Associations between self-reported and actual face recognition abilities are only evident in above- and below-average recognisers

Alejandro J. Estudillo Corresp., 1, 2, **Hoo Keat Wong** [2]

[1] Department of Psychology, Bournemouth University, Bournemouth, United Kingdom

[2] School of Psychology, University of Nottingham - Malaysia Campus, Semenyih, Selangor, Malaysia

Corresponding Author: Alejandro J. Estudillo
Email address: aestudillo@bournemouth.ac.uk

The 20-Item Prosopagnosia Items (PI-20) was recently introduced as a self-report measure of face recognition abilities and as an instrument to help the diagnosis of prosopagnosia. In general, studies using this questionnaire have shown that observers have moderate to strong insights into their face recognition abilities. However, it remains unknown whether these insights are equivalent for the whole range of face recognition abilities. The present study investigates this issue using the Mandarin version of the PI-20 and the Cambridge Face Memory Test Chinese (CFMT-Chinese). Our results showed a moderate negative association between the PI-20 and the CFMT-Chinese. However, this association was driven by people with low and high face recognition ability, but absent in people within the typical range of face recognition performance. The implications of these results for the study of individual differences and the diagnosis of prosopagnosia are discussed.

Associations between self-reported and actual face recognition abilities are only evident in above- and below-average recognisers

Alejandro J. Estudillo [1,2], Hoo Keat Wong[2]

[1] Department of Psychology, Bournemouth University, Bournemouth, UK
[2] School of Psychology, University of Nottingham Malaysia, Semenyih, Malaysia

Corresponding Author:
Alejandro J. Estudillo[1]
Department of Psychology, Bournemouth University, Poole House Talbot Campus, BH12 5BB UK.
Email address: aestudillo@bournemouth.ac.uk

Abstract

The 20-Item Prosopagnosia Items (PI-20) was recently introduced as a self-report measure of face recognition abilities and as an instrument to help the diagnosis of prosopagnosia. In general, studies using this questionnaire have shown that observers have moderate to strong insights into their face recognition abilities. However, it remains unknown whether these insights are equivalent for the whole range of face recognition abilities. The present study investigates this issue using the Mandarin version of the PI-20 and the Cambridge Face Memory Test Chinese (CFMT-Chinese). Our results showed a moderate negative association between the PI-20 and the CFMT-Chinese. However, this association was driven by people with low and high face recognition ability, but absent in people within the typical range of face recognition performance. The implications of these results for the study of individual differences and the diagnosis of prosopagnosia are discussed.

Introduction

Face recognition is a very important cognitive skill that enables successful social interactions with peers. Interestingly, despite being a remarkably common process, face recognition presents substantial variation among individuals, and this variation has important theoretical and practical consequences (Lander, Bruce, & Bindemann, 2018; Wilmer, 2017). On one side of the distribution, we find people with extraordinary abilities to identify faces, known as super-recognizers (Russell, Duchaine, & Nakayama, 2009). Super-recognizers present above normal performance in a variety of face identification tasks, including unfamiliar and familiar face recognition (Russell et al., 2009), and face matching (Robertson et al., 2016). Given their extraordinary abilities to identify faces, employing super-recognizers can be highly valuable in those applied scenarios whereby the identification of faces is of paramount importance, such as surveillance, eyewitness identification, and ID-verification settings (Ramon, Bobak, & White, 2019).

On the other side of the distribution, we find people with severe difficulties to recognize faces. These difficulties can arise following brain injury –as in the case of acquired prosopagnosia (Rossion, 2018)–, or as consequence of atypical brain development –as in the case of developmental prosopagnosia (Bowles et al., 2009; Dalrymple & Palermo, 2016; Duchaine & Nakayama, 2006)–. Although acquired prosopagnosia is an extremely rare disorder (Rossion, 2018), it has been estimated that the prevalence of developmental prosopagnosia is around 2-3% in general population (Barton & Corrow, 2016; Bate & Tree, 2017; Bowles et al., 2009; Dalrymple & Palermo, 2016; Kennerknecht, Ho, & Wong, 2008). As consequence of their difficulties identifying faces, people with prosopagnosia find social situations particularly stressful and are prone to depression, anxiety and social avoidance disorders (Dalrymple et al., 2014; Yardley, McDermott, Pisarski, Duchaine, & Nakayama, 2008).

The Cambridge Face Memory Test (CFMT) was introduced as an objective tool to study individual differences in face identification (Duchaine & Nakayama, 2006; Russell et al., 2009). This task can be completed in approximately 20 minutes and requires the identification of faces across different images of the same person, avoiding the limitations of simple pictorial recognition (Bruce, 1982; Estudillo, 2012; Estudillo & Bindemann, 2014; Longmore, Liu, & Young, 2008) and the use of non-facial cues (e.g., make up, clothing, hairstyle). Although the CFMT was initially introduced with Caucasian faces, more recent versions have adapted the face stimuli to Chinese and South East Asian populations: the CFMT-Chinese (McKone et al., 2012; McKone, Wan, Robbins, Crookes, & Liu, 2017). Remarkably, these two versions of the CFMT are psychometrically quite robust as they present internal reliability scores of between .85 and .90 (Bowles et al., 2009; Estudillo, Lee, Mennie, & Burns, 2020), which is an important requirement for measures of individual differences.

Although few researchers would disagree about the importance of objective measures to evaluate individual differences in face identification, phenomenological or self-reported measures have attracted the interest of researchers in recent years (Bobak, Mileva, & Hancock, 2019; Livingston & Shah, 2018; Palermo et al., 2017; Shah, Gaule, Sowden, Bird, & Cook, 2015; Shah, Sowden, Gaule, Catmur, & Bird, 2015). In self-reported measures of face identification, observers are, generally, asked to rate their level of agreement with a set of statements describing different situations involving face recognition abilities. It has been suggested that these self-reported measures can be used as screening or complementary tools to measure individual differences in face identification and, particularly, in the diagnosis of prosopagnosia (Shah, Gaule, et al., 2015; Shah, Sowden, et al., 2015). Although several self-reported measures of face identification has been built (see e.g., Bate & Dudfield, 2019; Bobak et al., 2019 Palermo et al., 2017), the 20-item prosopagnosia index (PI-20) is probably the most widely-used (Shah, Gaule, et al., 2015; Shah, Sowden, et al., 2015). This questionnaire is comprised of 20 items in a five-point Likert scale, describing different situations involving face identification (e.g., “My face recognition ability is worse than most people”). Higher scores in the PI20 index worse face recognition skills. Scores in the PI-20 are negatively associated with different objective face identification measures, such as the CFMT original (Livingston & Shah, 2018; Shah, Gaule, et al., 2015; Ventura, Livingston,

& Shah, 2018) and the CFMT-Chinese (Estudillo, 2020; Nakashima et al., 2020) versions, famous faces recognition tests (Shah, Gaule, et al., 2015; Ventura et al., 2018), and the Glasgow Face Matching Test (Shah, Sowden, et al., 2015). Importantly, this negative association is held in those participants who have not received formal feedback about their face recognition abilities (Gray, Bird, & Cook, 2017; Livingston & Shah, 2018). Therefore, it seems that the PI-20 is a fast and valid method that can be used as a complementary tool for studying individual differences in face identification.

However, despite these promising findings, the PI-20 and other self-reported measures of face identification are not free of criticisms. For example, it has been reported that the associations between objective and self-reported measures of face identification are only moderate (Bobak et al., 2019; Gray et al., 2017; Shah, Gaule, et al., 2015). This is such that PI-20 scores explain only around 5 to 15% of the variance in the scores of the CFMT in normal populations (Gray et al., 2017; Livingston & Shah, 2018; Matsuyoshi & Watanabe, in press; Nakashima et al., 2020). Interestingly, when developmental prosopagnosics are tested, the amount of explained variance increases to 46% (Shah, Gaule, et al., 2015), suggesting that compared to normal population, people with prosopagnosia might have more accurate insights into their face recognition abilities (Palermo et al., 2017). It has also shown that super-recognizers also seem to have better insights into their face recognition abilities compared to control participants, especially in target-present face matching trials (Bate & Dudfield, 2019), although this study did not use the PI-20. Thus, one question that arises is whether the moderate association usually found between objective and self-reported measures of face identification is merely driven by people with relatively low and high face recognition abilities.

The present study seeks to shed light on this question using the Mandarin version of the PI-20. Similar to other studies, our observers performed both the PI-20 and the CFMT. In addition to exploring individuals' insights into face recognition abilities on the entire distribution of scores, unlike other studies, we also explored whether these insights depend on observers' face recognition performance level. To achieve this, we divided our sample into four different quartiles according to their scores in the CFMT. This quartile-split approach is a standard approach in metacognition research that was firstly introduced by Dunning and colleagues (Dunning, Johnson, Ehrlinger, & Kruger, 2003). This method has been widely used since then to study metacognition in different cognitive processes, including reasoning (Pennycook, Ross, Koehler, & Fugelsang, 2017), intelligence (Unsworth & Engle, 2005), working memory (Adam & Vogel, 2017) and, more recently, face perception (Zhou & Jenkins, 2020). The aim of this approach is to have four subgroups of participants of approximately the same size, representing different degrees of performance in the task (i.e., Q1: low performance, Q2: low-average performance, Q3: average-high performance, Q4: high performance). We also applied the quartile-split approach to reanalyze the data of a published study that found a *robust* association between the CFMT and the PI20 in the general population (Gray et al., 2017). If observers have insights into their face recognition abilities, we would find a negative association between the PI20 and the CFMT in the whole sample. If these insights are presented across the whole range of face recognition abilities, this negative association between the PI20 and the CFMT will also be observed in each quartile separately.

Materials & Methods

We confirm that we report how all the measures, manipulations and data exclusions in this study. We also report how we have determined our sample size.

Participants

Our sample size was determined a priori based on other studies (e.g., Shah, Sowden, et al., 2015; Ventura et al., 2018). A total of 280 Chinese ethnicity students from HELP University and the University of Nottingham Malaysia took part in this study for course credits. Twenty-five participants were excluded due to performance at chance level and/or abnormally fast response times (< 500 ms), suggesting lack of engagement with the task. Our final web sample consisted of 255 participants (67 males). Observers' mean age was of 21 years (SD = 4.2). All participants reported having normal or corrected-to-normal vision. Observers were naïve regarding the aims of the study and were never tested before with either the CFMT or the PI-20. Participants provided written informed consent[1] and were debriefed at the end of the study. This study was approved by the university research ethics review committee (AJE271017).

Materials, Apparatus and Procedure

Participants were tested over the web using the application testable (www.testable.com) to present stimuli and to record observers' responses. This study involves an objective measure of face recognition (i.e., the CFMT-Chinese; McKone et al., 2012) and a self-reported measure of face recognition (i.e., the PI-20; Shah, Gaule, et al., 2015). The PI-20 was translated into Mandarin. The order of these tasks was randomized across participants.

The CFMT-Chinese. The paradigm of the CFMT-Chinese (McKone et al., 2012) is identical to the classical CFMT (Duchaine & Nakayama, 2006) but it contains Chinese-ethnic faces as stimuli. This task requires participants to learn and recognize different unfamiliar faces in three different stages: *same image*, *novel images* and *novel images with noise*. Observers are firstly required to study a target identity presented in frontal, mid-profile left, and mid-profile right orientations Each of these orientations is presented individually for three seconds. Observers are then presented with the target identity among two other filler face distractors and are required to identify the target, in each of the three orientations. The three face images are presented until response. This procedure is repeated for five additional target identities. The *same image* stage contains a total of 18 trials (three face orientation for each of the six identities). Observers then proceed to the *novel images* stage. In this stage, observers are required to study the same six target identities for 20 seconds. All the target identities are simultaneously presented in the same display. Observers are then presented with a new instance of the target identity among two filler face distractors and are asked to identify the target face. On each 3-item stimulus array, the target

[1]The consent form was provided in English language

face can be any one of the six learned targets, always presented in a novel image (i.e., different viewpoints, lighting condition or both). This second stage has a total of 30 trials. The *novel images with noise* stage is identical to the *novel images* stage, but target identities and filler faces distractors are presented with visual noise to make the task harder. This stage has 24 trials. The maximum total scores observers can get in the CFMT is 72 (i.e., one point for each correct trial). Internal reliability analysis showed an alpha value of 0.85 which is in agreement with previous research (e.g., Estudillo et al., 2020; Estudillo, 2020; McKone et al., 2012).

The Mandarin PI-20. In this stage, observers completed the Mandarin version of the PI-20 (see Appendix 1). The PI-20 (Shah, Gaule, et al., 2015) is a self-reported measure of face recognition. It contains 20 items describing daily life situations related with face recognition (e.g., My face recognition ability is worse than most people). Observers are required to rate their agreement with each statement on a five-point Likert-scale (1 = strongly agree, 5 = strongly disagree). Items 8, 9, 13, 17 and 19 were reverse scores. Lower scores in the PI-20 indicates lower face recognition abilities. Internal reliability analysis revealed an alpha value of 0.88, which is in agreement with previous research (e.g., Estudillo, 2020; Shah, Gaule, et al., 2015).

Results

We firstly explored observer's insights into their face recognition abilities. As shown in Figure 1A, observers scores in the CFMT-Chinese were negatively associated with their scores in the PI-20 [r = -0.35, $p < .001$]. This moderate correlation shows that around 12% of the variation in the CFMT scores can be explained by the scores in the PI-20.

Secondly, we explored whether the insights into face recognition abilities are stable across different levels of recognition performance. To achieve this aim, observers were grouped in four quartiles, following their score in the CFMT-Chinese (see Table 1). The range of scores were 32-50, for the first quartile; 51-56, for the second quartile; 57-63, for the third quartile; and 64-72, for the fourth quartile. As shown in Figure 1B, observers' scores in the CFMT-Chinese were negatively associated with their scores in the PI-20 for the first [r = -0.26, $p = .03$] and fourth [r = -0.28, $p = .02$] quartiles. Despite these reliable associations, only approximately 7% of the variation in the CFMT scores can be explained by the scores in the PI-20 . For the second and third quartiles, the association between the CFMT-Chinese and the PI-20 was not reliable [both rs ≤ -.06, $ps \geq .96$]. It is possible that the lack of correlation in the second and third quartiles is due to a lack of variation in the data. In fact, a closer inspection of Figure 1B reveals that this explanation is plausible, especially for the second quartile. To rule out this possibility, we increased the variability of the data by combining scores in these two quartiles. However, the association between CFMT-Chinese and the PI-20 was still not reliable [r = -.00, $p = .99$]. Altogether our results suggest that, at the best, only above- and below-average recognisers have insights into their face recognition abilities.

Re-analysis of Gray et al's (2017) study

Gray and colleagues' data (Gray, Bird, & Cook, 2017) are freely available (see their supplemental data). Their study presented the results of two independent samples (n = 142, and n = 283). We decided to reanalyse Gray and colleagues' results as their procedure is highly similar to ours. As the only remarkable difference between Gray and colleagues' samples is that they were collected in different cities of the UK, we decided to combine them (n = 425 participants, 162 males). This approach has two main advantages. First, it increases the power to detect a potential effect if that effect truly exists. This is particularly important for the quartile-split analysis, as the total sample size is reduced. In addition, as the quartile-split approach takes into consideration the whole range of scores to create the quartiles, the larger the sample size the more certain we are that a specific score corresponds to a specific quartile in the population. .
As Gray and colleagues reported (see Figure 2A), scores in the CFMT were negatively associated with scores in the PI-20 [r = -.39, $p < .001$]. This moderate correlation is consistent with our results and shows that around 15% of the variation in the CFMT scores can be explained by the scores in the PI-20. Interestingly, when their observers were grouped into quartiles according to their scores in the CFMT (see Figure 2B and Table 1), there was a negative association between the CFMT and the PI-20, for the first [r = -0.30, $p < .001$] and fourth [r = -0.21, $p = .03$] quartiles. Variation in the CFMT scores explains around 9% and 4% of the scores in the PI-20, for the first and fourth quartile, respectively. Although there was no association between the CFMT and the PI-20 for the second quartile [r = -.01, $p = .91$], there was a positive reliable association between the CFMT and the PI-20 for the third quartile [r = .21, $p = .02$]. This association, which is in the opposite direction to the expected if observers had insights into their recognition abilities, disappears when scores in the second and third quartiles are combined [r = -.00, $p = .63$]. Overall, the re-analysis of Gray and colleagues' data is in line with our hypothesis that only below- and above-average recognizers have insights into their face recognition abilities.

Discussion

This study investigated observers' insight into their face recognition abilities with the Mandarin version on the PI-20. We found a reliable negative association between observers' scores in the CFMT-Chinese and their self-reported face recognition abilities on the PI-20. We also explored whether these insights are consistent across different levels of actual face recognition performance. To achieve this, following previous research in metacognition (e.g., Dunning et al., 2003), we adopted a quartile-split approach. We found a weak but reliable negative association between the CFMT-Chinese and the PI-20 in the first and fourth quartiles, but not in the second and third quartiles. We also re-analysed a publicly available sample of 425 Caucasian participants (Gray et al., 2017). In the first and fourth quartile, we found a small but significant negative association between the CFMT and the PI20. In the second quartile, no association was found between both measures. Finally, although in the third quartile we found a positive association between the CFMT and the PI20, this association is in the opposite direction to that expected if participants had insights into their face recognition abilities. Thus, our results not only question previous findings that suggest that adults have moderate to strong insights into their face recognition (Gray et al., 2017; Livingston & Shah, 2018; Shah, Gaule, et al., 2015), but also suggest that only good and bad recognizers have (limited) insights into their face recognition abilities. It is important to note that the pattern of results found cannot be explained in terms of lack of variation in the scores in the CFMT in the second and third quartiles, as the same pattern of results was observed when the scores in these quartiles were combined. This is remarkable as the range of the CFMT scores in the combined quartiles is similar in size to that in the first quartile and larger than the range of scores in the fourth quartile. This combination of the scores in the second and third quartiles also rules out that our results are due to lack of power, as the number of observations is approximately twice compared to the first and the fourth quartiles.

Some authors have suggested that previously observed associations between objective and self-reported measures of face identification are inflated because those previous studies included developmental prosopagnosic patients in the sample (Bobak, Mileva, & Hancock, 2019; Palermo et al., 2017). More recent research showed that this association was held reliable –but much weaker– when developmental prosopagnosic patients were not included in the sample (Gray et al., 2017; Livingston & Shah, 2018). Our findings provide compelling evidence suggesting that this association is still mainly driven by people with above- and below-average face recognition abilities.

One question that arises, therefore, is why insights into face recognition abilities are only observed at the lower and upper end of the face recognition abilities distribution. One potential reason could be that these people have previously received formal feedback as part of their participation in face recognition studies (Bobak et al., 2019). Yet, in Gray et al.'s (2017) and the current study, observers were naïve regarding the aims of the study and did not complete formal testing of their face recognition ability. In addition, it could also be possible that people with low and high face recognition abilities receive more consistent social feedback about their recognition abilities (e.g., when not recognizing a close friend or when recognizing someone not seen in years). However, this explanation is inconsistent with some reported cases of people with developmental prosopagnosia who were largely unaware of their face recognition deficits (Bowles et al., 2009; Grueter et al., 2007). Thus, why only above- and below-average recognizers have insights into their face recognition abilities is a question for future research.

It must be noted that the aim of the PI-20 is to help the diagnosis of face recognition disorders and particularly prosopagnosia (Gray et al., 2017; Shah, Gaule, et al., 2015; Shah, Sowden, Gaule, Catmur, & Bird, 2015). In principle, this is further supported by our results. However, as also shown by our results, variation in the CFMT scores only explained around 7% of the scores in the PI-20, which suggests that even people within the lower range of face identification abilities have very limited insights into their face recognition abilities. In fact, it has been estimated that the PI-20 would fail to detect around 60% of developmental prosopagnosics who would be diagnosed with objective measures of face recognition (Arizpe et al., 2019). For this reason, it is recommended that the diagnosis of prosopagnosia should be mostly based on objective tests and complemented with self-reported measures of face identification (Arizpe et al., 2019; Bobak et al., 2019; Palermo et al., 2017).

Conclusions

In summary, the current study reports a moderate negative association between the CFMT and the Mandarin version of the PI-20. This association is in agreement with previous research (Bobak et al., 2019; Gray et al., 2017; Livingston & Shah, 2018; Shah, Sowden, et al., 2015; Ventura et al., 2018). However, a deeper analysis of our study and the reanalysis of publicly available data (Gray et al., 2017) suggest that this association is mainly driven by people below- and above-average face recognition abilities. Altogether our results suggest that the use of self-reported measures of face identification should be, when possible, complemented with objective measures.

References

Adam, K. C. S., & Vogel, E. K. (2017). Confident failures: Lapses of working memory reveal a metacognitive blind spot. *Attention, Perception, and Psychophysics, 79(5)*, 1506–1523. https://doi.org/10.3758/s13414-017-1331-8

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

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

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., & Tree, J. J. (2017). The definition and diagnosis of developmental prosopagnosia. *The Quarterly Journal of Experimental Psychology*, *70*(2), 193–200. https://doi.org/10.1080/17470218.2016.1195414

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*, *72*, 872–881. https://doi.org/10.1177/1747021818776145

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., … Yovel, G. (2009). Diagnosing prosopagnosia: effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive Neuropsychology*, *26*(5), 423–455. https://doi.org/10.1080/02643290903343149

Bruce, V. (1982). Changing faces: visual and non-visual coding processes in face recognition. *British Journal of Psychology*, *73*, 105–116.

Dalrymple, K. A., Fletcher, K., Corrow, S., das Nair, R., Barton, J. J. S., Yonas, A., & Duchaine, B. (2014). “A room full of strangers every day”: The psychosocial impact of developmental prosopagnosia on children and their families. *Journal of Psychosomatic Research*, *77*(2), 144–150. https://doi.org/10.1016/j.jpsychores.2014.06.001

Dalrymple, K. A., & Palermo, R. (2016). Guidelines for studying developmental prosopagnosia in adults and children. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(1), 73–87. https://doi.org/10.1002/wcs.1374

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576–585. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2005.07.001

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why People Fail to Recognize Their Own Incompetence. *Current Directions in Psychological Science, 12(3)*, 83–87. https://doi.org/10.1111/1467-8721.01235

Estudillo, A. J. (2012). Facial Memory: The Role of the Pre-Existing Knowledge in Face Processing and Recognition. *Europe’s Journal of Psychology*, *8*(2), 231–244. https://doi.org/10.5964/ejop.v8i2.455

Estudillo, A. (2020). Self-reported face recognition abilities for own and other-race faces. *PsyArXiv*. https://doi.org/10.31234/osf.io/ru973

Estudillo, A. J., & Bindemann, M. (2014). Generalization across view in face memory and face matching. *I-Perception*, *5*(7), 589–601. https://doi.org/10.1068/i0669

Estudillo, A. J., Lee, J. K. W., Mennie, N., & Burns, E. (2019). No evidence of other-race effect for Chinese faces in Malaysian non-Chinese population. *Applied Cognitive Psychology*, *34*, 270–276. https://doi.org/10.1002/acp.3609

Gray, K. L. H., Bird, G., & Cook, R. (2017). Robust associations between the 20-item prosopagnosia index and the Cambridge Face Memory Test in the general population. *Royal Society Open Science*, *4*(3), 160923. https://doi.org/10.1098/rsos.160923

Grueter, M., Grueter, T., Bell, V., Horst, J., Laskowski, W., Sperling, K., … Kennerknecht, I. (2007). Hereditary Prosopagnosia: the First Case Series. *Cortex*, *43*(6), 734–749. https://doi.org/10.1016/S0010-9452(08)70502-1

Kennerknecht, I., Ho, N. Y., & Wong, V. C. N. (2008). Prevalence of hereditary prosopagnosia (HPA) in Hong Kong Chinese population. *American Journal of Medical Genetics Part A*, *146A*(22), 2863–2870. https://doi.org/10.1002/ajmg.a.32552

Lander, K., Bruce, V., & Bindemann, M. (2018). Use-inspired basic research on individual differences in face identification: implications for criminal investigation and security. *Cognitive Research: Principles and Implications*, *3*(1), 26. https://doi.org/10.1186/s41235-018-0115-6

Livingston, L. A., & Shah, P. (2018). Article Commentary: People with and without Prosopagnosia Have Insight into Their Face Recognition Ability. *Quarterly Journal of*

Experimental Psychology, *71*, 1260–1262. https://doi.org/10.1080/17470218.2017.1310911

Longmore, C. a, Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(1), 77–100. https://doi.org/10.1037/0096-1523.34.1.77

Matsuyoshi, D., & Watanabe, K. (in press). People have modest, not good, insight into their face recognition ability: a comparison between self-report questionnaires. *Psychological Research*, *1*, 3. https://doi.org/10.1007/s00426-020-01355-8

McKone, E., Stokes, S., Liu, J., Cohan, S., Fiorentini, C., Pidcock, M., … Caldara, R. (2012). A Robust Method of Measuring Other-Race and Other-Ethnicity Effects: The Cambridge Face Memory Test Format. *PLoS ONE*, *7*(10), e47956. https://doi.org/10.1371/journal.pone.0047956

McKone, E., Wan, L., Robbins, R., Crookes, K., & Liu, J. (2017). Diagnosing prosopagnosia in East Asian individuals: Norms for the Cambridge Face Memory Test–Chinese. *Http://Dx.Doi.Org/10.1080/02643294.2017.1371682*. https://doi.org/10.1080/02643294.2017.1371682

Nakashima, S. F., Ukezono, M., Sudo, R., Nunoi, M., Kitagami, S., Okubo, M., … 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

Palermo, R., Rossion, B., Rhodes, G., Laguesse, R., Tez, T., Hall, B., … McKone, E. (2017). Do People Have Insight into their Face Recognition Abilities? *Quarterly Journal of Experimental Psychology*, *70*(2), 218–233. https://doi.org/10.1080/17470218.2016.1161058

Ramon, M., Bobak, A. K., & White, D. (2019). Super-recognizers: From the lab to the world and back again. *British Journal of Psychology*, *110*(3), 461–479. https://doi.org/10.1111/bjop.12368

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin and Review, 24(6)*, 1774–1784. https://doi.org/10.3758/s13423-017-1242-7

Robertson, D. J., Noyes, E., Dowsett, A., Jenkins, R., Burton, A. M., & Burton, M. (2016). Face recognition by Metropolitan Police Super-recognisers Correspondence to. *PLoS ONE*, *11*, e0150036. https://doi.org/10.1371/journal.pone.0150036

Rossion, B. (2018). Damasio's error - Prosopagnosia with intact within-category object recognition. *Journal of Neuropsychology*. https://doi.org/10.1111/jnp.12162

Russell, R., Duchaine, B., & 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), 1–11. https://doi.org/10.1098/rsos.140343

Shah, P., Sowden, S., Gaule, A., Catmur, C., & Bird, G. (2015). The 20 item prosopagnosia index (PI20): Relationship with the Glasgow face-matching test. *Royal Society Open Science*, *2*(11), 0–5. https://doi.org/10.1098/rsos.150305

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence, 33*(1), 67–81. https://doi.org/10.1016/j.intell.2004.08.003

Ventura, P., Livingston, L. A., & Shah, P. (2018). Adults have moderate-to-good insight into their face recognition ability: Further validation of the 20-item Prosopagnosia Index in a Portuguese sample. *Quarterly Journal of Experimental Psychology*, *71*, 2677–2679. https://doi.org/10.1177/1747021818765652

Wilmer, J. B. (2017). Individual Differences in Face Recognition: A Decade of Discovery. *Current Directions in Psychological Science*, *26*(3), 225–230. https://doi.org/10.1177/0963721417710693

Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008). Psychosocial consequences of developmental prosopagnosia: A problem of recognition. *Journal of Psychosomatic Research*, *65*(5), 445–451. https://doi.org/10.1016/j.jpsychores.2008.03.013

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

Figure 1

A) Associations between PI20 scores and performance on the CFMT-Chinese. (B) Associations between PI20 scores and performance on the CFMT-Chinese for each quartile.

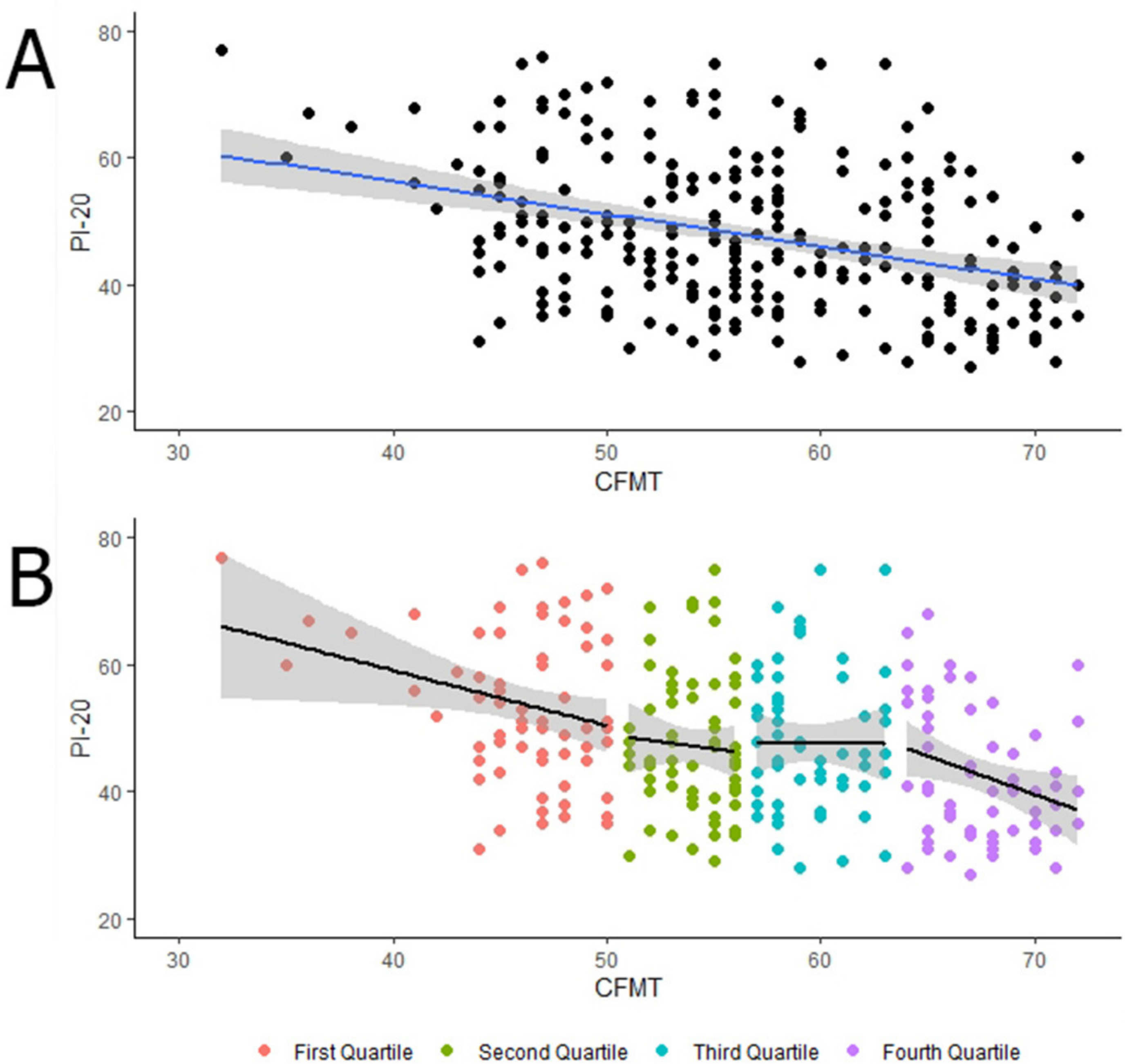

Figure 2

Reanalysis of Gray and Colleagues' results (A) Associations between PI20 scores and performance on the CFMT. (B) Associations between PI20 scores and performance on the CFMT for each quartile

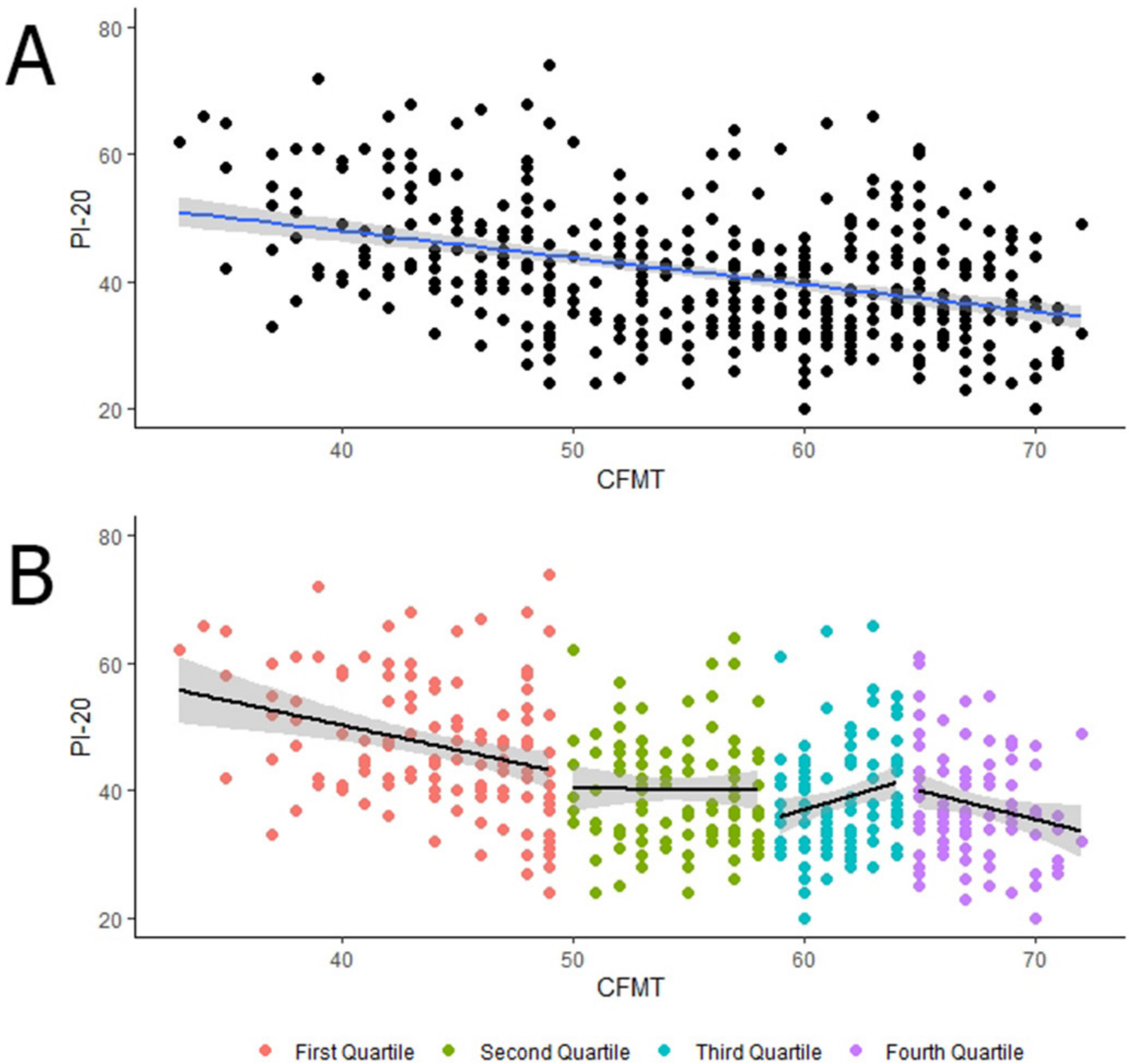

Table 1(on next page)

Descriptive statistics for the total sample and across each quartile in our study and Gray and colleagues' study.

	PRESENT STUDY						GRAY AND COLLEAGUES' STUDY					
	N	PI-20		CFMT-Chinese			N	PI-20		CFMT		
		Mean	SD	Mean	SD	Range		Mean	SD	Mean	SD	Range
Q1	66	53.81	11.87	46.07	3.62	32-50	120	47.16	10.46	43.96	4.08	33-49
Q2	68	47.10	10.97	54.04	1.65	51-56	102	40.14	8.94	54.48	2.48	50-58
Q3	60	47.60	11.35	59.51	2.07	57-63	110	38.63	8.38	61.57	1.71	59-64
Q4	61	42.65	10.35	67.40	2.46	64-72	93	37.51	37.51	67.67	2.01	65-72
TOTAL	255	47.89	11.78	56.46	8.19	32-72	425	41.16	9.92	56.23	9.34	33-72