE1 and VE6 (modification index = 94.649), and VE4 and VE7 (modification index = 48.054). Those were also judged to have very similar meaning, even more so than the same items in the English version (Innamorati et al., 2019). With these modifications, there were still some pairs of items (all within same factors) that could have correlated errors, but the modification indices suggested only minor improvement in model fit compared to the already modified item pairs. After modifications, the modified two-factor model showed acceptable fit according to $\chi^2/df$, SRMR, and RMSEA, and good fit for CFI. This model had fairly similar indices as the bifactor model, but with slightly better $\chi^2/df$. As the psychometric indices from the bifactor model did not support unidimensionality, we considered the modified two-factor model for assessment of measurement invariance.

Estimates for the multigroup tests (gender, nationality, and age) of measurement invariance are shown in Table 2. The results showed that configural invariance was mediocre for CFI, acceptable for SRMR, and good for RMSEA on all grouping variables. For subsequent steps of invariance testing comparing gender (with conservative cutoffs), and nationality and age (with less conservative cutoffs), changes of all fit indices (CFA, SRMR, and RMSEA) were within thresholds. Thus, the results of tests of measurement invariance showed that there was similar pattern of factor loadings (configural invariance), equal contribution from items to latent factors (metric invariance), equally shared variance of items from latent factors (scalar invariance), and comparable item and error variance (residual variance).

As shown in Table 3, both factors in the EES showed strong support for reliability as CR was substantially larger than .7. The standardized factor loadings of the items in Vicarious Experience were between .56 and .81, yielding convergent validity both for the criterion that factors loadings should be ≥ .5 and with AVE almost reaching the .5 cutoff. Intuitive Understanding met the criterion both for convergent validity with standardized factor loadings between .62 and .88, and AVE > .5. Finally, both factors showed good discriminant validity as AVE > MSV for both factors.

**Table 3.** Psychometric values stemming from the CFA of the modified two-factor model of the EES

| Factor | CR | AVE | MSV |
|---|---|---|---|
| Vicarious experience | .93 | .49 | .28 |
| Intuitive understanding | .96 | .63 | .28 |

*Note*. CR = composite reliability, AVE = average variance extracted, MSV = maximum shared variance.

## Full Sample Analysis: Correlations, Mean Values, and Nested Subsample Test

We combined the whole sample of the study ($N$ = 805) to assess correlations between the summated subscales as well as mean values of them. The correlation between

© 2024 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)


https://econtent.hogrefe.com/doi/pdf/10.1027/2698-1866/a000079 - Friday, September 20, 2024 6:23:20 AM - IP Address:158.162.233.152

https://econtent.hogrefe.com/doi/pdf/10.1027/2698-1866/a000079 - Friday, September 20, 2024 6:23:20 AM - IP Address:158.162.233.152

Vicarious Experience ($M$ = 2.89, $SD$ = 0.78) and Intuitive Understanding ($M$ = 3.65, $SD$ = 0.67) was $r$ = .50.

In line to test whether our two nested subsamples in the form of different nationalities (mainly from Sweden and Finland) differ in their ratings of the EES, we conducted intraclass correlation analyses using intraclass correlation coefficient (ICC; 2, k) models separately for the two nationalities on all 30 items. The ICC for the Swedish population ($N$ = 567) was .91 with a 95% confidence interval from .90 to .93, whereas the ICC for the Finnish population ($N$ = 221) was .92 with a 95% confidence interval from .90 to .94. As the confidence intervals of the two subsamples were overlapping, data suggest that we cannot assume that they differ in their ratings on the EES.

Finally, independent sample t-tests were used to compare means between genders for the two subscales, showing that for Vicarious Experience women ($M$ = 2.95, $SD$ = 0.77) scored higher compared with men ($M$ = 2.43, $SD$ = 0.68), $t(793)$ = 6.24, $p$ < .001, $d$ = 0.68. Also, for Intuitive Understanding, women ($M$ = 3.70, $SD$ = 0.66) scored higher than men ($M$ = 3.30, $SD$ = 0.70), $t(793)$ = 5.51, $p$ < .001, $d$ = 0.60.

## Discussion

Consistent with previous research (Innamorati et al., 2019), our findings supported a two-factor structure comprising Vicarious Experience and Intuitive Understanding. CFA in combination with psychometric indices of a bifactor model indicated that the EES did not support unidimensionality, and therefore, the two-factor model was favored. In contrast to Innamorati et al. (2019) who relied on judgment of item-factor correlations to reject the bifactor model, we utilized prespecified indices with cutoffs for assessing unidimensionality. Despite this methodological difference, we arrived at a similar conclusion as a reliable amount of variance of the EES does not stem from a general factor. Therefore, it may be more meaningful to consider the Vicarious Experience and Intuitive Understanding dimensions separately, and together with Hall and Schwartz (2019) suggesting the use of more specific, lower-level constructs rather than the overarching concept of empathy. By recognizing empathy as a multifaceted construct compromising unique components, we contribute to the ongoing advancement of empathy theory, paving the way for more nuanced and comprehensive research in the field. As earlier highlighted by Innamorati et al. (2019), the EES's distinctiveness in capturing the intuitive and effortless aspects of empathy is also emphasized through our study. With our replication of the findings from Innamorati et al. (2019) using Intuitive Understanding, an effortless cognitive component of empathy, the results underscore the importance of these automatic processes in empathic responses. Nonetheless, we acknowledge that the cognitive component can also be manifested as more deliberate and effortful, as demonstrated in other studies (Cuff et al., 2016).

We modified the two-factor model (based on modification indices and judgment of similar meaning of different items) by covarying residuals. We correlated residuals of the same two item pairs as in Innamorati et al.'s (2019) study, but also three other item pairs. This resulted in the modified two-factor model that showed a good fit for the CFI and acceptable fit for all other fit indices. Some caution of our results should be taken, as we used WLSMV estimator for the CFAs, which is known to yield better fit indices of both CFI and RMSEA compared to estimators based on Maximum Likelihood (Xia & Wang, 2019). However, we compared our CFA models with WLSMV estimator versus using robust Maximum Likelihood (MLR) estimator with Satorra–Bentler scaled test statistic, and the only substantial difference was for CFI which was better with WLSMV for all models (see the results from MLR in in the supplementary materials Table S4 and Table S5). Given that several individual items within the same latent factors of the EES shared both similar meaning and content, a potential avenue for improvement could involve evaluating the suitability of removing redundant items from the scale. For instance, item response theory (IRT), a methodology commonly employed in the development of shorter scales, could be utilized to assess the appropriateness of item removal (Clark & Watson, 2019). Removing redundant items has been demonstrated to increase model fit of scales (see, e.g., Brucato et al., 2023) could therefore possibly both make the scale more parsimonious and facilitate an increased fit.

It was expected that the two factors in the EES would be moderately correlated, but the magnitude of the correlation between the factors (.50) was somewhat stronger than the correlations (.32 and .37) found previously (Ebisch et al., 2022; Innamorati et al., 2019). This larger correlation can be attributed to either greater item similarity between the factors in the Swedish translation, or inherent variations within the population under study.

Women scored higher than men on Vicarious Experience (consistent with Ebisch et al., 2022). However, unlike the findings by Ebisch et al. (2022) who did not find a difference between sexes on Intuitive Experience, our results showed that women also scored higher on this dimension. Gender differences on both empathic dimensions are in line with previous findings that women have higher empathy levels than men, which can be explained by both biological and social factors (Christov-Moore et al., 2014).

 © 2024 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2698-1866/a000079 - Friday, September 20, 2024 6:23:20 AM - IP Address:158.162.233.152

Despite the evidence of differences in empathy levels between genders, the tests of measurement invariance showed that neither genders were invariant with respect to factor structure, nor the items' contribution, variance, and error variance on factors, which was also the case when comparing groups of both nationality and age. These findings differ from the Italian sample in Innamorati et al.'s study (2019), where comparisons between men and women showed configural invariance but not metric invariance, meaning that the Swedish scale shows more equal patterns between genders.

The modified two-factor model of the EES demonstrated high internal consistency, consistent with prior research (Innamorati et al., 2019). Unlike Innamorati et al. (2019), this study assessed the convergent and discriminant validity of the EES within the scale itself, rather than comparing it to other similar measures. Both discriminant and convergent validity were satisfactory for both factors. By establishing convergent and discriminant validity within the factors itself, this study provides stronger psychometric support for the factor structure rather than comparing it to other measures. However, it is important to acknowledge that our study did not directly assess the validity of the scale in relation to other similar measures, which could be a potential pathway for future research.

One limitation of the study is the use of convenience sampling to recruit participants, which may limit the generalizability of the results to the target populations (adults in Sweden and Finland). While using multiple recruitment channels to reach out to a wide group of people, some uneven distribution of participants was observed, such as a higher proportion of both women and Swedish participants. The use of stratified sampling could have addressed this limitation. However, the results from the ICCs of the two nationalities indicate no difference between the ratings, which mitigates this limitation to some extent. Also, while testing the EES in this study, the cross-sectional data only allowed reliability testing for the subscales at one point in time, and longitudinal data could use a test–retest method to expand knowledge about how reliable the individual score of the EES is over time.

To conclude, our findings from the Swedish version of the EES support a two-factor structure consistent with previous research (Innamorati et al., 2019), but also support for measurement invariance for gender, nationality, and age. Using standardized indices for assessing unidimensionality, our study provided more extensive psychometric support for the measure. We demonstrated high internal consistency, good discriminant validity, and good convergent validity within the measure itself, building on findings of validity compared to other measures (Innamorati et al., 2019). However, the modified two-factor model only reached the threshold for good according to one fit index (CFI) and gave three adequate fit indices ($\chi^2/df$, SRMR, and RMSEA), prompting a discussion to revise the scale to further increase fit, possibly by removing redundant items using IRT.

## References


Brauer, K., Ranger, J., & Ziegler, M. (2023). Confirmatory factor analyses in psychological test adaptation and development: A nontechnical discussion of the WLSMV estimator [Editorial]. *Psychological Test Adaptation and Development*, *4*(1), 4–12. https://doi.org/10.1027/2698-1866/a000034

Brucato, M., Frick, A., Pichelmann, S., Nazareth, A., & Newcombe, N. S. (2023). Measuring spatial perspective taking: Analysis of four measures using item response theory. *Topics in Cognitive Science*, *15*(1), 46–74. https://doi.org/10.1111/tops.12597

Chen, F. F. (2007). Sensitivity of goodness of git indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience & Biobehavioral Reviews*, *46*(4), 604–627. https://doi.org/10.1016/j.neubiorev.2014.09.001

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, *31*(12), 1412–1427. https://doi.org/10.1037/pas0000626

Coll, M-P., Viding, E., Rütgen, M., Silani, G., Lamm, C., Catmur, C., & Bird, G. (2017). Are we really measuring empathy? Proposal for a new measurement framework. *Neuroscience and Biobehavioral Reviews*, *83*, 132–139. https://doi.org/10.1016/j.neubiorev.2017.10.009

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. *Emotion Review*, *8*(2), 144–153. https://doi.org/10.1177/1754073914558466

Decety, J. (2010). The neurodevelopment of empathy in humans. *Developmental Neuroscience*, *32*(4), 257–267. https://doi.org/10.1159/000317771

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100. https://doi.org/10.1177/1534582304267187

Dueber, D. M. (2017). *Bifactor indices calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models*. https://doi.org/10.13023/edp.tool.01 [Available at http://sites.education.uky.edu/apslab/resources/]

Ebisch, S. J. H., Scalabrini, A., Northoff, G., Mucci, C., Sergi, M. R., Saggio, A., Aquino, A., Alparone, F. R., Perrucci, M. G., Gallese, V., & Di Plinio, S. (2022). Intrinsic shapes of empathy: Functional brain network topology encodes intersubjective experience and awareness traits. *Brain Sciences*, *12*(4), Article 477. https://doi.org/10.3390/brainsci12040477

Gerdes, K. E., Segal, E. A., & Lietz, C. A. (2010). Conceptualising and measuring empathy. *The British Journal of Social Work*, *40*(7), 2326–2343. https://doi.org/10.1093/bjsw/bcq048

Hair, J., Black, W., Babin, B., & Anderson, R. (2014). *Multivariate data analysis* (7th ed.). Pearson Education Limited.

Hall, J. A., & Schwartz, R. (2019). Empathy present and future. *The Journal of Social Psychology*, *159*(3), 225–243. https://doi.org/10.1080/00224545.2018.1477442

Innamorati, M., Ebisch, S. J. H., Gallese, V., & Saggino, A. (2019). A bidimensional measure of empathy: Empathic Experience Scale. *PLoS ONE*, *14*(4), Article e0216164. https://doi.org/10.1371/journal.pone.0216164

JASP Team. (2023). *JASP* (Version 0.16.4) [Computer software]. https://jasp-stats.org/


© 2024 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)

https://econtent.hogrefe.com/doi/pdf/10.1027/2698-1866/a000079 - Friday, September 20, 2024 6:23:20 AM - IP Address:158.162.233.152

Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, *9*(8), 2207–2230. https://doi.org/10.4236/psych.2018.98126

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team. (2021). *R: A language and environment for statistical computing (version 4.1.0)* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Rodriguez, A., Reise, S. P., & Haviland, M., G. (2015). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Sarling, A. (2024). *Psychometric evaluation of the Swedish version of the Empathic Experience Scale (EES)* [Open data]. https://osf.io/fr7wz/

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74. https://doi.org/10.23668/psycharchives.12784

Sperber, A. D. (2004). Translation and validation of study instruments for cross-cultural research. *Gastroenterology*, *126*(1), S124–S128. https://doi.org/10.1053/j.gastro.2003.10.016

Williams, M. J., Dalgleish, T., Karl, A., & Kuyken, W. (2014). Examining the factor structures of the Five Facet Mindfulness Questionnaire and the Self-Compassion Scale. *Psychological Assessment*, *26*(2), 407–418. https://doi.org/10.1037/a0035566

Xia, Y., & Wang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409–428. https://doi.org/10.3758/s13428-018-1055-2

**History**
Received July 25, 2023
Revision received March 18, 2024
Accepted May 17, 2024
Published online July 16, 2024

Section: Positive Psychology & Assessment

**Conflict of Interest**

The authors have no competing interests to declare.

**Open Science**

All relevant materials for Open Data, Open Materials, Open Analytic Code, and supplementary materials are available at https://osf.io/fr7wz/ (Sarling, 2024).

Open Data: The data necessary to reproduce all reported results are publicly accessible.

Open Materials: Detailed materials used in this study are available to verify the methodology.

Open Analytic Code: The codebook and all code in R for CFA, required to replicate the reported analyses, have been made available. Calculations for bifactor indices as well as reliability and validity are also accessible.

Study and Analysis Plan Preregistration: Neither the overall study nor the analysis plan was preregistered.

**ORCID**
Andreas Sarling
https://orcid.org/0000-0002-9033-6303
Örjan Sundin
https://orcid.org/0000-0002-4116-5501

**Andreas Sarling**
Department of Psychology and Social Work (PSO)
Mid Sweden University
Kunskapens väg 1
831 25 Östersund
Sweden
andreas.sarling@gmail.com


© 2024 The Author(s). Distributed as a Hogrefe OpenMind article under the license CC BY 4.0 (https://creativecommons.org/licenses/by/4.0)