Psychological Test and Assessment Modeling, 2026.68:53-76
DOI: https://doi.org/10.2440/001-0026

Self-reported face recognition ability – an adapted German version of the Stirling Face Recognition Scale

Lara Aylin Petersen[1,2], Anja Leue[1]

[1] Kiel University, Kiel, Germany
[2] Leibniz Institute for Science and Mathematics Education, Kiel, Germany

***Abstract**:*
With the discovery of Super-Recognizers, research on questionnaires measuring self-reported face recognition abilities has been inspired. By adapting and translating the Stirling Face Recognition Scale (SFRS; Bobak et al., 2018) into German we investigated psychometric properties, reliability, and factorial validity of the SFRS-German. The sample (female: n = 220 and male: n = 154) comprised mainly adult European Caucasian (95.70%) participants. Ethnicity was based on self-selection without any exclusion of other ethics groups. The SFRS-German shows excellent reliability coefficients. The best model fit has been observed for a bifactor Multiple-Indicator-Multiple-Cause (MIMIC) model including Face Processing and Face Memory as primary-order factors and Face Recognition ability as a second-order factor. The MIMIC factor Gender indicated measurement invariance. Medium correlations occurred between the SFRS-German and the CFMT+. Future research might address the nomological network of self-reported face recognition ability in relation to individual differences of stimulus processing, attentional abilities, or temperament.

Keywords: *Self-report, face recognition, factorial validity, psychometric properties*

Correspondence:

Prof. Dr. Anja Leue, Christian-Albrechts-Universität zu Kiel, Institute of Psychology, Chair of Psychological Assessment, Differential and Personality Psychology, Olshausenstr. 75, 24118 Kiel, Germany, E-Mail: leue@psychologie.uni-kiel.de, ORCID ID 0000-0002-2588-5226

Introduction

Face recognition performance has been often investigated by means of computer-based tests like the Cambridge Face Memory Test – Long Form (CFMT+, Russell et al., 2009; Petersen & Leue, 2021). Self-reported insights into face recognition abilities have been often discussed in the context of prosopagnosia (low face recognition performance; Matsuyoshi & Watanabe, 2021; Shah et al., 2015; Tsantani et al., 2021). Based on Russell et al. (2009), Super-Recognizers (SR) with their extraordinary face recognition abilities have been introduced in face recognition research for the first time (e.g., Bate et al., 2018; Davis et al., 2020; Fysh et al., 2020; McCaffery et al., 2018; Petersen & Leue, 2021). Although face recognition abilities have been reported for a low and a high end of performance, self-reported face recognition ability has been rarely investigated for the entire ability spectrum (i.e., as a continuum). Questionnaires like the Stirling Face Recognition Scale (SFRS; Bobak et al., 2018) have been developed to assess self-reported face recognition ability. A face recognition questionnaire that has been published in several languages might be useful for research on construct validity, for research on criterion validity (e.g., personnel recruitment of police officers), and individual differences in face recognition. Moreover, testing the reliability and the construct validity of a scale (Nunnally & Bernstein, 1994; Campbell & Fiske, 1959; Cronbach & Meehl, 1955) is important to ensure the quality of scores (American Educational Research Association et al., 2014). To extend the research for the SFRS in a German sample, the aim of our study was to investigate psychometric properties, reliability coefficients, and the factorial structure of an adapted German SFRS. In addition, the adapted German SFRS was correlated with an objective face recognition measurement (e.g., CFMT+, Russell et al., 2009).

Self-report questionnaires for face recognition ability

In face recognition research, questionnaires for measuring self-reported face recognition ability are used to explore how well individuals can describe their face recognition abilities (e.g., Livingston & Shah, 2018; Palermo et al., 2017). Other studies argued in favor of self-report questionnaires as a cost-effective assessment to identify Super-Recognizers (e.g., Arizpe et al., 2019; Bate & Dudfield, 2019). In prior research, single items (Bobak et al., 2016; Verhallen et al., 2017) and multiple item questionnaires (Arizpe et al., 2019; Bate & Dudfield, 2019; Palermo et al., 2017; Shah et al., 2015) were used to measure self-reported face recognition abilities. In the present study, we focus on the investigation of a German translation of the SFRS. Bobak et al. (2018) developed the 20-items SFRS including two scales entitled as face memory and face processing. The original SFRS included ten adapted items of the often used 20-item prosopagnosia index questionnaire (PI-20; Estudillo & Wong, 2021; Gehdu et al., 2024; Kramer, 2023; Nakashima et al., 2020; Shah et al., 2015; Stantic et al., 2022; Tsantani et al., 2021). The item content has been extended by means of ten items resulting in ten old and ten new items for the SFRS. The SFRS is suitable for measuring a continuum of self-reported face recognition abilities from prosopagnosia to SR ability (Bobak et al., 2018). Bobak et al. (2018) reported a significant positive correlation ($r(96) = .36$, $p < .01$, their Table 3) between the English SFRS and the CFMT+ (Russell et al., 2009) to discuss the relation of subjective and objective face recognition measurements. Face recognition ability as a construct can be described in multiple ways. For example, Bruce and Young (1986) used the term "face recognition" in "a broad sense, covering the derivation of any type of stored information from faces […] to include what might well be called identification or retrieval of personal information.". Therefore, face recognition ability would be the ability to store pictures of faces and to recall or recognize whether faces are known or unknown.

Table 1 gives an overview of validity studies reporting the factors or components of self-reported face recognition ability and their reliability. Further psychometric properties, such as mean inter-item correlations or part-whole corrected item total correlations have not yet been reported for self-report inventories of face recognition. For the adapted German version of the SFRS we aimed at investigating psychometric properties.

The factor structure of the Stirling Face Recognition Scale

Considering Table 1, exploratory and confirmatory factor analyses as well as principal component analyses were performed to investigate the factor structure of the SFRS and similar questionnaires. These analyses showed one to three factors or components depending on the questionnaire. The English (Shah et al., 2015) and Mandarin PI-20 (Sun et al., 2021) consistently showed an interpreted single factor solution of self-reported face recognition ability, whereas the Super-Recognizer Questionnaire (Bate & Dudfield, 2019) represents three components of face recognition ability (Table 1). Bobak et al. (2018) as well as Ritchie et al. (2022) reported a two-factor solution for the SFRS. Based on a PCA, Bobak et al. (2018) suggested that *Component 1* reflects poor face recognition (face memory and general face processing items) and *Component 2* reveals good face recognition (especially face memory items). A conference poster of Oren et al. (2019) reported also a two-factor solution for the Turkish version of the SFRS. As no details of the factor analysis are described on the poster and no journal article has been published for these results, we did not include the Turkish version of the SFRS in Table 1. Overall, the current results on factorial validity presume primary-order factor models with two factors for the SFRS, whereas one factor solution has been obtained for the PI-20 (ten identical items with the SFRS). As the present study investigated a German translation of the SFRS we started the factorial validation with two two-factor models in accordance with Bobak et al. (2018) who factorially investigated the English SFRS. In addition to the two factor solutions, the present study explores evidence of a bifactor model with Face Recognition as a second-order factor and Face Memory as well as Face Processing as primary-order factors.

Table 1.
Overview of studies reporting the factorial validity and reliability of self-reported face recognition ability questionnaires.

Name of questionnaire	Language	Authors, year	Sample size	Gender and age (in years) characteristics	No. of items	Type of factor analysis with type of rotation	No. of factors	Title of factors	Reliability
20-item prosopagnosia index (PI20), English Original Version	English	Shah, Gaule, Sowden, Bird & Cook, 2015	319	202 females, *M* age = 33.0 years	20	EFA with Varimax rotation	1	Face recognition ability (61% variance)	Cronbach's Alpha = .96
Mandarin 20-item prosopagnosia index (Mandarin PI20)	Mandarin	Sun, Wang, Wang & Luo, 2021	647	464 females, *M* age = 22.11 years (*SD* = 3.41), range 18-45	20	CFA single factor model	1	Face recognition ability	Cronbach's Alpha = .93; test-retest (one month interval) = .88

Table 1 continued

Super-Recognizer Questionnaire (SRQ)	English	Bate & Dudfield, 2019	264 (Exp. 1)	181 females, *M* age = 37.2 (*SD* = 7.7), range 18-50	20	PCA with Varimax rotation	3	Face memory (variance = 27.35%), Spotting faces in a crowd (variance = 8.23%), face matching (variance = 6.81 %)	Cronbach's Alpha = .85, Split-Half with Spearman-Brown = .79
Stirling Face Recognition Scale (SFRS)	English	Bobak, Mileva & Hanock, 2018	165 (3 samples combined)	118 females, *M* ages of the samples = 23-26; age range overall 18-58 years	20	PCA with oblimin rotation	2	Poor at Face recognition (variance 32.04 %), good at face recognition (variance 8.54 %; especially face memory questions)	Cronbach's Alpha = .88, Split-Half with Spearman-Brown = .89
Stirling Face Recognition Scale (SFRS)	English	Ritchie, Flack, Fuller, Cartledge & Kramer, 2022 Supplement	80 (2 samples combined)	S1: 29 females, *M* age = 23, range 18-48; S2: 20 females, *M* age = 20; range 18-24	20	PCA with oblimin rotation	2	Poor at Face recognition (variance 31.7%), good at face recognition (variance = 8.5%)	-

Aims and hypotheses

Regarding psychometric parameters we disentangle hypothesis 1 for the item difficulty index and for the part-whole corrected item total correlation: By investigating the item difficulty index (Moosbrugger & Kelava, 2020, formula 7.2) we presume moderate item difficulty indices between 20 and 80 (hypothesis 1-1). We expect moderate part-whole corrected item total correlations of at least .30 (hypothesis 1-2). We expect moderate (i.e., at least .80) internal consistency coefficients of Cronbach's Alpha and Split-half reliability (hypothesis 2). For benchmarks of the psychometric properties and reliability we refer to Bühner (2011) as well as Nunnally and Bernstein (1994). To investigate the factorial validity of the German translation of the SFRS (hypothesis 3), we expect the best model fit for a two-factor solution of the German SFRS including Face Processing and Face memory as reported in Bobak et al. (2018). As no intercorrelation between Face Processing and Face Memory has been reported in Bobak et al. (2018) for the oblimin rotation, we test a two-factor orthogonal model and a two-factor model with an intercorrelation of .30. In addition, we investigate a bifactor model. The bifactor model presumes self-reported face recognition ability as a second-order factor, whereas Face Processing and Face Memory are conceived as two primary-order factors. With regard to Bobak et al. (2018), we also calculated the correlation between the German SFRS and the CFMT+ (Russell et al., 2009).

Method

The authors report there are no competing interests to declare. The project had no special funding opportunities or a funding grant to describe. The study confirms the ethical principles of the Helsinki declaration (2013, 2024): https://www.wma.net/policies-post/wma-declaration-of-helsinki/. The data that support the findings of this study are available at OSF | Self-reported face recognition ability - German SFRS. The present manuscript fulfills the Data Transparency: Level 1. The study design had been previewed as a kind of pre-registration based on an exposé for a diploma thesis and a dissertation thesis at the Chair of the second author.

Participants

A total of $N = 377$ participants (220 females, 58.40 %) with a mean age of 44.41 years ($SD = 10.11$, *range* 25-81 years) took part in this study. The sample consisted mostly of people with a higher educational level (76.30%). The participants were recruited via email. They had previously participated in face recognition studies of our lab. Prior to examination, all participants gave written informed consent and said that their data (e.g., for the SFRS) could be included in an anonymized manner in further publications. Three participants showed more than five missing values for the SFRS and, thus, were excluded for statistical analysis leaving a final sample of $N = 374$ participants ($n = 219$

female, n = 155 male, aged: M = 44.34 years, SD = 10.09, range: 25 to 81 years). These participants mainly classified themselves as European Caucasian (95.70%). The examination took place online in March 2022 with SoSci Survey. For the present sample we provide an example of a post-hoc power calculation using G*Power. The included parameters were as follows: Exact - Correlation: Bivariate normal model, Options: exact distribution, Analysis: Post hoc: Compute achieved power, Input: Tail(s) = Two, Correlation ρ H1 = 0.3 (see Results section for the German SFRS and CFMT+), α err prob = 0.05, Total sample size = 374, Correlation ρ H0 = 0, Output: Lower critical r = -0.1014253, Upper critical r = 0.1014253, Power (1-β err prob) = 0.9999680.

Materials

Stirling Face Recognition Scale (SFRS; Bobak et al., 2018) – Adapted German SFRS

The scale comprises 20 face recognition statements assessing the self-reported face recognition ability. Bobak et al. (2018) adapted the statements of the 20-item Prosopagnosia-Index (PI-20; Shah et al., 2015) and developed the SFRS with two scales entitled as self-reported face memory and self-reported face processing abilities. The items of the English SFRS have been translated - with permission of Dr. Anna Bobak- by using a forward method into German language. In addition, the German items have been back translated by a professional translation office in January 2023 (see guidelines of the International Test Commission Version 2.4: guideline_test_adaptation_2ed.pdf, ITC: TD-1, section 4, TD-2, section 5). The translation of the English SFRS items into German language and their back translation was performed based on the guidelines of the International Test Commission (2017). In purpose of comparison, the original English items, the translated German items, and the back translation of the German items into the English items has been summarized in Table 2. Moreover, item labels and indications of items that should be recoded prior to statistical analysis are reported in Table 2. Brief descriptions in Table 2 highlight the variations between both translations. Overall, the differences of the item formulations between the forward translation and the backward translation (Table 2, column 2) can be summarized as slight variations. The response categories of the items were *strongly disagree* (numerical answer category = 1), *rather not agree* (numerical answer category = 2), *neutral* (numerical answer category = 3), *rather agree* (numerical answer category = 4), *strongly agree* (numerical answer category = 5). The answer format was, thus, applied with regard to ITC guideline TD-4, section 7. The minimum raw sum score across all 20 items is 20. The maximum raw sum score across all 20 items is 100. We used a 5-point rating scale to test psychometric properties and factorial validity with a smaller range of answer categories than reported in Bobak et al. (2018). Previous studies suggested that reliability coefficients can be excellent (i.e., ≥.90) with even five answer categories of a rating scale (Preston & Colman, 2000; Revilla et al., 2014). The original SFRS has a 7-point rating scale. The SFRS includes ten items formulated in the direction of the construct definition (i.e., higher face memory and face processing performance). Ten items mainly measuring face processing are formulated against the direction of the construct definition. These items are recoded prior to statistical analysis (Table 2). There was no time limit for answering the items. Participants were asked to answer the items as honestly as possible. The reliability of the SFRS was good in Bobak et al. (2018) with Cronbach's Alpha = .88 and Split-Half Spearman-Brown corrected reliability = .89 (Bühner, 2011; Nunnally & Bernstein, 1994).

Cambridge face memory test long (CFMT+; Russell et al., 2009)

The CFMT+ is a standardized computerized test for short-term face memory. First, participants learn six male target faces (gray-scaled). Afterwards, they have to recognize the target faces (retrieved from memory) out of three presented faces (forced choice). Each correct decision gets one point (102 items = maximal 102 points). The difficulty of the presented items increases over four sections. The CFMT+ is not time-limited and can be completed in 20 minutes on average. Cronbach's Alpha has been reported to be moderate with $\alpha = .89$ (Bate et al., 2021) to excellent with $\alpha = .92$ (Petersen & Leue, 2021). Further description and examples of the items are reported in Russell et al. (2009). Moreover, norms and psychometric properties for the CFMT+ in German samples have been reported in Petersen & Leue (2021) and Petersen & Leue (2022).

Procedure

The data were collected online via SoSci Survey, a web-based platform for surveys. First, participants gave written informed consent for their participation in a diploma thesis (Rieger, 2021) that tested a newly developed face recognition test. All participants answered demographic questions in the first part of the study, performed the new face recognition test in the second part of the study. Finally, all participants filled in the SFRS items. In the paradigm participants were asked to recognize faces in several contexts (data not presented in this study). The new face recognition test lasted about 25 minutes. The SFRS questionnaire could be answered in about 10 minutes. The participants were not paid. Their reward for study participation was the result of the correct face recognitions in the above-mentioned paradigm. As the participants were part of another study and had previously completed the CFMT+ (with agreement to use data and took part on further studies), we were able to match the new data from this study and their CFMT+ results via their participant code.

Table 2.
English and German forward translations as well as backward translations of SFRS items.

No.	German SFRS translation (2018)	Description of variations between forward and back translation	Variable labels based on Bobak et al. (2018, Tab. 2)
1	Ich weiß genau, wo ich jemanden zum ersten Mal getroffen habe (z.B. bei Arbeitstreffen, Partys etc.)	slight change in word sequence	F1_M
2	In meiner Schulzeit fiel es mir schwer, meine Klassenkammeraden wiederzuerkennen.	slight variation in wording	F2_P®
3	Es fällt mir sehr leicht, mir einzelne Gesichter detailreich in Gedanken vorzustellen.	slight change in word sequence, slight variation in wording	F3_M
4	Die Angst davor, Gesichter schlecht wiederzuerkennen, hat dazu geführt, dass ich bestimmte soziale oder berufliche Situationen vermeide.	slight variation in wording	F4_P®
5	Ich bin besser als die meisten anderen Personen darin, Namen Gesichtern zuzuordnen.	British-English to American-English writing style, quotation missing	F5_M
6	Ich halte Personen, die ich schon einmal getroffen habe, fälschlicherweise oft für Fremde.	slight variation in wording	F6_P®
7	Ich neige dazu, Gesichter sehr schnell wieder zu vergessen.	less contextualized in back translation	F7_M/P®
8	Mir fällt es sehr leicht, bekannte Personen wiederzuerkennen, wenn ich sie in einem ungewohnten Kontext treffe (z.B. Arbeitskollegen unerwartet im Fitnessstudio treffen).	slight variation in wording and American-English writing to British-English writing	F8_M
9	Gesichter genügen mir, um Personen wiederzuerkennen – ich muss nicht ihre Stimme hören oder ihren gesamten Körper sehen.	change from plural to singular object wording and American-English writing to British-English writing	F9_M/P
10	Wenn Personen ihren Haarschnitt verändern oder Hüte tragen, habe ich Schwierigkeiten, sie wiederzuerkennen.	slight variation in wording and American-English writing to British-English writing	F10_P®
11	Ich bin sehr gut darin, in einer Gruppe von unbekannten Personen miteinander verwandte Personen zu erkennen.	slight variation in wording and American-English writing to British-English writing	F11_M
12	Manchmal muss ich neu kennen gelernte Personen warnen, dass ich schlecht im Wiedererkennen von Gesichtern bin.	slight change in word sequence, British-English writing in back translation, quotation lost	F12_P®
13	Ich erkenne oft Personen wieder, die ich erst kurz zuvor getroffen habe, sie haben jedoch keine Idee, wer ich bin.	slight variation in wording	F13_ M

Table 2 continued.

No.	German SFRS translation (2018)	Description of variations between forward and back translation	Variable labels based on Bobak et al. (2018, Tab. 2)
14	Ich muss mich mehr anstrengen als andere Personen, um mir Gesichter zu merken.	slight variation in wording	F14_P®
15	Ich vergesse niemals ein Gesicht.	identical	F15_M/P®
16	Manchmal fällt es mir schwer, Filmen zu folgen, weil ich Schwierigkeiten habe, Charaktere wiederzuerkennen.	slight change in word sequence and word variation; British-English writing in back translation	F16_P®
17	Es fällt mir leicht, meine Familie und Freunde auf Kindheitsfotos wiederzuerkennen.	slight change in word sequence and less contextualized in forward translation	F17_M
18	Es fällt mir schwer, Menschen wiederzuerkennen ohne ihre Stimme zu hören.	slight change in word sequence, British-English writing in back translation, quotation lost	F18_P®
19	Ich erkenne oft jemanden nicht, der mich kennt.	slight variation in wording and American-English writing to British-English writing	F19_P®
20	Es fällt mir leicht, eine Person wiederzuerkennen, selbst wenn ich nur Teile ihres Gesichts sehen kann.	slight variation in wording and American-English writing to British-English writing; Singular to Plural change	F20_M

Note. Three items had secondary loadings in Bobak et al. (2018, Tab. 2: 7, 9, 15), therefore we labeled them M/P. ® indicates that an item should be recoded prior to statistical analysis because item content is not formulated in the direction of construct definition. For items that should be recoded initial answer categories are recoded as follows: 5 = 1, 4 = 2, 3 = 3, 2 = 4, 1 = 5. Items that have been coded as "_M/P" have been sorted to the latent factor FM when the factor loading in Table 2 of Bobak et al. (2018) was numerically higher compared to the factor loading on FP (see Table 3).

Statistical analysis

The descriptive item statistics were conducted with IBM SPSS Statistics (Version 28, for Mardia's test syntax in IBM SPSS 27 has been applied for Windows). Mplus 8.3 (Muthen & Muthen 1998-2019) was used to perform confirmatory factor analyses. Maximum Likelihood estimation (MLR) was applied as the estimator for the CFA models. Fit indices were evaluated based on Beauducel and Wittmann (2005) using LISREL as well as Hu and Bentler (1999) using EQS software. A very good model fit can be described for the Comparative Fit Index (CFI) with values between .90 and .95, for the Root Mean Square Error of Approximation (RMSEA) with values between .02 and .06, and for the Standardized Root Mean Square Residual (SRMR) with values of ≤ .05.

Results

Descriptive statistics and psychometric properties for the German SFRS items (hypotheses 1 and 2)

The item difficulty index indicates the tendency of a participant to agree with an item content given in a range between 0 and 100 (Moosbrugger & Kelava, 2020). The item difficulty index ranged between 61.25 (item 11 with lowest item difficulty index) to 95.75 (item 4 with highest item difficulty index, Table 3). A total of 14 out of the 20 items had a difficulty index between 20 and 80. That is, 70% of the items fulfill the criterion of a moderate item difficulty index (hypothesis 1-1). The part-whole corrected item total correlations ranged from .32 (item 2) to .78 (item 14, Table 3). Therefore, all items fulfill the cut-off criterion for a moderate part-whole corrected item total correlation (Bühner, 2011, hypothesis 1-2). Exploratorily, kurtosis and skewness of the items have been investigated. According to Moosbrugger and Kelava (2020), negative values for the kurtosis represent a platykurtic distribution (i.e., all or almost all answer categories have been chosen). Positive values for the kurtosis represent leptokurtic item distributions. In the present data, 4 of 20 items show a platykurtic item distribution. Negative values for the skewness suggest more frequent answers for right-sided answer categories (i.e., rather agreement). Positive values for the skewness reveal more frequent answers for left-sided answer categories (i.e., rather disagreement with item content). All items (including recoded ones) have negative values for the skewness, which suggests right-sided item distributions (i.e., rather agreement with item content).

Table 3.
Item statistics for the SFRS-German ($N = 374$).

No.	Abbreviation	Mean	*SD*	Item difficulty index†	Part-whole corrected item total correlations based on Pearson	Part-whole corrected item total correlation based on Spearman-Rank#	Kurtosis	Skewness
1	F1_M	3.89	.88	72.25	.53	.46	0.58	-0.86
2	F2_P®	4.78	.59	94.50	.32	.33	14.60	-3.52
3	F3_M	3.57	.95	64.25	.52	.49	-0.25	-0.42
4	F4_P®	4.83	.58	95.75	.47	.30	21.42	-4.38
5	F5_M	3.64	1.12	66.00	.50	.47	-0.44	-0.53
6	F6_P®	4.45	.77	86.25	.64	.55	2.66	-1.59
7	F7_M®	4.30	.88	82.50	.70	.67	1.80	-1.42
8	F8_M	4.17	.95	79.25	.60	.55	1.75	-1.39
9	F9_M	4.20	.84	80.00	.68	.61	1.91	-1.23
10	F10_P®	3.89	.88	72.25	.58	.55	0.45	-0.74
11	F11_M	3.45	.90	61.25	.43	.39	0.14	-0.35
12	F12_P®	4.70	.69	92.50	.72	.58	9.72	-2.92
13	F13_M	3.74	1.04	68.50	.54	.48	-0.21	-0.60
14	F14_P®	4.52	.79	88.00	.78	.65	5.42	-2.14
15	F15_M®	3.48	1.04	62.00	.55	.49	-0.27	-0.61
16	F16_P®	4.60	.71	90.00	.48	.38	6.21	-2.24
17	F17_M	4.12	.98	78.00	.32	.42	1.48	-1.30
18	F18_P®	4.36	.89	84.00	.47	.53	4.01	-1.89
19	F19_P®	4.25	1.01	81.25	.42	.47	2.41	-1.65
20	F20_M	3.54	.85	63.50	.63	.58	0.09	-0.32
Min.		3.45	.58	61.25	.32	.30	-0.44	-4.38
Max.		4.83	1.12	95.75	.78	.67	21.42	-0.32

Note. Abbreviation of the items with “F” referring to Face recognition. *Mean* represents a descriptive statistical parameter for item difficulty (range 1 to 5). † Item difficulty index based on the formula 7.2 in Moosbrugger and Kelava (2020). # Kolmogorov-Smirnov test indicated significant deviation from normal distribution for all 20 items. Thus, part-whole corrected item total correlations have been reported in addition to Spearman-Rank correlations.

Table 4 summarizes the reliability scores calculated for the present data. For the scale Face Memory (items 1, 3, 5, 7, 8, 9, 11, 13, 15, 17, 20; see Table 2 in Bobak et al., 2018) Cronbach's Alpha was .85 and the split-half reliability with Spearman-Brown correction was .84. For the calculation of the split-half reliability the first six items were considered as one half and the other five items as the second half. Furthermore, we calculated Omega as an additional reliability coefficient (Omega = .85) that is preferred by some researchers over Cronbach's Alpha because Omega shows robustness when factor loadings are not equal (Hayes & Coutts, 2020). All three reliability coefficients can be evaluated as good (Bühner, 2011). For the scale Face Processing (items 2, 4, 6, 10, 12, 14, 16, 18, 19; see Table 2 in Bobak et al., 2018) all coefficients were .83 or .84. For the calculation of the split-half reliability the first five items were considered as one half and the other four items as the second half. All three reliability coefficients can be evaluated as good (Bühner, 2011). Because the item assignments to the scales Face Memory and Face Processing vary depending on the result of the factorial analysis method in Bobak et al. (2018) and in our results (see Figure 1), we calculated the reliability coefficients for the different item assignments and scales (Table 4). In the Discussion, we give advises for the usability of the reliability scores.

In addition, we calculated the mean inter-item correlation (MIC) as the formula for Cronbach's Alpha integrates the MIC (Bernardi, 1994). MIC can be interpreted as an index of item homogeneity. The recommended range for a MIC is .15-.50 (Clark & Watson, 2016). The calculated MIC for all 20 German SFRS items was .33 which refers to the recommended range (Clark & Watson, 2016). The reliability results of at least ≥.83 are conceived as good. Thus, reliability coefficients confirmed hypothesis 2 presuming at least moderate reliability coefficients in terms of Cronbach's Alpha, Split-half reliability and Omega. Overall, psychometric properties have been tested with regard to ITC-guideline TD-5, section 8 and C-33, section 11.

Table 4.
Summary of reliability scores.

Scale name and items	Type of reliability	Reliability score	Comparability
Face memory 11 items: 1, 3, 5, 7_R†, 8, 9, 11, 13, 15, 17, 20	Cronbach's Alpha	.85	With Bobak et al. (2018, Table 2, results of the PCA)
	Split-half; 1st vs. 2nd half with Spearman-Brown Correction	.84	With Bobak et al. (2018, Table 2, results of the PCA)
	Omega	.85	
Face processing 9 items: 2_R, 4_R, 6_R, 10_R, 12_R, 14_R, 16_R, 18_R, 19_R	Cronbach's Alpha	.83	With Bobak et al. (2018, Table 2, results of the PCA)
	Split-half; 1st vs. 2nd half with Spearman-Brown Correction	.83	With Bobak et al. (2018, Table 2, results of the PCA)
	Omega	.84	
Face recognition 20 items	Cronbach's Alpha	.90	With Bobak et al. (2018) No evidence based on factorial validity
	Split-half; 1st vs. 2nd half with Spearman-Brown Correction	.91	
	Omega	.90	
Face memory 11 items (see Figure 1)	Hancock's H	.49	Items for the factorial validity, Figure 1
Face processing 9 items (see Figure 1)	Hancock's H	.50	Items for the factorial validity, Figure 1
Face recognition – 20 items (see Figure 1)	Hancock's H	.92	Items for the factorial validity, Figure 1

Note. † = _R indicates that an item was recoded for the calculation of the reliability coefficients.

Confirmatory Factor Analysis for the SFRS (hypothesis 3)

The result of Mardia's test (DeCarlo, 1997) for multivariate distribution suggested that the items were not normally distributed (Mardia = 577.68, $p < .05$). Thus, maximum likelihood estimator with robust standard errors entitled as MLR was applied in the CFA models (Luo, 2018). We performed three MIMIC models using gender as a factor tested for measurement invariance: (1) a two primary-order factor MIMIC model with orthogonal factors, (2) a two primary-factors MIMIC model with intercorrelated factors, and a (3) bifactor MIMIC model. The two primary-order factors MIMIC model presuming Face Processing and Face Memory as two orthogonal factors of the SFRS suggested a sub-optimal model fit (CFI = .78, RMSEA = .09, SRMR = .21). Similarly, the two primary-order factors MIMIC model presuming Face Processing and Face Memory as two intercorrelated factors of .30 showed a sub-optimal model fit (CFI = .84, RMSEA = .08, SRMR = .15). The bifactor MIMIC model including the two primary-order factors Face Processing and Face Memory and a second-order factor named as self-reported Face Recognition ability showed the best model fit (CFI = .94, RMSEA = .05, SRMR = .04). The standardized factor loadings (STDYX) of the bifactor model were mainly significant and positive (Figure 1).

The factor Face Processing incorporates exclusively inverted items (see item labeling with "R" in Figure 1). Positive significant factor loadings indicate that participants with higher self-described Face Processing ability reported that they do not often fail to recognize someone (item 19) or they do not have difficulties following movies because of recognizing characters (item 16). In contrast, the second factor entitled as Face Memory includes items that summarize higher self-described Face Memory with just one item being inverted (Figure 1: item FM_07R). That is, individuals with higher Face Memory agree with items like item 3: "I find it very easy to picture individual faces in my mind with great detail.". The second-order factor entitled as Face Recognition newly emerged in this CFA analysis. The second-order factor, Face Recognition, comprises a complete set of significant positive item loadings. This finding indicates that the German translation of the SFRS incorporates a second-order factor of self-reported Face Recognition. Self-reported Face Recognition means that individuals with higher and positive factor loadings describe themselves as recognizing faces, putting faces and names together, knowing where a person has been met first.

Finally, the MIMIC effect of the bifactor model was not significant for gender. That means, we did neither observe significant effects of gender (male, female) for both primary-order factors (Face Processing and Face Memory) nor for the second-order factor (Face Recognition) of the German SFRS (standardized beta coefficients: Face Processing = -0.02, Face Memory = -0.04, Face Recognition = 0.07, *p*s $\geq$.31).

Supplementary to the reliability coefficients Alpha, Split half, and Omega and specifically for the bifactor model, we report Hancock's *H* in Table 4. Hancock's *H* (also known as coefficient *H*) is based on the standardized loadings in a CFA and therefore, is used for scales in that the items should be optimally weighted. In those psychometric cases, Hancock's *H* can be an optimal estimator for reliability (for more information see Kalkbrenner, 2021). Following Kalkbrenner (2021), there is no consensus on recommended thresholds for acceptable Hancock's *H reliability* scores. However, in their simulation study Beauducel et al. (2016) mark a reliability threshold of .70 for variations of Hancock's H depending on salient loadings, secondary loadings, and factor intercorrelations. These results serve ITC-guidelines TD-5, section 8 and C-3, section 11.

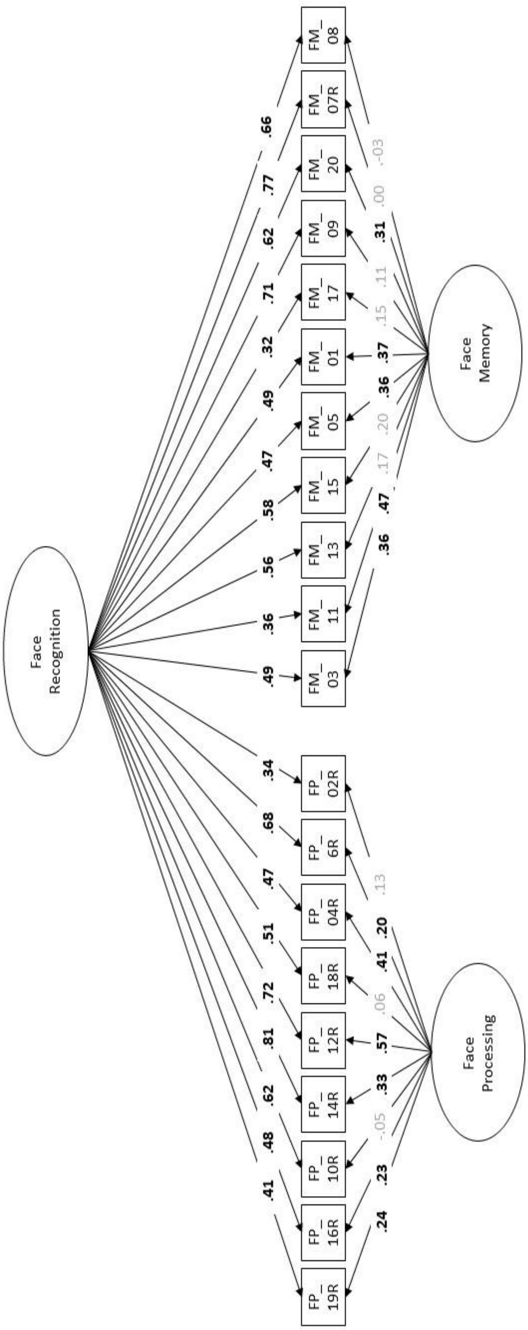

Figure 1.
Bifactor MIMIC model for German face recognition items of the Stirling Face Recognition Scale.

German SFRS and CFMT+

The descriptive statistics for the CFMT+ raw sum scores in our sample were M = 85.56, SD = 10.56, range = 44-100, skewness = -1.20. The descriptive statistics for the SFRS raw sum scores were M = 82.49, SD = 10.40, range 26-100, skewness = -1.67. Normal distribution of the SFRS und CFMT+ raw sum scores were tested with the Kolmogorov-Smirnov test (K-S). Both raw scores were not normally distributed (CFMT+ and SFRS: K-S(374) = .13, $p < .001$). Therefore, we calculated Spearman Rank correlations: CFMT+ with SFRS total sum score *Rho*(374) =.35 ($p < .001$, two-tailed); CFMT+ with SFRS–Face memory sum score *Rho*(374) = .37 ($p < .001$, two-tailed); CFMT+ with SFRS–Face processing sum score *Rho*(374) = .26 ($p < .001$, two-tailed). In terms of effect size r, the Spearman's Rho correlations can be interpreted as a positive medium correlation ($.30 > r < .50$; Cohen, 1988) and a as a low correlation (< .30; Cohen, 1988), respectively.

Spearman's Rho for the CFMT+ was significantly negative with the Face processing factor score (*Rho*(374) = -.18, $p < .01$, two-tailed) and significantly positive for the Face memory factor score (*Rho*(374) = .20, $p < .01$, two-tailed) and the Face Recognition factor score (*Rho*(374) = .34, $p < .01$, two-tailed). When controlling for Gender, partial correlations were significant exclusively for the CFMT+ with the Face matching factor score, $r(371) = .15$, $p < .01$, and with the second-order Face recognition factor score, $r(371) = .37$, $p < .01$, all p-values two-tailed. As for the sum scale values, effect size r for the factor scores also ranged between low and medium.

Discussion

This study investigated the psychometric properties (hypothesis 1-1, 1-2), the reliability (hypothesis 2) and the factorial structure of the adapted German SFRS (hypothesis 3) to extend the research on self-reported face recognition ability questionnaires in adult Caucasian participants. The present sample comprised adult female (n = 220) and male participants (n = 154). There was no preselection of any ethical group. Participants took part in this study based on their free and confirming decision. Neither for gender nor for nationality any subgroup had been excluded a-priori. Consequently, racial health equality has been ensured by means of a free self-selection of the sample (Boyd et al., 2020). The present findings have been obtained from a sample living in a Western, educated, industrialized, rich, and democratic (WEIRD) society. Thus, with regard to Simons et al. (2017) the findings of this study are restricted to the described sample characteristics.

In sum, the item difficulty indices show that 70% of the items have a moderate item difficulty between 20 and 80 (hypothesis 1-1 mainly confirmed). All part-whole corrected item total correlations were above .30 (hypothesis 1-2 confirmed). All reliability coefficients (Cronbach's Alpha, split-half reliability with Spearman-Brown correction, Omega) for the scales Face Memory and Face Processing were above .83 and can be rated as good, which supports hypothesis 2. The CFA model comparisons showed the best fit for a bifactor MIMIC model (*CFI* = .94, *RMSEA* = .05, *SRMR* = .04) with two primary-order factors *Face Processing* and *Face Memory* and a second-order factor named *Self-reported Face Recognition Ability*. Thus, hypothesis 3 presuming the best model fit for the two-factor model has not been confirmed. Instead, a hierarchical model could be demonstrated for the German SFRS. Furthermore, the Spearman Rank correlations between German SFRS sum scale values and CFMT+ sum scale value scores were calculated post-hoc as well as for the German SFRS factor scores.

The SFRS-German in comparison to the original English SFRS

Based on the psychometric properties and slight variations in item content in the forward and backward translation, we consider the translation and adaption of the SFRS from English into German language as successful. Our data extend previous research on SFRS items because this study initially reports psychometric properties for the SFRS-German (Table 3 and 4). We obtained good reliability coefficients for the SFRS-German scales Face Memory and Face Processing across different types of reliability coefficients. Bobak et al. (2018, p. 4, right column) evaluated reliabilities of .88 and .89 for all items as "very good" suggesting widely comparable internal consistencies for the English and German SFRS items. We recommend the use and citation of the reliability scores for the two factors Face Memory and Face Processing - when comparisons to Bobak et al. (2018) are preferred- as our factorial analyses did not show evidence of a one factor model. However, by applying reliability scores for Face Memory and Face Processing in assessment settings, users should be aware that the orthogonal and the oblique MIMIC CFA models did not show the best model fit. To apply reliability scores that correspond to the factorial model, reliability scores for Hancock's H have been reported.

Based on the factor loadings of a principal component analysis presented in Bobak et al. (2018, their Table 2) we investigated the model fit indices for three CFA models in Mplus. We tested a primary-order model with Face Memory and Face Processing as two orthogonal factors (model 1), a primary-order model with both factors to be correlated (model 2), and a bifactor model (model 3). The bifactor model with a second-order factor entitled as Face Recognition and both primary-order factors named as Face Processing and Face Memory revealed the best model fit. Thus, the CFA results of the present study rather suggest a hierarchical psychometric model of face recognition. Face processing comprises five significant factor loadings out of nine items. Face memory comprises six significant factor loadings out of 11 items. Gender effects have not been observed for the SFRS-German by means of the MIMIC bifactor model. This result suggests that the primary-order factors and the second-order factor were measurement equivalent in female and male participants. It is worth noting that the bifactor model extends previous results reporting mainly evidence of one to three primary-order factors in questionnaires on self-reported face recognition (Table 1).

Moreover, the relevance of factor scores for both primary-order factors and for the second-order factor has been shown for the SFRS-German, which extends prior research (Bobak et al., 2018; Ritchie et al., 2022).

In accordance with Bobak et al. (2018), we found significant low to medium correlations for the SFRS-German and the CFMT+. Similar to Bobak et al. (2018), our participants knew their CFMT+ results before they answered the SFRS items. Therefore, answering the SFRS items could be biased. Moreover, it should be mentioned that correlations between subjective measures and objective measures for face recognition ability are still under debate (Kramer & Tree, 2024). Till date it has not yet been tested whether metacognitive insights into face recognition abilities can be comparably assessed in laboratory settings and in real-world settings. Nevertheless, correlations between SFRS-German and CFMT+ contributed to the investigation of convergent validity (Campbell & Fiske, 1959).

Future directions and perspectives

Future research on the SFRS is promising to further investigate convergent and divergent validity (Campbell & Fiske, 1959) with the Conflict Monitoring Questionnaire using 44 or 28 items (CMQ-44, Leue & Beauducel, 2021). The CMQ-44 also comprises a bifactor structure including four primary order factors entitled as Cognitive demand, Anticipation of negative consequences, Response adaptation, and Uncertainty of reinforcement. Correlations between CMQ-44 scales and the SFRS might indicate whether people with better self-reported Face recognition invest higher cognitive demand in face processing (CMQ-44: Cognitive demand), anticipate more negative consequences when they would do wrong in applied face processing contexts (CMQ-44: Anticipation), and might be more careful in their response tendencies when they are less sure about face classifications (CMQ-44: Uncertainty of reinforcement). Initial data for the German SFRS and the German CMQ-44 (Leue & Beauducel, 2021) in a sample of $N = 107$ participants indicate that the response-related subscales of the CMQ-44 entitled as response adaptation and Uncertainty of reinforcement are significantly positive correlated with the German SFRS subscale: Face Processing based on Lamda-coefficients and Spearman Rank correlations (Missling, 2024, Tables 7 and 8).

As temperament inventories postulate attention-related dimensions, the Adult Temperament Questionnaire (Evans & Rothbart, 2007) would be worthwhile for the investigation of construct validity. Moreover, individual differences of self-reported face recognition might be an important option to assess pre-requisites for psychological reports on criteria-based content analyses of eye witnesses (Deckers & Köhnken, 2022; Granhag et al., 2015).

Moreover, a hierarchical structure for the construct "self-reported face recognition ability" fits to the associated construct of objectively measurable face recognition abilities. In the literature, the construct "face recognition abilities" is usually understood as a multidimensional construct that is represented by various tasks compatible to real-life situations (e.g., Bate et al., 2018; Ramon, 2021; Wilhelm et al., 2010). Conceptual similarities and differences between the constructs self-reported face recognition ability and objective face recognition ability should be further explored.

It should be noted that subjective methods like a self-reported ability questionnaire can be influenced by personal biases in answering a questionnaire. Effects like Dunning-Kruger exist also in face perception (Zhou & Jenkins, 2020). The Dunning-Kruger effect describes the misjudgment in metacognitive insights in which people might be less aware of subjective to systematic biases (Zhou & Jenkins, 2020). Research could still discuss the suitability of questionnaires as an economic screening method to assess face recognition abilities, e.g., in police applications. For this purpose, investigations of the criterion validity of self-reported face recognition abilities questionnaires would be necessary in future research.

Conclusion

Taken together, this study extends the research on psychometric properties for a self-reported face recognition ability questionnaire which assesses the whole ability spectrum. We successfully translated the Stirling Face Recognition Scale (SFRS) into German and showed that the factorial structure fits best to a bifactor model with two primary-order factors Face Processing and Face Memory and a second-order factor named as Self-reported Face Recognition Ability. We recommend using the SFRS-German in future research as the psychometric properties can be interpreted as satisfactory in summary.

Acknowledgement

We are grateful to Dipl.-Psych. Niels Rieger for his support during data collection.

References

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

Arizpe, J. M., Saad, E., Douglas, A. O., Germine, L., Wilmer, J. B., & DeGutis, J. M. (2019). Self-reported face recognition is highly valid, but alone is not highly discriminative of prosopagnosia-level performance on objective assessments. *Behavior Research Methods*, *51*(3), 1102–1116. https://doi.org/10.3758/s13428-018-01195-w

Bate, S., & Dudfield, G. (2019). Subjective assessment for super recognition: An evaluation of self-report methods in civilian and police participants. *PeerJ*, *7*, e6330. https://doi.org/10.7717/peerj.6330

Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Murray, E., Bobak, A. K., Wills, H., & Richards, S. (2018). Applied screening tests for the detection of superior face recognition. *Cognitive Research: Principles and Implications*, *3*, 22. https://doi.org/10.1186/s41235-018-0116-5

Bate, S., Portch, E., & Mestry, N. (2021). When two fields collide: Identifying "super-recognisers" for neuropsychological and forensic face recognition research. *Quarterly Journal of Experimental Psychology (2006)*, 17470218211027695. https://doi.org/10.1177/17470218211027695

Beauducel, A, & Wittmann, W. W. (2005). Simulation Study on Fit Indexes in CFA Based on Data With Slightly Distorted Simple Structure. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(1), 41–75. https://doi.org/10.1207/s15328007sem1201_3

Beauducel, A, Harms, C., & Hilger, N. (2016). Reliability Estimates for Three Factor Score Estimators. *International Journal of Statistics and Probability*, *5*(6), 94. https://doi.org/10.5539/ijsp.v5n6p94

Bernardi, R. A. (1994). Validating Research Results when Cronbach'S Alpha is Below.70: A Methodological Procedure. *Educational and Psychological Measurement*, *54*(3), 766–775. https://doi.org/10.1177/0013164494054003023

Bobak, A. K., Mileva, V. R., & Hancock, P. J. (2019). Facing the facts: Naive participants have only moderate insight into their face recognition and face perception abilities. *Quarterly Journal of Experimental Psychology (2006)*, *72*(4), 872–881. https://doi.org/10.1177/1747021818776145

Bobak, A. K., Pampoulov, P., & Bate, S. (2016). Detecting Superior Face Recognition Skills in a Large Sample of Young British Adults. *Frontiers in Psychology*, *7*, 1378. https://doi.org/10.3389/fpsyg.2016.01378

Boyd, R. W., Lindo, E. G., Weeks, L. D., & McLemore, M. R. (2020). *On Racism: A New Standard For Publishing On Racial Health Inequities*. 10.1377/forefront.20200630.939347

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology (London, England : 1953)*, *77 (Pt 3)*, 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3., aktualisierte und erw. Aufl.). *PS Psychologie*. Pearson Studium. http://lib.myilibrary.com/detail.asp?id=404890

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Clark, L. A., & Watson, D. (2016). Constructing validity: Basic issues in objective scale development. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research (4th ed.)* (pp. 187–203). American Psychological Association. https://doi.org/10.1037/14805-012

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge. https://doi.org/10.4324/9780203771587

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Davis, J. P., Bretfelean, L. D., Belanova, E., & Thompson, T. (2020). Super-recognisers: Face recognition performance after variable delay intervals. *Applied Cognitive Psychology*, *34*(6), 1350–1368. https://doi.org/10.1002/acp.3712

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*(3), 292–307. https://doi.org/10.1037/1082-989X.2.3.292

Deckers, R., & Köhnken, G. (Eds.). (2022). *Die Erhebung und Bewertung von Zeugenaussagen im Strafprozess: Juristische, aussagepsychologische und psychiatrische Aspekte* [5. Auflage]. Berliner Wissenschafts-Verlag.

Estudillo, A. J., & Wong, H. K. (2021). Associations between self-reported and objective face recognition abilities are only evident in above- and below-average recognisers. *PeerJ*, *9*(5), e10629. https://doi.org/10.7717/peerj.10629

Evans, D. E., & Rothbart, M. K. (2007). Developing a model for adult temperament. *Journal of Research in Personality*, *41*(4), 868–888. https://doi.org/10.1016/j.jrp.2006.11.002

Fysh, M. C., Stacchi, L., & Ramon, M. (2020). Differences between and within individuals, and subprocesses of face cognition: Implications for theory, research and personnel selection. *Royal Society Open Science*, *7*(9), 200233. https://doi.org/10.1098/rsos.200233

Gehdu, B. K., Press, C., Gray, K. L. H., & Cook, R. (2024). Autistic adults have insight into their relative face recognition ability. *Scientific Reports*, *14*(1), 17802. https://doi.org/10.1038/s41598-024-67649-8

Granhag, P. A., Vrij, A., & Verschuere, B. (Eds.). (2015). *Wiley series in the psychology of crime, policing and law. Detecting deception: Current challenges and cognitive approaches*. Wiley. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=882585 https://doi.org/10.1002/9781118510001

Hayes, A. F., & Coutts, J. J. (2020). Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But… *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

International Test Commission. (2017). *The ITC Guidelines for Translating and Adapting Tests: ITC Guidelines for Translating and Adapting Tests (Second edition)*. www.InTestCom.org

Kalkbrenner, M. T. (2021). A Practical Guide to Instrument Development and Score Validation in the Social Sciences: The MEASURE Approach. Advance online publication. https://doi.org/10.7275/svg4-e671

Kramer, R. S. S. (2023). Face matching and metacognition: Investigating individual differences and a training intervention. *PeerJ*, *11*, e14821. https://doi.org/10.7717/peerj.14821

Kramer, R. S., & Tree, J. J. (2024). Investigating people's metacognitive insight into their own face abilities. *Quarterly Journal of Experimental Psychology (2006)*, *77*(10), 1949–1956. https://doi.org/10.1177/17470218231218662

Leue, A., & Beauducel, A. (2021). A facet theory approach for the psychometric measurement of conflict monitoring. *Personality and Individual Differences*, *171*(e11), 110479. https://doi.org/10.1016/j.paid.2020.110479

Livingston, L. A., & Shah, P. (2018). People with and without prosopagnosia have insight into their face recognition ability. *Quarterly Journal of Experimental Psychology*, *71*(5), 1260–1262. https://doi.org/10.1080/17470218.2017.1310911

Luo, Y. (2018). A Short Note on Estimating the Testlet Model With Different Estimators in Mplus. *Educational and Psychological Measurement*, *78*(3), 517–529. https://doi.org/10.1177/0013164417717314

Matsuyoshi, D., & Watanabe, K. (2021). People have modest, not good, insight into their face recognition ability: A comparison between self-report questionnaires. *Psychological Research*, *85*(4), 1713–1723. https://doi.org/10.1007/s00426-020-01355-8

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive Research: Principles and Implications*, *3*, 21. https://doi.org/10.1186/s41235-018-0112-9

Missling, M. (2024). *Konstruktvalidierung von Fragebögen zum Konflikt-Monitoring, zur selbstberichteten Gesichtserkennungsfähigkeit und zum Selbst-Monitoring im Sinne der konvergenten und diskriminanten Validität [Construct validation of questionnaires: Conflict-monitoring, self-reported face recognition, and self-monitoring with regard to convergent and discriminant validity.]* [Unpublished Bachelor thesis], Kiel University, Germany.

Moosbrugger, H., & Kelava, A. (2020). *Testtheorie und Fragebogenkonstruktion*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-61532-4

Nakashima, S. F., Ukezono, M., Sudo, R., Nunoi, M., Kitagami, S., Okubo, M., Toriyama, R., Morimoto, Y., & Takano, Y. (2020). Development of a Japanese version of the 20-item prosopagnosia index (PI20-J) and examination of its reliability and validity. *The Japanese Journal of Psychology*, *90*(6), 603–613. https://doi.org/10.4992/jjpsy.90.18235

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3. Ed.). *McGraw-Hill series in psychology*. McGraw-Hill.

Oren, S., Aydinlik, A., Kapucu, A., & Amado, S. (2019). *Evaluating Your Own Face Recognition Ability: Validation of the Stirling Face Recognition Scale in a Turkish sample.* European Congress of Psychology. https://doi.org/10.26226/morressier.5cf632c5af72dec2b0554cce

Palermo, R., Rossion, B., Rhodes, G., Laguesse, R., Tez, T., Hall, B., Albonico, A., Malaspina, M., Daini, R., Irons, J., Al-Janabi, S., Taylor, L. C., Rivolta, D., & McKone, E. (2017). Do people have insight into their face recognition abilities? *Quarterly Journal of Experimental Psychology (2006)*, *70*(2), 218–233. https://doi.org/10.1080/17470218.2016.1161058

Petersen, L. A., & Leue, A. (2021). Extraordinary face recognition performance in laboratory and online testing. *Applied Cognitive Psychology*, *3*(2), 22. https://doi.org/10.1002/acp.3805

Petersen, L. A., & Leue, A. (2022). Face Memory and Face Matching. *Journal of Individual Differences*, *43*(3), 152–159. https://doi.org/10.1027/1614-0001/a000361

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15. https://doi.org/10.1016/S0001-6918(99)00050-5

Ramon, M. (2021). Super-Recognizers - a novel diagnostic framework, 70 cases, and guidelines for future work. *Neuropsychologia*, *158*, 107809. https://doi.org/10.1016/j.neuropsychologia.2021.107809

Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the Number of Categories in Agree–Disagree Scales. *Sociological Methods & Research*, *43*(1), 73–97. https://doi.org/10.1177/0049124113509605

Rieger, N. (2021). *Realistische alltagsnahe Stimuli in der Gesichtserkennungsforschung: Aufbau einer Gesichts-Datenbank und Entwicklung eines Kurzzeitpersonengedächtnistest [Realistic everyday stimuli in face recognition research: Building a face database and developing a short-term person memory test]*, Unpublished diploma thesis, Kiel University, Germany.

Ritchie, K. L., Flack, T. R., Fuller, E. A., Cartledge, C., & Kramer, R. S. S. (2022). The pairs training effect in unfamiliar face matching. *Perception*, *51*(7), 477–495. https://doi.org/10.1177/03010066221096987

Russell, R., Duchaine, B [Brad], & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257. https://doi.org/10.3758/PBR.16.2.252

Shah, P., Gaule, A., Sowden, S., Bird, G., & Cook, R. (2015). The 20-item prosopagnosia index (PI20): A self-report instrument for identifying developmental prosopagnosia. *Royal Society Open Science*, *2*(6), 140343. https://doi.org/10.1098/rsos.140343

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on Generality (COG): A Proposed Addition to All Empirical Papers. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Stantic, M., Brewer, R., Duchaine, B [Bradley], Banissy, M. J., Bate, S., Susilo, T., Catmur, C., & Bird, G. (2022). The Oxford Face Matching Test: A non-biased test of the full range of individual differences in face perception. *Behavior Research Methods*, *54*(1), 158–173. https://doi.org/10.3758/s13428-021-01609-2

Sun, W., Wang, Y., Wang, J., Luo, F., Zhu, T., Anpo, M., & Sharifi, A. (2021). Psychometric Properties of the Chinese version of the 20-item Prosopagnosia Index (PI20). *E3S Web of Conferences*, *271*(1), 1036. https://doi.org/10.1051/e3sconf/202127101036

Tsantani, M., Vestner, T., & Cook, R. (2021). The Twenty Item Prosopagnosia Index (PI20) provides meaningful evidence of face recognition impairment. *Royal Society Open Science*, *8*(11), 202062. https://doi.org/10.1098/rsos.202062

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research*, *141*, 217–227. https://doi.org/10.1016/j.visres.2016.12.014

Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces-One element of social cognition. *Journal of Personality and Social Psychology*, *99*(3), 530–548. https://doi.org/10.1037/a0019972

Zhou, X., & Jenkins, R. (2020). Dunning-Kruger effects in face perception. *Cognition*, *203*, 104345. https://doi.org/10.1016/j.cognition.2020.104345