Validation study of a novel approach for assessment of retronasal olfactory function with combination of odor thresholds and identification

Ayaho Yoshino[1,2] · Robert Pellegrino[1,3] · Curtis R. Luckett[3] · Thomas Hummel[1]

Abstract

Purpose The present study aimed to establish a test battery for the clinical assessment of retronasal odor thresholds and retronasal odor identification.

Study design Prospective case–control series.

Methods Sixty participants (36 women, 24 men, mean age 37.6 ± 19.4 years) were enrolled in this study; 36 were healthy controls and 24 were patients with olfactory dysfunction. Orthonasal olfactory function was assessed with the "Sniffin' Sticks" test battery. Retronasal olfaction was assessed with oral odorant delivery using special containers for threshold function, and with oral tasteless powders for odor identification.

Results Retronasal and orthonasal olfaction were significantly correlated for threshold scores, identification score, and the sum of threshold and identification score (TI score). Validity analyses showed that the retronasal TI score was able to discriminate between healthy controls and patients with olfactory dysfunction.

Conclusions Normosmic participants can be distinctly differentiated from patients with olfactory dysfunction using a valid test battery comprised of retronasal threshold and identification tests. Based on the current findings, we advocate a TI score of 16 as a cut-off between patients and controls. Therefore, TI scores of 17 and above would indicate retronasal normosmia.

Keywords Olfactory disorders · Retronasal olfaction · Retronasal threshold test · Retronasal identification test · Olfaction

Introduction

Olfactory dysfunction is known to be associated with reduced quality of life and major health issues such as neurodegenerative disease [1]. Recently, the coronavirus disease 2019 (COVID-19) has led to increase focus on olfactory and gustatory dysfunctions due to their prevalence as symptoms in European COVID-19 patients [2]. The angiotensin-converting enzyme II (ACE2) receptor has been reported to play a crucial role in the entry of the virus into the cell to cause the final infection [3]. ACE2 expressed on the support cells and stem cells in the olfactory epithelium, as well as mucosa of the oral cavity, and is especially highly enriched in the epithelial cells of the tongue [3, 4]. Therefore, it could be suggested that severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2) targets both olfactory and gustatory systems. Although gustatory dysfunction is reported to be common in COVID-19 [5], it is still unclear whether it is a genuine independent sequela of COVID-19 or if it is influenced to some extent by the co-existing olfactory loss [6]. In fact, in most cases of dysfunction not related to COVID-19 in which patients report distortion in their sense of taste, this symptom can be attributed to impaired retronasal olfaction (flavor) rather than impaired gustation [7]. This is due to "Smell-Taste-Confusion" [8] Retronasally perceived odors are often misinterpreted as "taste", whereas taste refers only to the basic categories of sweet, sour, salty, bitter and umami [9]. Therefore, the retronasal olfactory function should be assessed with psychophysically. However, retronasal olfaction receives much less attention than orthonasal olfaction

✉ Ayaho Yoshino
  yoshinoayaho@nms.ac.jp

1 Smell and Taste Clinic, Department of Otorhinolaryngology, TU Dresden, Dresden, Germany

2 Department of Otorhinolaryngology, Nippon Medical School Hospital, Tokyo, Japan

3 Department of Food Science, University of Tennessee, Knoxville, TN, USA

even though flavor identification which is heavily influenced by retronasal olfaction, is strongly associated with quality of life [10]. Methods for assessing orthonasal olfactory function such as the University of Pennsylvania Smell Identification Test (UPSIT) [11], Sniffin' Sticks [12], and T&T olfactometer [13] have been applied frequently in both clinical and research settings, whereas methods for assessing the retronasal flavor identification function such as taste powders test (Schmeckpulver) [14], and the candy smell test (CST) [15] are used relatively scarcely.

The "taste powders test" comprises 20 selected substances of similar texture (i.e. non-sticky, small grains) with each taste widely recognized by the general population [14]. Taste powders have been confirmed as an effective means to collect stable results among different countries and the test is easy to understand and can be quickly performed [16]. The CST consists of 23 hard candies with each containing 500 mg sorbitol and one unique aroma [15]. Taste powders or aromatized candies are placed on the subject's tongue. Subjects are then asked to select one of four possible choices (four- alternative, forced-choice procedure). A correct answer in either test confers one point to the subject's overall score, resulting in a maximum score of 20 points for the taste powders test or 23 points for the CST. Both tests assess retronasal olfactory identification function but are limited in that they have both taste and texture components. To overcome the issue with the taste component, a new method for assessing retronasal identification with "tasteless" powder has been developed [17]. The odor delivery container (ODC) procedure was also applied to provide a suprathreshold stimuli for the assessment of retronasal odor identification [18]. This procedure allows researchers to overcome the issue of accompanying taste and texture components. However, without threshold testing, it may be hard to determine patterns of olfactory impairment that may reflect an underlying disease etiology [19]. It has been reported that orthonasal odor threshold detection was more affected by COVID-19 than odor identification suggesting that the mechanism behind loss of sense of smell lies at the level of the olfactory neuroepithelium rather than in the central nervous system [20]. In other words, if only odor identification function was assessed, the COVID-19 patients could be misdiagnosed in terms of their olfactory status. It is not clear whether misdiagnosis could occur if only retronasal odor identification function was examined because there is no valid retronasal test battery for both, odor threshold and identification functions.

Thus, the aims of the present study were; (1) to develop a kit for the assessment of retronasal threshold olfactory function; and (2) to investigate the reliability and validity of a retronasal test battery comprising the assessment of retronasal odor thresholds and retronasal odor identification without taste stimulation. Thus, we combined two tests, a modified

ODC procedure to obtain retronasal olfactory thresholds and "tasteless" powder for identification to address the full retronasal function. We hereby demonstrate that this test battery is both reliable and valid for clinical use.

Materials and methods

Participants

All participants were recruited at the Smell and Taste outpatient Clinic at TU Dresden from September 2019 to December 2019. Patients with olfactory dysfunction due to various etiologies were enrolled in this study as the case. As the present study was not specifically designed to investigate retronasal function with a specific etiology, patients with olfactory dysfunction due to various etiologies were enrolled. Healthy volunteers, who were confirmed as normosmic following a medical history check and an examination of olfactory function with the quick identification test (Q-Sticks) [21, 22] (see the orthonasal olfactory function section), were also enrolled as control subjects and received moderate monetary compensation for their participation.

All participants aged under 18 years of age, patients with serious health conditions (e.g., diabetes mellitus), and women who were pregnant or lactating were excluded from the study.

All procedures described in the present study were conducted in accordance with the Declaration of Helsinki's guidelines on biomedical studies involving human subjects (Summerset West Amendment). The Ethics Committee of the Technical University of Dresden approved the study protocol (protocol number EK408102018), and all participants gave their written informed consent.

Sample size estimation

We determined the sample size by utilizing G*Power software [23]. Within the multivariate analysis of variance with two response variables (described in detail in the Statistical approach section) to obtain power of 0.8 with alpha level set to 0.05 to detect large effects size of $f^2 = 0.35$ [21], the projected sample size was at least 21 subjects. Thus, our proposed sample size of $n = 60$ serves as an adequate sample size for the purpose of this study.

Psychophysical examinations

Retronasal olfactory function

Odor threshold To assess retronasal olfactory threshold function, we modified the ODC procedure. Retronasal odor threshold function was evaluated using an originally

designed 60 mL Nalgene bottle (39 mm total diameter × 86 mm height with a 28 mm mouth) modified to have a one-way valve on its side and a plastic tube protruding from the top (Fig. 1). 4 mL of odor was placed in the bottle. Participants were blindfolded and asked to open the valve and inhale the odor through the tube on the top of the bottle. By opening the valve, the air is free to flow through the bottle and inhalation by the participant allows external air to enter the bottle, carrying the odor into the oral cavity of the participant. This procedure allows for retronasal odor delivery through active sampling on the participant's part.

2-Phenylethanol (dissolved in propylene glycol) was used as an odor. The odor threshold test comprises 8 triplets of these original bottles (a total of 24 bottles) numbered from 1 to 8. The highest concentration was a 4% odor solution labeled as "1" with the lowest was labeled as "8". Eight different concentrations were created by stepwise dilution of each previous concentration in a ratio of 1:4. Eight triplets were used instead of 16 triplets as sued in the Sniffin' Sticks battery to shorten the time of testing.

The threshold score was obtained in a three-alternative forced choice, reversal staircase paradigm similar to the Sniffin' Sticks test methodology [12, 24]. Participants were repeatedly presented with triplets of bottles in a randomized order starting with the lowest odor concentrations and were instructed to select one bottle containing an odor from two distracter bottles with the solvent. A staircase paradigm was used where two subsequent correct identifications of the odorous bottle or one incorrect answer were marked as the turning point and resulted in a decrease or increase of concentration in the next triplet, respectively. The inter-stimulus interval was approximately 30 s. The threshold score was the mean of the last four turning points in the seven staircases, with the final score ranging between 1 and 8 points.

Odor identification Retronasal odor identification function was also evaluated using "tasteless" powders [17]. This test comprised of 16 aroma powders which were selected to minimize gustatory stimulation. Each participant was blindfolded and instructed to block their nose with their fingers. A small sample of powder (approximately 0.05 g) was placed on the mid-dorsal section of the participant's tongue. After drawing their tongue back into their mouth, the subject's nostrils would be unblocked. Participants were asked to exhale through their nostrils and select a matching description of the odor from a list of four possible choices. Participants were asked to wash out their mouths with tap water between each trial. The inter-stimulus interval was approximately 30 s. For each odor, the participants would score 1 for a correct identification or 0 for an incorrect identification. The final score represents the sum of correctly identified odors and could range from 0 to 16.

The sum of odor threshold and identification scores were defined as TI score (maximum of 24).

Orthonasal olfactory function

"Sniffin' Sticks" Orthonasal olfaction was evaluated using the Sniffin' Sticks olfactory test battery (Burghart, Wedel, Germany) [12]. TDI (threshold-discrimination-identification) score (maximum of 48) can be used to give a diagnosis for anosmia, hyposmia, or normosmia [12, 24]. Patients also

(a) Odorant delivery container **(b)** Tasteless powders
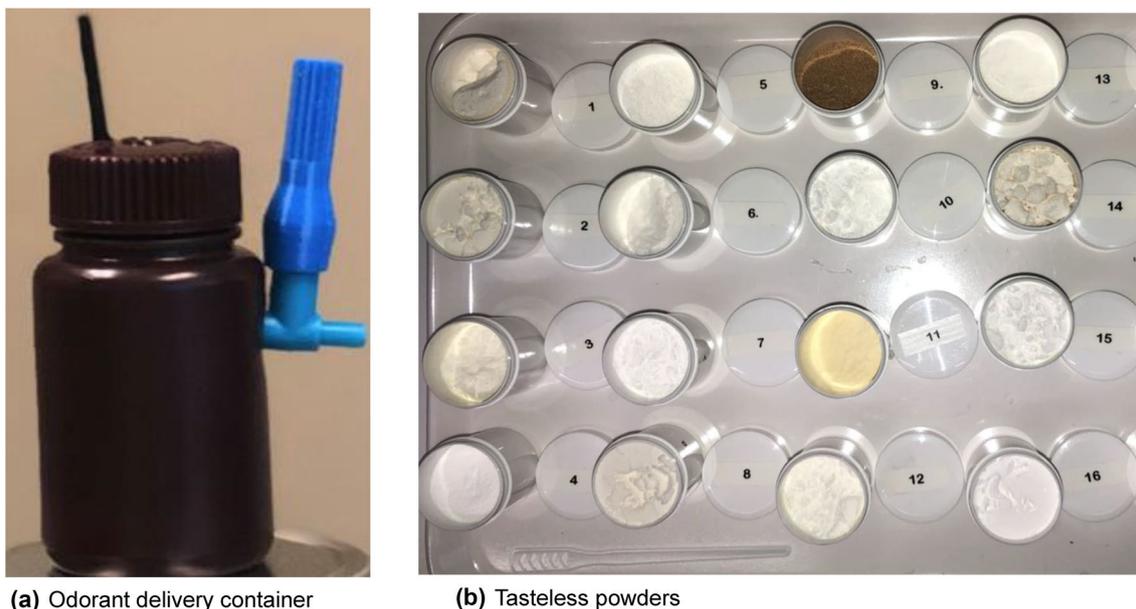
Fig. 1 Picture of the odorant delivery container (**a**), including the straw on top and the valve on the side. Picture of tasteless powders (**b**)

received a detailed orthonasal threshold test. On the other hand, the quick identification test (Q-Sticks) [21, 22] was conducted to confirm that healthy participants were normosmic. "Q-Sticks" were developed based on three odors, "cinnamon", "banana", and "fish", to allow olfactory testing in a very short period of time [22]. This quick test is a valuable tool in clinical screening of the sense of smell. With a score of 3, no further testing is required, and the subject is thought to be normosmic. Healthy participants who scored below 3 were excluded from this study.

Examinations were conducted twice with an interval between sessions ranging from 1 to 14 days. All tests from the main series of experiments were conducted by the same examiner. The measurements were carried out twice for healthy participants and once for patients.

Reliability analysis

Three reliability analyses were conducted with normosmic subjects: (1) the correlation between scores in test 1 and 2, (2) intraclass correlation coefficient (ICC), and (3) Bland–Altman plot. Despite the frequent use of correlation analyses in the assessment of test–retest reliability and for comparison between two methods, the use of these measurements has garnered controversy [25, 26]. The ICC is a frequently applied reliability index to investigate the association between test and retest [27]. The ICC reflects the average correlation between all possible ordering of pairs. It varies from 0 to 1, where 1 corresponds to perfect reliability. An ICC > 0.9 is considered to be excellent reliability; 0.75–0.9 is considered good; 0.5–0.75 is considered moderate; and < 0.5 is considered poor [27]. The Bland–Altman plot was also drawn to indicate systematic bias between repeated measurements [26, 28]. The Bland–Altman plot illustrates the agreement between the two test occasions (time 1 and time 2) and identifies possible outliers. The 95% CI of the mean difference was used to determine systematic bias. If zero is included within the 95% CI, no significant systematic bias between measurements can be inferred [26]. The 95% limits of agreement (LOA) were used to examine the natural variation over time, with a narrow LOA indicating higher stability [25, 28].

Validity analysis

To evaluate validity of the retronasal threshold and identification test, the following steps were conducted. First, the correlation between orthonasal and retronasal function was examined in the patient group. The sum of the threshold score and identification score (TI score) of the orthonasal olfactory test with Sniffin' Sticks was calculated to examine the correlation between orthonasal and retronasal olfactory function. Second, a two-way multivariate analysis of covariance (MANCOVA) was performed with groups based on orthonasal olfactory function (normosmic, hyposmic, anosmic), retronasal olfaction measures (retronasal threshold and retronasal identification score) as within-subject factors and using age as a covariate. Age was included due to its effects on both orthonasal [24, 29] and retronasal [14, 16] function. Gender was not included as a covariate as it was not significant in the initial analysis (f [2,54] = 0.73, $p = 0.49$, partial $\eta 2 = 0.03$). An analysis of covariance (ANCOVA) was performed separately with TI score as a within-subject factor and age as a covariate. A receiver operating characteristic (ROC) curve was used to show the diagnostic ability of the test. The ROC curve was drawn by plotting the sensitivity of a test on the y-axis against 1-specificity on the x-axis [30]. Additionally, the area under the curve (AUC) provides a measure to summarize the overall diagnostic accuracy of the test. AUC takes a value of 0 to 1, where 0 is interpreted as a perfectly inaccurate test and 1 indicates a perfectly accurate test [31].

All statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS version 25.0; SPSS Inc, Chicago, IL, USA). The level of significance was set at 0.05 for all tests.

Results

Descriptive data of participants

A total of 60 participants (36 females, 24 males, mean age 37.6 ± 19.4 years) were enrolled in this study. Thirty-six participants were healthy volunteers (27 females, 9 males; mean age 28.5 ± 12.1 years), and 24 participants were patients (12 females, 12 males; mean age 51.3 ± 20.4 years). The causes of olfactory loss were idiopathic dysfunction (4 females, 5 males; mean age 68.7 ± 12.1 years), congenital olfactory dysfunction (2 females, 2 males; mean age 21.5 ± 6.6 years), post-traumatic olfactory dysfunction (3 females, 2 males; mean age 40.4 ± 10.2 years), post-infectious olfactory dysfunction (2 females, 2 males; mean age 48.8 ± 13.4 years), Parkinson's disease (1 male; 71 years), sinonasal disease (1 female; 58 years) (Table 1). The retronasal olfactory test scores for each olfactory function group (i.e. normosmic, hyposmic, anosmic), are shown in Table 2.

Test–retest reliability [in normosmia ($n = 36$)]

To determine the reliability of the stimulus set, Pearson's correlation coefficients were measured using healthy controls. There was a significant correlation between test and retest values for the retronasal threshold test, identification test, and TI score [r (36) = 0.71, $p < 0.001$, r (36) = 0.67, $p < 0.001$, r (36) = 0.72, $p < 0.001$, respectively] (Fig. 2).

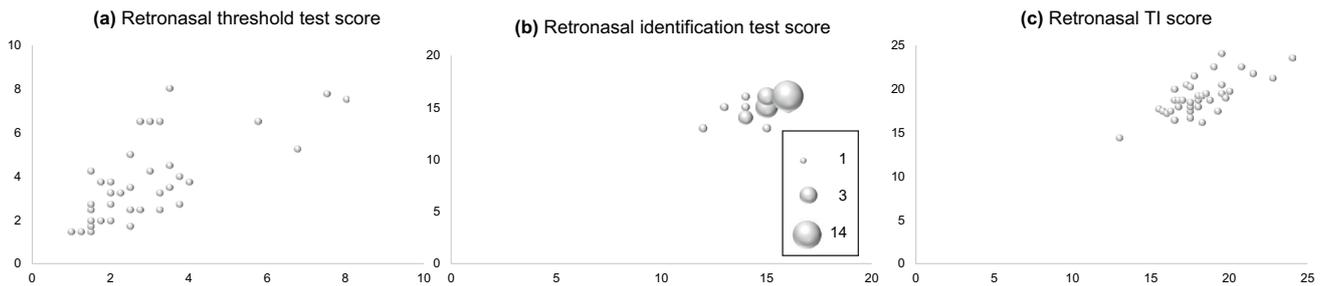**Table 1** Description of participant characteristics by orthonasal olfactory ability

	Normosmia n = 36	Hyposmia n = 7	Anosmia n = 17	p value
Female:male (n)	27:9	12:12		0.13
		3:4	9:8	
Age (years)				
Mean ± SD	28.5 ± 12.1	47.2 ± 16.1	52.9 ± 22.2	< 0.001
Minimum–maximum (in years)	19–68	29–70	17–82	
Cause of olfactory loss (n)				
Idiopathic		3	6	
Infection of upper respiratory tract		2	2	
Head trauma		2	3	
Sinonasal disease		0	1	
Congenital		0	4	
Parkinson		0	1	
Sniffin' Sticks score				
Mean score of thresholds (SEM)		3.51 (0.97)	1.18 (0.09)	0.001
Mean score of identification (SEM)		10.1 (1.37)	4.88 (0.49)	< 0.001
Mean score of TI score (SEM)		13.7 (1.9)	6.06 (0.55)	< 0.001

The between-group comparisons were based on analysis of variance

SD standard deviation, *SEM* standard error of the mean, *TI score* the sum score of threshold and identification in Sniffin' Sticks test

Table 2 Retronasal threshold score, identification score and TI score of all participants

Retronasal score	Threshold			Identification			TI score		
	Normosmia n = 36	Hyposmia n = 7	Anosmia n = 17	Normosmia n = 36	Hyposmia n = 7	Anosmia n = 17	Normosmia n = 36	Hyposmia n = 7	Anosmia n = 17
Mean (SEM)	2.95 (0.28)	1.71 (0.32)	1.12 (0.06)	15.17 (0.16)	10.00 (2.02)	5.77 (0.82)	18.12 (0.35)	11.71 (2.22)	6.88 (0.83)
Minimum	1	1	1	12	5	1	13	6	2
Maximum	8	3.5	1.75	16	16	14	24	18.5	15
Percentile									
5	1.21	1	1	12.85	5	1	15.13	6	2
10	1.5	1	1	14	5	1.8	15.93	6	2.8
25	1.75	1.25	1	15	5	4	16.75	6.25	5
50	2.5	1.5	1	15	7	5	17.88	8.75	6
75	3.5	1.75	1.13	16	16	7	19.44	17.75	8
90	6.05		1.55	16		12.4	20.98		14
95	7.58			16			22.94		

Fig. 2 Bubble chart of the retronasal threshold (**a**), identification (**b**), and TI score (**c**) obtained from test and retest. The number of participants per data point is indicated by the size of bubbles. *x*-axis represents test score, and *y*-axis represents retest score

As detailed in Table 3, the ICC for test–retest of retronasal threshold, retronasal identification and TI score were calculated (0.78, 0.80 and 0.78, respectively), indicating good reliability of the test–retest.

A Bland–Altman plot was also drawn, which can quantify agreement in test–retest paradigms. The difference in two measurements and 95% limits of agreement (LOA) for each test–retest difference were calculated as the mean difference $\pm 1.96 \times$ standard deviation (SD) of the difference. This means that in 95 out of 100 comparisons, the difference between two measurements can be expected to fall into this range. Narrower limits of agreement indicate a better agreement between two measurements [26]. The Bland–Altman plot suggested a good consistency between the results from the 2 sessions in the retronasal threshold, identification and TI score (Fig. 3).

Correlation between retronasal and orthonasal function for patients ($n = 24$)

To investigate the correlation between retronasal and orthonasal function, Pearson's correlation was calculated between

orthonasal olfactory function score (threshold, identification, and TI score) and retronasal olfactory function score (threshold, identification, and TI score). The correlation between retronasal olfactory function and orthonasal function in terms of the threshold was significant [$r(24) = 0.61$, $p = 0.002$] as well as identification [$r(24) = 0.71$, $p < 0.001$] and TI score [$r(24) = 0.76$, $p < 0.001$] (Fig. 4).

Comparison of each olfactory function group in terms of retronasal function

A MANCOVA was conducted with the two retronasal olfactory function test scores (threshold and identification score) as dependent variables, three groups based on orthonasal olfactory function (normosmic, hyposmic, anosmic) as independent variables, and using age as a covariate.

There was a statistically significant difference among the three groups on the combined dependent variable retronasal olfactory function when controlling for age ($f[4,110] = 26.17$, $p < 0.001$, Wilks' Lambda $= 0.26$, partial $\eta 2 = 0.49$). Significant differences among normosmic, hyposmic, and anosmic groups existed with regards to subset retronasal threshold

Table 3 Mean (standard error of the mean) of the test–retest score, 95% LOA, and ICC in 94 normosmic subjects

| | Mean score (SEM) | | ICC (95% CI) | Bland–Altman analysis | | | | | | | |
	Test	Retest		d	SD diff	SE of d	95% CI Lower	Upper	LOA Lower	Upper
Retronasal threshold	2.95 (0.28)	3.80 (0.32)	0.78 (0.46–0.90)	− 0.85	1.39	0.23	− 1.32	− 0.38	− 3.57	1.88
Retronasal identification	15.17 (0.16)	15.36 (0.14)	0.80 (0.60–0.90)	− 0.19	0.75	0.12	− 0.45	0.06	− 1.66	1.27
TI score	18.12 (0.35)	19.16 (0.35)	0.78 (0.43–0.90)	− 1.04	1.56	0.26	− 1.57	− 0.51	− 4.10	2.02

TI score score sum of retronasal threshold and identification scores, *ICC* intraclass correlation coefficient, *d* mean of the difference between the two test sessions (test–retest), *SD diff* standard deviation of mean difference, *SE of d* standard error, *95% CI* mean difference $\pm 1.96 \times$ SE, *LOA* 95% limits of agreement $= d \pm 1.96$ SD diff
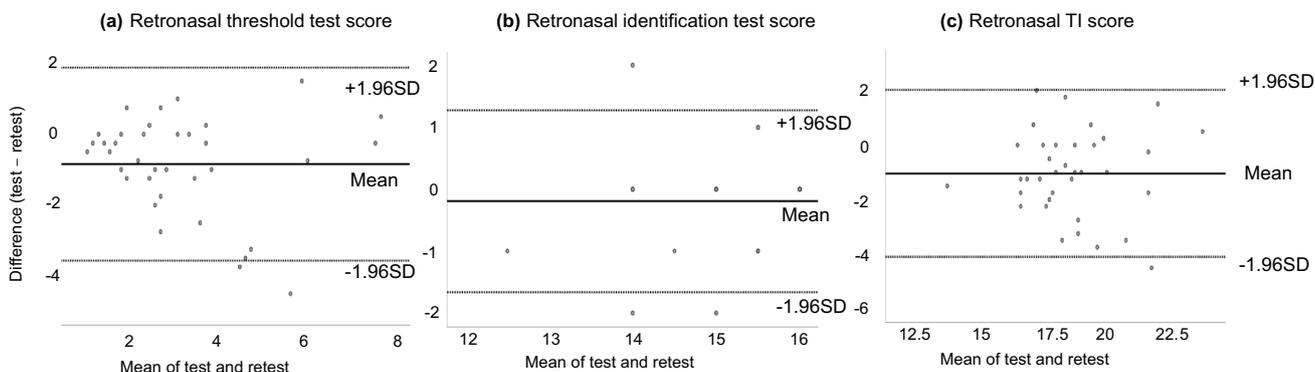
Fig. 3 Bland–Altman plots for the test–retest reliability. The mean differences between test and retest (*y*-axis) are plotted against the mean score of the two sessions (*x*-axis). Reference lines shows mean difference between test and retest (solid line), and 95% limits of agreement for the mean difference (dotted lines). Less variety was shown in retronasal identification score with regards to the matching of difference and mean score of test and retest (i.e. 14 participants scored 16 points at both sessions in identification test, resulting in one dot.) The Bland–Altman plot showed good consistency between the results from test and retest
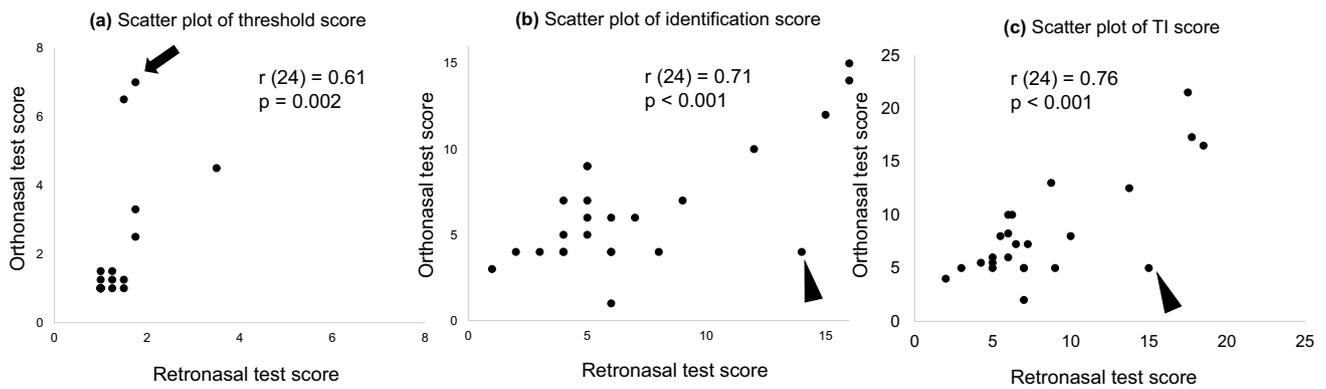
Fig. 4 Scatter plots of retronasal threshold (**a**), identification (**b**), and TI score (**c**) by orthonasal function (*x*-axis represents retronasal score and *y*-axis represents orthonasal score). Arrow represents the patient who had low retronasal threshold score with high orthonasal threshold score. Arrowhead represents the patient who had high retronasal identification score with low orthonasal identification score

Table 4 Comparison results of retronasal threshold, identification and TI score

	Mean score			*p* value	Pairwise comparisons		
	Normosmia *n* = 36	Hyposmia *n* = 7	Anosmia *n* = 17		Normosmia vs anosmia	Normosmia vs hyposmia	Hyposmia vs anosmia
Threshold	2.95	1.71	1.12	0.003	0.003	ns	ns
Identification	15.17	10.00	5.77	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
TI score	18.12	11.71	6.88	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

The between-group comparisons were based on analysis of variance for threshold and identification score. Analysis of covariance was performed with TI score

ns non-significant

score; $f [2,56] = 6.29$, $p = 0.003$, partial $\eta2 = 0.18$, retronasal identification score; $f [2,56] = 77.53$, $p < 0.001$, partial $\eta2 = 0.74$. A pairwise comparison, using a Bonferroni adjustment, showed there was a significant difference between normosmia and anosmia in retronasal threshold score ($p = 0.003$) while there was no significant difference between normosmia and hyposmia ($p = 0.19$), hyposmia and anosmia ($p = 1.0$). On the other hand, retronasal identification scores discriminated between the three groups significantly [normosmia vs hyposmia ($p < 0.001$), normosmia vs anosmia ($p < 0.001$), hyposmia vs anosmia ($p < 0.001$)]. An ANCOVA showed that TI scores were also significantly different between the three groups [normosmia vs hyposmia ($p < 0.001$), normosmia vs anosmia ($p < 0.001$), hyposmia vs anosmia ($p = 0.001$)] (Table 4).

The ROC curve of the TI scores is shown in Fig. 5. The ROC curve was drawn to determine the cut-off point between controls and patients. The AUC for the TI score was 0.93.
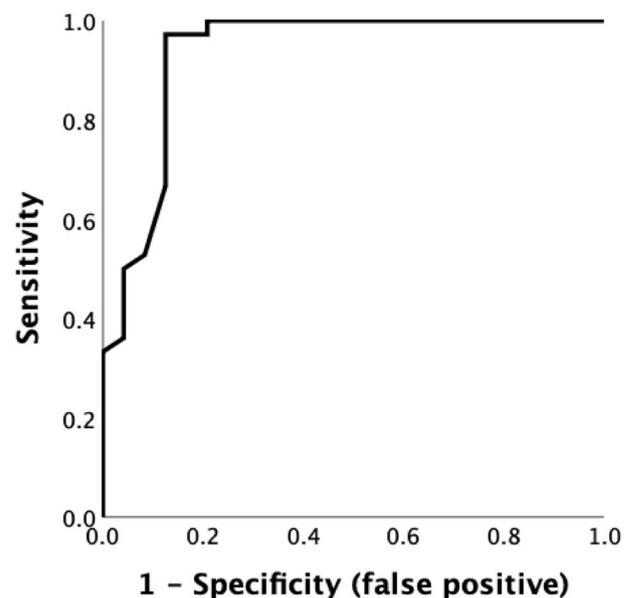
Fig. 5 Receiver operating characteristic (ROC) curve. *y*-axis represents sensitivity and *x*-axis represents the false positive rate of the retronasal test. The area under the curve for TI score is 0.93

Discussion

This study demonstrates the reliability and validity of a retronasal test that includes retronasal thresholds and retronasal odor identification with minimal taste stimulation. Gustatory stimulation is known for its influences on the processing of intranasal stimuli [32]. Hence, it was challenging to establish a retronasal test battery without gustatory stimulation. In previous studies, retronasal olfactory function test were mainly focused on the identification function [14, 15, 17]. The limitations of those tests were that they focused on the perception of odors at suprathreshold levels [33]. It was reported that decreased orthonasal threshold function in combination with relatively better orthonasal odor identification is a typical finding in chronic rhinosinusitis patients [34]. Therefore, testing methods that only report identification scores might underestimate the degree of olfactory loss present in chronic rhinosinusitis patients or COVID-19 patients [20, 34]. The ability to test both threshold and identification has implications for differentiating olfactory dysfunction in the periphery and that at the central processing level. As mentioned earlier these patterns in olfactory impairment can be connected with disease etiology [19].

Test–retest reliability was assessed using three approaches. Among normosmic subjects, there was a significant correlation between test and retest in retronasal threshold test, retronasal identification test, and TI score. The ICCs for the retronasal threshold test, retronasal identification test, and TI score were above 0.75, indicating good reliability. Bland–Altman plot showed good agreement between test and retest.

The results of the present study showed a significant correlation between retronasal and orthonasal olfaction in patients. This result was in line with a previous study that showed a significant correlation between the retronasal taste powders test and TDI score of the Sniffin' Sticks test [16]. However, there were some patients who demonstrated a good orthonasal function with bad retronasal function, or vice versa. One patient with post-infectious olfactory dysfunction had low retronasal threshold score (1.5 point) with relatively good orthonasal threshold score (6.5 point). On the other hand, a patient with post viral olfactory dysfunction patient scored 14 in the retronasal identification test and scored 4 in the orthonasal identification test. A previous study already showed similar discrepancies with retronasal function being better than orthonasal olfactory function in sinonasal disease patients [35, 36]. In comparisons of various etiologies, retronasal scores for nasal polyp patients were higher compared with a post-traumatic olfactory loss and postinfectious olfactory loss patients [37]. Due to the limited number of participants

in our research, the relationship between orthonasal and retronasal olfactory function by different etiologies is still unclear. Further study is required to investigate these cases of imbalance between orthonasal and retronasal olfactory function by different etiologies. Such detailed investigations could be carried out using the techniques described in the present study.

Both tests for retronasal olfactory threshold and identification function allowed discrimination between controls and patients. In order to clinically separate patients from normosmic participants, we used the 25th percentile of the distribution of the TI score for the normosmic group and the 75th percentile of the TI score for the hyposmic group to determine the cut-off point between the control and patient groups. At a cut-off point of 16, the sensitivity to detect patients (hyposmia and anosmia) was calculated at 89% with a false-positive rate of 13%.

Based on the current findings, we advocate the cut-off between patients and controls at a TI score of 16. Therefore, TI scores of 17 and above would indicate retronasal normosmia. Further research should seek to collect more normative data on retronasal olfaction. Also, of interest, this study found some orthonasally normosmic participants, who scored in an anosmic range retronasally. These individual differences in retronasal olfaction, in those with intact orthonasal sense of smell, should be further investigated. Furthermore, the investigation of the patterns of retronasal olfactory impairment with threshold and identification also should be considered in a future study.

In the era of COVID-19, as olfactory and gustatory dysfunction were reported in many patients, chemosensory studies should include assessment of retronasal olfactory function and extended gustatory investigation [20]. Future study for the patients with COVID-19 reported loss of taste should be considered to assess retronasal olfactory function including odor threshold and odor identification.

Conclusion

The results of the present study indicate that a reliable and valid assessment of retronasal olfaction is possible. Further study is required to investigate the cause for the imbalance between ortho-and retronasal olfactory function and the clinical associations of specific patterns of scores of retronasal odor thresholds and retronasal odor identification. In particular, the examination of retronasal function with this reliable and valid test battery should be considered for the patients with COVID-19 in future studies.

Author contributions AY: collection and analysis of the data, wrote the manuscript; RP: created the ODC, conceptualization of the study, critical revision of the manuscript; CRL: created the ODC, conceptualization of the study, critical revision of the manuscript; TH: conceptualization of the study; collection of data; analysis of the data, critical revision of the manuscript.

Compliance with ethical standards

Conflicts of interest The authors have no conflicts of interest to declare.

Ethics approval All the procedures performed in the study were in accordance with the ethical standards of the institutions at which the study was conducted.

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

Consent for publication Participants signed informed consent forms regarding the publication of their data.

References

1. Hummel T, Whitcroft KL, Andrews P, Altundag A, Cinghi C, Costanzo RM, Damm M, Frasnelli J, Gudziol H, Gupta N, Haehne A, Holbrook E, Hong SC, Hornung D, Huttenbrink KB, Kamel R, Kobayashi M, Konstantinidis I, Landis BN, Leopold DA, Macchi A, Miwa T, Moesges R, Mullol J, Mueller CA, Ottaviano G, Passali GC, Philpott C, Pinto JM, Ramakrishnan VJ, Rombaux P, Roth Y, Schlosser RA, Shu B, Soler G, Stjarne P, Stuck BA, Vodicka J, Welge-Luessen A (2017) Position paper on olfactory dysfunction. Rhinol Suppl 54(26):1–30

2. Lechien JR, Chiesa-Estomba CM, De Siati DR, Horoi M, Le Bon SD, Rodriguez A, Dequanter D, Blecic S, El Afia F, Distinguin L, Chekkoury-Idrissi Y, Hans S, Delgado IL, Calvo-Henriquez C, Lavigne P, Falanga C, Barillari MR, Cammaroto G, Khalife M, Leich P, Souchay C, Rossi C, Journe F, Hsieh J, Edjlali M, Carlier R, Ris L, Lovato A, De Filippis C, Coppee F, Fakhry N, Ayad T, Saussez S (2020) Olfactory and gustatory dysfunctions as a clinical presentation of mild-to-moderate forms of the coronavirus disease (COVID-19): a multicenter European study. Eur Arch Otorhinolaryngol 277(8):2251–2261. https://doi.org/10.1007/s00405-020-05965-1

3. Xu H, Zhong L, Deng J, Peng J, Dan H, Zeng X, Li T, Chen Q (2020) High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. Int J Oral Sci. https://doi.org/10.1038/s41368-020-0074-x

4. Brann DH, Tsukahara T, Weinreb C, Lipovsek M, Van Den Berge K, Gong B, Chance R, Macaulay IC, Chou H-J, Fletcher RB, Das D, Street K, De Bezieux HR, Choi Y-G, Risso D, Dudoit S, Purdom E, Mill J, Hachem RA, Matsunami H, Logan DW, Goldstein BJ, Grubb MS, Ngai J, Datta SR (2020) Non-neuronal expression of SARS-CoV-2 entry genes in the olfactory system suggests mechanisms underlying COVID-19-associated anosmia. Sci Adv 6(31):eabc5801. https://doi.org/10.1126/sciadv.abc5801

5. Tong JY, Wong A, Zhu D, Fastenberg JH, Tham T (2020) The prevalence of olfactory and gustatory dysfunction in COVID-19 patients: a systematic review and meta-analysis. Otolaryngol Head Neck Surg 163(1):3–11. https://doi.org/10.1177/0194599820926473

6. Walker A, Pottinger G, Scott A, Hopkins C (2020) Anosmia and loss of smell in the era of COVID-19. BMJ. https://doi.org/10.1136/bmj.m2808

7. Whitcroft KL, Hummel T (2020) Olfactory dysfunction in COVID-19. JAMA 323(24):2512. https://doi.org/10.1001/jama.2020.8391

8. Rozin P (1982) "Taste-smell confusions" and the duality of the olfactory sense. Percept Psychophys 31(4):397–401

9. Small DM, Gerber JC, Mak YE, Hummel T (2005) Differential neural responses evoked by orthonasal versus retronasal odorant perception in humans. Neuron 47:593–605

10. Oleszkiewicz A, Park D, Resler K, Draf J, Schulze A, Zang Y, Hahner A, Hummel T (2019) Quality of life in patients with olfactory loss is better predicted by flavor identification than by orthonasal olfactory function. Chem Senses 44(6):371–377. https://doi.org/10.1093/chemse/bjz027

11. Doty RL, Shaman P, Kimmelman CP, Dann MS (1984) University of Pennsylvania smell identification test: a rapid quantitative olfactory function test for the clinic. Laryngoscope 94(2 Pt 1):176–178

12. 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(1):39–52

13. Takagi SF (1987) A standardized olfactometer in Japan. A review over ten years. Ann N Y Acad Sci 510:113–118. https://doi.org/10.1111/j.1749-6632.1987.tb43476.x

14. Heilmann S, Strehle G, Rosenheim K, Damm M, Hummel T (2002) Clinical assessment of retronasal olfactory function. Arch Otolaryngol Head Neck Surg 128(4):414–418

15. Haxel BR, Bertz-Duffy S, Faldum A, Trellakis S, Stein B, Renner B, Kobal G, Letzel S, Mann WJ, Muttray A (2011) The candy smell test in clinical routine. Am J Rhinol Allergy 25:e145-148

16. Croy I, Hoffmann H, Philpott C, Rombaux P, Welge-Luessen A, Vodicka J, Konstantinidis I, Morera E, Hummel T (2014) Retronasal testing of olfactory function: an investigation and comparison in seven countries. Eur Arch Otorhinolaryngol 271(5):1087–1095. https://doi.org/10.1007/s00405-013-2684-9

17. Yoshino A, Goektas G, Mahmut MK, Zhu Y, Goektas O, Komachi T, Okubo K, Hummel T (2020) A new method for assessment of retronasal olfactory function. Laryngoscope. https://doi.org/10.1002/lary.28698

18. Chen V, Halpern BP (2008) Retronasal but not oral-cavity-only identification of "purely olfactory" odorants. Chem Senses 33(2):107–118. https://doi.org/10.1093/chemse/bjm069

19. Whitcroft KL, Cuevas M, Haehner A, Hummel T (2017) Patterns of olfactory impairment reflect underlying disease etiology. Laryngoscope 127:291–295

20. Le Bon S-D, Pisarski N, Verbeke J, Prunier L, Cavelier G, Thill M-P, Rodriguez A, Dequanter D, Lechien JR, Le Bon O, Hummel T, Horoi M (2020) Psychophysical evaluation of chemosensory functions 5 weeks after olfactory loss due to COVID-19: a prospective cohort study on 72 patients. Eur Arch Otorhinolaryngol. https://doi.org/10.1007/s00405-020-06267-2

21. Hummel T, Pfetzing U, Lotsch J (2010) A short olfactory test based on the identification of three odors. J Neurol 257(8):1316–1321. https://doi.org/10.1007/s00415-010-5516-5

22. Lotsch J, Ultsch A, Hummel T (2016) How many and which odor identification items are needed to establish normal olfactory function? Chem Senses 41(4):339–344. https://doi.org/10.1093/chemse/bjw006

23. Faul F, Erdfelder E, Lang A-G, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social,

behavioral, and biomedical sciences. Behav Res Methods 39(2):175–191. https://doi.org/10.3758/BF03193146

24. Oleszkiewicz A, Schriever VA, Croy I, Hähner A, Hummel T (2019) Updated Sniffin' Sticks normative data based on an extended sample of 9139 subjects. Eur Arch Otorhinolaryngol 276:719–728

25. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1(8476):307–310

26. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. Stat Methods Med Res 8(2):135–160. https://doi.org/10.1177/096228029900800204

27. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15(2):155–163. https://doi.org/10.1016/j.jcm.2016.02.012

28. Chuang LL, Lin KC, Hsu AL, Wu CY, Chang KC, Li YC, Chen YL (2015) Reliability and validity of a vertical numerical rating scale supplemented with a faces rating scale in measuring fatigue after stroke. Health Qual Life Outcomes 13:91. https://doi.org/10.1186/s12955-015-0290-9

29. 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 3000 subjects. Eur Arch Otorhinolaryngol 264(3):237–243. https://doi.org/10.1007/s00405-006-0173-0

30. Akobeng AK (2007) Understanding diagnostic tests 3: receiver operating characteristic curves. Acta Paediatr 96(5):644–647. https://doi.org/10.1111/j.1651-2227.2006.00178.x

31. Mandrekar JN (2010) Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 5(9):1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

32. Welge-Lussen A, Drago J, Wolfensberger M, Hummel T (2005) Gustatory stimulation influences the processing of intranasal stimuli. Brain Res 1038(1):69–75. https://doi.org/10.1016/j.brainres.2005.01.011

33. Ozay H, Cakir A, Ecevit MC (2019) Retronasal olfaction test methods: a systematic review. Balkan Med J 36(1):49–59. https://doi.org/10.4274/balkanmedj.2018.0052

34. Soler ZM, Kohli P, Storck KA, Schlosser RJ (2016) Olfactory impairment in chronic rhinosinusitis using threshold, discrimination, and identification scores. Chem Senses 41(9):713–719. https://doi.org/10.1093/chemse/bjw080

35. Landis BN, Giger R, Ricchetti A, Leuchter I, Hugentobler M, Hummel T, Lacroix JS (2003) Retronasal olfactory function in nasal polyposis. Laryngoscope 113:1993–1997

36. Ganjaei KG, Soler ZM, Storck KA, Rowan NR, Othieno FA, Schlosser RJ (2018) Variability in retronasal odor identification among patients with chronic rhinosinusitis. Am J Rhinol Allergy 32(5):424–431. https://doi.org/10.1177/1945892418793540

37. Rombaux P, Weitz H, Mouraux A, Nicolas G, Bertrand B, Duprez T, Hummel T (2006) Olfactory function assessed with orthonasal and retronasal testing, olfactory bulb volume, and chemosensory event-related potentials. Arch Otolaryngol Head Neck Surg 132(12):1346–1351. https://doi.org/10.1001/archotol.132.12.1346