Chemical Senses, 2019, Vol 44, 189–195
doi:10.1093/chemse/bjz006
Original Article
Advance Access publication 30 January 2019

Original Article

Development of the Chinese Smell Identification Test

Guo Feng[1,2,4,*], Yuan Zhuang[2,3,*], Fangshu Yao[2,3], Yuting Ye[2,3], Qiannong Wan[2,3] and Wen Zhou[2,3]

[1]CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang District, Beijing 100101, China, [2]Department of Psychology, University of Chinese Academy of Sciences, No.19(A), Yuquan Road, Shijingshan District, Beijing 100049, China and [3]State Key Laboratory of Brain and Cognitive Science, CAS Center for Excellence in Brain Science and Intelligence Technology, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang District, Beijing 100101, China

[4]Present address: Psychological Research and Counseling Center, Southwest Jiaotong University, Chengdu, Sichuan 611756, China.

*These authors contributed equally to the work.

Correspondence to be sent to: Wen Zhou, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang District, Beijing 100101, China. e-mail: zhouw@psych.ac.cn

Editorial Decision 15 January 2019.

Abstract

Smell identification ability reflects the functional integrity of the human olfactory system. Its deficit is a prodromal marker for Parkinson's disease and is also implicated in Alzheimer's disease and other neurological and psychiatric disorders. Considering the impact of cultural factors on odor identifiability, we have developed a smell identification test specifically for the Chinese population (CSIT), which includes 40 odor items that are familiar to this population, presented in a multiple-choice format. The CSIT has a test–retest reliability of 0.92 and is validated against the University of Pennsylvania Smell Identification Test (UPSIT) and the Sniffin' Sticks Identification Test 16 (SS-16). In terms of identification accuracy, Chinese participants on average score 15% higher on the CSIT than on the UPSIT or SS-16. The CSIT is also sensitive to age and gender differences in smell identification ability. As such, the CSIT provides an effective tool for the assessment of olfactory function in the Chinese population.

Key words: Chinese population, reliability, sensitivity, smell identification, validity

Introduction

The olfactory system enables us to detect food and hazards. Olfactory loss has substantial adverse effects on the quality of life (Blomqvist et al. 2004; Croy et al. 2014) and predicts mortality risk independent of dementia conversion (Ekström et al. 2017). It is also a prodromal marker for Parkinson's disease and is implicated in a range of other neurological and psychiatric disorders including Alzheimer's disease (Mesholam et al. 1998; Woodward et al. 2017), schizophrenia (Moberg et al. 1999), and major depression (Pause et al. 2001). However, individuals with olfactory loss are generally unaware of it until formal testing (Nordin et al. 1995), highlighting the necessity of assessment of olfactory function.

Tests of olfactory threshold, discrimination, and identification, while respectively tapping into low-, intermediate-, and high-order olfactory processing, pose different cognitive demands (Hedner et al. 2010). In particular, odor identification draws on an individual's knowledge of specific odors and is significantly influenced by semantic and cultural factors (Schab 1991; Ayabe-Kanamura et al. 1998; Kobayashi et al. 2006). On the other hand, assessment of olfactory identification is more sensitive than that of olfactory threshold or discrimination to Parkinson's and Alzheimer's diseases (Koss et al. 1988; Potagas et al. 1998; Boesveldt et al. 2008; Hummel et al. 2017) and is hence more widely used in clinical settings. Based on these considerations, several culturally specific tests of olfactory identification have been developed since the University of Pennsylvania Smell Identification Test (UPSIT, USA) (Doty et al. 1984b), including the Sniffin' Sticks Identification Test 16 (SS-16, Germany) (Hummel et al. 1997) and its extended version (Haehner et al. 2009), the Scandinavian Odor Identification Test (Sweden) (Nordin et al. 1998), and the Odor Stick Identification Test for the Japanese (Japan) (Kobayashi 2005). They are all presented in a multiple-choice format, as inaccessibility of odor names has been shown to severely impede identification (Cain 1979). Participants smell an odor and make a forced choice of its name from four descriptors. The UPSIT consists of 40 odor items and has the best test–retest reliability (r = 0.92) (Doty et al. 1984b). The other tests each comprises 32 (extended version of the Sniffin' Sticks Identification Test [Haehner et al. 2009]), 16 (the SS-16 and the Scandinavian Odor Identification Test [Nordin et al. 1998]), or 13 (the Odor Stick Identification Test for the Japanese [Saito et al. 2006]) odor items. Attempts have also been made to develop a culturally universal smell identification test. The 12-item Cross-Cultural Smell Identification Test (Doty et al. 1996), for example, selects odor items from the UPSIT that are comparatively familiar to most people from North American, European, South American, and Asian cultures. However, its suitability for Taiwanese people has been questioned (Jiang et al. 2002).

In mainland China, olfactory function is scantly tested in clinical practice, partially due to the lack of a culturally appropriate smell identification test. Despite a few reports suggesting that the SS-16 could be directly applied to Chinese patients (Yang et al. 2010; Yuan et al. 2010; Yang et al. 2013; Huang et al. 2016; Chen et al. 2018), it is apparent that some odors in this test are foreign to most Chinese (e.g., sauerkraut, raspberry, rum, etc). Other studies have sought to adapt the UPSIT or SS-16 for the Chinese population by removing/replacing some of the odor items or response alternatives (Shu and Yuan 2008; Chen et al. 2012; Jiang et al. 2014; Jiang and Liang 2016), yet the choices of which to remove and what to use as substitutions are often idiosyncratic and inconsistent. Critically, such slight modifications cannot change a smell identification test developed for the American (UPSIT) or German population (SS-16) into an optimized test for the Chinese population. To address this issue, we have developed the Chinese Smell Identification Test (CSIT)—a 40-item smell identification test specifically for the Chinese population. We report here the development, reliability, and validity of this test. We also show that it is sensitive to age and gender differences in smell identification (Doty et al. 1984a).

Materials and methods

Participants

A total of 736 nonsmokers with birthplaces across mainland China took part in the main study, including 296 (135 males [m], 135 females [f], 26 unreported; mean age ± standard deviation [SD] = 22.3 ± 3.6 years) in Experiment 1, 46 (25 m, 24.1 ± 2.1 years) in Experiment 2, 89 (53 m, 41.2 ± 25.2 years) in Experiment 3, 66 (36 m, 34.0 ± 16.2 years) in Experiment 4, 119 (55 m, 20.2 ± 1.9 years) in Experiment 5, and 120 in Experiment 6 (60 m, 34.4 ± 17.4 years). They reported to have no respiratory allergy or upper respiratory infection at the time of testing. Written informed consent and consent to publish was obtained from participants in accordance with ethical standards of the Declaration of Helsinki (1964). The study was approved by the Institutional Review Board at the Institute of Psychology, Chinese Academy of Sciences.

Olfactory stimuli

Odorants pertaining to the development of the CSIT were obtained from domestic flavor and fragrance companies (the majority were obtained from the Apple Flavor and Fragrance Group) and were used in Experiments 2–6. In Experiment 2, they were presented in 4-mL glass vials, each containing 1 mL liquid. In Experiments 4–6, they were presented in felt-tip pens (Hummel et al. 1997), each filled with 1 mL liquid. Experiment 3 involved both methods of presentation (see Procedure). For odor presentation, the lid of the vial or the cap of the pen was removed and the vial mouth or the pen tip was placed approximately 2 cm in front of the two nostrils for 2–3 s. The odorants were judged as clearly detectable and roughly matched in intensity in a pilot testing. Aside from these odorants, Experiment 4 also employed the UPSIT (simplified Chinese version, Sensonics International) and the SS-16 (Burghart Medical Technology). Experiment 5 also employed the SS-16.

Procedure

The goal of Experiments 1 and 2 was to find odor items that are familiar and identifiable to most Chinese people. In Experiment 1, we first generated a list containing the names of 105 odorous objects that are encountered in daily life and asked participants to rate the familiarity and identifiability of the smell of each item based on their experience on a 7-point Likert scale, with 7 signifying very familiar or very identifiable. The order of the items was pseudorandomized across participants. Based on the participants' ratings, we ranked the 105 items by their combined familiarity and identifiability (i.e., mean familiarity rating + mean identifiability rating) and obtained the highest ranked 45 odorants that were commercially available to use in Experiment 2.

Participants in Experiment 2 were presented with these 45 odorants, one at a time, and were asked to make a forced choice of each odorant's name from a list of four descriptors. They also provided ratings for the familiarity, intensity, pleasantness, and irritability of each odorant on a 7-point Likert scale, with 7 representing very familiar, very strong, very pleasant, and very irritating, respectively. For each odorant, the four response alternatives (one being the correct odor name) were chosen, mainly from the aforementioned list of 105 odorous items, to be roughly matched in familiarity and identifiability (Supplementary Table S1). We refrained from using only the 45 odor names as response alternatives so as to eliminate repetitions of response alternatives. Following (Doty et al. 1984b), we selected response alternatives to be as distinct from one another as possible in order to better differentiate between hyposmic and anosmic individuals (Gudziol and Hummel 2009). Five odor items that were misidentified by >30% participants were subsequently excluded, leaving 40 odor items in the CSIT.

Experiment 3 assessed the reliability of the 40-item CSIT. To quantify its test–retest reliability, participants were tested twice, with an interval of 3 or 6 months in between. Specifically, 28 participants performed the retest 3 months after the initial test, with the odorants presented in vials at the initial test and in felt-tip pens at the retest (this allowed us to assess whether the CSIT scores were also robust against a change in test format). Another 61 participants performed the retest 6 months after the initial test, with the odorants presented in felt-tip pens at both the initial test and the retest. These participants spanned the entire continuum of CSIT scores.

Experiment 4 compared participants' accuracies on the CSIT with those on two widely used smell identification tests—the UPSIT (simplified Chinese version) and the SS-16 (response alternatives translated to simplified Chinese by the authors). Each participant was tested on two consecutive days, with the CSIT on one day and with the UPSIT and SS-16 on the other day, in counterbalanced order.

Unlike the UPSIT (Doty et al. 1984b) or the CSIT, the distractors listed for each odor item in the SS-16 are typically similar (Gudziol and Hummel 2009). Experiment 5 quantified the similarities between distractors and targets in the CSIT and the SS-16, and analyzed whether the degree of similarity could have caused the difference between the identification accuracies for the two tests in Experiment 4. For each of the CSIT or the SS-16 odor items, participants were presented with the odorant, together with the corresponding label (target), and were asked to rate on a 7-point Likert scale the similarity of each of the three distractors (presented in random order) to that odor, with 7 representing very similar. All 119 participants provided ratings for the CSIT, whereas 50 (18 m, 21.2 ± 2.1 years) of them also provided ratings for the SS-16. For those that rated both the CSIT and the SS-16, the order of the two tests was balanced across participants.

Experiment 6 tested whether the CSIT was sensitive to the known age and gender differences in smell identification (Doty et al. 1984a). Four groups of healthy nonsmokers (30 per group), namely young females (17.2 ± 1.6 years), young males (16.9 ± 1.4 years), middle-aged females (51.7 ± 1.4 years), and middle-aged males (51.7 ± 1.3 years), who were matched in education (years of education = 12.9 ± 2.5, 12.0 ± 2.4, 12.0 ± 2.4, and 12.9 ± 2.5, respectively), were administered with the CSIT. Their scores were subsequently analyzed with univariate ANOVA, using age group and gender as fixed factors.

In Experiments 2–6, participants were free to resample the odorants. There was an interval of at least 30 s in between odor presentations to reduce olfactory fatigue. For young children and senior adults who had difficulty reading the response alternatives, the experimenter read those aloud immediately after the presentation of each odorant.

Results

Selection of odor items

As an initial step to develop a culturally appropriate smell identification test for the Chinese population, in Experiment 1, we conducted a survey to find odor items that are familiar and identifiable to most Chinese. The survey comprised a list of the names of 105 odorous objects encountered in daily life. Participants were asked to rate the familiarity and identifiability of the smell of each item based on their own experience on a 7-point Likert scale, with 7 denoting very familiar or very identifiable. A total of 296 nonsmokers from various parts of China (north: 95, northeast: 17, east: 70, southeast: 27, south-central: 46, northwest: 21, unreported: 20) completed the survey. The top 3 items rated as most familiar as well as identifiable were garlic (familiarity: 6.0, identifiability: 6.0), vinegar (6.0 and 6.0), and Florida water (pinyin: *hua lu shui*, used widely in China as a mosquito repellent and to ease skin irritation; 5.9 and 5.8). Overall, an item's identifiability strongly correlated with its familiarity ($r_{105} = 0.95$, $P < 0.0001$). We subsequently ranked the items by their combined familiarity and identifiability averaged across participants and selected the first 45 whose odors were commercially available to be included in Experiment 2. Some high-ranked items, like mosquito coil, for which we were unable to obtain appropriate odor substances, were not included. All the selected 45 items scored above 4 on the familiarity and the identifiability scales.

Experiment 2 assessed whether odorants corresponding to the selected items could be correctly identified by young adult nonsmokers aged between 20 and 29 years, who as a group has been shown to exhibit superior olfactory performance (Hubert et al. 1980; Hummel et al. 2007). Like most existing smell identification tests, the test adopted a multiple-choice format. For each odorant, participants were asked to make a forced choice from four pregenerated descriptors (Supplementary Table S1) the one that best matched with the perceived smell. Analyses of the participants' responses showed that the odorants differed significantly in their ability to be identified ($F_{44, 1980} = 19.61$, $P < 0.0001$). In particular, 5 of them were misidentified by over 30% of the participants. These 5 odorants, namely black pepper, alcohol, pine tree, licorice, and pan-roasted chestnut, were since removed, leaving 40 odor items (Table 1) in the CSIT. Overall, the 40 odorants were perceived as quite familiar (mean ± SD = 5.4 ± 0.7 on a 7-point Likert scale with 7 signifying very familiar), moderately strong (5.3 ± 0.3 on a 7-point Likert scale with 7 signifying very strong), slightly pleasant (4.6 ± 0.7 on a 7-point Likert scale with 7 signifying very pleasant), and minimally irritating (1.7 ± 0.6 on a 7-point Likert scale with 7 signifying very irritating). Supplementary Table S2 summarizes the ratings and identification accuracies for the 40 individual odorants.

CSIT's reliability and validity

The usefulness of a psychometric test depends on its reliability and validity. To determine whether the CSIT consistently measures one's smell identification ability over time, in Experiment 3, we quantified its test–retest reliability in a group of individuals with various degrees of olfactory function. Specifically, 89 participants aged between 7 and 92 (53 m) were administered with the CSIT twice, with an extended interval of 3 months (28 participants) or 6 months (61 participants) in between. Their CSIT scores at the initial test ranged from 11 to 40 (accuracy = 27.5–100% vs. chance = 25%). The retest scores (range: 12–40) strongly correlated with those obtained at the initial test, $r_{89} = 0.92$, $P < 0.0001$, and did not differ from the latter ($t_{88} = -1.60$, $P = 0.11$). In other words, the CSIT scores were highly stable (Figure 1A). Moreover, whether or not to exclude the participants tested with a 3-month interval had no influence on the correlation strength between the test and retest scores. The 6-month test–retest-reliability coefficient was also 0.92 ($r_{61} = 0.92$, $P < 0.0001$). We also calculated, based on the participants' initial CSIT scores, the test's split-half reliability. Overall, scores on the odd items strongly correlated with those on the even items ($r_{89} = 0.89$, $P < 0.0001$) and scores on the first half of the test strongly correlated with those on the second half ($r_{89} = 0.85$, $P < 0.0001$), indicating superior internal consistency.

In Experiment 4, we validated the CSIT against two widely used smell identification tests, the UPSIT (simplified Chinese version, 40 items) (Doty et al. 1984b) and the SS-16 (response alternatives translated to simplified Chinese, 16 items) (Hummel et al. 1997), in a group of 66 participants (36 m) aged between 20 and 68. We expected that CSIT scores would significantly correlate with UPSIT and SS-16 scores, as all three tests assess smell identification ability in a multiple-choice format. Indeed, the correlation between the participants' CSIT scores (range: 19–40, accuracy: 47.5–100%) and UPSIT scores (13–38, 32.5–95%) was 0.77 ($P < 0.0001$, Figure 1B); that between the CSIT scores and SS-16 scores (6–15, 37.5–93.8%) was 0.65 ($P < 0.0001$, Figure 1C). Both values were higher than the correlation between the UPSIT and SS-16 scores, which was 0.55 ($P < 0.0001$). Notably, the participants' average accuracy on the CSIT was 15.1% higher than that on the UPSIT (mean accuracy difference = 15.1%, $t_{65} = 16.13$, $P < 0.0001$) and 16.2% higher than

Table 1. Odor items in the 40-item CSIT and their familiarity and identifiability ratings obtained in Experiment 1.

	Odor item	Familiarity	Identifiability		Odor item	Familiarity	Identifiability
1	Garlic	5.95	6.01	21	Sichuan pepper	5.35	4.95
2	Vinegar	5.95	5.99	22	Lemon	5.38	5.22
3	Florida water[a]	5.93	5.83	23	Soybean milk	5.52	5.06
4	Banana	5.87	5.61	24	Hawthorn berry	5.38	5.09
5	Ginger	5.75	5.56	25	Dried jujube date	5.42	5.00
6	Apple	5.88	5.52	26	Mango	5.28	5.05
7	Baked sweet potato	5.78	5.56	27	Rose	5.36	5.32
8	Fish	5.74	5.67	28	Pomelo	5.26	4.92
9	Pineapple	5.76	5.53	29	Green tea	5.33	4.85
10	Orange	5.72	5.46	30	Honey	5.27	4.88
11	Cucumber	5.73	5.42	31	Cumin	5.01	5.04
12	Poached corn	5.71	5.43	32	Walnut	5.10	4.64
13	Coffee	5.59	5.52	33	Sesame oil	4.85	4.54
14	Coke	5.66	5.36	34	Aniseed	4.68	4.66
15	Chocolate	5.70	5.32	35	Almond	4.78	4.51
16	Grape	5.63	5.24	36	Peanut	4.73	4.45
17	Soy sauce	5.60	5.25	37	Longan	4.73	4.44
18	Milk	5.61	5.25	38	Coconut	4.66	4.48
19	Strawberry	5.62	5.17	39	Sandalwood	4.54	4.51
20	Pear	5.63	5.14	40	Vanilla	4.20	4.04

The ratings were on a 7-point Likert scale, with 7 denoting very familiar or very identifiable.
[a]Pinyin: *hua lu shui*, used widely in China as a mosquito repellent and to ease skin irritation.

that on the SS-16 (mean accuracy difference = 16.2%, t_{65} = 13.36, P < 0.0001), whereas their UPSIT and SS-16 accuracies did not differ (t_{65} = 0.75, P = 0.46) (Figure 1D). The SDs for the CSIT, UPSIT, and SS-16 accuracies were 10.0%, 11.9%, and 12.8%, respectively. Hence, on average, a participant's CSIT accuracy was more than 1 SD higher than his/her UPSIT or SS-16 accuracy. These sizeable differences spoke to the prominent role of culture in odorants' identifiability.

The simplified Chinese version of the UPSIT contains 11 of the 12 items in the Cross-Cultural Smell Identification Test (Doty et al. 1996), namely banana, onion, gasoline, chocolate, turpentine, pineapple, paint thinner, smoke, lemon, soap, and rose (cinnamon is not included in the simplified Chinese version of the UPSIT). We analyzed performances on these 11 cross-cultural items for 45 of the participants in Experiment 4 (raw records for the other 21 participants were missing) and found that the accuracies were in fact not significantly different from those on the other 29 UPSIT items (t_{44} = −0.18, P = 0.86). The participants' CSIT accuracies were again 16.4% higher than their accuracies on the 11 cross-cultural UPSIT items (mean accuracy difference = 16.4%, t_{44} = 7.68, P < 0.0001), reaffirming the cultural appropriateness of the CSIT.

Whereas both the CSIT and the UPSIT (Doty et al. 1984b) employ response alternatives as distinct from one another as possible, the distractors in the SS-16 are typically similar to the corresponding target odors (Gudziol and Hummel 2009). One could argue that the difference between the identification accuracies for the CSIT and the SS-16, as observed in Experiment 4, was due to the degree of similarity among response alternatives rather than cultural factors per se. To directly examine this possibility, Experiment 5 assessed in another panel of 119 participants (birthplaces in mainland China: north: 16, northeast: 6, east: 31, southeast: 27, south-central: 29, northwest: 10) the similarities between distractors and the corresponding odor items in the CSIT. Fifty of them also provided ratings for the SS-16. Overall, distractors and targets in the CSIT were rated as significantly less similar than those in the SS-16 (t_{49} = −10.50, P < 0.0001). But there was a significant main effect of odor item in the ratings for the CSIT ($F_{21.6, 2551.3}$ = 52.12, P < 0.0001) as well as for the SS-16 ($F_{10.5, 515.3}$ = 18.71, P < 0.0001). As shown in Supplementary Table S3, the mean distractor-to-target similarity ranged from 1.5 to 4.0 across the CSIT odor items. The top 10 CSIT items with the highest distractor-to-target similarities were in fact comparable to the SS-16 items in terms of distractor-to-target similarity (3.0 vs. 2.9, t_{49} = 1.20, P = 0.24). Nonetheless, a reexamination of the available raw records in Experiment 4 revealed that the participants' accuracies on these 10 CSIT items were still 13.1% higher than their SS-16 accuracies (t_{45} = 6.89, P < 0.0001). This led us to conclude that distractor-to-target similarity could not account for the sizeable difference between the participants' CSIT and SS-16 accuracies in Experiment 4.

Age, gender, and CSIT scores

It is well established that smell identification ability changes with age and differs between the two genders (Doty et al. 1984a; Hummel et al. 2007). In Experiment 6, we examined whether the CSIT is sensitive to such age and gender differences. To this end, we recruited four groups of healthy nonsmokers: young females (mean age ± SD: 17.2 ± 1.6 years), young males (16.9 ± 1.4 years), middle-aged females (51.7 ± 1.4 years), and middle-aged males (51.7 ± 1.3 years), who were matched in years of education (mean ± SD = 12.9 ± 2.5, 12.0 ± 2.4, 12.0 ± 2.4, and 12.9 ± 2.5 years, respectively). The participants were each administered with the CSIT independently. Analyses of their CSIT scores showed a significant main effect of age group ($F_{1, 116}$ = 23.92, P < 0.0001) as well as a significant main effect of gender ($F_{1, 116}$ = 12.03, P = 0.001) (Figure 2). Overall, the young participants scored 4 points (out of 40 total) higher than the middle-aged ones (t_{118} = 4.69, P < 0.0001). Within each age group, the females outperformed the males by approximately 3 points (young: t_{58} = 2.26, P = 0.028; middle-aged: t_{58} = 2.66, P = 0.010). Hence, the CSIT nicely captured the age and gender differences in smell identification among these four groups of participants. Notably, the same pattern of results was obtained when we limited our analyses to only the odd items (age: $F_{1, 116}$ = 25.79, P < 0.0001; gender: $F_{1, 116}$ = 9.56, P = 0.002), the

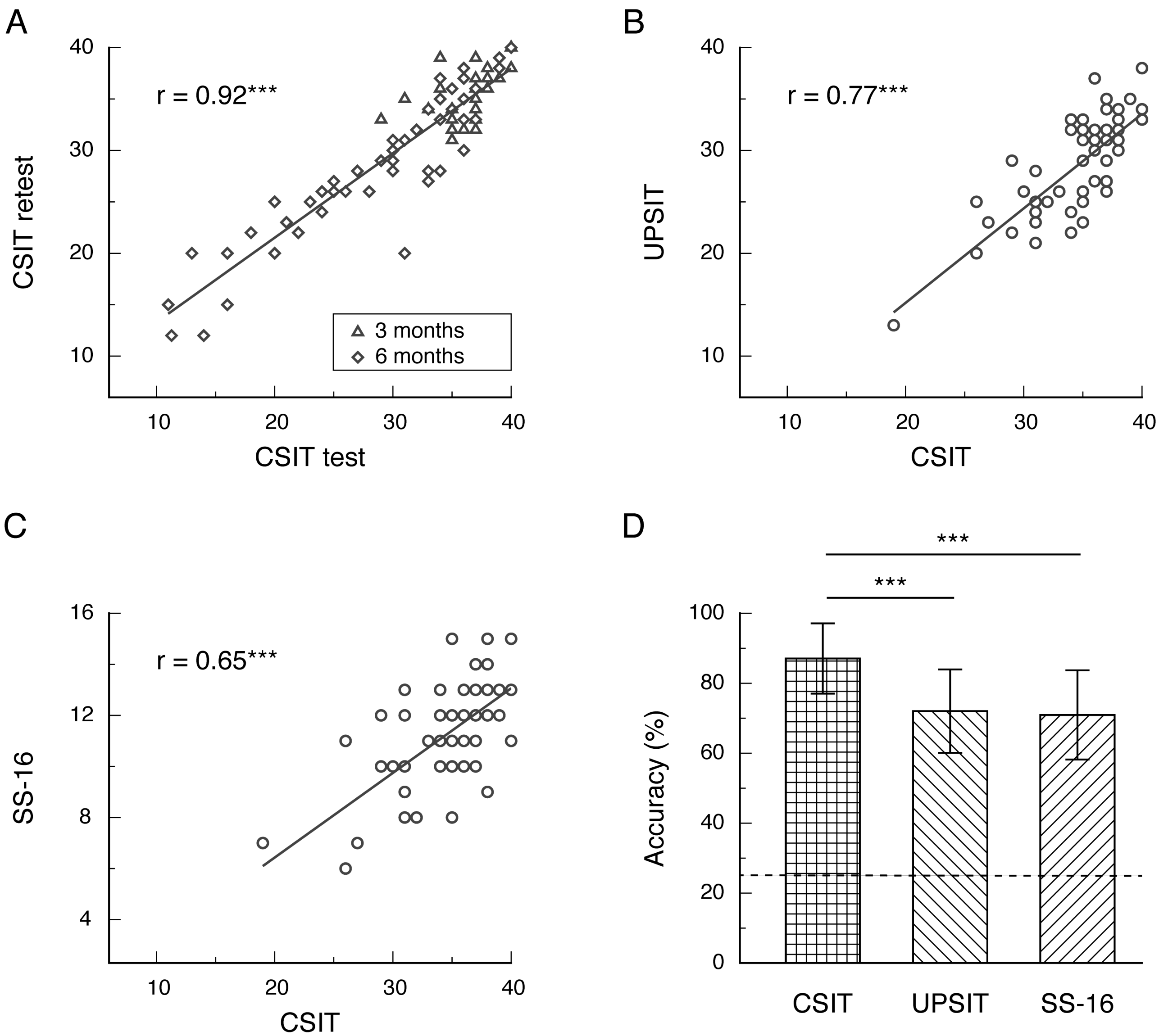

Figure 1. Reliability and validity of the CSIT. (**A**) Test–retest correlation between the CSIT scores of the same individuals on two administrations separated by 3 months (triangles) or 6 months (diamonds). (**B** and **C**) CSIT scores were significantly correlated with UPSIT scores (B) and SS-16 scores (C). (**D**) Chinese participants showed significantly higher smell identification accuracies on the CSIT as compared to the UPSIT or SS-16. Error bars: SD; dashed line: chance level; ***$P < 0.0001$.

even items ($F_{1, 116}$ = 14.04 and 9.33, $P < 0.001$ and = 0.003, respectively), the first half ($F_{1, 116}$ = 20.73 and 10.56, $P < 0.0001$ and = 0.002, respectively), or the second half ($F_{1, 116}$ = 16.82 and 8.36, $P < 0.0001$ and = 0.005, respectively) of the CSIT, which further underscored the internal consistency of the CSIT.

Discussion

We have described the development of a 40-item smell identification test for the Chinese—the CSIT. It adopts odor items that are familiar and identifiable to most Chinese people (Experiments 1 and 2), has a test–retest reliability of 0.92 (Experiment 3), a split-half reliability of no less than 0.85 (Experiment 3), and is validated against two widely used smell identification tests developed in the USA and Germany, namely the 40-item UPSIT (simplified Chinese version) (Doty et al. 1984b) and the 16-item SS-16 (Hummel et al. 1997) (Experiment 4). We also demonstrate that the CSIT is sensitive to age and gender differences in smell identification ability (Experiment 6). Whereas the CSIT odorants were mainly presented in felt-tip pens in the current study, we note that there are other possible means to deliver the odorants, e.g., in a scratch'n sniff format or via an olfactometer, and that the key to a useful culturally specific smell identification test is the cultural suitability of the odor items.

In terms of reliability, the CSIT (test–retest $r = 0.92$) is comparable to the UPSIT ($r = 0.92$) (Doty et al. 1984b) and the extended version of the Sniffin' Sticks Identification Test ($r = 0.88$) (Haehner et al. 2009), and outperforms the SS-16 ($r = 0.73$) (Hummel et al. 1997) and the Scandinavian Odor Identification Test ($r = 0.79$) (Nordin et al. 1998) that have fewer odor items. In terms of internal consistency, the CSIT (split-half reliability $r \geq 0.85$) is again comparable to the UPSIT ($r \approx 0.86$) (Doty et al. 1989). Importantly, the CSIT more faithfully measures smell identification ability for Chinese people than the UPSIT, SS-16, or the cross-cultural items of the UPSIT. As shown in Experiment 4, their accuracies on the CSIT exceeded those on the

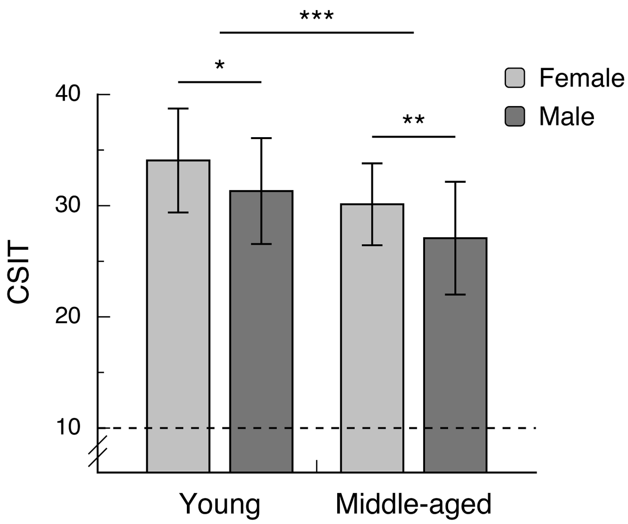

Figure 2. Sensitivity of the CSIT to age and gender differences in smell identification ability. Overall, young participants outperformed middle-aged ones and females outperformed males on the CSIT. Error bars: SD; dashed line: chance level; $*P < 0.05$, $**P \leq 0.01$; $***P < 0.0001$.

UPSIT, SS-16, or the cross-cultural items of the UPSIT by over 15%. This result, while highlighting the cultural appropriateness of the CSIT, also cautions that olfactory diagnoses made for Chinese patients based on the UPSIT, SS-16, or the Cross-Cultural Smell Identification Test and their established norms (Doty et al. 1996; Hummel et al. 2007; Doty 2008) could be inadequate, as these tests significantly underrate Chinese individuals' olfactory functions.

We have started testing the CSIT's effectiveness in clinical settings. Some earlier data indicate that it nicely differentiates between patients with Parkinson's disease and healthy age-matched controls (Hu et al. 2016). As such, we believe that the CSIT provides a useful tool for the assessment of olfactory function in the Chinese population, which accounts for about one-fifth of the world population and the world's disease burden.

Supplementary material

Supplementary data are available at *Chemical Senses* online.

Funding

This work was supported by the Science and Technology Service Network Initiative (grant number KFJ-STS-ZDTP-032), the Strategic Priority Research Program (grant number XDB32010200), and the CAS Key Laboratory of Mental Health (grant number 2011CB711000) of the Chinese Academy of Sciences.

Acknowledgments

We thank Kankan Wu and Zhengkui Liu for assistance with data collection.

References

Ayabe-Kanamura S, Schicker I, Laska M, Hudson R, Distel H, Kobayakawa T, Saito S. 1998. Differences in perception of everyday odors: a Japanese-German cross-cultural study. *Chem Senses*. 23:31–38.

Blomqvist EH, Brämerson A, Stjärne P, Nordin S. 2004. Consequences of olfactory loss and adopted coping strategies. *Rhinology*. 42:189–194.

Boesveldt S, Verbaan D, Knol DL, Visser M, van Rooden SM, van Hilten JJ, Berendse HW. 2008. A comparative study of odor identification and odor discrimination deficits in Parkinson's disease. *Mov Disord*. 23:1984–1990.

Cain WS. 1979. To know with the nose: keys to odor identification. *Science*. 203:467–470.

Chen W, Chen S, Kang WY, Li B, Xu ZM, Xiao Q, Liu J, Wang Y, Wang G, Chen SD. 2012. Application of odor identification test in Parkinson's disease in China: a matched case-control study. *J Neurol Sci*. 316:47–50.

Chen G, Pan H, Li L, Wang J, Zhang D, Wu Z. 2018. Olfactory assessment in the Chinese pediatric population. *Medicine (Baltimore)*. 97:e0464.

Croy I, Nordin S, Hummel T. 2014. Olfactory disorders and quality of life–an updated review. *Chem Senses*. 39:185–194.

Doty RL. 2008. *The smell identification test administration manual*. Haddon Heights, NJ: Sensonics Inc.

Doty RL, Frye RE, Agrawal U. 1989. Internal consistency reliability of the fractionated and whole University of Pennsylvania Smell Identification Test. *Percept Psychophys*. 45:381–384.

Doty RL, Marcus A, Lee WW. 1996. Development of the 12-item cross-cultural smell identification test (CC-SIT). *Laryngoscope*. 106:353–356.

Doty RL, Shaman P, Applebaum SL, Giberson R, Siksorski L, Rosenberg L. 1984a. Smell identification ability: changes with age. *Science*. 226:1441–1443.

Doty RL, Shaman P, Dann M. 1984b. Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. *Physiol Behav*. 32:489–502.

Ekström I, Sjölund S, Nordin S, Nordin Adolfsson A, Adolfsson R, Nilsson LG, Larsson M, Olofsson JK. 2017. Smell loss predicts mortality risk regardless of dementia conversion. *J Am Geriatr Soc*. 65:1238–1243.

Gudziol V, Hummel T. 2009. The influence of distractors on odor identification. *Arch Otolaryngol Head Neck Surg*. 135:143–145.

Haehner A, Mayer AM, Landis BN, Pournaras I, Lill K, Gudziol V, Hummel T. 2009. High test-retest reliability of the extended version of the "Sniffin' Sticks" test. *Chem Senses*. 34:705–711.

Hedner M, Larsson M, Arnold N, Zucco GM, Hummel T. 2010. Cognitive factors in odor detection, odor discrimination, and odor identification tasks. *J Clin Exp Neuropsychol*. 32:1062–1067.

Hu P, Feng G, Chen X, Yang Q, Fang J, Zhou W, Wang K. 2016. Application of smell identification test in Parkinson's disease in China. *Chinese J Neurol* 8:610–614.

Huang SF, Chen K, Wu JJ, Liu FT, Zhao J, Lin W, Guo SS, Wang YX, Wang Y, Luo SS, et al. 2016. Odor identification test in idiopathic REM-behavior disorder and Parkinson's disease in China. *PLoS One*. 11:e0160199.

Hubert HB, Fabsitz RR, Feinleib M, Brown KS. 1980. Olfactory sensitivity in humans: genetic versus environmental control. *Science*. 208:607–609.

Hummel T, Kobal G, Gudziol H, Mackay-Sim A. 2007. Normative data for the "Sniffin' Sticks" including tests of odor identification, odor discrimination, and olfactory thresholds: an upgrade based on a group of more than 3,000 subjects. *Eur Arch Otorhinolaryngol*. 264:237–243.

Hummel T, Sekinger B, Wolf SR, Pauli E, Kobal G. 1997. 'Sniffin' sticks': olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold. *Chem Senses*. 22:39–52.

Hummel T, Whitcroft KL, Andrews P, Altundag A, Cinghi C, Costanzo RM, Damm M, Frasnelli J, Gudziol H, Gupta N, et al. 2017. Position paper on olfactory dysfunction. *Rhinol Suppl*. 54:1–30.

Jiang RS, Kuo LT, Wu SH, Su MC, Liang KL. 2014. Validation of the applicability of the traditional Chinese version of the University of Pennsylvania Smell Identification Test in patients with chronic rhinosinusitis. *Allergy Rhinol (Providence)*. 5:28–35.

Jiang RS, Liang KL. 2016. Establishment of olfactory diagnosis for the traditional Chinese version of the University of Pennsylvania Smell Identification Test. *Int Forum Allergy Rhinol*. 6:1308–1314.

Jiang R-S, Lin T-K, Hsu C-Y. 2002. The applicability of the University of Pennsylvania smell identification test and cross-cultural smell identification test in Taiwan. *Taiwan Otolaryngol Head Neck Surg* 37:431–436.

Kobayashi M. 2005. The Odor Stick Identification Test for the Japanese (OSIT-J): clinical suitability for patients suffering from olfactory disturbance. *Chem Senses*. 30(Suppl 1):i216–i217.

Kobayashi M, Saito S, Kobayakawa T, Deguchi Y, Costanzo RM. 2006. Cross-cultural comparison of data using the Odor Stick Identification Test for Japanese (OSIT-J). *Chem Senses*. 31:335–342.
Koss E, Weiffenbach JM, Haxby JV, Friedland RP. 1988. Olfactory detection and identification performance are dissociated in early Alzheimer's disease. *Neurology*. 38:1228–1232.
Mesholam RI, Moberg PJ, Mahr RN, Doty RL. 1998. Olfaction in neurodegenerative disease: a meta-analysis of olfactory functioning in Alzheimer's and Parkinson's diseases. *Arch Neurol*. 55:84–90.
Moberg PJ, Agrin R, Gur RE, Gur RC, Turetsky BI, Doty RL. 1999. Olfactory dysfunction in schizophrenia: a qualitative and quantitative review. *Neuropsychopharmacology*. 21:325–340.
Nordin S, Brämerson A, Lidén E, Bende M. 1998. The Scandinavian Odor-Identification Test: development, reliability, validity and normative data. *Acta Otolaryngol*. 118:226–234.
Nordin S, Monsch AU, Murphy C. 1995. Unawareness of smell loss in normal aging and Alzheimer's disease: discrepancy between self-reported and diagnosed smell sensitivity. *J Gerontol B Psychol Sci Soc Sci*. 50:P187–P192.
Pause BM, Miranda A, Göder R, Aldenhoff JB, Ferstl R. 2001. Reduced olfactory performance in patients with major depression. *J Psychiatr Res*. 35:271–277.
Potagas C, Dellatolas G, Ziegler M, Leveteau J, Bathien N, Mac Leod P, Rondot P. 1998. Clinical assessment of olfactory dysfunction in Parkinson's disease. *Mov Disord*. 13:394–399.
Saito S, Ayabe-Kanamura S, Takashima Y, Gotow N, Naito N, Nozawa T, Mise M, Deguchi Y, Kobayakawa T. 2006. Development of a smell identification test using a novel stick-type odor presentation kit. *Chem Senses*. 31:379–391.
Schab FR. 1991. Odor memory: taking stock. *Psychol Bull*. 109:242–251.
Shu CH, Yuan BC. 2008. Assessment of odor identification function in Asia using a modified "Sniffin' Stick" odor identification test. *Eur Arch Otorhinolaryngol*. 265:787–790.
Woodward MR, Amrutkar CV, Shah HC, Benedict RH, Rajakrishnan S, Doody RS, Yan L, Szigeti K. 2017. Validation of olfactory deficit as a biomarker of Alzheimer disease. *Neurol Clin Pract*. 7:5–14.
Yang L, Wei YX, Ren YY, Yu D, Sun YX, Yang BB. 2013. [Clinical application of Sniffin' Sticks olfactory psychophysical measurements]. *Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi*. 48:741–745.
Yang L, Wei Y, Yu D, Zhang J, Liu Y. 2010. Olfactory and gustatory function in healthy adult Chinese subjects. *Otolaryngol Head Neck Surg*. 143:554–560.
Yuan BC, Lee PL, Lee YL, Lin SH, Shu CH. 2010. Investigation of the Sniffin' Sticks olfactory test in Taiwan and comparison with different continents. *J Chin Med Assoc*. 73:483–486.