A new Asian version of the CFMT: The Cambridge Face Memory Test – Chinese Malaysian (CFMT-MY)

Siew Kei Kho[1], Bryan Qi Zheng Leong[1], David R. T. Keeble[1], Hoo Keat Wong[1] and Alejandro J. Estudillo[1,2]

[1]*School of Psychology,*

University of Nottingham Malaysia, Semenyih, Malaysia.

[2]*Department of Psychology,*

Bournemouth University, Bournemouth, U.K.

Correspondence to:

Siew Kei, Kho, School of Psychology, University of Nottingham Malaysia, Jalan Broga, 43500 Semenyih, Selangor.

Telephone number: +601111434017 MY, e-mail address: khpy5ksk@nottingham.edu.my

or

Alejandro J. Estudillo, Bournemouth University, Department of Psychology, Poole House Talbot Campus, BH12.

Telephone number: +4401202962229 5BB UK, e-mail address: aestudillo@bournemouth.ac.uk

Abstract

The Cambridge Face Memory Test (CFMT) is one of the most important measures of individual differences in face recognition and for the diagnosis of prosopagnosia. Having two different CFMT versions using a different set of faces seems to improve the reliability of the evaluation. However, at the present time, there is only one Asian version of the test. In this study, we present the Cambridge Face Memory Test – Chinese Malaysian (CFMT-MY), a novel Asian CFMT using Chinese Malaysian faces. In Experiment 1, Chinese Malaysian participants (*N* = 134) completed two versions of the Asian CFMT and one object recognition test. The CFMT-MY showed a normal distribution, high internal reliability, high consistency and presented convergent and divergent validity. Additionally, in contrast to the original Asian CFMT, the CFMT-MY showed an increasing level of difficulties across stages. In Experiment 2, Caucasian participants (*N* = 135) completed the two versions of the Asian CFMT and the original Caucasian CFMT. Results showed that the CFMT-MY exhibited the other-race effect. Overall, the CFMT-MY seems to be suitable for the diagnosis of face recognition difficulties and could be used as a measure of face recognition ability by researchers who wish to examine face-related research questions such as individual differences or the other-race effect.

Keywords: Cambridge Face Memory Test, Asian, other-ethnicity effect, face memory, face recognition, prosopagnosia, neuropsychological test

A new version of the Asian CFMT: The Cambridge Face Memory Test-Chinese Malaysian (CFMT-MY)

Faces are one of the most critical stimuli for successful social interaction. However, despite its importance, face recognition abilities present substantial inter-individual variability (Bowles et al., 2009; Bruce et al., 2018; Wang et al., 2012; Wilmer, 2017) with some people showing superior face recognition (i.e., super-recognizers) (Russell et al., 2009) while others presenting difficulties in recognizing even highly familiar faces (i.e., prosopagnosics) (Rossion, 2014). Prosopagnosia, also known as face blindness, is a visual impairment that affects face recognition despite intact visual acuity and intelligence and can result from brain injury (i.e., acquired prosopagnosia) or abnormal development (i.e., developmental or congenital prosopagnosia). Remarkably, these difficulties in face recognition could contribute to negative social consequences (e.g., high anxiety in social situations) not only for adults (Yardley et al., 2008), but also for children (Dalrymple et al., 2014). Although the estimated prevalence of developmental prosopagnosia in the general population is around 2.5% (Bowles et al., 2009; Kennerknecht et al., 2006, 2008), many cases remain undiagnosed (Duchaine, 2000). Given the limited insights that people have into their own face recognition skills (Bate & Dudfield, 2019; Bobak et al., 2019; Estudillo, 2021; Estudillo & Wong, 2021; Palermo et al., 2017), objective measures of face identification are crucial for the study of individual differences in face recognition skills and the diagnosis of prosopagnosia.

One of the most prominent objective measures of face recognition abilities is the Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006). This test, which can be completed in about 15 minutes, provides a valid measure of face recognition, as it requires the identification of faces across different views (Bruce, 1982; Estudillo & Bindemann, 2014). The CFMT is poorly correlated with general intelligence (Shakeshaft & Plomin, 2015) and object recognition ability (Dennett et al., 2012; Shakeshaft & Plomin, 2015), which suggests that this test taps into face identification specific processes. The original version of the CFMT consists of a three-alternative forced choice paradigm subdivided into three stages of increasing difficulty. Participants are firstly asked to study six Caucasian target faces. Subsequently, during the recognition trials, the target faces are presented without any variation in the image (learning stage) with different lighting and viewpoint (novel stage) and with the addition of visual noise (novel-with-noise stage). The CFMT has been widely used to investigate different aspects of face recognition, including its heritability (Wilmer et al., 2010), development (Germine et al., 2011), relationship with holistic processing (DeGutis et al., 2013) and other group effects (Childs et al., 2021; Estudillo et al., 2020; McKone et al., 2012; Wan et al., 2017). Importantly, because it has high reliability (Cronbach's alpha (α) ≈ .90), the CFMT is also used to aid the diagnosis of prosopagnosia (e.g., Bowles et al., 2009; Duchaine & Nakayama, 2006; Estudillo et al., 2020; McKone et al., 2017). Specifically, individuals scoring two standard deviations below the mean CFMT performance are considered as possible prosopagnosia cases (Duchaine & Nakayama, 2006).

Despite the remarkable psychometric properties of the CFMT, several factors (i.e., problems understanding the instructions and inattentiveness) could influence the final test scores irrespective of actual face recognition skills (see e.g., Gamaldo & Allaire, 2016). Although repeating the same test could provide a more reliable score, this practice is not exempt from problems (McCaffrey & Westervelt, 1995). For example, due to face familiarity effects as a consequence of using the same face stimuli, an individual who scored below the cut-off value during the first assessment may score above the cut-off value in the next reassessment test (Murray & Bate, 2020). This familiarity effect could be easily avoided by using a complementary version of the CFMT containing a different set of face stimuli (Murray & Bate, 2020). In addition, with the increasing interest in face training protocols (Bate, Adams, et al., 2019; Corrow et al., 2019; Davies-Thompson et al., 2017), having complementary versions of the CFMT is also highly useful for rigorous pre-post training comparisons. For Caucasian participants, such a complementary version does exist, the CFMT-Australian (CFMT-Aus) (McKone et al., 2011). Importantly, the psychometric properties of the CFMT-Aus are comparable to those of the original CFMT, making this test not only an alternative to the original CFMT, but also a complementary assessment tool in the aforementioned situations.

People tend to be better recognizing faces from their own-race compared to other-race faces, the so-called other-race effect (Meissner & Brigham, 2001). Both the CFMT-original and the CFMT-Aus consist of Caucasian face stimuli and have shown strong other-race effects (see e.g., Estudillo et al., 2020; McKone et al., 2012; Wan et al., 2017), limiting their use to Caucasian populations. The CFMT-Chinese (McKone et al., 2012) was introduced to study individual differences in face recognition and aid the diagnosis of prosopagnosia in Asian populations. This test follows an identical format compared to the original version of the test and has comparable psychometric properties (McKone et al., 2017). However, at present, the CFMT-Chinese is the only Asian version of the CFMT which, as previously discussed, might present difficulties for the study of individual differences, the diagnosis of borderline cases of prosopagnosia and pre-post face training comparisons.

Although the CFMT-Chinese aims to explore individual differences in face recognition and aid the diagnosis of prosopagnosia in the Asian population, the other-race effect has also been found within the Asian population (Wong et al., 2020). However, other studies using the CFMT-Chinese found that the scores of the CFMT-Chinese were still higher than the CFMT-original where the Asian participants recruited comprised of a variety of Asian origins, some of which were not Chinese, such as Indonesian (McKone et al., 2012), Malay and Filipino participants (Bate, Bennetts, et al., 2019). Similarly, Estudillo et al. (2020) found that although Malaysian Malay and Malaysian Indian showed a clear other-race effect for Caucasian faces, they presented identical performance for Chinese faces compared to Malaysian Chinese participants in the CFMT-Chinese. Altogether, these findings suggest that non-Chinese Asians may perform better for Chinese faces as compared to Caucasian faces. Despite the fact that using the CFMT with Chinese faces for the diagnosis of prosopagnosia among the non-Chinese Asian population may not be ideal, currently, the CFMT-Chinese may still be a superior face recognition measure compared to the Caucasian CFMT versions for the diagnosis of prosopagnosia among the non-Chinese Asian population.

Present study

In the current study, we presented a novel Asian version of the CFMT, the CFMT-Chinese Malaysian (CFMT-MY). In Experiment 1, we determined the psychometric properties of the CFMT-MY using a Chinese Malaysian sample. Specifically, in Experiment 1 we explored the internal reliability, convergent validity and divergent validity of the CFMT-MY. Experiment 1 also tested whether the three stages of the CFMT-MY represent increasing levels of difficulty. The increasing levels of difficulty across stages is an important property of the CFMT-original (Duchaine & Nakayama, 2006) that has been overlooked in the CFMT-Chinese (e.g., Estudillo et al., 2020; McKone et al., 2012, 2017). After checking the psychometric properties of the CFMT-MY, Experiment 2 used a sample of Caucasian participants to explore whether the CFMT-MY captures an other-race effect of similar magnitude compared to that of the CFMT-Chinese.

Experiment 1

Experiment 1 aimed to investigate the psychometric properties of the CFMT-MY. In addition to measures of reliability (Cronbach's α) and internal consistency across stages, we explored the convergent and divergent validities of the test. Convergent validity was explored by correlating participants' performance in the CMFT-MY with their performance in the CFMT-Chinese. Divergent validity was explored by correlating participants' performance in the CFMT-MY and their performance in a general object recognition task that follows the same format as the CFMT: the Cambridge Car Memory Test (CCMT) (Dennett et al., 2012). If the CFMT-MY had appropriate convergent and divergent validity we would expect a stronger correlation between the CFMT-MY and the CFMT-Chinese than between the CFMT-MY and the CCMT. Additionally, we examined the increasing level of difficulty across the three stages of the CFMT-Chinese and the CFMT-MY. Differences in accuracy between the different stages of the CFMT-MY and CFMT-Chinese were assessed using repeated-measures ANOVA.

Methods

Participants

One hundred and thirty-nine participants took part in this experiment, but the final sample included 134 Chinese Malaysians (92 females and 42 males) with an age range of 18 to 66 years (*M* = 22.81 years, *SD* = 5.53 years). The age range for female participants was between 18 and 66 years (*M* = 22.50 years, *SD* = 6.24 years) while for male participants, the age range was from 18 to 35 years (*M* = 23.48 years, *SD* = 3.47 years). Data from participants of other-ethnicity (e.g., Malay, Indian, Eurasian, mixed) (four participants) and that had median reaction times less than 500ms (one participant) were removed from further analysis. Eight additional participants were excluded from the data analysis (except for internal reliability and internal consistency analyses) as their performance on the face memory tasks was indicative of possible prosopagnosia (Appendix A). The remaining participants were 126 Chinese Malaysians (87 females and 39 males) with an age range from 18 to 66 years (*M* = 22.92 years, *SD* = 5.64 years). The age range for female participants was between 18 and 66 years (*M* = 22.58 years, *SD* = 6.38 years) while for male participants, the age range was between 18 and 35 years (*M* = 23.69 years, *SD* = 3.40 years).

An a priori power analysis was conducted using G*Power 3.1 (Faul et al., 2009) for a repeated-measures ANOVA comparing the stages of the two Asian CFMT versions (CFMT-MY and CFMT-Chinese). The effect size for the CFMT stage was based on Murray and Bate (2020) where η_p^2

= .824, a large effect size. A large effect size estimate (η_p^2 = .14) was entered into the power analysis with the following parameters: α = .05, power = .95. The power analysis implied that *N* = 50 would be required to detect a difference between the CFMT versions with 95% probability. A priori power analysis was also conducted for correlation tests comparing two versions of face memory test (CFMT-Chinese and CFMT-MY). The correlations between two different versions of the CFMT reported in past studies were higher than .5 (e.g., *r* = .71 in Arrington et al., 2022; *r* = .61 in McKone et al., 2011). A medium correlation (*r* = .5) was entered into the power analysis with the following parameters: α = .05, power = .95. The power analysis suggested that *N* = 46 would be required to detect a correlation between the two CFMT versions with 95% probability.

All participants provided informed consent to participate in the study. Upon recruiting every 10 participants, a lucky draw was held with each participant given a chance to win RM20 or alternatively course credits were given for participation. The study has been reviewed and approved by the Science and Engineering Research Ethics Committee (SEREC) at the University of Nottingham Malaysia (approval code: KSK050320).

Cambridge Face Memory Test – Chinese (CFMT-Chinese)

The CFMT-Chinese was obtained from McKone et al. (2012). Fifty-two male identities were used in the task. The CFMT consists of three stages with increasing difficulty: learning stage (i.e., faces are presented in the same lighting and viewpoint condition), novel stage (i.e., faces are presented in different lighting and viewpoint condition) and novel-with-noise stage (i.e., faces are presented in different lighting and viewpoint condition with Gaussian noise applied). In total, there were 72 trials and six target faces to be memorized throughout the whole task.

Three practice trials with feedback were given before the experimental trials to familiarize participants with the procedure. The practice trials were identical to the procedure in the learning stage, but using cartoon images of Bart Simpson. In the learning stage, three study images (left 1/3 profile, frontal view and right 1/3 profile) of the same identity were presented sequentially for three seconds each with inter-trial interval of 500ms (Figure 1a). The target face was then presented with two distractor faces and participants were required to select the target face shown using the “1”, “2” or “3” key with no time limit (Figure 1b). In total, there were 18 trials in the learning stage (six target faces × three trials).

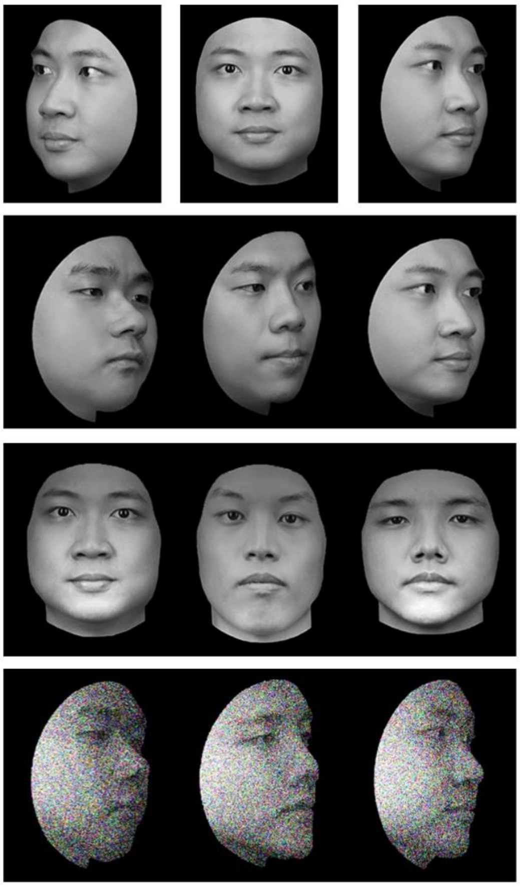
In the novel stage, participants were required to memorize the same six target faces in the learning stage which were presented simultaneously in frontal view for 20 seconds. Similar to the test phase of the previous stage, participants were required to select the target face presented with two distractor faces with no time limit. The images presented in this stage were different from the learning stage in terms of lighting and/or viewing angle (Figure 1c). In total, there were 30 trials in the novel stage (six target faces × five trials). The novel-with-noise stage was identical to the novel stage, except that noise was added to the test images to increase the difficulty level (Figure 1d). In total, there were 24 trials in the novel-with-noise stage (six target faces × four trials).

Cambridge Face Memory Test – Chinese Malaysian (CFMT-MY)

The stimuli used in the CFMT-MY were created using the University of Nottingham Malaysia face database, where photographs of students from the University of Nottingham Malaysia were obtained with informed consent before photographing. In total, 52 male Chinese Malaysian identities were used as stimuli. The faces had no piercings or glasses. Editing of images was conducted using Adobe Photoshop CS6. Blemishes, moles and facial hair were removed. Five different viewing angles of each identity were used (frontal, 45 degrees left, 45 degrees right, 90 degrees left and 90 degrees right). The face images were cropped to a size of 210 pixels in height while the width of the face images was resized according to the original proportion of the face. Each image was then placed onto a 200 x 250 pixels black canvas. Examples of the CFMT stimuli similar to the actual test stimuli used are shown below in Figure 1.

Figure 1

Sample CFMT-MY stimuli.



(a) Three study images in learning stage presented in different views

(b) Test trials in learning stage (faces are presented with same light and viewpoint condition as in study image)

(c) Test trials in novel stage (faces are presented with different lighting and cropping template)

(d) Test trials in novel-with-noise stage (faces are presented with different light and viewpoint condition with Gaussian noise applied)

Note. None of the faces shown in the sample figure were used in the actual task to avoid familiarity with the actual target faces used in the task.

The CFMT-MY was designed to replicate the original CFMT but using Chinese Malaysian faces. For the learning stage, the same cropping template was used for all targets and distractors. Frontal viewpoint, 45 degrees right and 45 degrees left were used. The distractors were matched to the target faces in the testing phase based on their similarity in appearance. Replicating the original

CFMT, target faces were never used as distractors and the distractors were presented repeatedly to ensure that participants could not use familiarity to decide if the faces were previously memorized or not.

In the novel stage of the original CFMT, images of the same identity were captured with different poses and physical lighting (i.e., the frontal view of the same identity was captured with lighting from the bottom or a slightly different frontal pose). However, such images did not exist in our face database. Thus, we used frontal viewpoint, 45 degrees right, 90 degrees right and 90 degrees left for the novel stage. We followed the procedure of CFMT-Aus (McKone et al., 2011) where instead of poses, different templates were used and lighting was added to the images using Adobe Photoshop CS6. For the frontal view and the 45 degrees right view, the images from the learning stage were used with modifications (i.e., the use of different external template shape and/or the addition of lighting). The external templates used were replicated based on CFMT-Aus. Point light was added using the function *Lighting effects*. The lighting was directed from the right for the 45 degrees right images, from the left for half of the frontal view images and from the bottom for the other half of the frontal view images. As the 90 degrees right and left images were not shown in the learning stage, only a template was used with no lighting changes made.

In the novel-with-noise stage, the viewpoints used were frontal, 45 degrees left and 90 degrees right. The lighting was directed from the right for half of the frontal view images and the 45 degrees left images. For the 90 degrees right images, the lighting was directed from the left. The other half of the front-facing images were made to appear lightly shadowed by adjusting the brightness and contrast (-30 brightness and +30 contrast). Different templates were applied to the frontal view and the 45 degrees left images. Next, 30% coloured Gaussian noise was added using the function *Add noise*. The CFMT-MY materials (stimuli and trial order) are available in the Open Science Framework repository, https://osf.io/gu4fy/?view_only=1ee5963c97664354a9d321057d22b641.

Cambridge Car Memory Test (CCMT)

CCMT was obtained from the authors of the task (Dennett et al., 2012). Fifty-two different cars were used in the CCMT. The CCMT follows the same procedure as CFMT, except the images presented were cars instead of faces.

Procedure

Testable (https://www.testable.org/) was used to run the online experiment (Rezlescu et al., 2020). To ensure that the stimuli size remained the same for different screen sizes, calibration was included before the start of the task where participants had to match the length of a line on the screen to the length of a bank card. The average vertical height of the face stimuli in the CFMT-Chinese and CFMT-MY was 4 cm while the average vertical height of the car stimuli in the CCMT was 3.5 cm. Participants completed all three tasks: Asian CFMT (Chinese and Malaysia) and CCMT in random order. The experiment took about 45 minutes to complete.

Results

All data analysis was conducted using JASP (JASP Team, 2022), except for the internal reliability analysis which was carried out using R software and R Studio (R Core Team, 2021; RStudio Team, 2021) including several R packages: dplyr (Wickham et al., 2021), tidyr (Wickham, 2021), data.table (Dowle & Srinivasan, 2021), psy (Falissard, 2012) and reshape (Wickham, 2007).

Normal distribution

The skewness (skew = −.397, *SE* = .216) and kurtosis (kurtosis = -.485, *SE* = .428) values for the CFMT-MY score were between ± 1 which indicates normal distribution (George & Mallery, 2019). Additionally, no significant skew was found for the scores of CFMT-MY (z = -1.838, p = .07). The mean score for CFMT-MY was 59.94/72, *SD* = 6.93.

Internal reliability

The internal reliability of the test was measured using Cronbach's α. For all trials, internal reliability was α = .86 for CFMT-MY. Results showed high internal reliability for CFMT-MY which was in line with previous work such as CFMT-Chinese, α = .86 (McKone et al., 2017) and CFMT-Aus, α = .88 (McKone et al., 2011).

Internal consistency

The internal consistency of the CFMT-MY at stage level (i.e., learning, novel and novel-with-noise) was measured using Pearson correlation (r). Results showed positive correlation between the learning and novel stage, $r(134) = .55$, $p < .001$, learning and novel-with-noise stage, $r(134) = .40$, $p < .001$ and novel and novel-with-noise stage, $r(134) = .68$, $p < .001$ showing that the scores were highly consistent across the different stages of CFMT-MY.

Validity

Convergent and divergent validity were measured using Pearson correlation (r). Convergent validity was measured by examining the correlation between the CFMT-Chinese and the CFMT-MY whereas divergent validity was measured by examining the correlation between the CCMT and the

CFMT-MY. Results showed positive correlation between the scores of the CFMT-MY and the CFMT-Chinese, $r(124) = .59$, $p < .001$. A weak positive correlation was found between the scores of the CCMT and the CFMT-MY, $r(124) = .26$, $p = .004$. The difference between the two correlation was further analyzed by comparing the dependent overlapping correlations (Diedenhofen & Musch, 2015; Hittner et al., 2003). The test showed that the correlation between the CFMT-MY and the CFMT-Chinese (i.e., convergent validity) was larger than the correlation between the CCMT and the CFMT-MY (i.e., divergent validity), $z = 3.62$, $p < .001$.

Repeated-measures ANOVA

A repeated-measures ANOVA was conducted to explore (1) potential differences between the CFMT-Chinese and the CFMT-MY and (2) the increasing levels of difficulty across the test stages. A 3 (stage: learning vs. novel vs. novel-with-noise) × 2 (test version: CFMT-MY vs. CFMT-Chinese) repeated-measures ANOVA was conducted on the accuracy (calculated by proportion correct scores). When the Mauchly's test indicated that the assumption of sphericity was violated, the degrees of freedom were corrected using the Greenhouse-Geisser method.

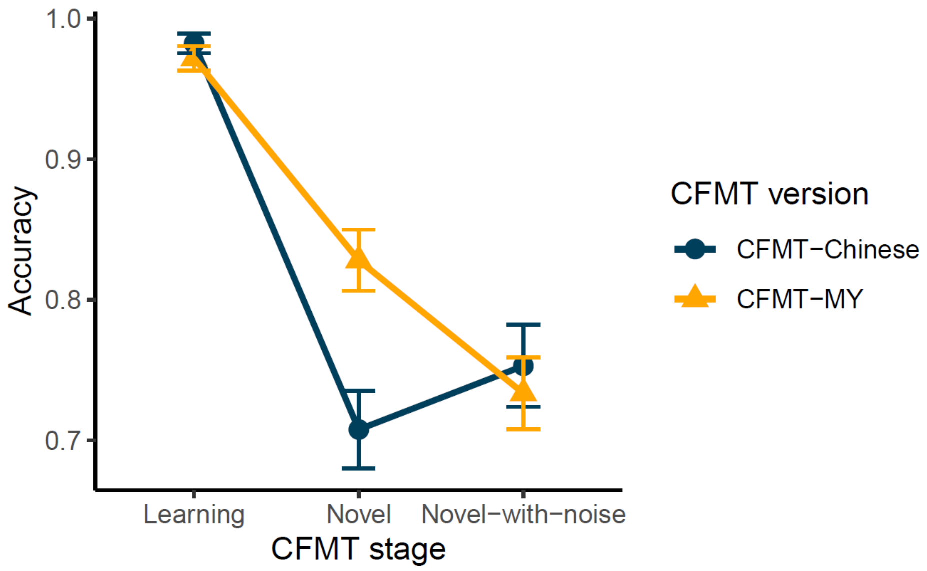
Analysis revealed a significant main effect of stage on accuracy, $F(1.70, 212.05) = 357.49$, $p < .001$, $\eta_p^2 = .74$. A post hoc Holm-Bonferroni test demonstrate that the accuracy of the learning stage ($M = 0.98$, $SD = 0.05$) was higher than the novel stage ($M = 0.77$, $SD = 0.15$), $p < .001$, $d = 1.94$. Similarly, the accuracy of the learning stage was higher than the novel-with-noise stage ($M = 0.74$, $SD = 0.16$), $p < .001$, $d = 2.17$. Accuracy was found to be higher in the novel stage compared to the novel-with-noise stage, $p = .01$, $d = 0.23$.

Results showed a significant main effect of test version on accuracy, $F(1, 125) = 14.03$, $p < .001$, $\eta_p^2 = .10$, where the accuracy of CFMT-MY ($M = 0.83$, $SD = 0.10$) was higher than CFMT-Chinese ($M = 0.79$, $SD = 0.12$). A significant interaction effect between stage and test version on accuracy was found, $F(2, 250) = 65.68$, $p < .001$, $\eta_p^2 = .34$ (Figure 2). Simple main effects analysis showed no differences between the test versions in the learning stage, $F(1, 125) = 3.497$, $p = .064$, $\eta^2 = .027$, and novel-with-noise stage, $F(1, 125) = 2.042$, $p = .156$, $\eta^2 = .016$. However, a significant effect was found in the novel stage, $F(1, 125) = 89.45$, $p < .001$, $\eta^2 = .417$, where the novel stage score for CFMT-MY ($M = 0.83$, $SD = 0.12$) was higher than CFMT-Chinese ($M = 0.71$, $SD = 0.16$).

Additional simple main effects analysis showed a significant main effect of stage on accuracy in the CFMT-MY, $F(1.86, 231.82) = 245.16$, $p < .001$, $\eta^2 = .66$. A post hoc Holm-Bonferroni test showed that the accuracy of the learning stage ($M = 0.97$, $SD = 0.05$) was higher than the novel stage ($M = 0.83$, $SD = 0.12$), $p < .001$, $d = 1.18$. Similarly, the accuracy of the learning stage was higher than the novel-with-noise stage ($M = 0.73$, $SD = 0.15$), $p < .001$, $d = 1.96$. Accuracy for the novel stage was found to be higher than the novel-with-noise stage, $p < .001$, $d = 0.78$. Results also showed a significant main effect of stage on accuracy in the CFMT-Chinese, $F(1.81, 226.17) = 270.01$, $p < .001$, $\eta^2 = .68$. A post hoc Holm-Bonferroni test demonstrated that the accuracy of the learning stage (M = 0.98, $SD = 0.04$) was higher than the novel stage ($M = 0.71$, $SD = 0.16$), $p < .001$, $d = 1.93$. Similarly, the accuracy of the learning stage was higher than the novel-with-noise stage ($M = 0.75$, $SD = 0.17$), $p < .001$, $d = 1.61$. Interestingly, the accuracy for the novel stage was found to be lower than the novel-with-noise stage, $p < .001$, $d = -0.318$.

Figure 2

Proportion correct scores of Chinese Malaysian participants in the three stages of CFMT.



Note. Error bars represent 95% confidence intervals.

Discussion

Overall, the results showed that the CFMT-MY seems to be suitable to study individual differences in face recognition and for the diagnosis of individuals with face recognition impairments. The scores of CFMT-MY were normally distributed when all trials were included in the analysis (72 trials). Hence, the standard method used to calculate the cut-off score, *M* – 2*SD* seems to be a suitable option for the diagnosis of face recognition impairments in the CFMT-MY. Additionally, the CFMT-MY was highly consistent and exhibited high internal reliability (α = .86) which was in line with those reported in previous work on CFMT-Chinese and CFMT-Aus (α = .86, .88; McKone et al., 2011, 2017). This high reliability further supports the suitability of the test to be used for diagnosis in clinical settings and for the measurement of individual differences in face recognition.

The findings also demonstrated convergent validity where the CFMT-MY was moderately correlated with the CFMT-Chinese. This suggest that both tests tap very similar cognitive processes. Results also demonstrated divergent validity where the CFMT-MY was weakly correlated with the

CCMT which measures object recognition, despite both tests having similar procedures and formats. Additionally, the correlation between the Asian CFMT versions was larger compared to the correlation between CCMT and CFMT-MY. Hence, there is strong evidence that the CFMT-MY taps face-recognition-specific processes rather than general visual memory.

Our results showed that the difficulty of the CFMT-MY increases across stages. Specifically, the learning stage achieved the highest accuracy followed by the novel and finally the novel-with-noise stage. The CFMT-Chinese showed a similar pattern of results where the learning stage achieved higher accuracy compared to the novel and novel-with-noise stages. However, the novel stage had lower accuracy compared to the novel-with-noise stage. This finding is surprising and contradicted the intended higher level of difficulty for the novel-with-noise stage (Duchaine & Nakayama, 2006).

In summary, the analysis revealed that the CFMT-MY seems to be suitable to use for diagnosis in clinical settings and the measurement of individual differences in face recognition ability with high consistency and high internal reliability scores. The CFMT-MY also shows appropriate convergent and divergent validity. In addition, the CFMT-MY scores show an increasing level of difficulty stages which is important for the assessment of a wide range of face recognition abilities.

Experiment 2

In Experiment 2, we aim to investigate if the CFMT-MY would be sensitive to a classical effect in face recognition literature: the other-race effect. We also aim to explore if Caucasian participants would present similar levels of other-race effect for the CFMT-MY and the CFMT-Chinese. Differences in accuracy between the CFMT-MY, CFMT-Chinese and CFMT-original would be assessed using a repeated-measures ANOVA. Additionally, the CFMT-MY scores of Chinese Malaysian participants in Experiment 1 and the Caucasian participants in Experiment 2 would be compared using an independent samples t-test to determine if the CFMT-MY is sensitive to the other-race effect.

Methods

Participants

One hundred and fifty participants took part in this experiment, but the final sample included 135 Caucasians (108 females, 25 males and 2 non-binary) with ages ranging between 18 to 52 years (*M* = 22.04 years, *SD* = 6.62 years). The age range for female participants was between 18 and 52 years (*M* = 21.64 years, *SD* = 6.54 years) while for male participants, the age range was between 18 and 49 years (*M* = 23.32 years, *SD* = 6.58 years). The age range for non-binary participants was between 19 and 36 years (*M* = 27.50 years, *SD* = 12.02 years). Data from participants who had a median reaction time of less than 500ms (nine participants) or scored below chance level (24/72) (one participant) for any one of the CFMT versions were removed from further analysis. Five participants of other ethnicities (e.g., Asian, Other) other than White/Caucasian were also excluded. Ten participants were excluded from the data analysis (except for internal reliability and internal consistency analysis) as their performance on the face memory tasks was indicative of possible prosopagnosia (Appendix B). The remaining participants were 125 Caucasians (100 females, 24 males and 1 non-binary) with ages ranging between 18 and 52 years (*M* = 21.53 years, *SD* = 5.66 years). The age range for female participants was between 18 and 52 years (*M* = 21.38 years, *SD* = 6.03 years) while for male participants, the age range was between 18 and 34 years (*M* = 22.25 years, *SD* = 3.92 years). The age for the non-binary participant was 19 years.

An a priori power analysis was conducted using G*Power 3.1 (Faul et al., 2009) for a repeated-measures ANOVA comparing the three CFMT versions (CFMT-MY, CFMT-Chinese and CFMT-original). The effect size for the other-race effect was estimated from two studies which had used CFMT in measuring the other-race effect (McKone et al., 2012; Wan et al., 2017) where the average η_p^2 = .44, large effect size (effect size (η_p^2) in the papers was calculated using formula 13 in Lakens (2013)). Additionally, a meta-analysis study has reported a large effect size for the other-race effect, Hedge's *g* = .82. Therefore, a large effect size estimate (η_p^2 = .14) was entered into the power analysis with the following parameters: α = .05, power = .95. The power analysis implied that 50 participants would be required to detect a difference between the CFMT versions with 95% probability.

All participants provided informed consent to participate in the study. Course credits were given for participation. The study has been reviewed and approved by the Science and Engineering Research Ethics Committee (SEREC) at the University of Nottingham Malaysia (approval code: KSK050320).

Materials and procedure

Three versions of CFMT were used: the CFMT-original, the CFMT-Chinese and the CFMT-MY. The CFMT-original was obtained from the authors of the task (Duchaine & Nakayama, 2006). Fifty-two male identities were used in the task. The CFMT-original follows the same procedure as the CFMT-Chinese and the CFMT-MY (refer to Experiment 1 for full procedure). Testable (https://www.testable.org/) was used to run the online experiment (Rezlescu et al., 2020). The average vertical height of the face stimuli in the CFMT was 4 cm. Participants completed all three CFMT versions in random order. The experiment took about 45 minutes to complete.

Results

The analysis of internal reliability, internal consistency and validity were consistent with Experiment 1 and are available in Appendix C. A repeated-measures ANOVA was conducted to examine if there were any differences between the scores of the different test versions and the different stages of the test. A 3 (test version: CFMT-MY vs. CFMT-Chinese vs. CFMT-original) x 3 (stage: learning vs. novel vs. novel-with-noise) repeated-measures ANOVA was conducted on the accuracy (calculated by proportion correct scores). When the Mauchly's test indicated that the assumption of sphericity was violated, the degrees of freedom were corrected using the Greenhouse-Geisser method.

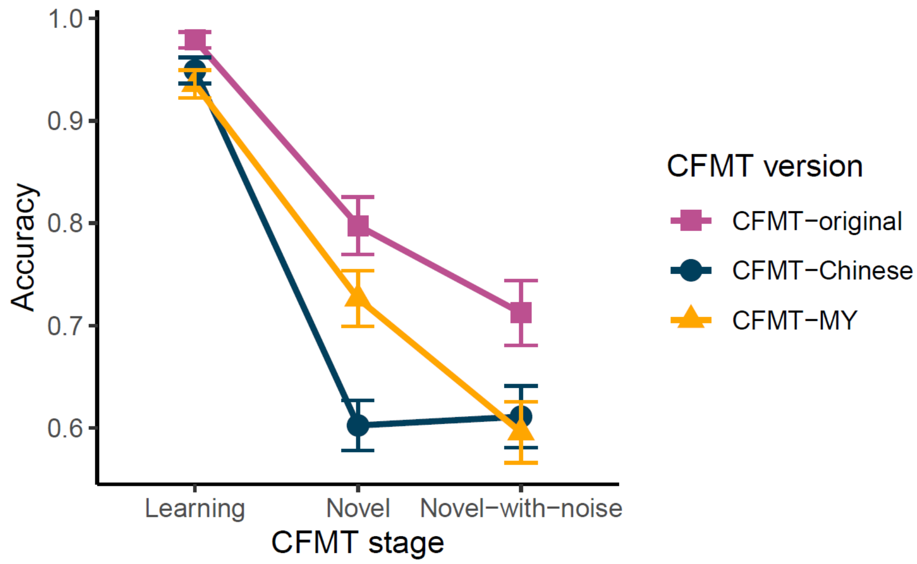
Results showed a significant main effect of test version on accuracy, $F(2, 248) = 70.49$, $p < .001$, $\eta_p^2 = .36$. A post hoc Holm-Bonferroni test demonstrated that the accuracy of CFMT-original ($M = 0.82$, $SD = 0.12$) was higher than CFMT-MY ($M = 0.74$, $SD = 0.12$), $p < .001$, $d = .73$. Similarly, the accuracy of CFMT-original was higher than CFMT-Chinese ($M = 0.69$, $SD = 0.11$), $p < .001$, $d = 1.03$. Accuracy for CFMT-MY was also found to be higher than CFMT-Chinese, $p < .001$, $d = -0.30$. To further demonstrate the other-race effect, we ran an additional analysis comparing the CFMT-MY scores of Chinese Malaysian participants in Experiment 1 and the Caucasian participants in Experiment 2. Independent samples t-test revealed that Chinese Malaysian participants ($M = 0.83$, $SD = 0.10$) scored higher than the Caucasian participants ($M = 0.74$, $SD = 0.12$) on the CFMT-MY, $t(249) = -7.20$, $p < .001$, $d = -.91$.

Analysis revealed a significant main effect of stage on accuracy, $F(1.64, 203.00) = 628.13$, $p < .001$, $\eta_p^2 = .84$. A post hoc Holm-Bonferroni test demonstrate that the accuracy of the learning stage ($M = 0.96$, $SD = 0.05$) was higher than the novel stage ($M = 0.71$, $SD = 0.12$), $p < .001$, $d = 2.35$. Similarly, the accuracy of the learning stage was higher than the novel-with-noise stage ($M = 0.64$, $SD = 0.14$), $p < .001$, $d = 3.02$. Accuracy was found to be higher for the novel compared to the novel-with-noise stage, $p < .001$, $d = .66$.

A significant interaction effect between test version and stage was found, $F(4, 496) = 38.47$, $p < .001$, $\eta_p^2 = .24$ (Figure 3). Results showed a significant main effect of test version on accuracy in the learning stage, $F(2, 248) = 17.48$, $p < .001$, $\eta^2 = .12$. A post hoc Holm-Bonferroni test demonstrated that the accuracy of CFMT-original ($M = 0.98$, $SD = 0.04$) was higher than CFMT-MY ($M = 0.94$, $SD = 0.07$) in the learning stage, $p < .001$, $d = .52$. Similarly, the accuracy of the CFMT-original was higher than the accuracy in the CFMT-Chinese ($M = 0.95$, $SD = 0.07$) in the learning stage, $p < .001$, $d = .36$. No difference was found between the accuracy for the CFMT-MY and the CFMT-Chinese in the learning stage, $p = .08$, $d = .16$. Analysis also revealed a significant main effect of test version on accuracy in the novel stage, $F(2, 248) = 96.96$, $p < .001$, $\eta^2 = .44$. A post hoc Holm-Bonferroni test demonstrate that the accuracy of the CFMT-original ($M = 0.80$, $SD = 0.16$) was higher than the CFMT-MY ($M = 0.73$, $SD = 0.15$) in the novel stage, $p < .001$, $d = .45$. Similarly, the accuracy of the CFMT-original was higher than the CFMT-Chinese ($M = 0.60$, $SD = 0.14$) in the novel stage, $p < .001$, $d = 1.23$. Accuracy for the CFMT-MY was also found to be higher than the accuracy in the CFMT-Chinese in the novel stage, $p < .001$, $d = -0.78$. A significant main effect of test version on accuracy in the novel-with-noise stage was found, $F(2, 248) = 30.57$, $p < .001$, $\eta^2 = .20$. A post hoc Holm-Bonferroni test demonstrated that the accuracy of the CFMT-original ($M = 0.71$, $SD = 0.18$) was higher than the accuracy in the CFMT-MY ($M = 0.60$, $SD = 0.17$) in the novel-with-noise stage, $p < .001$, $d = .64$. Similarly, the accuracy of the CFMT-original was higher than the accuracy in the CFMT-Chinese ($M = 0.61$, $SD = 0.17$) in the novel-with-noise stage, $p < .001$, $d = .56$. No difference was found for the accuracy in the CFMT-MY and the CFMT-Chinese in the novel-with-noise stage, $p = .35$, $d = .09$.

Figure 3

Proportion correct scores of Caucasian participants in the three stages of CFMT.



Note. Error bars represent 95% confidence intervals.

Simple main effects analysis also revealed differences on accuracy across stages in the CFMT-MY, $F(2, 248) = 330.81$, $p < .001$, $\eta^2 = .73$. A post hoc Holm-Bonferroni test demonstrated that the accuracy of the learning stage ($M = 0.94$, $SD = 0.08$) was higher than accuracy in the novel stage ($M = 0.73$, $SD = 0.15$), $p < .001$, $d = 1.4$. Similarly, the accuracy of learning stage was higher than the accuracy of the novel-with-noise stage ($M = 0.6$, $SD = 0.17$), $p < .001$, $d = 2.28$. Accuracy for the novel stage was found to be higher than the novel-with-noise stage, $p < .001$, $d = .88$. Analysis revealed a significant main effect of stage on accuracy in the CFMT-Chinese, $F(1.85, 228.85) = 480.27$, $p < .001$, $\eta^2 = .8$. A post hoc Holm-Bonferroni test demonstrate that the accuracy of the learning stage ($M = 0.95$, $SD = 0.07$) was higher than the novel stage ($M = 0.6$, $SD = 0.14$), $p < .001$, $d = 2.43$. Similarly, the accuracy of the learning stage was higher than the novel-with-noise stage ($M = 0.61$, $SD = 0.17$), $p < .001$, $d = 2.37$. No difference was found for the accuracy of the novel stage and novel-with-noise stage, $p = .5$, $d = -.06$. A significant main effect of stage on accuracy in the CFMT-original was found, $F(1.65, 205.15) = 212.68$, $p < .001$, $\eta^2 = .63$. A post hoc Holm-Bonferroni test demonstrated that the accuracy of the learning stage ($M = 0.98$, $SD = 0.04$) was higher than the novel stage ($M = 0.8$, $SD = 0.16$), $p < .001$, $d = 1.23$. Similarly, the accuracy of the learning stage was higher than the accuracy in the novel-with-noise stage (*M* = 0.71, *SD* = 0.18), $p < .001$, $d = 1.81$. Accuracy for the novel stage was found to be higher than the accuracy in the novel-with-noise stage, $p < .001$, $d = .58$.

Discussion

Our findings revealed that the CFMT-MY was sensitive to the other-race effect. Although the accuracy of the CFMT-original was higher than the CFMT-MY and CFMT-Chinese in all three stages, this could not adequately demonstrate the other-race effect since the other-race CFMT (i.e., CFMT-Chinese and CFMT-MY) may be more difficult compared to the own-race CFMT (i.e., CFMT-original). Hence, we ran an additional analysis which showed that Caucasian participants scored lower compared to the Chinese Malaysian participants from Experiment 1 on the CFMT-MY. This indicated that Chinese Malaysian participants had superior recognition of own-race faces (i.e., Chinese Malaysian faces) as compared to Caucasian participants, replicating the other-race effect in face recognition (Meissner & Brigham, 2001).

Our results also showed that the scores of CFMT-MY and CFMT-original clearly represent the increasing difficulty of the three stages, with the learning stage achieving close-to-ceiling scores, followed by the novel stage and the novel-with-noise stage with the highest difficulty. Interestingly, while the learning stage had higher accuracy compared to the novel and novel-with-noise stage in the CFMT-Chinese, accuracies for the novel and novel-with-noise stages were similar. Additionally, higher accuracy was found for the CFMT-MY compared to CFMT-Chinese in our Caucasian sample, replicating our findings from Experiment 1.

To summarize, our results show that the CFMT-MY is sensitive to the other-race effect as Caucasian participants scored lower on the CFMT-MY compared to the Chinese Malaysian participants from Experiment 1. Interestingly, with a Caucasian sample, we have shown that the difficulty of the CFMT-MY increases across stages. This result was, however, not found in the CFMT-Chinese.

General discussion

The current study aimed to develop a new version of the Asian CFMT using Chinese Malaysian faces, the CFMT-MY, as a standardized test of face recognition ability. Overall, results indicated that the CFMT-MY has high consistency and high reliability and so is suitable for the diagnosis of individuals with difficulty in face recognition in clinical settings and also for research measuring individual differences in face recognition ability. The CFMT-MY also showed convergent validity with the CFMT-Chinese and divergent validity with the CCMT. Scores for the CFMT-MY corresponded to the increasing level of difficulty intended for the CFMT stages (see Duchaine & Nakayama, 2006), where the learning stage achieved the highest accuracy followed by the novel and finally the novel-with-noise stage. The CFMT-MY scores were also normally distributed when all trials were included in the analysis (72 trials). Thus, the standard method used to calculate the cut-off score, $M - 2SD$ can be used for the diagnosis of impairments related to face recognition. The CFMT-MY was also sensitive to the other-race effect. Our results revealed that Caucasian participants scored lower on the CFMT-MY compared to the Chinese Malaysian participants from Experiment 1. Chinese Malaysian participants showed superior recognition of own-race faces compared to Caucasian participants, supporting the other-race effect in face recognition (Meissner & Brigham, 2001; Wong et al., 2020, 2021).

The results of both experiments also revealed an interesting pattern: Chinese Malaysian and Caucasian participants showed higher accuracy in the CFMT-MY compared to the CFMT-Chinese. This result seems to be explained by a surprisingly low performance in the novel stage of the CFMT-Chinese. In fact, performance on this stage was lower (Experiment 1) or identical (Experiment 2) to that of the novel-with-noise stage. This pattern of results, which was only found in the CFMT-Chinese, is problematic as it shows no linear increment of difficulty across stages. The increment of difficulty across stages is important for the assessment of a wide range of face recognition abilities (Duchaine & Nakayama, 2006). Because in our first experiment we used a sample of Chinese Malaysian participants, it could be argued that these results could be explained by the other-ethnicity effect (McKone et al., 2012). However, this hypothesis cannot explain why in Experiment 2, with a Caucasian sample, we found no differences between the novel and the novel-with-noise stages in the CFMT-Chinese, but clear differences across these stages in the CFMT-MY. More importantly, previous studies have also revealed a similar percentage of correct responses across the novel and the novel-with-noise stage in the CFMT-Chinese (i.e., 72.13% and 71.58%, see McKone et al. 2017, table 3 and 79.24% and 80.11%, see McKone et al. 2012, table 1). Past research has suggested that the CFMT could be shortened by including only the first two stages (i.e., learning and novel stage) for diagnosis of prosopagnosia (Corrow et al., 2018; Murray & Bate, 2020). In this case, including only the first two stages of the CFMT-Chinese may be problematic, as in this test the novel stage seems to be identical or even more difficult than the novel-with-noise stage, which could potentially result in more individuals scoring below the cut-off.

It is important to note here that, compared to the CFMT-Chinese and CFMT-original, we used a different method to create the novel stage. Images of the same identity that were captured with different poses and physical lighting were used for the CFMT-Chinese novel stage, but such images did not exist in our face database, and hence we followed the procedure of CFMT-Aus (McKone et al., 2011) where instead of poses, different cropping templates were used and lighting was added into the images using photo editing software. Despite these differences, the CFMT-MY showed a clear increment in difficulty across stages. In addition, these differences in the stimuli cannot explain the discrepancy in difficulty levels between the novel stage of the CFMT-Chinese and CFMT-MY as both the CFMT-original and the CFMT-Chinese use the same method in the novel stage.

In this sense, higher accuracy for the novel stage compared to the novel-with-noise stage was found in the CFMT-original while these differences were not found in the CFMT-Chinese. Thus, we conclude that the lower accuracy for CFMT-Chinese compared to CFMT-MY among Chinese Malaysian and Caucasian participants was due to differences in test difficulty, specifically in the novel stage.

Malaysia is a multiracial country with Malays constituting 57.93% of the population; followed by Chinese at 22.58%; Indians at 6.7%; indigenous people (i.e., Orang Asli) at 3.95%; others at 0.64%; and non-citizen at 8.2% (Department of Statistics Malaysia, 2011). In this case, the use of CFMT-MY in Malaysia may be limited to Chinese Malaysian participants due to the presence of the other-race effect in face recognition (Meissner & Brigham, 2001; Wong et al., 2020). However, recent research showed no differences in the recognition of Chinese faces between Chinese Malaysian and non-Chinese Malaysian (i.e., Malays and Indians) (Estudillo et al., 2020) suggesting that the CFMT-MY may also be suitable to use for diagnosis of face recognition difficulties in the non-Chinese population. Because Chinese is Malaysia's second most populous race, non-Chinese Malaysian may have developed greater expertise with Chinese faces due to extensive experience with the Chinese Malaysian population (Tanaka et al., 2013; Wan et al., 2015). However, some of the states in Malaysia have majority Malay populations (e.g., Kelantan, Terengganu and Perlis) (Saravanamuttu, 2010) and hence, the population in those states may not be as familiar with Chinese faces, hindering the use of CFMT-MY for diagnosis of face recognition difficulties in those regions.

Conclusion

In summary, we report that the CFMT-MY is a highly consistent and reliable test for diagnosing individuals with difficulty in face recognition in clinical settings, measurement of individual differences in face recognition ability and measurement of the other-race effect. The standard method to calculate the cut-off score (*M* – 2*SD*) seems to be appropriate for the diagnosis of impairments related to face recognition. Additionally, the lower end of the norm scores was far from the chance level (24/72 trials) which permits a range of scores for the diagnosis of impairments related to face recognition such as prosopagnosia. Although the psychometric properties of the CFMT-MY has been shown to be appropriate for diagnosis of face recognition impairments, future research involving Asian prosopagnosics participants would be required to further validate the use of CFMT-MY for diagnosis of prosopagnosia. The CFMT-MY scores corresponded to the increasing level of difficulty intended for the CFMT stages where the learning stage achieved the highest accuracy followed by the novel and finally the novel-with-noise stage. Finally, the current availability of two Asian CFMT versions could lead to improvement of diagnosis for face recognition difficulties and is beneficial for use in pre-post face recognition ability assessments.

Declarations

Funding

This study was funded by the Fundamental Research Grant Scheme (FRGS) from the Ministry of Education (MOE) Malaysia (Grant number: FRGS/1/2018/SS05/UNIM/02/4).

Conflicts of interest/Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval

Approval was obtained from the Science and Engineering Research Ethics Committee (SEREC) at the University of Nottingham Malaysia (approval code: KSK050320). The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated during the current study and the CFMT-MY materials (stimuli and trial order) are available in the Open Science Framework repository, https://osf.io/gu4fy/?view_only=1ee5963c97664354a9d321057d22b641.

Code availability

Not applicable.

Authors' contributions

The authors confirm contribution to the paper as follows: study conception and design: Siew Kei Kho, Bryan Qi Zheng Leong, and Alejandro J. Estudillo; data collection: Siew Kei Kho and Alejandro J. Estudillo; analysis and interpretation of results: Siew Kei Kho and Alejandro J. Estudillo; draft manuscript preparation: Siew Kei Kho, David R. T. Keeble, Hoo Keat Wong, and Alejandro J. Estudillo. All authors reviewed the results and approved the final version of the manuscript.

Open Practices Statements

The data for all experiments and the CFMT-MY materials (stimuli and trial order) are available at https://osf.io/gu4fy/?view_only=1ee5963c97664354a9d321057d22b641. None of the experiments was preregistered.

References

Arrington, M., Elbich, D., Dai, J., Duchaine, B., & Scherf, K. S. (2022). Introducing the female Cambridge face memory test – long form (F-CFMT+). *Behavior Research Methods*, *17*(10), 841. https://doi.org/10.3758/s13428-022-01805-8

Bate, S., Adams, A., & Bennetts, R. J. (2019). Guess Who? Facial Identity Discrimination Training Improves Face Memory in Typically Developing Children. *Journal of Experimental Psychology: General*, 1–47. https://doi.org/10.1037/xge0000689

Bate, S., Bennetts, R., Hasshim, N., Portch, E., Murray, E., Burns, E., & Dudfield, G. (2019). The limits of super recognition: An other-ethnicity effect in individuals with extraordinary face recognition skills. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(3), 363–377. https://doi.org/10.1037/xhp0000607

Bate, S., & Dudfield, G. (2019). Subjective assessment for super recognition: An evaluation of self-report methods in civilian and police participants. *PeerJ*, *2019*(1), 1–17. https://doi.org/10.7717/peerj.6330

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

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., Rivolta, D., Wilson, C. E., & 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*(1), 105–116. https://doi.org/10.1111/j.2044-8295.1982.tb01795.x

Bruce, V., Bindemann, M., & Lander, K. (2018). Individual differences in face perception and person recognition. *Cognitive Research: Principles and Implications*, *3*(1), 10–12. https://doi.org/10.1186/s41235-018-0109-4

Childs, M. J., Jones, A., Thwaites, P., Zdravković, S., Thorley, C., Suzuki, A., Shen, R., Ding, Q., Burns, E., Xu, H., & Tree, J. J. (2021). Do individual differences in face recognition ability moderate the other ethnicity effect? *Journal of Experimental Psychology: Human Perception and Performance*, *47*(7), 893–907. https://doi.org/10.1037/xhp0000762

Corrow, S. L., Albonico, A., & Barton, J. J. S. (2018). Diagnosing Prosopagnosia: The Utility of Visual Noise in the Cambridge Face Recognition Test. *Perception*, *47*(3), 330–343. https://doi.org/10.1177/0301006617750045

Corrow, S. L., Davies-Thompson, J., Fletcher, K., Hills, C., Corrow, J. C., & Barton, J. J. S. (2019). Training face perception in developmental prosopagnosia through perceptual learning. *Neuropsychologia*, *134*, 107196. https://doi.org/10.1016/j.neuropsychologia.2019.107196

Crawford, J. R., Garthwaite, P. H., & Slick, D. J. (2009). On percentile norms in neuropsychology: Proposed reporting standards and methods for quantifying the uncertainty over the percentile ranks of test scores. *Clinical Neuropsychologist*, *23*(7), 1173–1195. https://doi.org/10.1080/13854040902795018

Dalrymple, K. A., Fletcher, K., Corrow, S., 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

Davies-Thompson, J., Fletcher, K., Hills, C., Pancaroglu, R., Corrow, S. L., & Barton, J. J. S. (2017). Perceptual Learning of Faces: A Rehabilitative Study of Acquired Prosopagnosia. *Journal of Cognitive Neuroscience*, *29*(3), 573–591. https://doi.org/10.1162/jocn_a_01063

DeGutis, J., Wilmer, J., Mercado, R. J., & Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition*, *126*(1), 87–100. https://doi.org/10.1016/j.cognition.2012.09.004

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, *44*(2), 587–605. https://doi.org/10.3758/s13428-011-0160-2

Department of Statistics Malaysia. (2011). *Population Distribution and Basic Demographic Characteristic Report 2010 (Updated: 05/08/2011)*. https://www.dosm.gov.my/v1/index.php?r=column/ctheme&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09&bul_id=MDMxdHZjWTk1SjFzTzNkRXYzcVZjdz09

Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE*, *10*(4), 1–12. https://doi.org/10.1371/journal.pone.0121945

Dowle, M., & Srinivasan, A. (2021). *data.table: Extension of `data.frame`*. https://cran.r-project.org/package=data.table

Duchaine, B. (2000). Developmental prosopagnosia with normal configural processing. *NeuroReport*, *11*(1), 79–83. https://doi.org/10.1097/00001756-200001170-00016

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

Estudillo, A. J. (2021). Self-reported face recognition abilities for own and other-race faces. *Journal of Criminal Psychology*, *11*(2), 105–115. https://doi.org/10.1108/JCP-06-2020-0025

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. (2020). No evidence of other-race effect for Chinese faces in Malaysian non-Chinese population. *Applied Cognitive Psychology*, *34*(1), 270–276. https://doi.org/10.1002/acp.3609

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*, 1–12. https://doi.org/10.7717/peerj.10629

Falissard, B. (2012). *psy: Various procedures used in psychometry*. https://cran.r-project.org/package=psy

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Gamaldo, A. A., & Allaire, J. C. (2016). Daily Fluctuations in Everyday Cognition: Is It Meaningful? *Journal of Aging and Health*, *28*(5), 834–849. https://doi.org/10.1177/0898264315611669

George, D., & Mallery, P. (2019). IBM SPSS Statistics 26 Step by Step. In *IBM SPSS Statistics 26 Step by Step*. Routledge. https://doi.org/10.4324/9780429056765

Germine, L. T., Duchaine, B., & Nakayama, K. (2011). Where cognitive development and aging meet: Face learning ability peaks after age 30. *Cognition*, *118*(2), 201–210. https://doi.org/10.1016/j.cognition.2010.11.002

Hittner, J. B., May, K., & Silver, N. C. (2003). A monte carlo evaluation of tests for comparing dependent correlations. *Journal of General Psychology*, *130*(2), 149–168. https://doi.org/10.1080/00221300309601282

JASP Team. (2022). *JASP (Version 0.16.3)[Computer software]*. https://jasp-stats.org/

Kennerknecht, I., Grueter, T., Welling, B., Wentzek, S., Horst, J., Edwards, S., & Grueter, M. (2006). First report of prevalence of non-syndromic hereditary prosopagnosia (HPA). *American Journal of Medical Genetics Part A*, *140A*(15), 1617–1622. https://doi.org/10.1002/ajmg.a.31343

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

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*(863). https://doi.org/10.3389/fpsyg.2013.00863

McCaffrey, R. J., & Westervelt, H. J. (1995). Issues associated with repeated neuropsychological assessments. *Neuropsychology Review*, *5*(3), 203–221. https://doi.org/10.1007/BF02214762

McKone, E., Hall, A., Pidcock, M., Palermo, R., Wilkinson, R. B., Rivolta, D., Yovel, G., Davis, J. M., & O'Connor, K. B. (2011). Face ethnicity and measurement reliability affect face recognition performance in developmental prosopagnosia: Evidence from the Cambridge Face Memory Test–Australian. *Cognitive Neuropsychology*, *28*(2), 109–146. https://doi.org/10.1080/02643294.2011.616880

McKone, E., Stokes, S., Liu, J., Cohan, S., Fiorentini, C., Pidcock, M., Yovel, G., Broughton, M., & Pelleg, M. (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. *Cognitive Neuropsychology*, *34*(5), 253–268. https://doi.org/10.1080/02643294.2017.1371682

Meissner, C. A., & Brigham, J. C. (2001). Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review. *Psychology, Public Policy, and Law*, *7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3

Murray, E., & Bate, S. (2020). Diagnosing developmental prosopagnosia: repeat assessment using the Cambridge Face Memory Test. *Royal Society Open Science*, *7*(9), 200884. https://doi.org/10.1098/rsos.200884

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*, *70*(2), 218–233. https://doi.org/10.1080/17470218.2016.1161058

R Core Team. (2021). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. https://www.r-project.org/

Rezlescu, C., Danaila, I., Miron, A., & Amariei, C. (2020). More time for science: Using Testable to create and share behavioral experiments faster, recruit better participants, and engage students in hands-on research. *Prog. Brain Res*, *253*, 243–262.

Rossion, B. (2014). Understanding face perception by means of prosopagnosia and neuroimaging. *Frontiers in Bioscience*, *6*(2), 706. https://doi.org/10.2741/e706

RStudio Team. (2021). *RStudio: Integrated Development Environment for R*. http://www.rstudio.com/

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

Saravanamuttu, J. (2010). Malaysia: Multicultural society, Islamic state, or what? *State and Secularism: Perspectives from Asia*, 279–300. https://doi.org/10.1142/9789814282383_0016

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

Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), 12887–12892. https://doi.org/10.1073/pnas.1421881112

Tanaka, J. W., Heptonstall, B., & Hagen, S. (2013). Perceptual expertise and the plasticity of other-race face recognition. *Visual Cognition*, *21*(9–10), 1183–1201. https://doi.org/10.1080/13506285.2013.826315

Wan, L., Crookes, K., Dawel, A., Pidcock, M., Hall, A., & McKone, E. (2017). Face-blind for other-race faces: Individual differences in other-race recognition impairments. *Journal of Experimental Psychology: General*, *146*(1), 102–122. https://doi.org/10.1037/xge0000249

Wan, L., Crookes, K., Reynolds, K. J., Irons, J. L., & McKone, E. (2015). A cultural setting where the other-race effect on face recognition has no social-motivational component and derives entirely from lifetime perceptual experience. *Cognition*, *144*(0010), 91–115. https://doi.org/10.1016/j.cognition.2015.07.011

Wang, R., Li, J., Fang, H., Tian, M., & Liu, J. (2012). Individual differences in holistic processing predict face recognition ability. *Psychological Science*, *23*(2), 169–177. https://doi.org/10.1177/0956797611420575

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12). http://www.jstatsoft.org/v21/i12/paper

Wickham, H. (2021). *tidyr: Tidy Messy Data*. https://cran.r-project.org/package=tidyr

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A Grammar of Data Manipulation*. https://cran.r-project.org/package=dplyr

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

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., Nakayama, K., & Duchaine, B. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(11), 5238–5241. https://doi.org/10.1073/pnas.0913053107

Wong, H. K., Estudillo, A. J., Stephen, I. D., & Keeble, D. R. T. (2021). The other-race effect and holistic processing across racial groups. *Scientific Reports*, *11*(1), 1–15. https://doi.org/10.1038/s41598-021-87933-1

Wong, H. K., Stephen, I. D., & Keeble, D. R. T. (2020). The Own-Race Bias for Face Recognition in a Multiracial Society. *Frontiers in Psychology*, *11*, 208. https://doi.org/10.3389/fpsyg.2020.00208

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

Appendix A

Exclusion of possible prosopagnosia cases for Experiment 1

Possible prosopagnosia cases were excluded to provide calculation representing “norm” participants in order to be able to use the test for diagnosing prosopagnosia cases (see Bowles et al., 2009 and McKone et al., 2017 for a similar procedure). Percentile ranks (Crawford et al., 2009) were calculated to determine the bottom 2% of the sample using the formula $(m + 0.5k) / N \times 100$ where m is the number of participants scoring below a given score, k is the number of participants which have obtained the given score and N is the total sample size. Using this formula, CFMT-MY score of 39/72 was equivalent to a percentile rank of 1.87% of the total sample size ($N = 134$) and the score after, 42/72 was equivalent to 2.61%. Three participants (participant ID: 71, 102 and 75) which scored ≤ 39 were excluded. Based on the scores in Table A1, these participants scored quite well in the learning stage of CFMT-Chinese and CFMT-MY, showing that the low scores were not attributable to lack of effort. Similarly, the scores for CFMT-Chinese were at the lower end of the normal distribution. Raw data file showed no indication of repeated same key pressing.

The standard method was used to calculate the cut-off value of the CFMT-Chinese for prosopagnosia, $M - 2SD$. The cut-off score was 36.46/72. Five participants (participant ID: 78, 105, 123, 83 and 45) which scored ≤ 36 were excluded. Based on the scores in Table 1A, these participants scored quite well in the learning stage of CFMT-Chinese and CFMT-MY, showing that the low scores were not attributable to lack of effort. However, they unexpectedly scored in the average to high range for CFMT-MY, except for participant 105. Raw data file showed no indication of repeated same key pressing, however, participant 123 had 16/72 trials (22.22%) with response time < 500ms and four trials with abnormally long response time (12194-168585ms) in the CFMT-Chinese block showing that the participation may be distracted during the task. Additionally, all five participants completed the CFMT-Chinese as the last block as per randomization, hence, the low performance could be attributed to fatigue or loss of attention/effort towards the end of the experiment.

It is unclear if these cases presented are prosopagnosia as some of the participants may have scored on the lower end due to fatigue or loss of attention towards the end of the experiment. Other measures such as the 20-item prosopagnosia index (PI20) (Shah et al., 2015), famous face test and a clinical interview are needed to confirm these cases.

Table A1

Possible prosopagnosia cases based on CFMT-MY and CFMT-Chinese scores.

	CFMT-MY		CFMT-Chinese		CCMT		
Participant ID	All trials (/72)	Learning stage (/18)	All trials (/72)	Learning stage (/18)	All trials (/72)	Learning stage (/18)	Order of tasks
71	39	18	37	14	49	14	CCMT > CFMT-MY > CFMT-Chinese
102	31	11	40	17	37	8	CCMT > CFMT-Chinese > CFMT-MY
75	36	15	43	18	41	17	CCMT > CFMT-MY > CFMT-Chinese
78	61	18	30	18	47	13	CCMT > CFMT-MY > CFMT-Chinese
105	44	15	33	17	44	15	CFMT-MY > CCMT > CFMT-Chinese
123	53	17	33	15	32	10	CFMT-MY > CCMT > CFMT-Chinese

83	53	17	34	14	46	17	CCMT > CFMT-MY > CFMT-Chinese
45	56	18	35	17	51	18	CCMT > CFMT-MY > CFMT-Chinese

Note. Participant 71, 102 and 75 scored below percentile rank of 2% for CFMT-MY and participant 78, 105, 123, 83 and 45 scores below cut-off score (*M* - 2*SD*).

Appendix B

Exclusion of possible prosopagnosia cases for Experiment 2

Possible prosopagnosia cases were excluded to provide calculation representing "norm" participants (see Bowles et al., 2009 and McKone et al., 2017 for a similar procedure). The standard method, *M* – 2*SD,* was used to calculate the cut-off value of the CFMT-original for prosopagnosia. The cut-off score was 37.79/72. Four participants (ID: 49, 87, 43 and 96) which scored ≤ 38 were excluded. Based on the scores in Table B1, these participants scored quite well in the learning stage of all test versions of CFMT, showing that the low scores were not attributable to lack of effort (except for participant 43 in the CFMT-MY learning stage). Similarly, the scores for CFMT-Chinese and CFMT-MY were at the lower end of the normal distribution. Raw data file showed no indication of repeated same key pressing, however, participant 43 had 32/72 trials (44.44%) for the CFMT-Chinese block, 17/72 trials (23.61%) for the CFMT-MY block and 34/72 trials (47.22%) for the CFMT-original block with response time < 500ms indicating that the participation may be pressing some of the keys randomly during the task.

The standard method to calculate the cut-off value, *M* – 2*SD,* was used to determine participants which had average score ranked in the bottom 2% of the sample for CFMT-Chinese (for similar procedure, see Wan et al., 2017). The cut-off value was 31.49/72. Four participants (ID: 129, 115, 37 and 9) which scored ≤ 31 were excluded. Based on the scores in Table B1, these participants scored quite well in the learning stage of all test versions of CFMT, showing that the low scores were not attributable to lack of effort. Similarly, the scores for CFMT-MY were at the lower end of the normal distribution (except for participant 9). Raw data file showed no indication of repeated same key pressing. It is unclear if the participants were possible prosopagnosia cases or if they were severely affected by the other-race effect (ORE). For example, participant 37 scored on the lower end for both CFMT-Chinese and CFMT-MY (other-race) but scored on the average range for CFMT-original (own-race).

Cut-off value was also calculated using the standard method, *M* – 2*SD*, to determine participants which had average score ranked in the bottom 2% of the sample for CFMT-MY. The cut-off value was 32.8/72. Two participants (ID: 71 and 104) which scored ≤ 33 were excluded. Based on the scores in Table 7, these participants scored quite well in the learning stage of all test versions of CFMT, showing that the low scores were not attributable to lack of effort. Similarly, the scores for CFMT-Chinese were at the lower end of the normal distribution. Raw data file showed no indication of repeated same key pressing.

All 10 participants were excluded from the data analysis (except for internal reliability analysis). As in Experiment 1, it is unclear if the participants scoring below the cut-off value on the CFMT-original were indicative of possible prosopagnosia as further diagnosis using other measures are needed to confirm these cases. It is also unclear if the participants which had average score ranked in the bottom 2% of the sample for CFMT-MY and CFMT-Chinese were possible prosopagnosia cases, or if they were severely affected by the other-race effect with average face recognition ability for own-race faces (Wan et al., 2017).

Table B1

Possible prosopagnosia cases based on CFMT-original*, CFMT-Chinese and CFMT-MY scores.*

	CFMT-MY		CFMT-Chinese		CFMT-original		
Participant ID	All trials (/72)	Learning stage (/18)	All trials (/72)	Learning stage (/18)	All trials (/72)	Learning stage (/18)	Order of tasks
49	35	15	41	18	34	17	CFMT-Chinese > CFMT-original > CFMT-MY
87	39	15	39	18	34	16	CFMT-MY > CFMT-Chinese > CFMT-original
43	26	6	37	18	36	11	CFMT-Chinese > CFMT-original > CFMT-MY
96	36	17	37	16	38	18	CFMT-Chinese > CFMT-original > CFMT-MY
129	38	15	30	18	39	16	CFMT-MY > CFMT-Chinese > CFMT-original
115	34	12	31	14	43	15	CFMT-original > CFMT-Chinese > CFMT-MY
37	39	13	31	15	56	18	CFMT-MY > CFMT-original > CFMT-Chinese
9	46	17	31	12	41	16	CFMT-Chinese > CFMT-original > CFMT-MY
71	30	14	40	13	48	15	CFMT-Chinese > CFMT-MY > CFMT-original
104	31	17	44	16	45	18	CFMT-original > CFMT-Chinese > CFMT-MY

Note. Participants scored below percentile rank of 2% for CFMT-original (49, 87, 43 and 96), CFMT-Chinese (129, 115, 37 and 9) and CFMT-MY

Appendix C

Internal reliability

The internal reliability of the test was measured using Cronbach's α. For all trials, internal reliability was α = .87 for CFMT-MY. Results showed high internal reliability for CFMT-MY which was in line with previous work such as CFMT-Chinese, α = .86 (McKone et al., 2017).

Internal consistency

The internal consistency of the CFMT-MY at stage level (i.e., learning, novel and novel-with-noise) was measured using Pearson correlation (r). Results showed positive correlation between the learning and novel stage, $r(133) = .45$, $p < .001$, learning and novel-with-noise stage, $r(133) = .34$, $p < .001$ and novel and novel-with-noise stage, $r(133) = .67$, $p < .001$ showing that the scores were highly consistent across the different stages of CFMT-MY.

Validity

Convergent validity was measured using Pearson correlation (r). Convergent validity was measured by examining the correlation between the CFMT-MY and CFMT-Chinese and between the CFMT-MY and CFMT-original. Results showed positive correlation between the scores of CFMT-MY and CFMT-Chinese, $r(123) = .57$, $p < .001$ and of CFMT-MY and CFMT-original, $r(123) = .52$, $p < .001$. The difference between the two correlation was further analyzed by comparing the dependent overlapping correlations (Diedenhofen & Musch, 2015; Hittner et al., 2003). The test showed that the correlation between CFMT-MY and CFMT-Chinese was no different compared to the correlation between CFMT-MY and CFMT-original ($z = 0.7$, $p = .49$).