Huge intrinsic correlation between developmental prosopagnosia questionnaires: A comment on Shah et al. (2015)

Running title: Prosopagnosia Questionnaires

Daisuke Matsuyoshi [1,2,3], Katsumi Watanabe [1,2]

[1] Faculty of Science and Engineering, Waseda University, 3-4-1 Ohkubo, Shinjuku, Tokyo 169-8555, Japan

[2] Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro, Tokyo 153-8904, Japan

[3] Araya Inc., 2-8-10 Toranomon, Minato, Tokyo 105-0001, Japan

Corresponding Author:

Daisuke MATSUYOSHI

Research Institute for Science and Engineering,

Faculty of Science and Engineering, Waseda University

Rm 409-1B, 59th Bldg, 3-4-1 Ohkubo, Shinjuku, Tokyo 169-8555, Japan

Phone/FAX: +81-3-5286-3335 E-mail: matsuyoshi@aoni.waseda.jp

Abstract

This paper complements the 20-item prosopagnosia index (PI20), which is a self-report measure of developmental prosopagnosia (DP) developed by Shah et al. (2015, R. Soc. Open. Sci. 2, 150305. [doi: 10.1098/rsos.140343]). Although they validated PI20 in several ways and it can serve as a quick and cost-effective measure for estimating DP risk, they did not formally evaluate its validity against a pre-existing alternative questionnaire (Kennerknecht et al., Am. J. Med. Genet. A 146A, 2863-2870. [doi: 10.1002/ajmg.a.32552]). Thus, we administered the questionnaires to a large population (N = 855) and found a very strong correlation (r = 0.82 [95% confidence interval: 0.80, 0.84]), a principal component that accounted for more than 90% of the variance, and comparable reliability between the questionnaires. These results suggest unidimensionality and equivalence between the two questionnaires, or at least, a very strong common latent factor underlying them.

Keywords:

developmental prosopagnosia, face recognition, self-report, questionnaire, sex differences

1. Materials and Methods

1.1 Participants

Eight hundred and fifty-five young Japanese adults (427 female, 428 male; mean age: 20.9 ± 2.2 [±1 SD] years; range: 18–36 years) participated in the study. All had normal or corrected-to-normal vision and none reported a history of neurological or developmental disorders.

1.2 Procedure

We asked participants to complete the questionnaires using an 8-inch touchscreen tablet PC. They were required to indicate the extent to which 36 items (15 from the pre-existing Hong Kong (HK) prosopagnosia questionnaire [1], and 20 from PI20 [2], and an additional item pertaining self-confidence in face recognition ability: "I am confident that I can recognize faces well compared to others") described their face recognition experiences. Responses were provided using a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The participants were instructed to complete the questionnaires at their own pace.

1.3 Data analysis

Because the HK prosopagnosia questionnaire developed by Kennerknecht et al. [1] contains four dummy questions (HK#10, #11, #12, and #13) that are irrelevant with respect to face recognition, we excluded these items and calculated the total scores ranging 11 to 55, using the remaining 11 items (hereafter, 'HK11') [score range: 11-55].

PI20 scores were calculated using all 20 items and ranged from 20 to 100. As women have exhibited superior performance in behavioral face recognition studies [3], we examined sex differences between the questionnaire scores. In addition, we used polychoric correlation coefficients to infer latent Pearson correlations between individual items from the ordinal data. The polychoric correlation matrix was estimated using two-step approximation [3].

Cronbach's α and Revelle & Zinbarg's omega total coefficients were calculated to assess the scale reliability of both HK11 and PI20. Omega total coefficients were estimated using a maximum likelihood procedure [4]. CIs for the coefficients were estimated using a bootstrap procedure (10,000 replications) with a bias-corrected and accelerated approach [5, 6]. As it was possible that higher reliability coefficients merely reflected the higher number of items in the PI20, relative to that in the HK11 [7], we performed a brute-force calculation of reliability coefficients for all 167,960 (${}_{20}C_{11}$) possible combinations of PI20 items, 11 items at a time (i.e., subsets of the PI20 generated by choosing 11 of the 20 items), which allowed us to compare reliability coefficients between the questionnaires with a virtual match of the numbers of items.

2. Results and Discussion

2.1 Total scores and score distribution

Table 1 shows descriptive statistics for the total HK11 and PI20 scores. Independent two sample t tests showed no significant differences between in HK11 (t_{853} = 0.0511, p = 0.9592, Cohen's d = 0.0035 [95% confidence interval (CI): −0.1306, 0.1376] or PI20 (t_{810} = 0.9578, p = 0.3384, Cohen's d = 0.0655 [95% CI: −0.0686, 0.1996]) scores between males and females. In addition, Bayesian analysis using a JZS prior (r scaling = 1) [8] showed strong evidence for the null hypothesis (i.e., no sex difference) for both HK11 (Bayes factor BF_{10} = 0.0544) and PI20 (BF_{10} = 0.0856) scores. In addition, two-sample Kolmogorov-Smirnov tests showed no significant sex differences between the distributions (Figure 1) of HK11 (p = 0.9982) and PI20 (p = 0.7554) scores. These results indicate that females and males showed almost identical mean HK11 and PI20 scores and score distributions, suggesting that sex was not a significant factor.

2.2 Correlations between total scores

The results showed a very strong significant correlation between the total scores for the two questionnaires (Figure 1, r = 0.8228 [95% CI: 0.7999, 0.8433], $p = 1.6510 \times 10^{-211}$), suggesting a significant overlap of prosopagnosia traits assessed via each measure. It should be noted that Fisher’s z test showed no significant sex difference in the correlation between total scores (difference: r_{diff} = 0.0065 [95% CI: −0.0371, 0.0502], z = 0.2917, p = 0.7705; females: r = 0.8200 [95% CI: 0.7863, 0.8489], $p = 4.7087 \times 10^{-105}$; males: r = 0.8265 [95% CI: 0.7939, 0.8543], $p = 2.3806 \times 10^{-108}$). Principal component analysis with singular value decomposition of the correlation matrix between total scores showed that the first principal component (PC1) accounted for 91.1% (using standardized scores) and 94.2% (using raw scores) of the total variance in scores.

2.3 Correlations between individual item scores

The correlation matrix (Figure 2) generally showed correlations between individual items *across* the two scales; however, some items were not correlated with other items to the extent that they would reduce the reliability or internal consistency of a single measure pertaining to a single construct (i.e., DP risk). In fact, hierarchical clustering using the unweighted pair group method with arithmetic mean showed that eight out of 36 items were distant from a cluster to which most items belonged (shaded areas in Figure 2, dendrogram). These eight items consisted of: the four items that is already known to be irrelevant with respect to DP (HK#10, HK#11, HK#12, and HK#13), and two items from the HK prosopagnosia questionnaire (HK#2 and HK#7), and two items from the PI20 (PI#3 and PI#13). Previous studies reported that five of the eight item-score differences were marginal (score difference [DP − control]: <1) between individuals with suspected DP and typically developed control individuals (0.45 for HK#10, −0.39 for HK#11, 0.11 for HK#12, −0.45 for HK#13, and 0.62 for PI#3) [1, 2]. However, it should be noted that the score difference exceeded 1 for the remaining three items (1.46 for HK#2, 1.12 for HK#7, and 1.16 for PI#13), suggesting that these three items could actually measure prosopagnosia traits that differ from those measured via the other 28 items.

2.4 Scale reliability

We found that the reliability coefficients for the PI20 were higher relative to those for the HK11 (HK11: α = 0.8449 [95% CI: 0.8273, 0.8633], ω_t = 0.8767 [95% CI: 0.8571, 0.8880]; PI20: α = 0.9174 [95% CI: 0.9102, 0.9249], ω_t = 0.9368 [95% CI: 0.9300, 0.9424]). Follow-up Feldt paired tests [9] confirmed significant differences in reliability coefficient between the HK11 and PI20 (difference in α: t_{853} = 16.4437, $p = 5.5132 \times 10^{-53}$; difference in ω_t: t_{853} = 17.4868, $p = 9.4696 \times 10^{-59}$).

The brute-force calculation of reliability coefficients showed that the coefficients for the eleven-item PI20 subsets were almost comparable (within 1 SD) to those for the HK11 (α: mean = 0.8530 ± 0.0392 [±1 SD], median = 0.8474, range = 0.7495–0.9324; ω_t: mean = 0.8914 ± 0.0225 [±1 SD], median = 0.8933, range = 0.8122–0.9438). This indicated that the HK11 and PI20 demonstrated almost equivalent reliability at the individual-item level.

3. Conclusion

In conclusion, the results indicated that the two representative DP questionnaires [1, 2] essentially measured the same prosopagnosia-related face recognition traits. In addition, the very strong correlation and robust principal component demonstrated a common latent factor between the two measures. This putative unidimensionality could be intrinsic to insight into one's face recognition ability, rather than behavioral performance per se [10]. Although the PI20 can serve as a reliable measure for estimating DP risk and face recognition ability in general populations [11], our findings suggest that its reliability and validity may be almost equivalent to that of the pre-existing questionnaire [1] (precisely, the reduced subset, HK11) at the individual-item level. Given the current state of DP, where neither objective diagnostic criteria nor biological markers have been established [12, 13], more exploratory research not only using HK11 and PI20 together, but also a range of other face processing measures could aid the extraction of latent prosopagnosia traits/dimensions and the development of a valid DP taxonomy.

Ethics

The study procedure was conducted in accordance with the Declaration of Helsinki and approved by the Committee of Ethics, Waseda University, Japan. All participants provided written informed consent prior to participation.

Data accessibility

The datasets and code are available at https://github.com/dicemt/matsuyoshi2018dp.

Authors' contributions

DM and KW designed the study and wrote the manuscript. DM collected and analyzed the data.

Competing interests

The authors have no competing interests to declare

Funding

This study was supported by grants from the Japan Society for the Promotion of Science (#26540061 to DM and #17H06344 to KW) and Core Research for Evolutional Science and Technology (CREST) at the Japan Science and Technology Agency (to KW).

References

1. Kennerknecht I., Ho N.Y., Wong V.C.N. 2008 Prevalence of hereditary prosopagnosia (HPA) in Hong Kong Chinese population. *Am. J. Med. Genet. A* **146A**(22), 2863-2870. (doi:10.1002/ajmg.a.32552).

2. 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. *R. Soc. Open Sci.* **2**(6), 150305. (doi:10.1098/rsos.140343).

3. Olsson U. 1979 Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* **44**(4), 443-460. (doi:10.1007/bf02296207).

4. Revelle W., Zinbarg R.E. 2008 Coefficients alpha, beta, omega, and the GLB: Comments on Sijtsma. *Psychometrika* **74**(1), 145. (doi:10.1007/s11336-008-9102-z).

5. DiCiccio T.J., Efron B. 1996 Bootstrap confidence intervals. *Stat. Sci.* **11**(3), 189-212.

6. Kelley K., Pornprasertmanit S. 2016 Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychol. Methods* **21**(1), 69-92. (doi:10.1037/a0040086).

7. Cortina J.M. 1993 What is coefficient alpha? An examination of theory and applications. *J. Appl. Psychol.* **78**(1), 98-104. (doi:10.1037/0021-9010.78.1.98).

8. Rouder J.N., Speckman P.L., Sun D., Morey R.D., Iverson G. 2009 Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* **16**(2), 225-237. (doi:10.3758/PBR.16.2.225).

9. Feldt L.S. 1980 A test of the hypothesis that Cronbach's alpha reliability

coefficient is the same for two tests administered to the same sample. *Psychometrika* **45**(1), 99-105. (doi:10.1007/bf02293600).

10. Palermo R., Rossion B., Rhodes G., Laguesse R., Tez T., Hall B., Albonico A., Malaspina M., Daini R., Irons J., et al. 2017 Do people have insight into their face recognition abilities? *Q. J. Exp. Psychol.* **70**(2), 218-233. (doi:10.1080/17470218.2016.1161058).

11. Livingston L.A., Shah P. in press People with and without prosopagnosia have insight into their face recognition ability. *Q. J. Exp. Psychol.*, 1-3. (doi:10.1080/17470218.2017.1310911).

12. Susilo T., Duchaine B. 2013 Advances in developmental prosopagnosia research. *Curr. Opin. Neurobiol.* **23**(3), 423-429. (doi:10.1016/j.conb.2012.12.011).

13. Barton J.J.S., Corrow S.L. 2016 The problem of being bad at faces. *Neuropsychologia* **89**, 119-124. (doi:https://doi.org/10.1016/j.neuropsychologia.2016.06.008).

Table

Table 1. Descriptive statistics of total scores for the questionnaires. A higher score indicates higher prosopagnosic face recognition traits. Note that females and males showed similar total scores in terms of not only summary statistics, but also distribution, as shown in Figure 1.

Questionnaire	Sex	Mean	SD	Min	Median	Max
HK11	Female	24.05	6.72	11	23	48
	Male	24.03	6.72	11	23	49
	Total	24.04	6.71	11	23	49
PI20	Female	48.33	13.00	26	46	89
	Male	47.49	12.70	24	46	85
	Total	47.91	12.85	24	46	89

Figure Legends

Figure 1. Correlation between total scores for the two prosopagnosia questionnaires. Scatter plot with color-coded transparent density curves of total scores for the 20-item prosopagnosia index (x-axis) and 11 items from Hong Kong prosopagnosia questionnaire (y-axis). Each circle indicates individuals' data, and color represents sex (red = female scores; blue = male scores). The gray transparent line represents a linear orthogonal regression line (first principal component, PC1 axis), which accounts for more than 90% of the total variance in scores in principal component analysis with a singular value decomposition.

Figure 2. Polychoric correlation matrix and hierarchical clustering (dendrogram) for individual item scores. Polychoric correlation coefficients are color-coded using the color key shown at the top left histogram. The dendrogram was obtained by a hierarchical clustering based on Pearson correlation distances using the unweighted pair group method with arithmetic mean. CONF, the question pertaining self-confidence in face recognition ability.

Figures

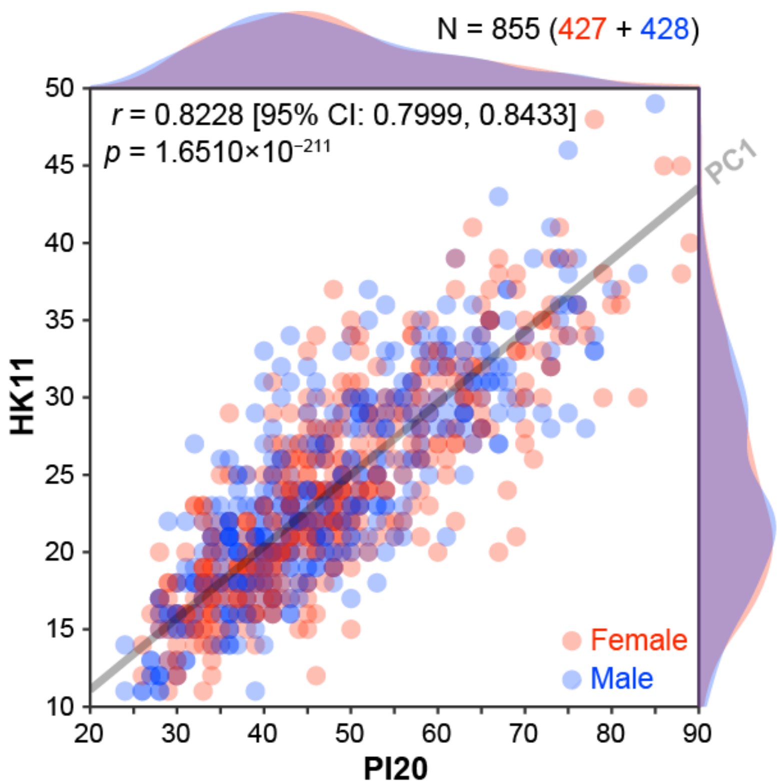

Fig. 1

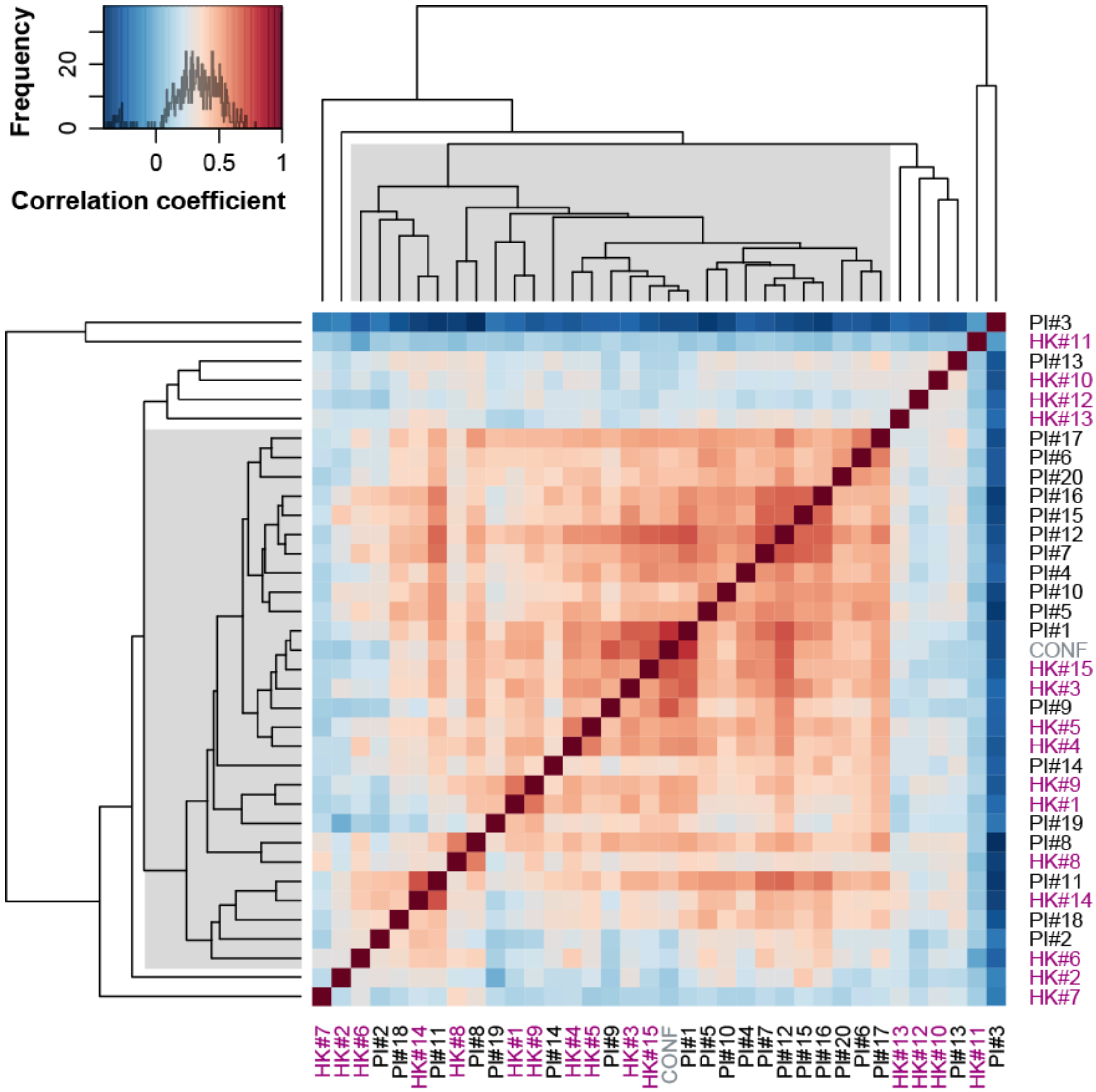
Frequency
0
20
0
0.5
1
Correlation coefficient
PI#3
HK#11
PI#13
HK#10
HK#12
HK#13
PI#17
PI#6
PI#20
PI#16
PI#15
PI#12
PI#7
PI#4
PI#10
PI#5
PI#1
CONF
HK#15
HK#3
PI#9
HK#5
HK#4
PI#14
HK#9
HK#1
PI#19
PI#8
HK#8
PI#11
HK#14
PI#18
PI#2
HK#6
HK#2
HK#7

Fig. 2.