



Cultural Adaptation and Validity of the Sniffin' Sticks Psychophysical Test for the UK Setting

Lorna Langstaff¹ · Allan Clark² · Mahmoud Salam³ · Carl M. Philpott^{1,2}

Received: 30 October 2020 / Accepted: 1 June 2021 / Published online: 19 June 2021
© The Author(s) 2021

Abstract

Introduction Olfactory testing must be culturally adapted to be relevant to the target population. This study aimed to validate the Sniffin' Sticks test for the UK setting.

Methods A cohort study was conducted at a tertiary olfactory dysfunction clinic. Phase 1—healthy volunteers underwent the original German identification test followed by a UK adapted version. Phase 2—patients with olfactory dysfunction underwent the extended smell test (TDI) including the new descriptors. Outcome measures included differences in identification test (phase 1), retest reliability and differences in scores before and after treatment.

Results A total of 31 healthy volunteers and 87 patients were recruited (6 and 31 males, respectively). Phase 1—mean identification scores showed a small improvement after descriptor adaptations (13.77 and 14.57, $p=0.0029$). Phase 2—41 untreated participants had a mean identification score of 7.31 at both intervals (95% CI: -1.15 to 1.15 , $p>0.999$). The mean change in treated participants was 1.88 (0.70 to 3.06, $p=0.0224$). TDI score difference between treated and untreated groups was 6.63 (2.48 to 10.79, $p=0.0023$). The intraclass correlation coefficient for untreated patients was high for both TDI score (ICC = 0.82, 95% CI 0.57 to 0.93) and identification score (ICC = 0.80, 0.52 to 0.93); CIs suggest the reliability is moderate to excellent.

Conclusions This study confirms the validity of the descriptor adaptations of the identification component of the Sniffin' Sticks test to distinguish between health and disease.

Implications The Sniffin' Sticks test can now reliably be used for clinical assessment of British patients, modifying only the descriptors.

Keywords Olfaction · Anosmia · Smell · Taste · Odour · Quality of Life · Nasal · Language · Reliability

Background

Olfactory disorders have a wide spectrum of quality of life impact with anxiety, depression and isolation as common sequelae (Philpott and Boak 2014; Erskine and Philpott 2020). There are many causes for people to lose their sense of smell, although the most common ones include chronic

rhinosinusitis, post-viral and post-traumatic with some cases being idiopathic. Patients' self-perception of olfactory function and their performance on psychophysical testing often bear little resemblance, prompting the need to ensure they receive an appropriate olfactory assessment in the clinic setting (Hummel 2017). The Sniffin' Sticks test is commonly used for this purpose and the test kit comprises three parts: odour threshold, odour discrimination and odour identification giving a total (TDI) score out of 48. The odour identification (I) part of the test uses 16 common odorants and asks the subject to verbally identify the smell from a selection of four options. The three parts of the test are combined to give a total TDI score. The Sniffin' Sticks test was initially developed and validated on large numbers of patients in Germany (Hummel 2007, 1997; Kobal 1996). Currently, the kit is used by many clinicians around the world and has been validated for various countries and populations (Australia

✉ Carl M. Philpott
C.Philpott@uea.ac.uk

¹ The Norfolk Smell & Taste Clinic, James Paget University Hospital NHS Foundation Trust, Gorleston, Norfolk NR31 6LA, UK

² Norwich Medical School, University of East Anglia, Norfolk NR4 7TJ, UK

³ The Ipswich Hospital NHS Foundation Trust, Suffolk IP4 5PD, UK

(Mackay-Sim 2004), Greece, Taiwan, Italy (Eibenstein 2005), Netherlands (Boesveldt 2008), Sri Lanka (Silveira-Moriyama 2009), Brazil (Silveira-Moriyama 2008)).

We reported initial results of a validation study of the Sniffin' Sticks test for a British population previously (Neumann 2012). This initial work successfully found the threshold and discrimination tasks of the test to be suitable for use in the UK setting but that there could be improvement in the identification task section of the test with cultural or language adaptation. A normal score for identification is ≥ 12 odorants correctly identified. Odour identification is strongly dependent on familiarity with the odours presented and the language used in the descriptors provided. Cultural differences and inadequate translation might prevent odour identification and thus limit the applicability of this olfactory test in the UK. Our initial study showed that in the tested population, the odour most commonly mistaken by subjects with normal olfactory function were apple (35%), turpentine (30%), lemon (30%), and cloves (26%). For the subjects reporting either anosmia or hyposmia, the most commonly mistaken odorants were apple (72%), cinnamon (67%), turpentine (63%), pineapple, and liquorice (both 55%). The possible reasons for these results could be the similarity of distracters with the true odorant or unfamiliarity with an odorant. Also, some distractor descriptors may not be very familiar to a UK-based population, for example, sauerkraut or gummy bears. Adaptations outlined in our previous study have therefore been applied in this study.

Aims and Objectives

This study aims to validate the Sniffin' Sticks test for the UK setting.

Primary Objectives

- To demonstrate that the adaptations of the test resulted in higher identification scores in phase 1 (control) participants
- To assess the test–retest reliability of the identification score in phase 2 participants who did not receive any medical or surgical intervention

Secondary Objectives

- To assess for the difference in the identification score in phase 2 participants between those who received treatment and those who did not
- To assess for the difference in the total TDI score in phase 2 participants between those who received treatment and those who did not

Methods and Materials

Setting

The study was conducted in a tertiary care setting at the James Paget University Hospital and the Ipswich Hospital in East Anglia in the UK, between February 2014 and June 2015.

Study Design

This prospective cohort study was conducted in two phases as follows:

Phase 1

Healthy volunteers were recruited from hospital staff by response to posters and leaflets regarding the study. Volunteers (hospital staff and non-affected visitors to the clinic) were checked for eligibility (see below) including anterior rhinoscopy and following consent, and invited to perform the identification part of the Sniffin' Sticks test only. The 16 odours were presented in turn and for each odour the participant was forced to choose from 1 of 4 options displayed on a computer monitor as each pen was presented (see Table 1). The test was revised by changing the distractors for 5 of the odours, as per the previous study (as mentioned above) to produce a British version which was then applied to the participants at the second visit (3 to 4 weeks later).

Phase 2

Patients attending the Smell & Taste Clinic at the James Paget University Hospital and also at the Ipswich Hospital ENT Department were invited to participate in the study. Patient information leaflets were posted along with their appointment letter for the clinic. Previous clinic visitors were also invited by making the consent form available through the patient support charity Fifth Sense's website (www.fifthsense.org.uk/research). Participants in phase 2 of the study were asked to perform the extended Sniffin' Sticks test on two occasions with the adapted identification test included. These participants also completed a modification of the olfactory disorders questionnaire (reported elsewhere (Langstaff 2019)). All participants received a patient information sheet and signed a valid consent form prior to their participation in the study.

Participants

Inclusion Criteria

- Subjects aged 18–60 years

Table 1 Odours (bold) and their distractors (original option in italics)

1	Orange	Blackberry	Strawberry	Pineapple
2	Smoke	Leather	Glue	Grass
3	Honey	Vanilla	Chocolate	Cinnamon
4	Chives	Peppermint	Wood	Onion
5	Coconut	Banana	Walnut	Cherry
6	Peach	Melon (<i>apple</i>)	Lemon	Orange (<i>grapefruit</i>)
7	Liquorice	Chocolate (<i>gummy bears</i>)	Chewing gum	Cookies
8	Mustard	Smoke (<i>rubber</i>)	Menthol	Turpentine
9	Onion	Pickle (<i>sauerkraut</i>)	Garlic	Carrot
10	Cigarette	Coffee	Wine	Candle smoke
11	Blackcurrant (<i>melon</i>)	Strawberry (<i>peach</i>)	Vanilla (<i>orange</i>)	Apple
12	Cloves	Pepper	Cinnamon	Mustard
13	Pear	Plum	Peach	Pineapple
14	Camomile	Raspberry	Rose	Cherry
15	Aniseed	Rum	Honey	Wood
16	Bread	Fish	Cheese	Ham

- In phase 1, any subject *without* reported olfactory dysfunction
- In phase 2, any patient *with* an olfactory disorder regardless of cause as determined by their history, examination and psychophysical test result (TDI < 31)

Exclusion Criteria

- In the healthy volunteer group, any subject with the following:
 - Active sino-nasal disease, e.g. chronic rhinosinusitis
 - Systemic disease such as Alzheimer disease
 - Liver disease
 - Uncompensated thyroid disease
 - Active B12 deficiency
- In both groups, subjects that do not understand the English language

Variables

The primary outcome measure was the identification score of the Sniffin' Sticks. For all phase 2 participants, data was recorded on their TDI scores for the extended Sniffin' Sticks along with their demographics, their diagnosis, and whether they had received any treatment between visits. Those in the non-treatment group had not yet received any active intervention from the clinic but may have had some treatment in primary care or from a referring secondary care provider. This group may have had investigations such as imaging or serology before the second visits. Visits in phase two coincided with clinic appointments, and due to the nature of

the wide geographic area from which participants came, the time interval varied between 3 and 12 months.

Data Sources/Management

To record the results of the Sniffin' Sticks test, the free "olaf" software download available from the Dresden Smell & Taste Clinic was used (Hummel 2018); an adapted UK setting was made available for the modified identification test. Electronic health records were used to confirm details of the diagnosis.

Sample Size

No formal sample size calculation was made for the purposes of the study; however, an indicative target of 30 healthy volunteers and 100 patients with olfactory disorders was set out at the beginning of the study.

Statistical Methods

Results were logged to a secure database and analysed with Stata/SE 14 (StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, Texas: StataCorp LP). Age and gender were compared using a two-sample *t* test and a Chi-squared test, respectively. Due to the normal distribution of the data, mean score values and standard deviations were calculated for each of the phases. A paired student's *t* test was used to measure the difference in mean identification scores between the first and second tests as well as between the mean TDI scores. In phase two, mean change between visits of both the identification and TDI scores in the treated and untreated cohorts was calculated and compared using unpaired *t* tests. The assumptions of the *t* tests were assessed

by applying a nonparametric test to compare the results, as the results and conclusions were almost identical, and only the *t* test results are reported here. The main aims of the statistical analysis were the following:

1. To demonstrate that the adaptations of the test resulted in higher identification scores in phase 1 (control) participants
2. To assess the test–retest reliability of the identification score in phase 2 participants who did not receive any medical or surgical intervention
3. To assess for the difference in the identification score in phase 2 participants between those who received treatment and those who did not
4. To assess for the difference in the total TDI score in phase 2 participants between those who received treatment and those who did not

Results

Participants

A total of 31 healthy volunteers were recruited to phase one; 30 of these performed the identification section of the Sniffin' Sticks test; one failed to complete the second test. Eighty-seven patients reporting olfactory dysfunction were recruited during phase 2; the target of 100 was not met due to the allocated duration of the study ending. All of those participants completed the British adapted Sniffin' Sticks test at first presentation. There were 31 male and 56 female participants in total. Due to dropouts, only 57 patients returned for a follow-up visit; 41 of these had received treatment and 16 had not. All returning participants completed the extended Sniffin' Sticks test again.

Descriptive Data

In phase 1, the time between the tests ranged from 14 to 52 days. There were 6 males and 25 females with a mean age of 42.29 years (range 28–59); all were non-smokers. Table 2 shows the key participant characteristics; on average, the phase 2 participants were 5.2 years older ($p=0.012$), but no difference was found in the percentage of males. In phase 2 of the study, the aetiology of the 87 participants was varied as characterised in Table 3.

Main Results

Phase One

Table 4 gives the summary statistics for changes between the first and second visits for phase 1 participants. The mean

Table 2 Demographics of control and patients

	Healthy volunteers	Patients	<i>p</i> value
Age, mean (SD)	42.29 (8.44)	47.53 (12.86)	0.0.012
Males, n (%)	6/31 (19.35)	31/87 (35.63)	0.093 ²

¹based on a two-sample *t* test (unequal variances); ²based on a chi-squared test

Table 3 Aetiology of participants in phase 2

Diagnosis for olfactory disorder	Frequency
Chronic rhinosinusitis and subtypes	25
Post-viral olfactory loss	18
Idiopathic olfactory loss	13
Post-traumatic olfactory loss	9
Congenital olfactory deficiency	9
Olfactory cleft disease	6
Iatrogenic olfactory disturbance	3
Nasal septal deviation	2
Other	2

identification score of the healthy volunteers from the first test was 13.77 (SD 1.45). After changes to the 5 distractor descriptors, there was a statistically significant change in mean identification score to 14.57 (SD 1.10, $p=0.0029$).

Phase Two

For the participants that did not receive treatment between the two tests, there was no significant change in the first and second mean identification scores (see Table 4). Table 5 describes the differences at visit 2 and the change between visits 1 and 2. The repeatability of the identification score was good with an intraclass correlation coefficient of 0.8 (0.52, 0.93) for patients not on treatment. Similarly, no significant difference was seen in TDI scores ($p=0.1671$) and the repeatability was also good (ICC=0.82, CI=0.57, 0.93).

For participants that received treatment, Table 5 demonstrates that there is a significant difference in the change between those participants who received an intervention and those who did not of 1.88 ($p=0.0224$); a similar significant difference was seen in TDI scores of 6.63 ($p=0.0023$).

Discussion

Key Results

The significant improvement in the mean identification score of the healthy volunteer groups demonstrates that the changes to the distractors have improved the cultural

Table 4 Summary statistics for phase 2 participants

Group	Visit 1	Visit 2	Mean difference (95% CI)	<i>p</i> value	ICC (95% CI)
Identification component only					
Healthy participants*	13.77 (1.45)	14.57 (1.10)	−0.80 (−1.30, −0.30)	0.0029	0.39 (0.04,0.65)
Patients not on treatment*	7.31 (3.22)	7.31 (3.50)	0.00 (−1.15, 1.15)	1.0000	0.80 (0.52,0.93)
Patients on treatment*	7.56 (3.80)	9.44 (3.87)	−1.88 (−3.06, −0.70)	0.0026	0.47 (0.18,0.69)
TDI score					
Patients not on treatment*	17.94 (10.77)	15.81 (9.20)	2.13 (−0.99, 5.24)	0.1671	0.82 (0.57,0.93)
Patients on treatment*	16.33 (7.16)	20.84 (8.44)	−4.51 (−6.85, −2.17)	0.0004	0.48 (0.15,0.70)

* only reported for individuals with data at both visits

TDI cut-offs: < 16 anosmia, 16–31 hyposmia, > 31 normosmia

Table 5 Mean change between visits at visit 2 (visit 1–visit 2) between participants grouped as with treatment and those without

Group	Mean change between visits 1 and 2		Mean difference (95% CI) between groups	<i>p</i> value
	No treatment	Treatment		
Identification score	0 (2.16)	−1.88 (3.74)	1.88 (0.28,3.48)	0.0224
TDI score	2.13 (5.85)	−4.51 (7.42)	6.63 (2.48,10.79)	0.0023

A positive difference implies a larger score at visit 1

suitability of this part of the Sniffin' Sticks test for use in a UK population. This is advantageous as it does not require the replacement of any of the existing odours in the identification test, merely an adaptation of the test software to include the new descriptors which are an important to set correctly (Gudziol and Hummel 2009). Our initial results from 30 healthy volunteers performing the task found that the five odours most commonly misidentified were lemon, liquorice, turpentine, garlic and apple. It was the distractors associated with these odorants that were changed, for example, sauerkraut was changed to pickle. In total, the descriptions of 8 distractors were changed (see Table 1) (Neumann 2012). In phase two of our study, the adapted version of the Sniffin' Sticks test was used to score patients; the untreated group showed no significant difference and a good reliability and hence showing stability of the test. The cultural adaptations we have made are now accepted as one of the many international cultural adaptations available for the Sniffin' Sticks software available from the Dresden Smell & Taste Clinic.

Limitations

In phase 1, participants were screened through a medical history and anterior rhinoscopy only; therefore, it is possible that underlying pathology was missed such as more discrete inflammatory disease, neurodegenerative disorders, mineral deficiencies, and other rare causes of olfactory loss; however, the first scores were all as expected for healthy subjects and it is unlikely that this was a significant limitation. There is a possibility that the improvement

in scores from test 1 to test 2 was due to a learning effect, but we believe this is highly unlikely due to the interval between the two tests and the adaptations made to the descriptors. The participants in phase 2 included those with various aetiologies and the period of time between the two tests performed by participants varied between 3 and 12 months. The number of participants in the non-treatment group was only 16. Additionally, 30 participants were lost to follow-up and did not complete the test twice and so could not be included in the full analysis. However, these participants did not differ in age, gender or baseline score from those who completed the study. Given the specialist nature of the main clinic in which the study was performed, many of the patients were not local and travelled from around the country to attend, hence the high dropout rate. East Anglia is not a culturally diverse part of the UK and so areas with significant ethnic groups may wish to refer to cultural normative values for reference countries; however, with patients coming from several parts of the UK, the participants were reasonably representative of the British population. An advantage is that the free download software mentioned above now contains various international settings, so the test can be adapted easily without changing the actual equipment. Although no formal power calculation was done a priori, the 95% confidence intervals show that the mean difference in TDI score for those participants not on treatment is lower than the recognised minimum clinically important difference in the TDI score (5.5 points) and the confidence interval for the reliability was also sufficient to enable the study to demonstrate a good reliability.

Generalisability

Our results reflect similar aforementioned validation studies performed in other countries where the Sniffin' Sticks test has been shown to work well following cultural adaptation. The mean scores are in keeping with the original German reference group where the means were around 13.5 (SD 1.6–1.9) in the corresponding age and sex groups (Hummel 2007). This now means the identification component of the Sniffin' Sticks test can be used for the assessment of olfactory disorders in a clinical setting in the UK without concern over any cultural biases and without a need to change any of the odours themselves; this also facilitates equivalence in multinational studies. One of the key advantages of the Sniffin' Sticks test is not only the comprehensive nature that the 3 parts of the test provide in assessing olfaction, but also that the threshold and discrimination components do not rely on prior knowledge of the odours and therefore any verbalisation around this (Gudziol 2006).

The test battery is readily available to purchase for clinical use, is reusable and has a shelf life of 12 months (threshold) and 18 months (discrimination and identification) and it is easily and quickly administered using the software available and is suitable for the outpatient environment. Initial cost of the extended set of Sniffin' Sticks is €918.09 (~£800) and for the complete refill set €546.21 (~£475) (<https://1485253724.jimdo.com/englisch/>); with the volume of patients seen in the Norfolk Smell & Taste Clinic (approximately 20 new referrals per month on average), this works out at about £3 per patient. If compared to the UPSIT test (Doty and Agrawal 1989) which apart from its limitation of being single use and only an identification test, it costs \$26.95 per patient (approximately £20). The Zurich Smell Diskettes are an alternative option but again have the limitation of being an identification only test; however, they do have data on sensitivity and specificity and are quick to use, but perhaps more suited to screening (Briner and Simmen 1999). Ultimately, the choice of test may depend on the setting with considerations given to ease of use, staffing requirements, throughput of patients, research requirements (Ta et al. 2021) and budget. To facilitate smell testing in a National Health Service clinic setting, an enhanced tariff has been agreed for referrals to account for the smell test to be administered by a trained member of nursing staff as part of the clinic visit. Coding for smell tests includes MI033 and MI049 available through the Clinical Coding & Schedule Development Group (Clinical Coding Schedule Development Group 2018).

Conclusion

This study has demonstrated validity of the revised odour distractors for a British version of the Sniffin' Sticks test. We have shown that the identification component of the Sniffin'

Sticks test can now be reliably used in British patients presenting with olfactory complaints to discriminate between normal olfaction and olfactory dysfunction and monitor the change in olfactory function in response to treatment.

Acknowledgements We would like to thank Jane Woods and Teresa Ferreira for their invaluable contribution to study recruitment and Fifth Sense for their support for the study.

Funding This study was funded by a pump priming grant from the British Otorhinolaryngology and Allied Sciences Research Society (formerly the Otorhinolaryngological Research Society).

Data Availability The data that support the findings of this study are available on request from the corresponding author.

Declarations

Conflict of Interest Professor Philpott is Director of Research and Medical Affairs on the board of trustees for the charity Fifth Sense.

Ethical Approval Ethical approval for the study was obtained from East of England, Cambridge East Research Ethics Committee (ref 14/EE/1010), and sponsored by the University of East Anglia. Strobe guidelines have been followed in reporting this study.

Informed Consent All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Boesveldt S et al (2008) Odour identification and discrimination in Dutch adults over 45 years. *Rhinology* 46(2):131–136
- Briner HR, Simmen D (1999) Smell diskettes as screening test of olfaction. *Rhinology* 37(4):145–148
- Clinical Coding & Schedule Development Group (2018) Available from: (<http://www.ccsd.org.uk/ccsdschedule/CodePrinciples?CodeID=8628&ctype=1>). Accessed 01.02.2021
- Doty RL, Agrawal U (1989) The shelf life of the University of Pennsylvania Smell Identification Test (UPSIT). *Laryngoscope* 99(4):402–404
- Eibenstein A et al (2005) Olfactory screening test: experience in 102 Italian subjects. *Acta Otorhinolaryngol Ital* 25(1):18–22

- Erskine SE, Philpott CM (2020) An unmet need: Patients with smell and taste disorders. *Clin Otolaryngol* 45(2):197–203
- Gudziol V et al (2006) Clinical significance of results from olfactory testing. *Laryngoscope* 116(10):1858–1863
- Gudziol V, Hummel T (2009) The Influence of Distractors on Odor Identification. *Arch Otolaryngol-Head Neck Surg* 135(2):143–145
- Hummel T (2018) Dresden Smell & Taste Clinic. Available from: <https://www.uniklinikum-dresden.de/de/das-klinikum/kliniken-polikliniken-institute/hno/forschung/interdisziplinaeres-zentrum-fuer-riechen-und-schmecken/neuigkeiten/downloads>. Accessed 01.02.2021
- Hummel T et al (1997) “Sniffin” sticks’: olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold. *Chem Senses* 22(1):39–52
- Hummel T et al (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 et al (2017) Position paper on olfactory dysfunction. *Rhinol Suppl* 54(26):1–30
- Kobal G et al (1996) “Sniffin” sticks”: screening of olfactory performance.” *Rhinology* 34(4):222–226
- Langstaff L et al (2019) Validation of the olfactory disorders questionnaire for English-speaking patients with olfactory disorders. *Clin Otolaryngol* 44(5):715–728
- Mackay-Sim A et al (2004) Australian norms for a quantitative olfactory function test. *J Clin Neurosci* 11(8):874–879
- Neumann C et al (2012) Validation study of the “Sniffin’ Sticks” olfactory test in a British population: a preliminary communication.” *Clin Otolaryngol* 37(1):23–27
- Philpott CM, Boak D (2014) The impact of olfactory disorders in the United Kingdom. *Chem Senses* 39(8):711–718
- Silveira-Moriyama L et al (2008) The use of smell identification tests in the diagnosis of Parkinson’s disease in Brazil. *Mov Disord* 23(16):2328–2334
- Silveira-Moriyama L et al (2009) Adapting the Sniffin’ Sticks to diagnose Parkinson’s disease in Sri Lanka. *Mov Disord* 24(8):1229–1233
- Ta NH, Gao J, Philpott C (2021) A systematic review to examine the relationship between objective and patient-reported outcome measures in sinonasal disorders: recommendations for use in research and clinical practice. *Int Forum Allergy Rhinol* 11(5):910–923. <https://doi.org/10.1002/alr.22744>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.