Time, Age, Gender, and Test Practice Effects on Children's Olfactory Performance: a Two-Year Longitudinal Study

Lenka Martinec Nováková^{1,2} • Jan Havlíček^{2,3}

Received: 20 August 2018 / Accepted: 11 February 2019 / Published online: 23 February 2019
© The Author(s) 2019

Abstract

Introduction Changes in olfactory perception observed in cross-sectional studies may not reflect actual ongoing change within individuals. The aim of the present study was to assess intra-individual as well as inter-individual variation in olfactory scores in pre-schoolers across five waves over a 2-year period.

Methods The participants were 157 children (79 boys) aged 5.8 ± 0.6 years at initial testing. We repeatedly examined the effects of time, age, gender, test practice, operationalised as the number of sessions attended and the intervals between them, and influence of school entry on identification, discrimination, and threshold Sniffin' Sticks scores. Data imputation was performed due to missing data.

Results In non-imputed data, odour identification and discrimination were higher in girls. More odours were also correctly identified by children who had attended fewer sessions in shorter intervals. In imputed data, in addition to these effects, odour identification and discrimination increased further into the study and were higher in children who were older at initial testing and those who had started attending school. Schoolchildren also had lower thresholds than pre-schoolers. However, both the significant and non-significant effects were generally small.

Conclusions We observed mainly small effects of gender and test practice on odour identification and discrimination, whereas intra-individual variation appeared only after data imputation.

Implications It is likely that olfactory development over time needs to be observed for longer than 2 years.

Keywords Abilities · Linear mixed models · Odour · Pre-school · Repeated measures · Smell · Sniffin' Sticks

Introduction

Inter-individual variability in olfactory performance is substantial and very well documented (e.g. Doty et al. 1984; Hummel et al. 2007b; Kobal et al. 2000). Nevertheless, there

are still knowledge gaps in our understanding of how this variability develops. Cross-sectional studies show that the most important demographic predictor is age (Doty and Kamath 2014; Hawkes and Doty 2009) although one must bear in mind that age itself or the passing of time is not what drives developmental changes across the human lifespan and should only be considered a proxy for the actual causal mechanisms. For instance, a child's olfactory performance improves with age, but this progress is most likely driven by growing experience with odours and improving linguistic abilities (Monnery-Patris et al. 2009; Stevenson et al. 2007), broadening working memory span (Larjola and von Wright 1976), improving recognition memory (Frank et al. 2011; Hvastja and Zanuttini 1989), changes in nasal aerodynamics or more effective inhalation of olfactory stimuli (Mennella and Beauchamp 1992). To this day, there are very few longitudinal studies on olfactory development in general and children in particular. Thus, our knowledge derives almost exclusively from cross-sectional studies even on the level of this proxy

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12078-019-09260-0>) contains supplementary material, which is available to authorized users.

✉ Lenka Martinec Nováková
lenka.novakova@nudz.cz; lenka.novakova@fhs.cuni.cz

¹ Department of Anthropology, Faculty of Humanities, Charles University, U Kříže 8, 158 00 Prague 5 - Jinonice, Czech Republic

² National Institute of Mental Health, Topolová 748, 250 67 Klecany, Czech Republic

³ Department of Zoology, Faculty of Science, Charles University, Viničná 7, 128 44 Prague 2, Czech Republic

(Monnery-Patris et al. 2009; Renner et al. 2009; Richman et al. 1992; Richman et al. 1995a; Schriever et al. 2014). Besides, there is an unsettled general debate about the extent to which variation observed in cross-sectional studies reflects actual ongoing changes within individuals or is caused by methodological artefacts. These may include, for example, discrepancies between cross-sectional and longitudinal modelling of age-related effects, that is, whether they are estimated by performing ordinary least squares on all data or by summarising each individual's data by a slope and then averaging these slopes. To be specific, when the age effect is modelled as linear when it is, in fact, non-linear, linear trends estimated cross-sectionally and longitudinally may differ (e.g. Louis et al. 1986). Furthermore, in olfactory research, children have been an underrepresented cohort and often treated indiscriminately as a single demographic group regardless of age (e.g. Monnery-Patris et al. 2009; Renner et al. 2009). This is despite the number of changes that take place over childhood in cognition, self-regulatory capacity and other factors that may affect olfactory performance considerably (e.g. Goswami 2011). Thus, a more discerning approach to developmental staging in olfactory studies must be embraced. This knowledge gap warrants further, longitudinally oriented research.

An area of study that calls for a longitudinal approach is the gender variation in olfaction and how it develops. A great body of literature reports the superior performance of women on odour identification tests (e.g. Hummel et al. 2007b; Larsson et al. 2004), although not uniformly so (e.g. Kobal et al. 2000). The results for odour discrimination and memory are mixed, with studies indicating small, if any, differences (e.g. Choudhury et al. 2003; Kobal et al. 2000; Öberg et al. 2002). Research assessing olfactory thresholds revealed either no gender difference or greater sensitivity of women to some odorants, but the differences were also small (e.g. Cometto-Muniz and Abraham 2008; Chopra et al. 2008; Koelega and Köster 1974). Some researchers argue that gender differences are present at birth (Doty and Laing 2015), but this proposition is based on cross-sectional data, as studies observing a single sample of children over any period of time are missing. Moreover, some studies indicate that gender variation in olfaction may not be reliably present in infancy and pre-school (Bastos et al. 2015; Cameron and Doty 2013; Martinec Nováková and Vojtušová Mrzílková 2016a, 2016b; Saxton et al. 2014; Schriever et al. 2014). Also, there is little clarity about the magnitude of the gender effect on olfaction in early childhood. These questions invite testing in longitudinal studies.

Studies in which repeated measurements are taken require controlling for test practice effects (that is, remembering the previous responses and proficiency in handling the research tasks encountered before; see Heilbronner et al. 2010 for a review). Repeated olfactory testing itself can be seen as an opportunity for the participant for perceptual learning, which

is a relatively long-term, learned increase in perceptual acuity (Wilson et al. 2009). Olfactory performance benefits from repeated odour exposures, which can be relatively brief (Dalton et al. 2002; Wysocki et al. 1989). It is, therefore, necessary to account for the number of sessions attended and their spacing. Another experiential effect to consider is school entry. Formal education modulates and facilitates performance on cognitive tasks across the human lifespan (Ceci 1991; Steinbrink et al. 2014); for review, see e.g. Rogoff (1981); Rosselli and Ardila (2003). Nevertheless, its potential effect on children's olfactory performance, in general, has been largely overlooked in research on the chemical senses.

Here, we repeatedly tested children's olfactory abilities across five waves over a 2-year period. We focused on pre-school children who were aged between 3.5 and 7 years at the beginning of the study. Even though some of the previous studies show that 6 years is the youngest age at which it is feasible to carry out psychophysical olfactory tests (e.g. Hummel et al. 2007a), in our study, data from younger children did not manifest themselves as outliers or influential cases. This age group is also of special interest because of the possible effect of experiential factors related to school entry. We hypothesised that olfactory scores would be higher in older children and would increase with time, that girls would outperform boys and that children who attended more testing occasions in shorter intervals and schoolchildren would outperform those who participated in fewer sessions over a wider time span, and pre-schoolers. As this was an initial study, we did not delve into the actual factors behind the proxies of age and gender.

Materials and Methods

Participants

The participants were 157 children of Czech origin: 79 boys, mean age at study commencement 5.87 ± 0.71 years, range 3.58–7.08 years, and 78 girls, mean age at study commencement 5.75 ± 0.57 years, range 4.42–7.00 years. At the start of the study, the children attended one of six public mixed-sex kindergartens in Prague and its suburbs. The kindergartens were attended by children from varied social backgrounds. Kindergarten principals were contacted via telephone, e-mail, and in person to inform them about the planned study. Those who had provided permission to perform the study on the kindergarten's premises were asked to pass the information on to the teachers, who distributed leaflets to children's parents. We kept the e-mail addresses and phone numbers the parents had provided to contact them later with an invitation for their children to take part in subsequent sessions. After the given child entered school, the session took place either on the school premises or at our department. The sessions were

scheduled in approximately 6-month intervals. Namely, data was collected in the late autumn and early winter of 2010 (Wave 1), late spring and early summer of 2011 (Wave 2), late autumn and early winter of 2011 (Wave 3), late spring and early summer of 2012 (Wave 4), and late autumn and early winter of 2012 (Wave 5). Half the children were recruited and tested for the first time during Wave 1 (48.41%, $N = 76$, 33 boys), whilst the other half entered the study and were first tested during Wave 2 (51.59%, $N = 81$, 47 boys). Because on the first testing occasion, these two groups did not differ in age, $t(153.72) = 0.45$, $p = 0.65$ or baseline olfactory performance (identification: $t(153) = -1.17$, $p = 0.24$; discrimination: $t(153) = -1.96$, $p = 0.30$; threshold: $t(138) = 1.12$, $p = 0.28$), group membership was disregarded. All the children were last tested during Wave 5, meaning that there was data from the maximum of five testing occasions available for those who entered the study during Wave 1, but a maximum of four for those who were recruited during Wave 2.

Due to the longitudinal nature of the study, missing data was a concern. However, the majority of the missing data was the result of participant absence on the day of data collection rather than attrition from the study. The number of tests administered was 76 (100% pre-schoolers) in Wave 1, 135 (100% pre-schoolers) in Wave 2, 91 (49% pre-schoolers) in Wave 3, 34 (94% pre-schoolers) in Wave 4, and 81 (100% school age) in Wave 5. For logistic reasons, we were unable to collect data from school children in Wave 4, resulting in lower N , a greater proportion of pre-schoolers and a drop in mean age within this wave. Participant-wise, complete data was available from two children recruited to the study during Wave 1 and 21 children recruited during Wave 2.

Thus, 85.35% of participants had some missing data due to skipping at least one testing occasion. Online Resource 1 shows that prior to Wave 5, children with complete and incomplete data did not differ on any measures. Because the children entered the study at different times (Wave 1 vs. Wave 2) and the pattern of skipping the sessions was highly individual (i.e. the total number of testing occasions already attended by the given wave varied), experiential effects on the children's reports could not be simply operationalised in terms of the waves. For example, of the 39 children recruited to the study during Wave 1 who participated in Wave 5, which would have been, in theory, their fifth testing occasion, in reality only two children had actually attended all four of the previous sessions, whilst 19 children had been tested only three times and 16 children only twice, and two had thus far attended only a single session. Therefore, the level of individual experience with the tasks within the specific wave was expressed as the cumulative total of sessions actually attended by the given child thus far. Only a single session was attended by 31 children (19.7%). The number of children who attended two, three, or four sessions was $N = 33$ (21.0%), 51 (32.5%), and 40 (25.5%), respectively. Additionally, we also took into

account the time interval since any previous testing. For instance, for children recruited to the study during Wave 1 who participated in Wave 5, this time interval ranged from 5 to 25 months. The number of children (boys, girls, and total) participating in each wave is given in Table 1, along with the descriptive data on their age, experience level, and olfactory performance. Figure 1 shows the frequencies of boys and girls participating within Waves 2–5 relative to the cumulative total of sessions already attended (newly recruited children within Wave 2 are not shown).

Binomial tests showed that the proportion of boys and girls participating in each wave was not significantly different from the 50:50 ratio: $p = 0.30$ (Wave 1), $p = 0.39$ (Wave 2), $p = 1.00$ (Wave 3), $p = 0.18$ (Wave 4), and $p = 0.65$ (Wave 5), nor did the proportions of boys and girls differ across the waves, $p = 0.81$. T -tests confirmed that children participating in each wave were significantly older than those who attended the preceding one: $p = 0.003$ (Wave 2 vs. Wave 1), $p < 0.001$ (Wave 3 vs. Wave 2), and $p < 0.001$ (Wave 5 vs. Wave 4), with the exception of the age difference between Wave 3 and 4, which was non-significant ($p = 0.93$). This means that the effect of different children participating in the individual waves was in general overshadowed by the effect of passing time. This also held true for boys and girls, as well as the recruitment groups analysed separately. The boys and girls within each group participating in each of the waves did not differ in age.

Olfactory Measures

General Considerations

The Sniffin' Sticks test (Hummel et al. 1997), manufactured by Burghart Messtechnik GmbH, was used to assess odour identification, discrimination, and threshold. This is one of the most widely used tests of olfactory performance, based on pen-like odour-dispensing devices. The Sniffin' Sticks test has been widely used by clinicians and researchers across Europe to test olfactory abilities in adults (Hummel et al. 2007b) and children (Ferdenzi et al. 2008b; Renner et al. 2009; Sorokowska et al. 2015), including Czech ones (Martinec Nováková et al. 2018a; Martinec Nováková et al. 2018b; Martinec Nováková et al. 2015; Martinec Nováková and Vojtušová Mrzílková 2016a, 2016b; Saxton et al. 2014).

Odour Identification

The 16-item "blue" identification test consists of odours familiar to the general European population, namely orange, leather, cinnamon, mint, banana, lemon, liquorice, turpentine, garlic, coffee, apple, clove, pineapple, rose, anise, and fish (exact chemicals are not specified by the manufacturer). In the original version of the test, cued identification is employed in which participants select a verbal label for the target odour

Table 1 Mean $\pm$ SD (range) and valid N of the children's age, interval in months since any previous testing occasion, olfactory scores (identification, discrimination, and threshold), and median of sessions already attended across the individual waves in boys, girls, and the total sample.

	Boys	Girls	Total	Gender difference	
				<i>t</i> (df)	<i>p</i>
Wave 1					
Age (months)	70.59 ± 8.40 (43.00–82.00), 33	69.61 ± 6.07 (53.00–80.00), 43	70.03 ± 7.14 (43.00–82.00), 76	0.59 (74)	0.55
Identification	7.03 ± 1.94 (2.00–10.00), 33	7.42 ± 1.94 (3.00–11.00), 43	7.25 ± 1.94 (2.00–11.00), 76	− 0.86 (74)	0.39
Discrimination	8.33 ± 2.51 (2.00–12.00), 33	9.09 ± 2.28 (4.00–14.00), 43	8.76 ± 2.39 (2.00–14.00), 76	− 1.38 (74)	0.17
Threshold	4.87 ± 3.17 (1.00–11.50), 28	6.37 ± 3.36 (1.50–14.00), 41	5.76 ± 3.35 (1.00–14.00), 69	− 1.87 (67)	0.07
Wave 2					
Age (months)	73.75 ± 8.98 (54.00–90.00), 74	73.01 ± 8.22 (54.00–87.00), 63	73.41 ± 8.62 (54.00–90.00), 137	0.50 (135)	0.62
Sessions attended	0	1	0		
Interval since last testing	6.82 ± 0.64 (6.00–8.00), 27	6.86 ± 0.67 (6.00–8.00), 30	6.84 ± 0.65 (6.00–8.00), 57	− 0.25 (55)	0.80
Identification	7.35 ± 1.88 (2.00–11.00), 71	7.83 ± 1.90 (3.00–11.00), 59	7.57 ± 1.90 (2.00–11.00), 130	− 1.44 (128)	0.15
Discrimination	9.51 ± 2.36 (4.00–15.00), 72	10.02 ± 2.47 (3.00–15.00), 63	9.75 ± 2.41 (3.00–15.00), 135	− 1.21 (133)	0.23
Threshold	5.71 ± 2.69 (1.00–14.00), 59	5.00 ± 2.57 (1.00–12.00), 57	5.36 ± 2.65 (1.00–14.00), 116	1.46 (114)	0.15
Wave 3					
Age (months)	77.76 ± 9.92 (60.00–96.00), 46	77.74 ± 9.03 (60.00–94.00), 45	77.75 ± 9.44 (60.00–96.00), 91	0.01 (89)	0.99
Sessions attended	1	1	1		
Interval since last testing	6.75 ± 1.31 (5.00–10.00), 46	7.13 ± 2.05 (5.00–14.00), 44	6.94 ± 1.71 (5.00–14.00), 90	− 1.05 (88)	0.30
Identification	7.89 ± 2.00 (4.00–12.00), 46	8.11 ± 1.80 (4.00–12.00), 45	8.00 ± 1.90 (4.00–12.00), 91	− 0.55 (89)	0.58
Discrimination	10.74 ± 2.27 (5.00–15.00), 46	11.16 ± 2.66 (5.00–16.00), 45	10.95 ± 2.47 (5.00–16.00), 91	− 0.80 (89)	0.42
Threshold	5.62 ± 2.26 (2.00–11.50), 41	5.60 ± 2.15 (2.00–9.00), 39	5.61 ± 2.19 (2.00–11.50), 80	0.04 (78)	0.97
Wave 4					
Age (months)	78.65 ± 7.07 (67.00–94.00), 22	75.85 ± 3.65 (68.00–83.00), 13	77.61 ± 6.12 (67.00–94.00), 35	1.54 (32.60)	0.13
Sessions attended	2	2	2		
Interval since last testing	7.61 ± 1.76 (5.00–12.00), 22	6.96 ± 2.04 (4.00–12.00), 13	7.37 ± 1.86 (4.00–12.00), 35	0.99	0.33
Identification	7.09 ± 1.72 (4.00–9.00), 22	7.67 ± 1.23 (6.00–10.00), 12	7.29 ± 1.57 (4.00–10.00), 34	− 1.02 (32)	0.31
Discrimination	9.72 ± 2.82 (5.00–13.00), 18	11.85 ± 1.72 (8.00–15.00), 13	10.61 ± 2.62 (5.00–15.00), 31	− 2.59 (28.37)	0.02
Threshold	4.63 ± 2.10 (2.00–7.50), 8	4.79 ± 1.82 (2.00–7.00), 7	4.70 ± 1.91 (2.00–7.50), 15	− 0.16 (13)	0.88
Wave 5					
Age (months)	91.36 ± 8.86 (73.00–106.00), 38	90.87 ± 7.83 (76.00–105.00), 43	91.10 ± 8.28 (73.00–106.00), 81	0.26 (79)	0.79
Sessions attended	2	2	2		
Interval since last testing	11.00 ± 4.56 (5.00–19.00), 37	11.65 ± 5.18 (5.00–25.00), 43	11.35 ± 4.88 (5.00–25.00), 80	− 0.59 (78)	0.56
Identification	7.47 ± 1.59 (3.00–10.00), 38	8.09 ± 1.74 (5.00–12.00), 43	7.80 ± 1.69 (3.00–12.00), 81	− 1.66 (79)	0.10
Discrimination	10.16 ± 2.63 (5.00–16.00), 38	11.12 ± 2.31 (5.00–15.00), 43	10.67 ± 2.50 (5.00–16.00), 81	− 1.75 (79)	0.08

Table 1 (continued)

	Boys	Girls	Total	Gender difference	
				<i>t</i> (df)	<i>p</i>
Threshold	6.13 ± 1.98 (2.00–10.00), 34	6.24 ± 1.89 (2.50–9.50), 41	6.19 ± 1.92 (2.00–10.00), 75	− 0.25 (73)	0.80

from a candidate list of four alternatives. The resulting score is the sum of correct answers (maximum of 16).

In the present study, the test was adapted to children who could not yet read or were only beginning to learn how to read. This was done by presenting both the targets and distractors in the form of colour pictures instead of verbal labels. Selection of pictures was based on a pilot study ($N = 31$, 16 boys, 5.85 ± 0.89 years) carried out in one of the kindergartens. It was split over two sessions. In the first session, children were interviewed about their understanding of the individual verbal descriptors, both the targets and the distractors. Prompted by the question “What do you think...is, what does it look like, when or where can you find it?,” children were encouraged to elaborate on the meanings of the individual labels. Based on these accounts, images depicting items most frequently associated with the given verbal label were selected. For instance, since the majority of children tended to associate “mint” with

chewing gum and mints, a picture of confectionery was used (see e.g. Bastos et al. 2015 for a similar approach). These interviews had also revealed that a majority of children expressed uncertainty about what most of the spices, menthol, and turpentine were, looked like and smelled like, and in which foods or products they could be found. Therefore, items 3, 8, 12, and 15 (targets: cinnamon, turpentine, clove, and anise) were excluded from the odour set. In the following session of the pilot study, four colour pictures arranged in a 2×2 square on an A4 page were presented for each item. The researcher pointed at each picture, one by one, and encouraged the child to describe it. On average, children produced the veridical label in 88% of cases and a near miss in 12% of cases. The veridical label is the commonly used name of the given stimulus (Dubois and Rouby 2002). The children often started off by explaining where and when they last encountered the depicted item. In that case, the researcher would

Fig. 1 The absolute frequencies of boys and girls participating within Waves 2–5 relative to the cumulative total of sessions already attended

engage in a conversation with the child to see whether he or she could produce the veridical label in the end. A near-miss referred to an incorrect label, though one that was precise and that represented an item readily confusable with the stimulus. Specifically, children tended to confuse blackberry with raspberry or blueberry (item 1), chives with parsley and fir with spruce (item 4), grapefruit with orange or clementine (item 6), peach with apricot (items 6, 11, and 13), wine with beer (item 10), and chamomile with daisy (item 14). For at least some of the items, there is evidence for an age-related increase in familiarity (e.g. Noll et al. 1990). Hence, some near-misses arose more as a result of children's limited experience with certain items, such as alcoholic beverages, not the depictions per se. The final set of pictures employed in the study has been published elsewhere (see Supplementary Material to Martinec Nováková and Vojtušová Mrzílková 2016a). Results of the pilot study, given in Online Resource 2, show that the scores were close to those reported in studies on children's Sniffin' Sticks performance that were available at the time.

In the study, at the beginning of each trial, children were first asked to describe each picture and were given immediate feedback about the correct answer. Stimuli were presented one by one, in the order recommended by Hummel (2004). The researcher removed the cap of the pen and placed the tip approximately 2 cm in front of both nostrils during several respiratory cycles (at least 5 s). At the same time, she asked the child to choose a label for the given odour by pointing at one of the four pictures. The interval between odour presentations was between 20 and 30 s to prevent olfactory adaptation. The theoretical score range was 0–12, with higher scores indicating better identification performance.

Odour Discrimination

The pilot study (see Online Resource 2 for details) showed that children's discrimination scores came close to those reported by Kobal et al. (2000), Hummel et al. (2007b), Hummel et al. (2007a), and Ferdenzi et al. (2008b), albeit those studies were carried out in broader age groups. Hence, no alterations were made to the discrimination test aside from masking the colour bands on the Sniffin' Sticks pens with opaque Sellotape because most children participating in the pilot study found the blindfold uncomfortable. The test of odour discrimination assesses the degree to which an individual can differentiate between odours in suprathreshold concentrations. The set comprises 16 triplets of odourised pens (marked with a blue, green, and red colour band), of which two are identical, and the participant is asked to indicate the odd one, which is always the green one. The odorants used in the test and the order of presentation (which was followed) are given in Hummel et al. (1997) and in Online Resource 3c. Presentation of triplets was separated by circa 20 s. The score is the total of correct trials (0–16), with higher scores

indicating a better odour discrimination ability. Prior to the first trial, the task was demonstrated using three colour pencils, of which two were of the same colour and the third one was different, explaining to the children that they had to identify the odd pen by smell, not colour. Next, task comprehension was determined by presenting the child with another three pencils of different colours, upon which they were reminded once again about the olfactory, not visual, nature of the task.

Odour Threshold

The olfactory threshold refers to the minimum concentration of a tested odorant that an individual is able to reliably differentiate from a blank sample. The set employed in the present study consisted of 16 dilution steps of n-butanol (targets), each of which formed a triplet with two blanks. As recommended by Hummel et al. (1997), a single-staircase, three-alternative forced-choice (3-AFC) method was used, in which, starting with the lowest concentration (dilution number 16), an ascending (low to high concentration) series of even-numbered triplets was presented, with successful trials prompting another presentation of the same triplet in a random order. Two successful trials in a row marked a turning point; starting with the nearest lower concentration, a descending series of triplets was presented until the child failed to detect the target. This marked a reversal towards the higher concentrations and, starting with the next higher concentration, an ascending series of triplets was presented until two correct trials occurred, marking another reversal. The testing was finished after a total of seven reversals was reached. To sustain the children's attention throughout the assessment, they were rewarded with a candy for each response, regardless of whether it was correct or not. They were not given any feedback on their performance during or after the test. The threshold score was computed as the arithmetic mean of the dilution number at the last four reversals. Ranging from 1 to 16, higher scores indicate greater olfactory sensitivity (i.e. lower threshold).

Procedure

Parents and teachers were instructed to only encourage their children to attend the testing sessions, scheduled between 9 a.m. and 3 p.m., when in good respiratory health. We relied on parental reports and absence of any condition potentially affecting the sense of smell was not further verified. Testing took place in a secluded, well-ventilated room without strong ambient odours. Firstly, children were briefly familiarised with the tasks, which were presented as a game, and ensured that they could stop or quit at any time. The order of the olfactory tests was randomised across children. However, within the olfactory tests, the stimuli were presented in the order recommended by Hummel et al. (1997). Since this study was part of a larger project (see Martinec Nováková et al.

2018a; Martinec Nováková et al. 2018b for other studies on this sample), the children were also interviewed about their olfactory behaviours using the Children's Olfactory Behaviours in Everyday Life (COBEL) questionnaire (Ferdenzi et al. 2008a). It includes questions designed to evaluate self-reported awareness and reactivity to odours in significant everyday contexts: food (e.g. "When you smell a food odour, do you try to guess for fun what it is (never/sometimes/often)?"), social (e.g. "Do you happen to smell parts of your body (never/sometimes/often)?"), and environment (e.g. "Imagine someone is smoking next to you. Do you care/like or not like/love or hate this odour?"). The sheer number of the various olfactory tests and the interview presented a cognitive load that could only be alleviated by splitting them over two sessions. Therefore, each child was tested on two consecutive days or within a week at the very latest. Each session took circa 30 min. Parents or teachers were never present in the room during the testing session.

Statistical Analysis

SPSS 24.0 (IBM Corp 2016) was used to carry out the majority of analyses except for the calculations of Cohen's f^2 , which were performed with the SAS University Edition software (SAS 2017). Plots were produced in SPSS and R 3.4.0 (R Development Core Team 2008).

The sample size was calculated using formulae for continuous outcomes given in Kim and Seo (2013). Data in a cohort that was close to the age group we planned to recruit had only been reported by Ferdenzi et al. (2008b) for children aged 7 to 11. Because information was only available on gender differences in odour identification, we calculated a sample size for this effect. At a significance level (α) of 0.05 and power ($1-\beta$) of 0.6, which is typical of studies published in major psychology journals, a sample size required per group in each wave was $N=40$. Despite the high dropout rates, we were able to achieve those sizes, except for Wave 4.

The normality of the raw data was checked, firstly, by visually examining the individual histograms of all relevant variables, secondly, by producing skewness and kurtosis values and their respective standard errors, from which z-scores were computed and compared to the value of 1.96, as suggested by Field (2005), and, thirdly, with multiple Shapiro-Wilk's W tests. Two approaches to analysis of longitudinal data were adopted. Both utilised linear mixed models (LMM), were run using the SPSS syntax MIXED command and yielded very similar results. The first data analytic strategy consisted in fitting individual growth curve (IGC) models. In so doing, we followed the procedure recommended by Shek and Ma (2011). One of the advantages of IGC models is that they allow the irregularity of number and spacing of waves by means of a time-structured predictor ("time") (Singer and Willett 2003). Thus, at Wave 1, the values of time were set

at 0, and the number of months from the date of data collection within Wave 1 was calculated for each subsequent wave (i.e. Waves 2–5). In order to be able to perform these calculations for the children who did not take part in Wave 1 and were first tested during Wave 2, for these participants this date was set to November 1, which was the average date of data collection in Wave 1. As an alternative, an arbitrary date in the past was established, at which time was set at 0 and the number of months was calculated from this date. This approach to treating the time variable led to very similar results. The continuous variables of initial age and time interval since any previous testing were treated by a grand mean centring method (i.e. by subtracting the mean, which is generally recommended in order to simplify the interpretation of the results (Hox 2002)). Next, following the strategy suggested by Singer and Willett (2003), several models were fitted and then compared by means of $-2 \log$ likelihood (i.e. a likelihood ratio test/deviance test) and Akaike Information Criterion (AIC, "smaller is better") in order to select the best model. Namely, to compare models, we calculated delta AIC (Δ_i) as follows: $AIC_i - AIC_{\min}$, where AIC_i is the AIC value for model i , and $AIC_{\min}$ is the AIC value of the "best" model. Then we followed the rule of thumb, whereby a $\Delta_i < 2$ indicates substantial evidence for the given model, values between 3 and 7 suggest that the model has considerably less support, whilst a $\Delta_i > 10$ says that the model is very unlikely (Burnham and Anderson 2002). Firstly, two unconditional models were generated to examine mean differences in the outcome variable across individuals and to compare the fit of models estimated by means of the restricted maximum likelihood method (REML; default option) and the maximum likelihood (ML) method. The methods yielded similar model fits ($\Delta_i \leq 1$), therefore the default REML method was used to estimate all subsequent models. Secondly, an unconditional growth model was tested, which served as a baseline model to explore whether the growth curves were linear or curvilinear, and thirdly, two higher-order polynomial models (quadratic and cubic, respectively) were fitted to investigate whether the rate of change accelerated or decelerated across time. In this way, we found that models including only the linear term were by far superior to those with the quadratic and cubic growth curve parameters ($\Delta_i > 10$). Fourthly, a conditional model was formed to determine whether the variables of initial age, gender, the cumulative total of sessions actually attended by the given child thus far, the time interval since any previous testing, and kindergarten/primary school attendance were related to the growth parameters (i.e. initial status and linear growth). It transpired that the best model fit was obtained when only the main effects of all of these independent variables (IVs) were retained in the model (Δ_i in the order of hundreds). Thus, the interactions of time and IVs, as well as those among the IVs were omitted from the subsequent models. Finally, several covariance structure models were explored to assess the error

covariance structure of the longitudinal data, whereby we determined that there were several, which yielded the best model fit, namely heterogeneous compound symmetry (CSH), diagonal (DIAG), Huynh-Feldt (HF), unstructured (UN), and variance components (VC). Results using the UN model are reported but the other models yielded similar results both in terms of statistical significance and effect size. The intercept and linear slope were allowed to vary across individuals. Missing data was handled through pairwise/listwise deletion. This procedure was followed to model the effects on each of the three olfactory scores (identification, discrimination, and threshold). For *t*-tests, a standardised measure of effect size, Cohen's *d*, was calculated after Cohen (1988). It is the difference between two means expressed in units of standard deviations. Hence, when $d = 1$, the two groups' means differ by one standard deviation. For mixed models, Cohen's f^2 was computed using SAS PROC MIXED according to Selya et al. (2012). Cohen's f^2 for a given IV is a ratio of the proportion of the variance in the DV uniquely explained by the IV to the proportion of the variance in the DV unexplained by any variable in the model. Global effect sizes across the waves (i.e. for the overall model) are reported, as well as local ones within the individual waves. According to Cohen's (1988) guidelines, $f^2 \geq 0.02$, $f^2 \geq 0.15$, and $f^2 \geq 0.35$ represent small, medium, and large effect sizes, respectively. Cohen's $f^2 < 0.02$ are below the recommended minimum effect size representing a "practically" significant effect for social science data (Ferguson 2009), which is why the exact values are not reported. Ninety-five percent confidence intervals (95% CIs) for the estimates were taken from the SPSS output. CIs can be interpreted in various ways (e.g. Cumming 2014). Here, we favoured the one stating that a 95% CI is an 83% prediction interval for the effect size estimate of a replication experiment (Cumming and Maillardet 2006).

The other data analytic strategy involved a repeated-measures analysis with time-dependent (time-varying) covariates. Waves represented the repeated-measures effect, gender was treated as a fixed factor and the child's age on the given testing occasion, the cumulative total of sessions actually attended by the given child thus far, the time interval since any previous testing, and kindergarten/primary school attendance as individual-level covariates that were also measured across the waves. Again, the model with the best fit included only the main effects of all the IVs. The residual covariance matrix structure was diagonal with heterogeneous variance, which is the default covariance structure for repeated effects. The model was again run separately for each of the three olfactory scores. Since the results matched those obtained with the first analytic strategy both in terms of statistical significance and effect size, they are not reported in the paper.

To model the relationships on a larger dataset, we then reran the analyses on imputed data ($N = 587$). Data imputation was performed with the missForest package (Stekhoven and

Bühlmann 2012) available from the Comprehensive R Archive Network (CRAN) and run in the R (R Development Core Team 2008). Recommended particularly for conducting multiple imputation of mixed data (numeric and factor variables in one data frame) (Starkweather 2014), it has been compared to other imputation methods and found to have the least imputation error for both continuous and categorical variables and the smallest prediction difference (error) (Waljee et al. 2013). Default settings were used. Imputation of the DVs was only carried out when the IVs were available. Explorations of imputed data revealed that the only difference from non-imputed data was in the link between the cumulative total of sessions attended and the time interval from any previous testing (non-imputed: $r = 0.08$ [$-0.03, 0.19$], $p = 0.23$, $N = 262$, $R^2 < 0.01$; imputed: $r = 0.27$ [$0.18, 0.33$], $p < 0.001$, $N = 587$, $R^2 = 0.07$).

Finally, as some of the previous studies suggested that olfactory testing only becomes meaningful in children around 6 years of age (e.g. Hummel et al. 2007a), we arbitrarily filtered out data on children under 50, 60, and 70 months of age respectively and reanalysed the data. It transpired that the results did not change in terms of statistical significance or effect size. Also, these olfactory data of these children did not manifest themselves as outliers or influential cases. Therefore, they were retained in the study.

Results

Individual Growth Curve Models: Influence of Time, Age, Gender, and Experiential Factors on Olfactory Measures

Non-imputed Data

The sample sizes for non-imputed data were $N = 246$ for odour identification and discrimination, and $N = 204$ for odour threshold. As detailed in Table 2, higher odour identification scores were linked to being a girl, having attended fewer testing sessions, and shorter intervals between the sessions. Across the waves, global sizes of all the effects, both significant and non-significant, barely qualified as small (Cohen's $f^2 < 0.02$). Within the waves, the effect sizes were also small, varying between < 0.02 and 0.08 . To explore the effect of the number of sessions already attended, we ran paired-sample *t*-tests comparing baseline and session 2, session 2 and 3, 3 and 4, and 4 and 5 (i.e. baseline scores and scores in children who had attended one, two, three, or four sessions, irrespective of wave). These revealed a significant initial increase in odour identification scores between baseline (7.33 ± 1.91) and session 2 (7.82 ± 1.94), $t(118) = 2.47$, $p = 0.02$, Cohen's $d = 0.25$. This rise was followed by fluctuations which were non-significant, with a mean difference ranging

Table 2 F-statistics, *p* values, estimates, 95% confidence intervals (95% CIs), and effect sizes global (i.e. across waves) and local (i.e. within waves) for the effects of time, age at initial testing, gender, cumulative total of sessions attended, interval since last testing, and kindergarten/school attendance on olfactory measures in non-imputed and imputed data

	Time	Initial age	Gender	Sessions attended	Interval	Pre-school/school attendance
Non-imputed data						
Identification						
F	2.50	1.99	4.61	5.50	3.82	3.24
(df ₁ , df ₂)	(1, 208.48)	(1, 157.34)	(1, 118.11)	(1, 177.48)	(1, 186.86)	(1, 136.92)
<i>p</i>	0.12	0.16	0.03	0.02	0.05	0.07
Estimate	0.09	0.03	0.60	− 0.90	− 0.12	0.55
95% CI	[− 0.02, 0.21]	[− 0.01, 0.07]	[0.05, 1.15]	[− 1.66, − 0.14]	[− 0.24, 0]	[− 0.05, 1.15]
Cohen's <i>f</i> ²						
Global	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02
Wave 1	< 0.02	0.02	< 0.02	< 0.02	< 0.02	< 0.02
Wave 2	< 0.02	< 0.02	< 0.02	< 0.02	0.08	< 0.02
Wave 3	< 0.02	0.03	< 0.02	< 0.02	0.02	< 0.02
Wave 4	0.03	0.04	0.03	0.02	0.04	0.04
Wave 5	< 0.02	< 0.02	0.03	0.05	0.02	< 0.02
Discrimination						
F	1.09	0.09	4.39	3.31	< 0.01	0.51
(df ₁ , df ₂)	(1, 203.60)	(1, 162.54)	(1, 115.69)	(1, 165.33)	(1, 186.98)	(1, 150.28)
<i>p</i>	0.30	0.76	0.04	0.07	0.97	0.48
Estimate	− 0.07	− 0.07	0.47	− 0.24	0.24	0.66
95% CI	[− 0.33, 0.19]	[− 0.17, 0.03]	[− 0.82, 1.76]	[− 2.10, 1.61]	[− 0.04, 0.52]	[− 0.39, 1.71]
Cohen's <i>f</i> ²						
Global	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02
Wave 1	< 0.02	0.02	< 0.02	< 0.02	< 0.02	< 0.02
Wave 2	0.02	0.02	< 0.02	< 0.02	0.29	< 0.02
Wave 3	< 0.02	< 0.02	< 0.02	< 0.02	0.09	< 0.02
Wave 4	0.03	0.03	0.13	< 0.02	0.03	< 0.02
Wave 5	0.06	0.04	0.05	0.03	0.01	< 0.02
Threshold						
F	< 0.01	1.36	0.26	0.23	0.32	0.05
(df ₁ , df ₂)	(1, 145.92)	(1, 108.96)	(1, 90.91)	(1, 111.20)	(1, 111.68)	(1, 105.81)
<i>p</i>	0.96	0.25	0.61	0.63	0.58	0.83
Estimate	− 0.01	0.03	− 0.15	0.26	0.05	0.11
95% CI	[− 0.16, 0.15]	[− 0.02, 0.08]	[− 0.82, 0.51]	[− 0.70, 1.23]	[− 0.11, 0.20]	[− 0.87, 1.10]
Cohen's <i>f</i> ²						
Global	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02
Wave 1	< 0.02	0.02	0.03	< 0.02	< 0.02	< 0.02
Wave 2	0.02	0.10	0.19	< 0.02	0.36	< 0.02
Wave 3	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02
Wave 4	0.11	0.07	0.08	0.10	0.09	0.10
Wave 5	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02
Imputed data						
Identification						
F	7.96	11.20	5.50	12.27	18.82	22.58
(df ₁ , df ₂)	(1, 522.51)	(1, 161.32)	(1, 137.62)	(1, 533.27)	(1, 473.16)	(1, 372.27)
<i>p</i>	0.005	0.001	0.02	< 0.001	< 0.001	< 0.001
Estimate	0.07	0.03	0.30	− 0.60	− 0.13	0.60
95% CI	[0.02, 0.12]	[0.01, 0.05]	[0.05, 0.56]	[− 0.93, − 0.26]	[− 0.20, − 0.07]	[0.35, 0.85]

Table 2 (continued)

	Time	Initial age	Gender	Sessions attended	Interval	Pre-school/school attendance
Cohen's f^2						
Global	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02	0.06
Wave 1	< 0.02	0.02	< 0.02	< 0.02	< 0.02	< 0.02
Wave 2	< 0.02	< 0.02	0.02	0.04	< 0.02	< 0.02
Wave 3	< 0.02	0.04	< 0.02	< 0.02	0.02	0.02
Wave 4	< 0.02	< 0.02	< 0.02	0.12	0.12	0.23
Wave 5	< 0.02	< 0.02	< 0.02	0.02	0.02	< 0.02
Discrimination						
F	4.08	7.46	9.83	0.39	< 0.01	5.62
(df ₁ , df ₂)	(1, 412.19)	(1, 77.88)	(1, 67.46)	(1, 408.99)	(1, 343.12)	(1, 441.01)
<i>p</i>	0.04	0.008	0.003	0.53	0.96	0.02
Estimate	0.07	0.04	0.68	− 0.16	0.01	− 0.47
95% CI	[0, 0.15]	[0.01, 0.07]	[0.25, 1.11]	[− 0.67, 0.35]	[− 0.09, 0.10]	[− 0.86, − 0.08]
Cohen's f^2						
Global	< 0.02	0.03	0.06	< 0.02	0.04	0.04
Wave 1	< 0.02	0.02	< 0.02	< 0.02	< 0.02	< 0.02
Wave 2	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02
Wave 3	< 0.02	< 0.02	< 0.02	< 0.02	0.07	< 0.02
Wave 4	0.20	< 0.02	0.09	< 0.02	0.15	0.05
Wave 5	0.06	< 0.02	0.03	0.02	< 0.02	< 0.02
Threshold						
F	0.35	6.80	< 0.01	2.04	1.33	6.96
(df ₁ , df ₂)	(1, 502.23)	(1, 182.02)	(1, 151.95)	(1, 487.53)	(1, 475.02)	(1, 364.32)
<i>p</i>	0.55	0.01	0.95	0.15	0.25	0.009
Estimate	− 0.02	0.02	0.01	0.27	0.04	0.39
95% CI	[− 0.07, 0.04]	[0.01, 0.04]	[− 0.25, 0.27]	[− 0.10, 0.65]	[− 0.03, 0.11]	[0.10, 0.67]
Cohen's f^2						
Global	< 0.02	0.06	< 0.02	< 0.02	< 0.02	0.08
Wave 1	< 0.02	0.02	0.03	< 0.02	< 0.02	< 0.02
Wave 2	< 0.02	0.04	0.05	< 0.02	< 0.02	< 0.02
Wave 3	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02
Wave 4	0.02	0.04	< 0.02	0.20	< 0.02	0.22
Wave 5	0.03	< 0.02	< 0.02	< 0.02	< 0.02	< 0.02

from − 0.07 to 0.11, Cohen's d = 0.04 to 0.06. This means that between baseline and session 2, there was a 90% overlap between the two distributions of odour identification scores and a 98% overlap between the other distributions.

Girls also outperformed boys on the test of odour discrimination. The effect sizes across the waves did not exceed Cohen's f^2 = 0.02 and within the waves varied between < 0.02 and 0.29, indicating small to medium effects. Odour threshold was not affected by any of the IVs, with Cohen's f^2 < 0.02 across the waves, and mostly varying between 0.01 and 0.19 within the individual waves. Figures 2, 3, and 4 show component plus residual plots in non-imputed data modelling the residuals of time, centred initial age, the cumulative total of sessions attended, and the time interval since any previous

testing session against odour identification, discrimination, and threshold, respectively. Online Resource 3 gives item-by-item performance on odour identification and discrimination across the testing occasions, irrespective of wave.

Imputed Data

In imputed data (N = 587), odour identification scores linearly increased with time and were higher in children who entered the study at a higher age than their peers, in girls, in children who had attended fewer sessions in shorter intervals, and in those who had already started school. However, effect sizes both across and within the waves were only small, varying between 0.02 and 0.12, or barely qualified as such (< 0.02).

Fig. 2 Component residual plots for the regression of odour **identification** on the residuals of time, centred initial age, the cumulative total of sessions attended, and the time interval since any previous testing in non-imputed data. The time and age data are given in months. The fitted black dashed lines represent a linear relationship and the loess fit lines are shown in grey. Deviations of the loess fit lines from the linear ones indicate non-linearity. Effect size for all of the effects was small (Cohen's $f^2 < 0.15$)

The only exception was the effect of pre-school/school attendance within Wave 4 (Cohen's $f^2 = 0.23$). Odour discrimination performance, that also increased further into the study, was better in older children, in girls, and in schoolchildren. Effect sizes across the waves remained small, not exceeding Cohen's $f^2 = 0.06$, and those within the waves ranged between < 0.02 and 0.20 , indicating small to medium effects. Finally, older children and those already attending school were more sensitive than younger ones and pre-schoolers. Effect sizes across the waves varied between < 0.02 and 0.08 and those within the individual waves between < 0.02 and 0.22 .

Discussion

To the best of our knowledge, the present study represents the first longitudinal examination of development of children's olfactory abilities. We expected that older children would outperform younger ones and that olfactory scores would increase with time, would be higher in girls than in boys, and that children who had attended more testing occasions in shorter intervals would outperform those who had participated in fewer sessions over a wider time span. Overall, we

observed that girls outperformed boys on the odour identification and discrimination tests but those effects were small in Cohen's (1988) terms. Also, having attended fewer sessions in shorter intervals improved odour identification performance but only to a small effect. This was true for both non-imputed and imputed data. The other small effects of time, initial age, and pre-school/school attendance gained statistical significance only after data imputation. Hence, in terms of effect size, which should be the focus of interpretation (e.g. Cumming 2012; Cumming 2014), non-imputed and imputed data produced very similar results.

The present study adds to the bulk of data provided by cross-sectional studies that show an age-related increase in olfactory performance in the first two decades of life (e.g. Doty et al. 1984; Hummel et al. 2007b; Hummel et al. 1997; Kobal et al. 2000; Sorokowska et al. 2015). However, the effects were rather minute or barely qualified as small. A possible explanation is that a greater variation in age at the commencement of the study, a longer time span than 2 years, and/or longer, possibly irregularly spaced intervals between the sessions were needed to capture any changes in the olfactory scores. Indeed, one of the issues to consider when designing a longitudinal study is to allow for enough

Fig. 3 Component residual plots for the regression of odour **discrimination** on the residuals of time, centred initial age, the cumulative total of sessions attended, and the time interval since any previous testing in non-imputed data. The time and age data are given in months. The fitted black dashed lines represent a linear relationship and the loess fit lines are shown in grey. Deviations of the loess fit lines from the linear ones indicate non-linearity. Effect sizes were small (Cohen's $f^2 < 0.15$)

repeated observations to recognise the change. Despite the analytical implications associated with the unequal spacing of observations, under certain circumstances, it is actually preferable (Ployhart and Vandenberg 2010). An argument in favour of irregularly spaced repeated olfactory measures would be that children's cognitive development, in general, is anything but gradual and linear. Rather, it seems analogous to overlapping waves (Siegler 1996), whereby development is characterised by changes in the distributions of strategies children use for problem solving. It appears that specific cognitive abilities mutually cause each other across the developmental life course under a particular profile of environmental constraints, as posited, for example, by the dynamical model of specific abilities (van der Maas et al. 2006). If olfactory abilities can be thought of as a specific category of cognitive abilities (e.g. McGrew 2005, 2009), their development should follow similar general principles. Hence, future studies should carefully consider major developmental milestones in cognitive abilities that are known to affect olfactory perception, co-examine other sensory modalities that could be expected to influence olfaction the most at the given age, and design the duration and spacing of intervals accordingly. Also, development over time likely needs a longer observation period than 2 years.

The effect of gender on odour identification and discrimination is another finding routinely reported in the literature on the development of the sense of smell (Ferdenzi et al. 2008a; Ferdenzi et al. 2008b; Renner et al. 2009; Richman et al. 1992; Stevenson et al. 2007; van Spronsen et al. 2013). There are nonetheless many studies in which it was not observed (Cameron and Doty 2013; Dzaman et al. 2013; Martinec Nováková and Vojtušová Mrzílková 2016a, 2016b; Richman et al. 1995a; Saxton et al. 2014; Schriever et al. 2014; Sorokowska et al. 2015). This inconsistency may stem from different sample sizes on which statistical significance depends (e.g. Cumming 2012). The effect of gender nevertheless tends to be quite small across studies. Perhaps even more importantly, sex or gender take on many meanings, e.g., chromosomal, hormonal or endocrine, gonadal, genital, body-type sex, sex of assignment and rearing, brain sex/gender, social and psychological gender (Fausto-Sterling 2012; Karkazis 2008; Zderic et al. 2002), and individuals may or may not develop in sex-typical ways and behave in a gender-conforming manner. This absence of a clear-cut distinction between males and females has been reported in olfaction as well (Nováková et al. 2013). Hence, the focus should shift from mere search of the so-called gender or sex differences to specific factors that may influence olfactory performance

Fig. 4 Component residual plots for the regression of odour threshold on the residuals of time, centered initial age, the cumulative total of sessions attended, and the time interval since any previous testing in non-imputed data. The time and age data are given in months. The fitted black dashed lines represent a linear relationship and the loess fit lines are shown in grey. Deviations of the loess fit lines from the linear ones indicate non-linearity. Effect sizes were small (Cohen's $f^2 < 0.15$)

and actually contribute to this aspect of inter-individual variation. Research indicates that verbal fluency is one of the factors in children (Monnery-Patris et al. 2009; Richman et al. 1992; Richman et al. 1995b) and adults (Öberg et al. 2002). For example, Monnery-Patris et al. (2009) reported that the gender effect vanished when verbal proficiency (verbal age and olfactory verbal fluency) was controlled for. Superior performance of females on verbal fluency tasks has been demonstrated in adults (Halari et al. 2006) and children (Anderson et al. 2001b), although there is also considerable within-gender variability (Rahman et al. 2003). Individuals with higher verbal fluency scores tend to outperform low-scoring ones on the task of odour identification (Larsson et al. 2000). Nevertheless, it is important to recognise that sources of variation, particularly in repeated assessments, are likely to be multifactorial and usually do not relate to a single construct. As this was an initial study, we did not delve into the actual factors behind the proxy of gender, but there is clearly a knowledge gap that needs to be addressed for us to fully understand the development of gender- or sex-related inter-individual variation in olfactory perception.

In longitudinal studies, the key element requiring careful consideration is the practice effect, whereby performance on re-testing improves as a result of previous exposure to the

same or similar neuropsychological measure rather than an actual change in the individual's ability (e.g. Collie et al. 2003). Tests with a single solution, such as the odour identification and discrimination tasks in the present study, are more likely to exhibit significant practice effects upon repeated testing (Basso et al. 1999; McCaffrey et al. 1992). This is particularly the case when content is repeated from the original test to the next (Krumboltz and Christal 1960; Kulik et al. 1984). However, scores may increase on re-test even when different items are employed (Benedict and Zgaljardic 1998; Wilson et al. 2000) because participants learn how to handle the task as such more effectively.

Further, numerous studies have shown the critical role of repeated testing in consolidating learning (Karpicke and Roediger 2008; Roediger and Butler 2011), including several that are now considered classics (Carrier and Pashler 1992; Gates 1917; Glover 1989; Jones 1923-1924; Spitzer 1939; Tulving 1967). Thus, a testing occasion may, in fact, represent a learning opportunity with a potentially greater impact than that of olfactory perceptual learning within the everyday olfactory environment we originally meant to address. Even though in the present study children were not corrected upon making an incorrect choice or given any feedback on their performance, the powerful mnemonic benefits of retrieval

practice during the testing cannot be ruled out. This is because retrieval practice is often effective even without feedback (for review see Roediger and Butler 2011). Hence, there was an expectation based on multiple lines of evidence from previous studies that multiple testing might positively affect children's scores on subsequent occasions.

We, however, observed the exact opposite as the children were performing worse after having attended more sessions, although the size of the effect was small. The decrease in performance may have arisen because young children often exhibit poor self-regulatory capacities (for review see McCabe et al. 2004) and thus are more heavily influenced by external environment than older children (López et al. 2005; Mahone 2005). Familiarity with test stimuli may have negatively affected performance in subsequent waves if the stimuli were perceived as less novel, interesting, and stimulating, resulting in lesser attention and interest (Courage et al. 2006; Sheese et al. 2008).

To analyse the practice effect in greater detail, one must acknowledge that gains (or losses) in test scores may not be equivalent across multiple testing sessions and may be modulated by the length of intervals between the occasions. For instance, it has been found that with brief intervals, gains in scores stabilise after an initial practice effect on measures of attention and processing speed (Falletti et al. 2006). In tests with a ceiling effect, the strongest practice effects have been found between the first and second testing with little to no improvement thereafter (Benedict and Zgaljardic 1998; Ivnik et al. 1999). Here, we did not observe any ceiling or bottom effects in any of the tests.

In the present study, odour identification scores, which showed a significant main effect of the cumulative total of sessions attended in the LMM, initially increased and then fluctuated, but the mean differences were less than half a point and the effects were small or barely qualified as such. This effect was statistically non-significant in the LMM analyses on non-imputed odour discrimination data. However, additional explorations using paired sample *t*-tests revealed that it also significantly improved by almost two points at the second testing occasion and that the effect bordered on large. It stagnated after that (session 2 vs. 3, mean difference < 0.5 point) and eventually dropped (session 3 vs. 4, small effect, mean difference = 0.97 point). Odour threshold also showed no main effect of the number of sessions in the LMM analyses but explorations with paired sample *t*-tests yielded an initial increase by about half a point in odour sensitivity that was not statistically significant (but note that $N = 70$) followed by a stagnation (a mean difference of less than 0.1 point), with Cohen's $d < 0.02$. This means a 99% overlap of the distributions. Similar findings were obtained when the covariate of the time interval between the testing occasions was included in the models. Thus, it can be cautiously concluded that the positive influence of repeated exposure to the same testing format

and stimuli was only limited to the session immediately following the baseline. In the following sessions, the scores reached a plateau or slightly dropped. Finally, the positive influence of schooling on cognitive development is well known (Rutter 1985), although in children who are not socially challenged, such as those in the present sample, the effects are far less noticeable as opposed to lower-socioeconomic status children (Downey et al. 2004; Raudenbush 2009).

An additional consideration for repeated cognitive testing is that the degree of gain may vary for different tests. To explore this possibility, we calculated differences in scores between baseline and sessions 2, 2 and 3, 3 and 4, and 4 and 5 irrespective of wave for each olfactory task. It transpired that between the baseline and the second session, children improved significantly more on the test of odour discrimination than on that of odour identification (Cohen's $d = 0.61$) or threshold (Cohen's $d = 0.40$), respectively. This means an overlap of 76 and 84%, respectively. Also, between sessions 3 and 4, the drop was more marked for odour discrimination scores compared to those for identification (Cohen's $d = 0.47$ (81% overlap)).

Further, the factors contributing to change may vary across tasks. Also, whilst certain variables may affect baseline performance on a given measure, *change* in performance on that measure can be influenced by others (Attix et al. 2009). The present sample size only allowed us to explore these ideas with the simplest univariate models. It transpired that when change was operationalised as a difference between baseline and after having attended one session, one vs. two sessions, and so on, the time elapsed from the preceding session seemed the most important factor in each of the three olfactory measures, albeit after a different number of exposures. Yet another issue, which we were unable to explore within the present study, is that with repeated assessments, a given task may begin to target a different cognitive function than initially intended. We present these explorations more as ideas to be addressed as proper hypotheses in future studies.

In paediatric populations, such considerations are of particular consequence because, in children, variability in performance may be inversely related to age, that is, variability can decrease as children grow older. At the very least, this would be true for measures of cognitive functions that can be modelled as linear and exhibit limited floor and ceiling effects. Nevertheless, to complicate matters, linear models of cognitive development are often an over-simplification because brain maturation proceeds in rapid developmental progression during growth spurts. Hence, during periods of accelerated development, a temporary rise in variance may be observed in at least some measures (Anderson et al. 2001a; Huizinga et al. 2006). In the present study, variability dropped significantly in odour threshold scores between the first and

last wave, which was probably attributable to children's better ability to handle this demanding task. Nevertheless, threshold testing, along with the odour identification and discrimination tasks, could be carried out effectively enough even in children under the age of 6. Their data did not represent outliers and additional analyses from which they were omitted led to results similar in both statistical significance and effect size to those reported above.

Future studies may also find it useful to employ statistical methods that may facilitate the interpretation of change, such as reliable-change index (RCI) scores (Iverson 2011) or standardised regression-based (SRB) change scores (McSweeney et al. 1993). Moreover, to correct for the initial practice effects on odour identification and discrimination, two or more baseline assessments may be considered in future studies (McCaffrey and Westervelt 1995).

One of the limitations of the present study was the amount of missing data and related issues, such as the higher proportion of pre-schoolers in Wave 4 compared to Wave 3, similar mean age of children participating in these two waves, and the low *N* in Wave 4. Collection of data within Wave 4 took place in May and June 2012, which was towards the busy end of the school year. As school principals, teachers, and parents each time provided a one-time permission only and had to be asked again within the next wave, sometimes they responded in the negative to our request. This was mostly because school teachers had busy curricula to follow and children's absence from classes due to testing would have been a major complication. This posed less of a problem for kindergarten teachers, who were generally more relaxed about the testing and more interested or even enthusiastic about co-operating. Besides, pre-schoolers tended to have similar schedules and could be approached within the kindergarten on almost any given day, as opposed to school children. When the first participants started school during Wave 3, they turned out to be difficult to reach because of their busy curricula and after-school activities. Also, after the majority of children in a given kindergarten or school had been tested, we did not have the luxury of coming once again to test those children who were previously absent or ill because the principals would not allow it. Parents of children who could not be tested at school were then invited to visit our laboratory, but they often found it logistically inconvenient. Future studies could overcome many of these issues by recruiting children from institutions with more relaxed curricula, such as outdoor kindergartens and elementary schools. This approach might also allow ecologically (externally) valid observation of olfactory behaviours, even though the generalisation of findings may be limited as such institutions are generally attended by children from specific backgrounds.

Conclusion

In the present study, we investigated the development of children's olfactory abilities over five waves which took place every 6 months. Odour identification and discrimination but not sensitivity were on average higher in girls, but the effect was small. Further, children's performance on the odour identification task was affected by the cumulative total of sessions attended and the time elapsed since any previous testing. After data imputation, other small effects gained statistical significance, namely those of time, initial age, and pre-school/school attendance. In terms of effect sizes, non-imputed and imputed data yielded very similar results. Despite the small magnitude of the effects reported in this paper, the unexpected findings, particularly regarding the practice effects, warrant replication and extension in longitudinal studies carried over a broader time span.

Acknowledgments The authors would like to express their gratitude to Jitka Fialová and Markéta Sobotková for their help with data collection and Lydie Kubíčková for her assistance with maintaining the participant database. A special thanks goes to David Le Sage for proofreading. We are very grateful to the children and their parents for their participation and school principals and teachers for allowing us to perform the study in school premises.

Author Contributions Conceived and designed the study: LMN and JH. Performed the study: LMN. Analysed the data: LMN. Wrote the paper: LMN and JH.

Compliance with Ethical Standards

Funding This study is a result of research funded by the project LO1611 with financial support from the Ministry of Education, Youth, and Sports (MEYS) under the NPU I program. LMN was supported by the Czech Science Foundation (GA17-14534S), the PROGRES program Q22 at Faculty of Humanities, Charles University within the Institutional Support for Long-Term Development of Research Organizations from MEYS, and Specific Academic Research project (SVV) number 260 469 realised at Faculty of Humanities, Charles University. LMN and JH are members of the Charles University Research Centre "Příroda a kultura: Historické, kulturní a biologické koncepty lidské přirozenosti" (UNCE/HUM/025, 204 056). The funding sources had no involvement in study design, in the collection, analysis, and interpretation of data, in the writing of the paper, or in the decision to submit the article for publication. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Conflict of Interest Both authors declare that they have no conflict of interest.

Ethical Approval All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the IRB of the Faculty of Science, Charles University.

Informed Consent The children's parents provided written informed consent. The children provided oral informed consent in the presence of a teacher employed by the school. There were no cases of children refusing to give their consent to participate. The child-parent pairs each received 300 CZK (approx. 12 EUR) in compensation.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

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

References

- Anderson P, Anderson V, Garth J (2001a) Assessment and development of organizational ability: the Rey complex figure organizational strategy score (RCF-OSS). *Clin Neuropsychol* 15:81–94. <https://doi.org/10.1076/clin.15.1.81.1905>
- Anderson VA, Anderson P, Northam E, Jacobs R, Catroppa C (2001b) Development of executive functions through late childhood and adolescence in an Australian sample. *Dev Neuropsychol* 20:385–406. https://doi.org/10.1207/S15326942DN2001_5
- Atix DK, Story TJ, Chelune GJ, Ball JD, Stutts ML, Hart RP, Barth JT (2009) The prediction of change: normative neuropsychological trajectories. *Clin Neuropsychol* 23:21–38. <https://doi.org/10.1080/13854040801945078>
- Basso MR, Bornstein RA, Lang JM (1999) Practice effects on commonly used measures of executive function across twelve months. *Clin Neuropsychol* 13:283–292. <https://doi.org/10.1076/clin.13.3.283.1743>
- Bastos LOD, Mantovani Guerreiro M, Lees AJ et al (2015) Effects of age and cognition on a cross-cultural paediatric adaptation of the Sniffin' Sticks identification test. *PLoS One* 10:e0131641. <https://doi.org/10.1371/journal.pone.0131641>
- Benedict RHB, Zgaljardic DJ (1998) Practice effects during repeated administrations of memory tests with and without alternate forms. *J Clin Exp Neuropsychol* 20:339–352. <https://doi.org/10.1076/jcen.20.3.339.822>
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer-Verlag, New York
- Cameron EL, Doty RL (2013) Odor identification testing in children and young adults using the smell wheel. *Int J Pediatr Otorhinolaryngol* 77:346–350. <https://doi.org/10.1016/j.ijporl.2012.11.022>
- Carrier M, Pashler H (1992) The influence of retrieval on retention. *Mem Cogn* 20:633–642. <https://doi.org/10.3758/BF03202713>
- Ceci SJ (1991) How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Dev Psychol* 27:703–722. <https://doi.org/10.1037/0012-1649.27.5.703>
- Chopra A, Baur A, Hummel T (2008) Thresholds and chemosensory event-related potentials to malodors before, during, and after puberty: differences related to sex and age. *Neuroimage* 40:1257–1263. <https://doi.org/10.1016/j.neuroimage.2008.01.015>
- Choudhury ES, Moberg P, Doty RL (2003) Influences of age and sex on a microencapsulated odor memory test. *Chem Senses* 28:799–805. <https://doi.org/10.1093/chemse/bjg072>
- Cohen JE (1988) Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ
- Collie A, Maruff P, Darby DG et al (2003) The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *J Int Neuropsychol Soc* 9: 419–428. <https://doi.org/10.1017/S1355617703930074>
- Cometto-Muniz JE, Abraham MH (2008) Human olfactory detection of homologous n-alcohols measured via concentration-response functions. *Pharmacol Biochem Behav* 89:279–291. <https://doi.org/10.1016/j.pbb.2007.12.023>
- Courage ML, Reynolds GD, Richards JE (2006) Infants' attention to patterned stimuli: developmental change from 3 to 12 months of age. *Child Dev* 77:680–695. <https://doi.org/10.1111/j.1467-8624.2006.00897.x>
- Cumming G (2012) Understanding the new statistics: effect sizes, confidence intervals, and meta-analysis. Routledge, New York
- Cumming G (2014) The new statistics: why and how. *Psychol Sci* 25:7–29. <https://doi.org/10.1177/0956797613504966>
- Cumming G, Maillardet R (2006) Confidence intervals and replication: where will the next mean fall? *Psychol Methods* 11:217–227. <https://doi.org/10.1037/1082-989X.11.3.217>
- Dalton P, Doolittle N, Breslin PAS (2002) Gender-specific induction of enhanced sensitivity to odors. *Nat Neurosci* 5:199–200. <https://doi.org/10.1038/nm803>
- Doty RL, Kamath V (2014) The influences of age on olfaction: a review. *Front Psychol* 5:20. <https://doi.org/10.3389/fpsyg.2014.00020>
- Doty RL, Laing DG (2015) Psychophysical measurement of human olfactory function. In: Doty RL (ed) *Handbook of olfaction and gustation*. John Wiley & Sons, Hoboken, NJ, pp 227–260
- Doty RL, Shaman P, Applebaum SL, Giberson R, Sikorski L, Rosenberg L (1984) Smell identification ability: changes with age. *Science* 226: 1441–1443. <https://doi.org/10.1126/science.6505700>
- Downey DB, von Hippel PT, Broh BA (2004) Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *Am Sociol Rev* 69:613–635. <https://doi.org/10.1177/000312240406900501>
- Dubois D, Rouby C (2002) Names and categories for odors: the veridical label. In: Rouby C, Schaal B, Dubois D et al (eds) *Olfaction, taste, and cognition*. Cambridge University Press, Cambridge, pp 47–66
- Dzaman K, Zielnik-Jurkiewicz B, Jurkiewicz D et al (2013) Test for screening olfactory function in children. *Int J Pediatr Otorhinolaryngol* 77: 418–423. <https://doi.org/10.1016/j.ijporl.2012.12.001>
- Falletti MG, Maruff P, Collie A, Darby DG (2006) Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *J Clin Exp Neuropsychol* 28:1095–1112. <https://doi.org/10.1080/13803390500205718>
- Fausto-Sterling A (2012) *Sex/gender: biology in a social world*. Routledge, New York
- Ferdenzi C, Coureaud G, Camos V, Schaal B (2008a) Human awareness and uses of odor cues in everyday life: results from a questionnaire study in children. *Int J Behav Dev* 32:422–431. <https://doi.org/10.1177/0165025408093661>
- Ferdenzi C, Mustonen S, Tuorila H, Schaal B (2008b) Children's awareness and uses of odor cues in everyday life: a Finland-France comparison. *Chemosens Percept* 1:190–198. <https://doi.org/10.1007/s12078-008-9020-6>
- Ferguson CJ (2009) An effect size primer: a guide for clinicians and researchers. *Prof Psychol-Res Prac* 40:532–538. <https://doi.org/10.1037/a0015808>
- Field A (2005) *Discovering statistics using SPSS*. SAGE Publications, London
- Frank RA, Brearton M, Rybalsky K, Cessna T, Howe S (2011) Consistent flavor naming predicts recognition memory in children and young adults. *Food Qual Prefer* 22:173–178. <https://doi.org/10.1016/j.foodqual.2010.09.009>
- Gates AI (1917) Recitation as a factor in memorizing. *Arch Psychol* 40: 1–104
- Glover JA (1989) The “testing” phenomenon: not gone but nearly forgotten. *J Educ Psychol* 81:392–399. <https://doi.org/10.1037/0022-0663.81.3.392>

- Goswami U (2011) The Wiley-Blackwell handbook of childhood cognitive development. Wiley, Oxford, UK
- Halari R, Sharma T, Hines M, Andrew C, Simmons A, Kumari V (2006) Comparable fMRI activity with differential behavioural performance on mental rotation and overt verbal fluency tasks in healthy men and women. *Exp Brain Res* 169:1–14. <https://doi.org/10.1007/s00221-005-0118-7>
- Hawkes CH, Doty RL (2009) Anatomy and physiology. In: The neurology of olfaction. Cambridge University Press, New York, NY, pp 1–58
- Heilbronner RL, Sweet JJ, Attix DK, Krull KR, Henry GK, Hart RP (2010) Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: the utility and challenges of repeat test administrations in clinical and forensic contexts. *Clin Neuropsychol* 24:1267–1278. <https://doi.org/10.1080/13854046.2010.526785>
- Hox JJ (2002) Multilevel analysis: techniques and applications. Erlbaum, Hillsdale, NJ
- Huizinga M, Dolan CV, van der Molen MW (2006) Age-related change in executive function: developmental trends and a latent variable analysis. *Neuropsychologia* 44:2017–2036. <https://doi.org/10.1016/j.neuropsychologia.2006.01.010>
- Hummel T. (2004) Sniffin' Sticks tutorial. Dresden
- Hummel T, Bensafi M, Nikolaus J et al (2007a) Olfactory function in children assessed with psychophysical and electrophysiological techniques. *Behav Brain Res* 180:133–138. <https://doi.org/10.1016/j.bbr.2007.02.040>
- Hummel T, Kobal G, Gudziol H, Mackay-Sim A (2007b) 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. <https://doi.org/10.1007/s00405-006-0173-0>
- 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. <https://doi.org/10.1093/chemse/22.1.39>
- Hvastja L, Zanuttini L (1989) Odor memory and odor hedonics in children. *Perception* 18:391–396. <https://doi.org/10.1068/p180391>
- IBM Corp (2016) IBM SPSS statistics for Windows. IBM Corp, Armonk
- Iverson GL (2011) Reliable change index. In: Kreutzer JS, DeLuca J, Caplan B (eds) Encyclopedia of clinical neuropsychology. Springer New York, New York, NY, pp 2150–2153
- Ivnik RJ, Smith GE, Lucas JA, Petersen RC, Boeve BF, Kokmen E, Tangalos EG (1999) Testing normal older people three or four times at 1- to 2-year intervals: defining normal variance. *Neuropsychology* 13:121–127. <https://doi.org/10.1037/0894-4105.13.1.121>
- Jones HE (1923–1924) The effects of examination on the performance of learning. *Arch Psychol* 10:1–70
- Karkazis K (2008) Fixing sex: intersex, medical authority, and lived experience. Duke University Press, Durham
- Karipke JD, Roediger HL (2008) The critical importance of retrieval for learning. *Science* 319:966–968. <https://doi.org/10.1126/science.1152408>
- Kim J, Seo BS (2013) How to calculate sample size and why. *Clin Orthop Surg* 5:235–242. <https://doi.org/10.4055/cios.2013.5.3.235>
- Kobal G, Klimek L, Wolfensberger M, Gudziol H, Temmel A, Owen CM, Seeber H, Pauli E, Hummel T (2000) Multicenter investigation of 1,036 subjects using a standardized method for the assessment of olfactory function combining tests of odor identification, odor discrimination, and olfactory thresholds. *Eur Arch Otorhinolaryngol* 257:205–211. <https://doi.org/10.1007/s004050050223>
- Koelega HS, Köster EP (1974) Some experiments on sex differences in odor perception. *Ann N Y Acad Sci* 237:234–246. <https://doi.org/10.1111/j.1749-6632.1974.tb49859.x>
- Krumboltz JD, Christal RE (1960) Short-term practice effects in tests of spatial aptitude. *Pers Guid J* 38:385–391. <https://doi.org/10.1002/j.2164-4918.1960.tb02568.x>
- Kulik JA, Kulik CLC, Bangert RL (1984) Effects of practice on aptitude and achievement test scores. *Am Educ Res J* 21:435–447. <https://doi.org/10.2307/1162453>
- Larjola K, von Wright J (1976) Memory of odors: developmental data. *Percept Mot Skills* 42:1138–1138. <https://doi.org/10.2466/pms.1976.42.3c.1138>
- Larsson M, Finkel D, Pedersen NL (2000) Odor identification: influences of age, gender, cognition, and personality. *J Gerontol B Psychol* 55:304–310. <https://doi.org/10.1093/geronb/55.5.P304>
- Larsson M, Nilsson LG, Olofsson JK, Nordin S (2004) Demographic and cognitive predictors of cued odor identification: evidence from a population-based study. *Chem Senses* 29:547–554. <https://doi.org/10.1093/chemse/bjh059>
- López F, Menez M, Hernández-Guzmán L (2005) Sustained attention during learning activities: an observational study with pre-school children. *Early Child Dev Care* 175:131–138. <https://doi.org/10.1080/0300443042000230384>
- Louis TA, Robins J, Dockery DW et al (1986) Explaining discrepancies between longitudinal and cross-sectional models. *J Chronic Dis* 39:831–839. [https://doi.org/10.1016/0021-9681\(86\)90085-8](https://doi.org/10.1016/0021-9681(86)90085-8)
- Mahone EM (2005) Measurement of attention and related functions in the preschool child. *Ment Retard Dev Disabil Res Rev* 11:216–225. <https://doi.org/10.1002/mrdd.20070>
- Martínez Nováková L, Fialová J, Havlíček J (2018a) Development of children's olfactory abilities and odor awareness is not predicted by temperament: a longitudinal study. *Chemosens Percept* 11:59–71. <https://doi.org/10.1007/s12078-017-9240-8>
- Martínez Nováková L, Fialová J, Havlíček J (2018b) Effects of diversity in olfactory environment on children's sense of smell. *Sci Rep-UK* 8:2937. <https://doi.org/10.1038/s41598-018-20236-0>
- Martínez Nováková L, Plotěná D, Roberts SC et al (2015) Positive relationship between odor identification and affective responses of negatively valenced odors. *Front Psychol* 6:607. <https://doi.org/10.3389/fpsyg.2015.00607>
- Martínez Nováková L, Vojtušová Mrzilková R (2016a) Children's exposure to odors in everyday contexts predicts their odor awareness. *Chemosens Percept* 9:56–68. <https://doi.org/10.1007/s12078-016-9205-3>
- Martínez Nováková L, Vojtušová Mrzilková R (2016b) Temperamental influences on children's olfactory performance: the role of self-regulation. *Chemosens Percept* 9:153–173. <https://doi.org/10.1007/s12078-016-9216-0>
- McCabe LA, Cunningham M, Brooks-Gunn J (2004) The development of self-regulation in young children: individual characteristics and environmental contexts. In: Baumeister RF, Vohs KD (eds) Handbook of self-regulation: research, theory, and applications. Guilford Press, New York, NY, pp 340–356
- McCaffrey RJ, Ortega A, Orsillo SM et al (1992) Practice effects in repeated neuropsychological assessments. *Clin Neuropsychol* 6:32–42. <https://doi.org/10.1080/13854049208404115>
- McCaffrey RJ, Westervelt HJ (1995) Issues associated with repeated neuropsychological assessments. *Neuropsychol Rev* 5:203–221. <https://doi.org/10.1007/BF02214762>
- McGrew KS (2005) The Cattell–Horn–Carroll theory of cognitive abilities. In: Flanagan DP, Harrison PL (eds) Contemporary intellectual assessment: theories, tests, and issues. Guilford Press, New York, pp 136–181
- McGrew KS (2009) Editorial: CHC theory and the human cognitive abilities project: standing on the shoulders of the giants of psychometric intelligence research. *Intelligence* 37:1–10. <https://doi.org/10.1016/j.intell.2008.08.004>
- McSweeney AJ, Naugle RI, Chelune GJ et al (1993) “T scores for change”: an illustration of a regression approach to depicting change in clinical neuropsychology. *Clin Neuropsychol* 7:300–312. <https://doi.org/10.1080/13854049308401901>
- Mennella JA, Beauchamp GK (1992) Developmental changes in nasal airflow patterns. *Acta Otolaryngol (Stockh)* 112:1025–1031. <https://doi.org/10.3109/00016489209137505>

- Monnery-Patris S, Rouby C, Nicklaus S, Issanchou S (2009) Development of olfactory ability in children: sensitivity and identification. *Dev Psychobiol* 51:268–276. <https://doi.org/10.1002/dev.20363>
- Noll RB, Zucker RA, Greenberg GS (1990) Identification of alcohol by smell among preschoolers: evidence for early socialization about drugs occurring in the home. *Child Dev* 61:1520–1527. <https://doi.org/10.1111/j.1467-8624.1990.tb02880.x>
- Nováková L, Valentová JV, Havlíček J (2013) Olfactory performance is predicted by individual sex-atypicality, but not sexual orientation. *PLoS One* 8:e80234. <https://doi.org/10.1371/journal.pone.0080234>
- Öberg C, Larsson M, Backman L (2002) Differential sex effects in olfactory functioning: the role of verbal processing. *J Int Neuropsychol Soc* 8:691–698. <https://doi.org/10.1017/S1355617702801424>
- Ployhart RE, Vandenberg RJ (2010) Longitudinal research: the theory, design, and analysis of change. *Aust J Manag* 36:94–120. <https://doi.org/10.1177/0149206309352110>
- R Development Core Team (2008) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Rahman Q, Abrahams S, Wilson GD (2003) Sexual-orientation-related differences in verbal fluency. *Neuropsychology* 17:240–246. <https://doi.org/10.1037/0894-4105.17.2.240>
- Raudenbush SW (2009) The Brown legacy and the O'Connor challenge: transforming schools in the images of children's potential. *Aust Educ Res* 38:169–180. <https://doi.org/10.3102/0013189X09334840>
- Renner B, Mueller CA, Dreier J, Faulhaber S, Rascher W, Kobal G (2009) The candy smell test: a new test for retronasal olfactory performance. *Laryngoscope* 119:487–495. <https://doi.org/10.1002/lary.20123>
- Richman RA, Post EM, Sheehee PR, Wright HN (1992) Olfactory performance during childhood. I. Development of an odorant identification test for children. *J Pediatr* 121:908–911. [https://doi.org/10.1016/S0022-3476\(05\)80337-3](https://doi.org/10.1016/S0022-3476(05)80337-3)
- Richman RA, Sheehee PR, Wallace K, Hyde JM, Coplan J (1995a) Olfactory performance during childhood. II. Developing a discrimination task for children. *J Pediatr* 127:421–426. [https://doi.org/10.1016/S0022-3476\(95\)70074-9](https://doi.org/10.1016/S0022-3476(95)70074-9)
- Richman RA, Wallace K, Sheehee PR (1995b) Assessment of an abbreviated odorant identification task for children: a rapid screening device for schools and clinics. *Acta Paediatr* 84:434–437. <https://doi.org/10.1111/j.1651-2227.1995.tb13666.x>
- Roediger HL, Butler AC (2011) The critical role of retrieval practice in long-term retention. *Trends Cogn Sci* 15:20–27. <https://doi.org/10.1016/j.tics.2010.09.003>
- Rogoff B (1981) Schooling and the development of cognitive skills. In: Triandis HC, Heron A (eds) *Handbook of cross-cultural psychology: developmental psychology*. Allyn & Bacon, Boston, pp 233–294
- Rosselli N, Ardila A (2003) The impact of culture and education on non-verbal neuropsychological measurements: a critical review. *Brain Cogn* 52:326–333. [https://doi.org/10.1016/S0278-2626\(03\)00170-2](https://doi.org/10.1016/S0278-2626(03)00170-2)
- Rutter M (1985) Family and school influences on cognitive development. *J Child Psychol Psychiatry* 26:683–704. <https://doi.org/10.1111/j.1469-7610.1985.tb00584.x>
- SAS (2017) SAS University Edition. SAS Institute Inc., Cary, NC
- Saxton TK, Martinec Nováková L, Jash R, Šandová A, Plotěná D, Havlíček J (2014) Sex differences in olfactory behavior in Namibian and Czech children. *Chemosens Percept* 7:117–125. <https://doi.org/10.1007/s12078-014-9172-5>
- Selya AS, Rose JS, Dierker LC, Hedeker D, Mermelstein RJ (2012) A practical guide to calculating Cohen's $f(2)$, a measure of local effect size, from PROC MIXED. *Front Psychol* 3:111. <https://doi.org/10.3389/fpsyg.2012.00111>
- Sheese BE, Rothbart MK, Posner MI, White LK, Fraundorf SH (2008) Executive attention and self-regulation in infancy. *Infant Behav Dev* 31:501–510. <https://doi.org/10.1016/j.infbeh.2008.02.001>
- Shek DTL, Ma CMS (2011) Longitudinal data analyses using linear mixed models in SPSS: concepts, procedures and illustrations. *Sci World J* 11:42–76. <https://doi.org/10.1100/tsw.2011.2>
- Schriever VA, Mori E, Petters W, Boerner C, Smitka M, Hummel T (2014) The “Sniffin’ kids” test - a 14-item odor identification test for children. *PLoS One* 9:e101086. <https://doi.org/10.1371/journal.pone.0101086>
- Siegler RS (1996) *Emerging minds: the process of change in children's thinking*. Oxford University Press, New York
- Singer JD, Willett JB (2003) *Applied longitudinal data analysis*. Oxford Press, New York
- Sorokowska A, Schriever VA, Gudziol V, Hummel C, Hähner A, Iannilli E, Sinding C, Aziz M, Seo HS, Negoias S, Hummel T (2015) Changes of olfactory abilities in relation to age: odor identification in more than 1400 people aged 4 to 80 years. *Eur Arch Otorhinolaryngol* 272:1937–1944. <https://doi.org/10.1007/s00405-014-3263-4>
- Spitzer HF (1939) Studies in retention. *J Educ Psychol* 30:641–656. <https://doi.org/10.1037/h0063404>
- Starkweather J (2014) A new recommended way of dealing with multiple missing values: using missForest for all your imputation needs. *Benchmarks RSS Matters* July 2014
- Steinbrink C, Zimmer K, Lachmann T, Dirichs M, Kammer T (2014) Development of rapid temporal processing and its impact on literacy skills in primary school children. *Child Dev* 85:1711–1726. <https://doi.org/10.1111/cdev.12208>
- Stekhoven DJ, Bühlmann P (2012) MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28:112–118. <https://doi.org/10.1093/bioinformatics/btr597>
- Stevenson RJ, Mahmut M, Sundqvist N (2007) Age-related changes in odor discrimination. *Dev Psychol* 43:253–260. <https://doi.org/10.1037/0012-1649.43.1.253>
- Tulving E (1967) The effects of presentation and recall of material in free-recall learning. *J Verb Learn Verb Behav* 6:175–184. [https://doi.org/10.1016/S0022-5371\(67\)80092-6](https://doi.org/10.1016/S0022-5371(67)80092-6)
- van der Maas HLJ, Dolan CV, Grasman R et al (2006) A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychol Rev* 113:842–861. <https://doi.org/10.1037/0033-295X.113.4.842>
- van Spronsen E, Ebbens FA, Fokkens WJ (2013) Olfactory function in healthy children: normative data for odor identification. *Am J Rhinol Allergy* 27:197–201. <https://doi.org/10.2500/ajra.2013.27.3865>
- Waljee AK, Mukherjee A, Singal AG, Zhang Y, Warren J, Balis U, Marrero J, Zhu J, Higgins PDR (2013) Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open* 3:e002847. <https://doi.org/10.1136/bmjopen-2013-002847>
- Wilson BA, Watson PC, Baddeley AD et al (2000) Improvement or simply practice? The effects of twenty repeated assessments on people with and without brain injury. *J Int Neuropsychol Soc* 6:469–479. <https://doi.org/10.1017/S1355617700644053>
- Wilson DA, Bell H, Chen C-F (2009) Olfactory perceptual learning. In: Hirokawa N, Windhorst U (eds) *Binder MD. Encyclopedia of neuroscience*. Springer, Berlin, Heidelberg, pp 3010–3013
- Wysocki CJ, Dorries KM, Beauchamp GK (1989) Ability to perceive androstenone can be acquired by ostensibly anosmic people. *Proc Natl Acad Sci U S A* 86:7976–7978. <https://doi.org/10.1073/pnas.86.20.7976>
- Zderic SA, Canning DA, Carr MC et al (2002) *Pediatric gender assignment: a critical reappraisal*. Kluwer Academic/Plenum, New York