Journal of Abnormal Psychology
2004, Vol. 113, No. 2, 207–216

0021-843X/04/$12.00 DOI: 10.1037/0021-843X.113.2.207

The Smell Identification Test as a Measure of Olfactory Identification Ability in Schizophrenia and Healthy Populations: A Rasch Psychometric Study

Kelly L. Minor
Northwestern University

Benjamin D. Wright
University of Chicago

Sohee Park
Vanderbilt University

This study examines University of Pennsylvania Smell Identification Test (UPSIT; R. L. Doty, 1995) performance in 133 controls and 54 chronic, medicated outpatients with schizophrenia (SZ) using item-response theory modeling. Results show that UPSIT items contribute to 1 factor, cover a range of 8 standard errors of measurement, and articulate 3 ability levels. Although it is not difficult enough to discriminate among persons of above-average ability, the test has diagnostic utility in detecting moderate impairment. Independent of item difficulty, 13 items differentiate patients from controls. When 45 patients and 45 controls were matched on gender and age, patient accuracy remained significantly reduced. The findings support the test's utility and demonstrate how traditional data analysis is insensitive to complexities in test performance.

In recent years, olfactory dysfunction has been recognized as an important clinical indicator of neurologic and psychiatric disorders. Olfactory assessment is especially promising in the field of schizophrenia research for several reasons. First, olfactory pathways are distributed throughout brain areas implicated in the pathophysiology of the disorder (Potter & Butters, 1980; Rausch & Serafetinides, 1975). Second, olfactory impairment in schizophrenia may be associated with hypoactivation of frontal brain regions (Bertollo, Cowen, & Levy, 1996; Clark, Kopala, Hurwitz, & Li, 1991; Malaspina et al., 1996; Wu et al., 1993) and negative symptoms (Brewer et al., 1996; Geddes, Huws, & Pratt, 1991; Malaspina et al., 1994; Pantelis & Brewer, 1995). Third, the olfactory system is unique in that only one synapse lies between the primary olfactory receptors and primary olfactory cortex; these receptors travel ipsilaterally (Dunn & Weller, 1989; Pansky & Allen, 1980). Thus, olfactory assessment may be one of the most direct, lateralized, and noninvasive measures of brain functioning (Eslinger, Damasio, & van Hoesen, 1982; Price, 1990).

The assessment of olfactory ability is complex, traditionally organized on four hierarchical levels: detection, discrimination, identification, and memory for odors. Odor identification refers to the ability to identify and name an odor, either verbally or through recognition format. Even for healthy controls, confrontation naming of odors may be difficult when identification is based solely on free recall (Cain, 1979). Recognition format significantly decreases the demands of this task (Doty, Shaman, & Dann, 1984).

The University of Pennsylvania Smell Identification Test (UPSIT) is a standardized, recognition-format measure of odor-identification ability (Doty, 1995; Doty et al., 1984). It is reported to have the highest reliability of any olfactory test (Doty, McKeown, Lee, & Shaman, 1995; Doty, Newhouse, & Azzalina, 1985; Doty et al., 1984). Its test–retest reliability in healthy participants is reported as high as .92 for 6-month (Doty et al., 1984) and .95 for 2-week (Doty et al., 1985) intervals. Split-half reliability coefficients (i.e., comparing accuracy on odd- versus even-numbered items) are reported as $r > .93$ for healthy participants (Doty et al., 1985). Thus, the UPSIT is a highly reliable and internally consistent measure among healthy respondents. Further, it is portable, has a long shelf life, and can be self-administered by the respondent. The UPSIT's favorable psychometric properties, low cost, and ease of use have contributed to its popularity in clinical research.

Patients with schizophrenia demonstrate impaired performance on the UPSIT (see Moberg et al., 1999; Rupp, 2003, for reviews), which is attributable neither to task complexity (Kopala, Good, Martzke, & Hurwitz, 1995) nor to executive–attentional components (Seidman et al., 1997; Seidman et al., 1992). A recent

Kelly L. Minor, Department of Psychology, Northwestern University; Benjamin D. Wright, MESA Psychometric Laboratory, University of Chicago; Sohee Park, Department of Psychology, Vanderbilt University.

This study is part of Kelly L. Minor's doctoral dissertation. The research was supported in part by Grant MH-058406, a Scottish Rite Schizophrenia Research Grant, and a Young Investigator Award from the National Alliance for Research on Schizophrenia and Depression (NARSAD) to Sohee Park. We are grateful to the research participants. Kelly L. Minor wishes to extend her gratitude to J. Michael Bailey, David Uttal, and Michael B. Miller for their assistance in various stages of preparation of this article.

Correspondence concerning this article should be addressed to Kelly L. Minor, who is now at Stanford University, Department of Psychology, Jordan Hall, Building 420, Stanford, CA 94305–2130. E-mail: kminor@psych.stanford.edu

meta-analysis determined that gender, smoking status, and medication status had no significant effects on odor-identification ability among patients with schizophrenia (Moberg et al., 1999). Taken together, the findings suggest that reduced UPSIT performance in schizophrenia is representative of discrete olfactory dysfunction. However, the test has not been psychometrically assessed in schizophrenia patients. It is therefore uncertain whether the UPSIT is, in fact, an adequate measure of odor-identification ability in schizophrenia.

There is no doubt that patients with schizophrenia show reduced performance on the UPSIT (Brewer et al., 1996; Good, Martzke, Frankland, & Kopala, 1998; Good, Martzke, Honer, & Kopala, 1998; Houlihan et al., 1994; Hurwitz, Kopala, Clark, & Jones, 1988; Kopala, Clark, & Hurwitz, 1989; Kopala, Good, Martzke, & Hurwitz, 1995; Kopala, Clark, & Hurwitz, 1992; Kopala, Good, & Honer, 1994; Kopala, Good, & Honer, 1995; Malaspina et al., 1994; Moberg, Doty, Mahr, et al., 1997; Moberg, Doty, Turetsky, & Arnold, 1997; Purdon, 1998; Saoud, Hueber, Mandran, Dalery, & d'Amato, 1998; Seidman et al., 1997; Serby, Larson, & Kalkstein, 1990; Wu et al., 1993). However, these findings are based on between-group comparisons of raw UPSIT scores for patients versus controls. Raw scores do not have optimal psychometric properties (i.e., unidimensional interval comparisons among persons and items). Traditional data analysis, such as between-groups raw-score comparison, is therefore lacking. The solution is simple. By transforming the data to an interval metric, from raw scores to logit measures, a continuum is constructed that has verifiable equal-interval units from one end of the scale to the other.[1] In this way, the data achieve linearity and the properties of true measurement. Item difficulty and person ability can be measured independently along this shared metric, and we can therefore assess the test's capacity to measure ability level in a particular sample.

When measuring a test's utility, there are several points of inquiry. First, it is important to understand whether the test items contribute to a single variable. Second, it is necessary to determine whether the items are arranged hierarchically with sufficient spread to measure the full range of that variable. Third, it is meaningful to know whether the test items and respondents behave in a predictable manner. For example, suppose that respondents are committing errors on an item that should be easy for them on the basis of their ability levels. It is important to be aware of such an item, that is, one that behaves unpredictably, so that we can investigate whether some outside influence interfered with the application of the subjects' ability to that item. Perhaps those persons were distracted, rushed, or bored. Perhaps the test item was biased against them. Clearly, analysis at the item level affords the discovery of important information that would otherwise remain hidden in the raw-score data.

The current study uses the one-parameter item-response theory (IRT) model, also known as the Rasch model, to provide the first psychometric analysis of the UPSIT in schizophrenia. The study has three major aims. The first aim is to verify that the items of the UPSIT contribute to a single factor and are sufficiently spread along this factor to define a recognizable olfactory hierarchy. The second aim is to evaluate whether the UPSIT separates patients with schizophrenia and healthy controls into the five distinct levels of olfactory diagnosis, as detailed in the test manual. The third aim is to contrast perceived item difficulties for patients versus controls and discuss the potential for a "schizophrenia profile" on the UPSIT. More broadly, the study demonstrates how traditional data analysis techniques can be insensitive to subtle, yet meaningful, aspects of patient performance in clinical research.

Method

Participants

Participants included 54 outpatients with schizophrenia (43 men) and 136 healthy controls (64 men). Questionnaires, chart reviews, and interviews were used to screen all participants for current substance abuse (within the past 6 months), upper respiratory problems, or medical history that might lead to olfactory impairment (e.g., nasal congestion, past head trauma with loss of consciousness, upper respiratory tract infection, chronic rhinitis). All participants spoke English fluently, gave informed consent, were tested at similar times of day, and were paid for their participation. All aspects of this study complied with the APA ethical standard for treatment of human subjects.

Patient group. Chronic, medicated outpatients with schizophrenia were recruited from a local residential care facility. *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*; American Psychiatric Association, 1994) diagnoses of schizophrenia were confirmed by medical chart review and clinical evaluations with the Schedule for Affective Disorders and Schizophrenia, Lifetime Version (SADS–L; Spitzer & Endicott, 1978) by a master's level psychologist. The patient group was characterized by a mean education level of 12.1 years ($SD = 1.9$) and mean age of 34.2 years ($SD = 10.2$). There were no significant gender differences in education (women, $M = 11.8$ years, $SD = 2.1$; men, $M = 12.1$ years, $SD = 1.8$). With regard to age, however, female patients ($M = 40.0$ years, $SD = 10.9$) were significantly older than their male counterparts ($M = 32.7$ years, $SD = 9.6$), $t(52) = 2.19$, $p < .05$. Fifty percent of patients were smokers (45% women, 51% men).

Control group. Healthy participants were recruited from the Chicago area and were screened for history of mental illness in the family or self. The control group was characterized by an average education level of 13.7 years ($SD = 1.3$) and a mean age of 25.8 years ($SD = 7.8$). There were no significant gender differences in terms of education (women, $M = 13.6$, $SD = 1.2$; men, $M = 13.7$, $SD = 1.4$), or age (women, $M = 24.9$, $SD = 7.9$; men, $M = 26.8$, $SD = 7.7$). Ten percent of controls were smokers (11% women, 9% men). Patients were significantly older, $t(188) = -6.06$, $p < .01$, and less educated, $t(188) = 6.67$, $p < .01$, than controls. Further, they were significantly more likely to smoke than controls, $Z = 5.99$, $p < .01$.

Although the primary aim of this study is to examine UPSIT performance at the item level, the current sample affords an opportunity for a between-groups comparison of test-level performance while controlling for the effects of age and gender. Thus, a subsample was created in which controls ($n = 45$) were matched to patients ($n = 45$) on gender and age (within 12 months). The matched sample included 72 men (36 patients) and 18 women (9 patients). There was no significant difference in age between diagnostic groups (controls, $M = 32.3$, $SD = 9.6$; patients, $M = 32.2$, $SD = 9.6$); however, women across diagnostic groups were significantly older than their male counterparts (women, $M = 39.1$, $SD = 11.4$; men, $M = 30.5$, $SD = 8.2$), $F(1, 86) = 13.1$, $p < .01$. Although there were no gender differences, controls had significantly greater educational attainment than patients (controls, $M = 13.6$ years, $SD = 1.5$; patients, $M = 12.0$ years, $SD = 1.5$), $F(1, 86) = 10.0$, $p < .01$.

[1] The Rasch model, through its mathematical form, defines a general unit called a *logit*. A person's ability in logits is his or her natural log odds for succeeding on items of the kind chosen to define the scale origin or "zero."

Materials: UPSIT

The UPSIT is a standardized, 40-item measure of odor-identification ability that is detailed elsewhere (Doty, 1995; Doty et al., 1984). Each item presents a scent-impregnated patch and a list of four answer choices. For example, one item reads, "This odor smells most like: (a) chocolate, (b) banana, (c) onion, (d) fruit punch." The UPSIT is scored dichotomously, with the overall score reflecting the sum total of all correct responses up to a maximum of 40 points. This raw score is then compared with age- and gender-matched standardization norms provided by the test manufacturer (Doty, 1995).

Procedure

All participants were administered the UPSIT under supervision, as outlined in the test manual (Doty, 1995). The examiner sat with each participant during completion of the UPSIT and demonstrated proper scratching of the first odor patch. Participants were instructed to scratch, sniff, and identify each scent by choosing one of four typed alternatives. In addition to circling their answer choices, participants were requested to verbalize their responses so that the examiner could check for accurate recording. In the event that participants were unable to initially select a response, they were asked to rescratch and smell the stimuli. In accordance with standardized procedure, that is, forced-choice paradigm, participants were told to select one answer choice regardless of whether they could perceive or identify the odor.

Statistical Analysis

Data from three female control participants were excluded from Rasch analysis on the basis of their having had perfect scores (maximum estimated measure), thus decreasing the sample size of the control group to 133 participants (69 women).

Of the three IRT models for items scored dichotomously, the one-parameter IRT model (named after its founder, Georg Rasch) is the simplest. The Rasch model consists of just one ability parameter (β_n) for each person (n) and one difficulty parameter (δ_i) for each item (i). These parameters represent the positions of persons and items on the latent variable they share. They are used in the model to determine the probability of person n succeeding on item i (Rasch, 1960, 1966a, 1966b; Wright, 1968). Raw-score data were analyzed with the Rasch model for dichotomies using WINSTEPS software (Linacre & Wright, 1998). The model was as follows:

$$\phi_{ni} = [\exp(\beta_n - \delta_i)]/[1 + \exp(\beta_n - \delta_i)],$$

where ϕ_{ni} is person n's probability of scoring 1 rather than 0 on item i, β_n is the ability of person n, and δ_i is the difficulty of item i (Wright & Masters, 1982).

Put into words, the ability (β_n) and difficulty (δ_i) parameters are combined by forming their difference ($\beta_n - \delta_i$). This difference governs the probability of what should happen when a respondent (n) pits his ability against the difficulty of an item (i). Given that either parameter can vary from minus infinity to plus infinity, so can their difference. However, probability must stay between zero and one. To accommodate this, one applies the difference ($\beta_n - \delta_i$) as the exponent of a base, $\exp(\beta_n - \delta_i)$, and this term is used in the ratio $\exp(\beta_n - \delta_i)/[1 + \exp(\beta_n - \delta_i)]$, which is the Rasch probability for a correct answer.

The Rasch model, through its mathematical form, defines a general unit called a *logit*. Logits are convenient to work with and are easily transformed into whatever applied units are subsequently defined. (Ultimately, the particular units applied to persons and items depend on how the variable is operationalized.) A person's ability in logits is the natural log odds for succeeding on items of the kind chosen to define the scale origin or "zero." Thus, the person's probability for succeeding on an item with difficulty $\delta_i = 0$ is $\exp\beta_n/(1 + \exp\beta_n)$, from which the success odds are $\phi/(1 - \phi) = \exp\beta_n$, the natural log of which is β_n. Similarly, an item's difficulty in logits is the natural log odds for failure on that item by persons with abilities at the scale origin. The probability of persons with abilities at $\beta = 0$ of succeeding on an item with difficulty δ is $\exp(-\delta_i)/[1 + \exp(-\delta_i)]$, from which their odds for failure are $(1 - \phi)/\phi = \exp\delta_i$, the natural log of which is δ_i.

The fit of data to the Rasch model can be evaluated by calculating how much is "left over" after the data have been used to estimate item difficulties δ_i and person abilities $\beta_n = \beta_r$, where r is the test score of person n. The standardized square of this residual after fitting the model is $\exp(\beta_n - \delta_i)$ for a wrong answer and $\exp(\delta_i - \beta_n)$ for a right one. The average degrees of freedom of each residual are $(L - 1)(N - 1)/LN$, where L is number of test items and N is number of respondents. These squared residuals can be summed over persons or items to form approximate chi-square-distributed variables for testing the fit of any particular item to any group of persons or of any individual person to any set of items.

A more extensive analysis of the response pattern of each person can be implemented by evaluating the way in which the person's residuals correlate with item difficulty, position, and type. For this, we can use standardized residuals in their unsquared form: $-\exp + (\beta_n - \delta_i)/2$ for a wrong answer and $\exp + (\beta_n - \delta_i)/2$ for a right one. Because these residuals are standardized—that is, centered at their expected mean and scaled by their expected standard deviation—their expected distribution can be modeled as approximately normal and their expected error variance as one.

Mathematical analysis shows the Rasch model to be statistically strong. It has estimators for its parameters, β_n and δ_i, that are sufficient, consistent, efficient, and unbiased (Andersen, 1970, 1972, 1973; Rasch, 1968). Numerical analysis supports simple approximations for estimating these parameters that are accurate enough for all practical purposes (Wright & Douglas, 1975). Experience has shown the model to be easy to apply in a wide variety of situations (Connolly, Nachtman, & Pritchett, 1971; Mead, 1976; Wilmott & Fowles, 1974; Woodcock, 1974). Technical details can be found in Wright and Panchapakesan (1969), Wright and Mead (1975, 1977), and Linacre and Wright (1998).

Results

Rasch item and person separation statistics are signal-to-noise ratios obtained by dividing the true spread of the measures by their measurement error (Linacre & Wright, 1998). They provide a statistical tool by which to evaluate the successful development of a variable (Wright & Stone, 1979). More specifically, item separation measures whether the item calibrations of the UPSIT are sufficiently separated along a line of increasing difficulty of odor identification. The greater the separation, the greater the range of odor-identification ability detected by the test. Item separation indices are 1.70 and 2.49 for patients and controls, respectively. These findings indicate adequate separation of the 40 test items to define a distinct construct of odor identification for both groups (see Table 1 for summary statistics).

Person-separation statistics indicate the UPSIT's capacity to discriminate odor-identification ability on the basis of the total number of items answered correctly. Person separation is the signal-to-noise ratio obtained by dividing the error-corrected sample standard deviation by the root mean square measurement error. Person-separation indices are 1.81 for patients and 0.96 for controls. Corresponding reliability indices, equivalent to Cronbach's alpha, are .77 and .48. However, these traditional reliability estimates are discounted by the skewed distributions of UPSIT scores. Therefore, standard error units have been used in Figure 1 to mark off the levels of significant difference in smell ability for each

Table 1
Summary Statistics for Patients With Schizophrenia (n = 54) and Healthy Controls (n = 133) on the University of Pennsylvania Smell Identification Test

	Group	
Summary statistic	Control	Schizophrenia
M (logits)	2.92	2.03
SD	1.05	1.26
SEM	0.09	0.17
Real RMSE[a]	0.76	0.61
Adjusted *SD*[b]	0.72	1.10
Person separation[c]	0.96	1.81
Person reliability	0.48	0.77
Item separation[c]	2.49	1.70
Item reliability	0.86	0.74

[a] Real root-mean-square error (RMSE) is the modeled standard error of measurement (*SEM*) inflated by the misfit encountered in these data. [b] Adjusted *SD* is the observed standard deviation corrected for the variance due to measurement error. [c] Separation is the signal-to-noise ratio of the adjusted *SD* divided by the RMSE.

sample. A distance of 3 standard errors implies a significant difference between any pair of estimates at the 95% confidence level. In Figure 1, the two bold-faced horizontal lines indicate these significant differences. It is clear that the UPSIT discriminates three levels of odor-identification ability among both schizophrenic and healthy respondents. According to the UPSIT manual, respondents can be separated into five levels of olfactory ability on the basis of raw UPSIT score, age, and gender. In Figure 1, four horizontal lines mark these cutoffs; corresponding olfactory diagnoses are listed.

With regard to the five olfactory diagnoses, that is, relative to published age- and gender-matched normative data for the UPSIT, 26 participants (9 patients, 17 controls) qualified for mild microsmia (decreased smell ability), 14 (8 patients, 6 controls) for moderate microsmia, 9 (7 patients, 2 controls) for severe microsmia, and 2 (1 patient, 1 control) for anosmia (complete inability to perceive qualitative odors). The remaining 136 participants (29 patients, 107 controls) qualified for normal smell identification ability. Thus, 20% of controls and 46% of patients would be classified as olfactory deficient according to age- and gender-matched UPSIT norms.

Figure 1 illustrates the extent to which controls ($M = 2.92$ logits, $SD = 1.05$) manifest better odor-identification ability than patients ($M = 2.03$ logits, $SD = 1.26$). The two distributions overlap substantially, such that no single cutoff point excludes all controls while also detecting most patients. Even the patient mean (2.03 logits) subsumes 26 supposed controls. Nevertheless, the skewed distribution among controls suggests that the UPSIT is clinically efficacious for detecting profound odor-identification deficits in otherwise healthy populations.

Hawkes and colleagues (Hawkes, Shephard, & Daniel, 1997) administered the UPSIT to patients with Parkinson's disease and found that some items (lemon, pizza, wintergreen, rose, clove) were more readily misidentified than others. Pizza was reported as the single most discriminant item for these patients. Table 2 lists UPSIT items in descending order of item difficulty as perceived by the control group, with Hawkes et al.'s (1997) findings listed in the rightmost column. The frequency of correct responses per item, along with the total number of responses, is indicated for the schizophrenia sample. Patients with schizophrenia experienced the following items as most difficult (1 *SD* above the mean): cheese, turpentine, fruit punch, lime, clove, gingerbread, lemon, cinnamon, menthol, and soap. It is clear from Table 2 that those items reported to be most difficult for patients with Parkinson's disease are not consistently difficult for patients with schizophrenia. Instead, their item difficulties range from 1.16 (hard) to −1.10 (easy) for the schizophrenia sample.

When item difficulties were contrasted for patients versus controls, 13 items showed significant displacement from normal performance, indicating schizophrenic deficit. However, significant displacement from normal performance could be attributable to individual variation among patients with schizophrenia, rather than a general deficit of the patient group regarding a specific item. Misfit can assess idiosyncratic variation within the patient group. Misfit occurs when there is a discrepancy between actual response patterns and those that are expected on the basis of the model (i.e., when an item accumulates unlikely responses across participants or when a person accumulates unlikely responses across items). Its expected value is 1.0, with larger values indicating greater individual variation.

To examine the influence of misfit on patient displacement from controls for each item, we adjusted displacement for differences in means between groups. Table 3 lists UPSIT items in descending order of displacement, so that patient performances most unlike those of controls are at the top. Thirteen items, indicated by asterisked displacement values, show statistically significant deviations from the performance of healthy respondents. Significant misfit characterizes most of these items, indicated by asterisked values in the infit column. However, two items show statistically significant displacement in the absence of significant misfit, indicating diagnostic agreement among patients with schizophrenia about the amount of deficit. These items (i.e., banana, cheese) have the greatest discriminating potential.

Patients ($n = 54$, $M = 31.5$, $SD = 6.4$) showed significantly reduced raw-score accuracy on the UPSIT relative to controls ($n = 136$, $M = 35.7$, $SD = 3.8$), $F(1, 186) = 23.16$, $p < .01$. There was a diagnosis-specific effect of smoking on UPSIT ability, such that patients who smoked ($n = 27$, $M = 28.0$, $SD = 6.2$) showed significantly reduced UPSIT performance relative to those who did not ($n = 27$, $M = 35.1$, $SD = 4.4$), $F(1, 186) = 11.04$, $p = .001$. Raw UPSIT score and age were correlated at $r = -.51$ for patients ($n = 54$) and $r = .01$ for controls ($n = 136$). With regard to raw-score UPSIT accuracy, the main effect of diagnosis remained significant when controls ($n = 45$, $M = 35.6$, $SD = 3.7$) were matched to patients ($n = 45$, $M = 32.6$, $SD = 5.7$) on gender and age, $F(1, 86) = 5.03$, $p < .05$.

Discussion

The results of the present study verify that the items of the UPSIT measure a single, recognizable construct of olfactory ability in healthy controls and chronic, medicated outpatients with schizophrenia (see Table 1). Figure 1 provides a graphical illustration of this variable as measured by the UPSIT. Item difficulty and person ability are independently delineated along a shared continuum, with items spread over this line from easy to hard and

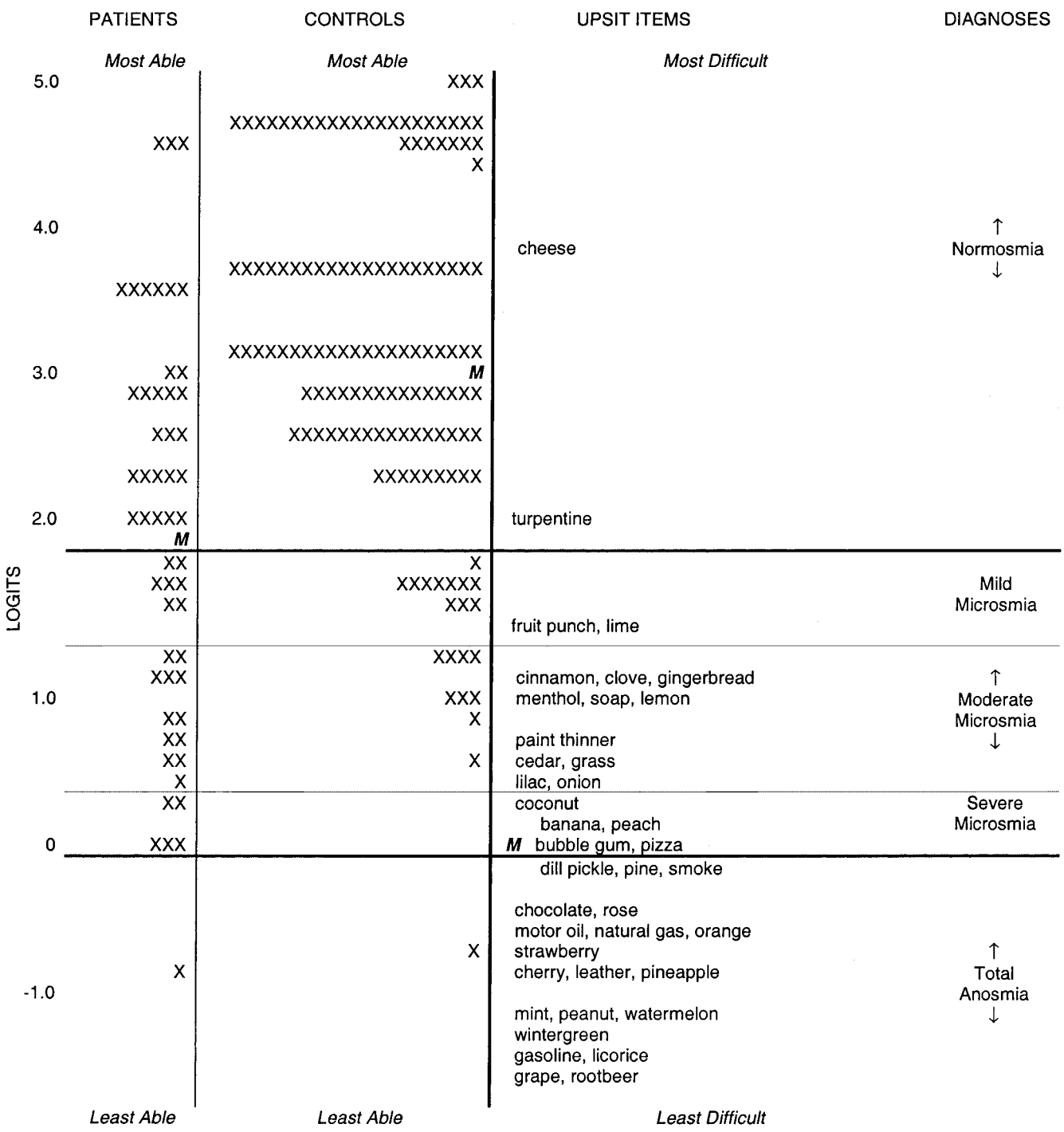

Figure 1. Person ability and item difficulty on the University of Pennsylvania Smell Identification Test (UPSIT) for patients with schizophrenia (n = 54) and healthy controls (n = 133). Person ability and item difficulty are measured in logits and plotted relative to one another. The right-most distribution is that of the 40 UPSIT items, with higher logits (at top) representing greater difficulty. The left-most columns are ability distributions for each group, with higher logits (at top) representing greater ability level. "X" represents one person. "M" marks the mean of each distribution. The horizontal lines are Doty's (1995) diagnostic cutoffs; corresponding diagnoses are listed in the right-most column. The measurement error of 0.8 justified three, rather than five, significantly different strata of olfactory identification ability in these data (represented by bold horizontal lines).

Table 2
Item Difficulty for Control (CL), Schizophrenia (SZ), and Parkinson's Disease (PD) Groups

	CL (n = 133)		SZ (n = 54)		PD (n = 96)
UPSIT item[a]	Item difficulty[b]	*SEM*	Score[c]	Count[d]	Most difficult[e]
Cheese	3.48	0.36	10	53	
Turpentine	1.99	0.32	33	53	
Fruit punch	1.43	0.32	37	54	
Lime	1.43	0.32	40	53	
Clove	1.16	0.33	43	54	Clove
Gingerbread	1.16	0.33	39	54	
Lemon	1.16	0.33	36	53	Lemon
Cinnamon	1.10	0.33	36	54	
Menthol	1.04	0.33	36	54	
Soap	1.04	0.34	40	53	
Paint thinner	0.78	0.35	46	53	
Cedar	0.63	0.35	40	54	
Grass	0.55	0.36	39	53	
Onion	0.38	0.37	49	54	
Lilac	0.38	0.37	44	53	
Coconut	0.28	0.38	46	54	
Peach	0.18	0.38	41	53	
Banana	0.07	0.39	42	54	
Pizza	−0.04	0.40	46	54	Pizza
Bubble gum	−0.04	0.40	46	54	
Dill pickle	−0.17	0.42	46	53	
Smoke	−0.17	0.42	40	53	
Pine	−0.17	0.42	39	52	
Chocolate	−0.46	0.45	44	54	
Rose	−0.46	0.45	38	53	Rose
Orange	−0.63	0.47	45	53	
Motor oil	−0.64	0.49	45	52	
Strawberry	−0.64	0.47	49	54	
Natural gas	−0.64	0.47	50	53	
Leather	−0.84	0.50	48	54	
Cherry	−0.85	0.50	47	54	
Pineapple	−0.85	0.51	44	53	
Mint	−1.10	0.55	44	54	
Wintergreen	−1.10	0.55	49	53	Wintergreen
Watermelon	−1.10	0.55	47	52	
Peanut	−1.10	0.55	45	53	
Licorice	−1.41	0.61	50	54	
Gasoline	−1.41	0.62	47	52	
Grape	−1.85	0.73	46	53	
Root beer	−2.58	0.99	50	53	
M (*SD*)	0 (1.15)	0.44 (.13)	43 (7)	53 (1)	

Note. UPSIT = University of Pennsylvania Smell Identification Test.
[a] Arranged in descending order of item difficulty according to CL group (i.e., cheese is hardest). [b] Average item difficulty = 0. [c] Score = number of correct responses per item. [d] Count = total number of responses (n = 54). [e] Items most frequently misidentified by PD patients (Hawkes et al., 1997).

persons from inept to superior. The more items there are positioned near a person's ability, the more precisely that person can be measured. Uniform coverage of the continuum without much space between items would be optimal. Instead, the item map shows that the UPSIT does not provide sufficient coverage of olfactory identification at the high end. There are two wide item gaps (between "cheese" and "turpentine"; "turpentine" and "fruit punch") in the region where the majority of patients and controls fall. These gaps show that the test does not incorporate enough difficult items to obtain accurate measurement in the upper range of olfactory identification.

The UPSIT is recommended as a tool to screen sensory panels, such as those in the food and beverage industries, requiring highly developed olfactory sense (Doty, 1995). Considering the ceiling effect in Figure 1, along with the wide item gaps, the utility of the UPSIT in making such distinctions seems limited. Indeed, more than half of the controls have an ability estimate above that of the hardest item. This limits the capacity of the UPSIT to distinguish average from superior olfactory ability. On the other hand, the skewed distribution among controls supports the test's utility in detecting odor-identification deficit in otherwise healthy populations.

Construct utility is supported by evidence that a test's items define a variable with sufficient coverage to separate persons on the basis of number correct. According to the UPSIT manual, persons can be separated into five ability levels on the basis of raw

Table 3
Item-by-Item Accuracy for Patients With Schizophrenia (SZ) Versus Controls (CL) on the University of Pennsylvania Smell Identification Test (UPSIT)

	CL (n = 133)	SZ (n = 54)		
UPSIT item	Item difficulty	*SE*	Infit MNSQ[a]	Displacement from CL
Grape	−1.85	0.73	3.05*	2.34*
Rose	−0.46	0.45	2.11*	2.14*
Mint	−1.10	0.55	2.26*	2.10*
Root beer	−2.58	0.99	2.80*	2.03
Peanut	−1.10	0.55	1.83*	1.80*
Pineapple	−0.85	0.51	1.59*	1.72*
Pine	−0.17	0.42	1.51*	1.64*
Smoke	−0.17	0.42	1.31*	1.62*
Gasoline	−1.41	0.62	1.74*	1.53
Chocolate	−0.46	0.45	1.26*	1.48*
Cherry	−0.85	0.50	1.20*	1.36
Orange	−0.63	0.47	1.30*	1.33
Cheese	3.48	0.36	1.02	1.31*
Motor oil	−0.64	0.49	1.09	1.26
Banana	0.07	0.39	1.04	1.24*
Licorice	−1.41	0.61	1.13	1.21
Watermelon	−1.10	0.55	0.99	1.18
Leather	−0.84	0.50	1.19	1.15
Peach	0.18	0.38	1.53*	1.13*
Grass	0.55	0.36	1.21*	1.02*
Menthol	1.04	0.33	1.39*	0.98*
Cedar	0.63	0.35	0.99	0.93
Cinnamon	1.10	0.33	1.30*	0.92
Wintergreen	−1.10	0.55	0.85	0.89
Lemon	1.16	0.33	0.75	0.76
Pizza	−0.04	0.40	0.80	0.72
Bubble gum	−0.04	0.40	0.94	0.72
Strawberry	−0.64	0.47	0.83	0.70
Dill pickle	−0.17	0.42	0.84	0.67
Gingerbread	1.16	0.33	1.08	0.51
Fruit punch	1.43	0.32	1.04	0.47
Lilac	0.38	0.37	0.87	0.46
Soap	1.04	0.34	1.18	0.38
Coconut	0.28	0.38	0.75	0.38
Turpentine	1.99	0.32	1.40*	0.26
Natural gas	−0.64	0.47	0.81	0.04
Lime	1.43	0.32	0.89	−0.01
Clove	1.16	0.33	0.80	−0.03
Paint thinner	0.78	0.35	0.92	−0.35
Onion	0.38	0.37	0.59	−0.41
Mean (*SD*)	0 (1.15)	0.44 (.13)	1.26 (.53)	

Note. After adjusting for differences in group means, UPSIT items are arranged in descending order of displacement from normal performance (i.e., patients' perception of grape was most unlike that of controls). Asterisks indicate statistically significant values. Infit MNSQ = measure of amount of distortion, or randomness, in the measurement system. Expected values are 1.0. Values ranging from .5 to 1.5 are productive for measurement.
[a] Expected value = 1.0
* $p < .05$.

UPSIT score, age, and gender (Doty, 1995). In Figure 1, horizontal lines mark these cutoffs; corresponding olfactory diagnoses are listed. For the current sample, the UPSIT did not separate respondents into five statistically significant levels of olfactory diagnosis. Rather, it identified three statistically distinct strata with cutoffs that are indicated in Figure 1 by bold-faced horizontal lines. These ability levels roughly correspond to Doty's normosmia (normal olfactory ability), microsmia (decreased olfactory ability), and anosmia (complete inability to perceive qualitative odors).

When patient performance on the UPSIT was contrasted with that of controls, thirteen items showed statistically significant deviation independent of item difficulty. For eleven of these items, displacement is attributable to individual variation among patients (i.e., misfit). However, two items show significant displacement without significant misfit. These items differentially discriminate patients with schizophrenia from controls (see Table 3).

It is interesting to note that those items reported to be most problematic for patients with Parkinson's disease (Hawkes et al.,

1997) are not equally problematic for schizophrenia patients in our sample (see Table 2). This suggests that patients with schizophrenia perceive some items of the UPSIT differently than do patients with Parkinson's disease, despite the fact that the two groups show raw-score reductions within the same range. The reasons underlying this group discrepancy are unclear; there are many possibilities. The UPSIT is multiple-choice format with distractors repeated throughout. Target scents differ in intensity, pleasantness, and familiarity. Therefore, olfactory acuity, attention, memory, and executive function might each contribute to unique UPSIT profiles for individual patient groups. This begs the question as to whether schizophrenia patients' UPSIT errors represent a pure odor-identification deficit. Clearly, a thorough examination of the qualitative nature of patient performance on this test is warranted. UPSIT items must be reevaluated for accuracy of presentation and relevance of "incorrect" choices before reduced scores can be considered a specific deficit of odor identification in schizophrenia.

The present study found significantly reduced UPSIT performance in patients with schizophrenia when groups were compared for (a) raw-score accuracy and (b) olfactory diagnoses relative to published age- and gender-matched normative data (Doty, 1995). It should be noted that our control sample's raw-score average is approximately one point below that typically reported (Houlihan et al., 1994; Hurwitz et al., 1988; Kopala et al., 1994; Malaspina et al., 1994; Moberg, Doty, Turetsky, & Arnold, 1997; Seidman et al., 1992; Seidman et al., 1995; Wu et al., 1993). Our control sample is considerably larger, and somewhat older, than those in previous studies. However, age does not appear to be a mediating factor with regard to the current findings. First, age-related decline in olfactory ability is reported to begin in the 7th decade of life, well beyond the age of any of our participants (Doty et al., 1984). Second, age and UPSIT accuracy are correlated at $r = .01$ for controls ($n = 136$) and $r = -.51$ for patients ($n = 54$). Third, when a subset of controls ($n = 45$) and gender- and age-matched schizophrenia patients ($n = 45$) were compared, patient performance on the UPSIT remained significantly impaired.

Although the UPSIT deficit is not attributable to advanced age, it may be secondary to other factors inherent to the patient condition. A recent meta-analytic review found no significant effects of gender ($N = 18$ studies), medication status (medicated, unmedicated, mixed groups; $N = 18$ studies), or smoking status (smokers, nonsmokers; $N = 11$ studies) on UPSIT performance in schizophrenia (Moberg et al., 1999). Our patient sample included 43 men and 11 women—all were medicated. Results of the original (136 controls, 54 patients) and matched-sample (45 controls, 45 patients) analyses support Moberg et al.'s (1999) findings of no gender differences in UPSIT ability. However, in contrast to the findings of Moberg et al. (1999), our results show significantly reduced UPSIT performance in patients who smoke relative to those who do not. Clearly, further examination of this potential moderator variable is warranted.

There is some question regarding the specificity of the olfactory deficit in schizophrenia and whether it might reflect generalized cognitive impairment (Serby et al., 1990). Previous findings suggest that task complexity (Kopala, Good, Martzke, & Hurwitz, 1995), fatigue (Hurwitz et al., 1988), and executive–attentional components (Seidman et al., 1992; Seidman et al., 1997) are not responsible for reduced odor-identification performance among patients with schizophrenia. Similarly, our findings suggest that patient impairment on the UPSIT is not secondary to reduced attentional capacity, lack of knowledge, or fatigue. All participants were administered the UPSIT under supervision, on an item-by-item basis, thus limiting the possibility of errors due to inattentiveness. General knowledge was likely not a complicating factor considering that the patient group had an average education of 12.1 years. Finally, item difficulties for patients and controls do not change as a function of presentation sequence; this error pattern is inconsistent with a fatigue effect.

Although UPSIT impairment appears to be unrelated to attentional or executive factors in schizophrenia, it may nonetheless reflect something other than pure olfactory identification deficit. Although the UPSIT is considered a suprathreshold measure of olfactory ability, raw scores are reported to be highly associated with olfactory acuity in healthy controls (Doty et al., 1995). Few studies have evaluated odor acuity (Bradley, 1984; Geddes et al., 1991; Isseroff et al., 1987; Kopala et al., 1989; Kopala et al., 1992; Serby et al., 1990) or discrimination (Dunn & Weller, 1989; Sreenivasan, Abraham, & Verghese, 1987) in schizophrenia. It is therefore uncertain to what extent peripheral olfactory dysfunction accounts for reduced patient performance in previous findings as well as those of the current study. Future research is necessary to establish absence of impairment at the lower levels of olfactory processing in schizophrenia.

There are a number of methodological limitations of this study that warrant consideration. First, we did not assess olfactory threshold. We are therefore unable to empirically assess whether some items are measuring odor detection rather than identification. Considering that many respondents said they could not perceive the hardest item's odorant, this item (No. 14, cheese) may be biased against persons with reduced olfactory acuity. Second, our findings support an effect of smoking status on UPSIT ability in schizophrenia. It would therefore have been interesting to examine dose-effects of this variable by calculating pack-years for all current and former smokers (see Frye, Schwartz, & Doty, 1990). Third, because we did not record medication history, medication type, and dose equivalents (e.g., mean daily chlorpromazine equivalent doses), we cannot evaluate the impact of specific medications, dose-effects, and drug interactions on olfactory ability in schizophrenia. However, evidence of reduced UPSIT performance in healthy schizotypal individuals who are medication free suggests that antipsychotic medication may not be a major factor in UPSIT deficit (Park & Schoppe, 1997). Finally, given that there is evidence that illness duration may be an important moderating variable in UPSIT performance for patients with schizophrenia (Moberg, Doty, Turetsky, & Arnold, 1997), we plan to collect this information in future studies.

In summary, the present findings confirm that the UPSIT adequately defines a single construct of olfactory ability for healthy and schizophrenic respondents. It spans 8 standard errors of measurement along this construct, thus defining three statistically distinct ability levels for both groups. It is interesting to note that some items were found to discriminate patients with schizophrenia from controls, as well as from patients with Parkinson's disease. It is therefore uncertain to what extent patient errors on the UPSIT represent a specific deficit in olfactory identification. A complete understanding of reduced performance on this task is necessary before it can be accepted as a diagnostic indicator of olfactory

deficit in any patient group. Future research is warranted to investigate the nature of UPSIT errors committed by diverse patient populations.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society, 32,* 283–301.

Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society, 34*(1), 42–54.

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38*(1), 123–140.

Bertollo, D. N., Cowen, M. A., & Levy, A. V. (1996). Hypometabolism in olfactory cortical projection areas of male patients with schizophrenia: An initial positron emission tomography study. *Psychiatry Research, 60*(2–3), 113–116.

Bradley, E. A. (1984). Olfactory acuity to a pheromonal substance and psychotic illness. *Biological Psychiatry, 19,* 899–905.

Brewer, W. J., Edwards, J., Anderson, V., Robinson, R., & Pantelis, C. (1996). Neuropsychological, olfactory, and hygiene deficits in men with negative symptom schizophrenia. *Biological Psychiatry, 40,* 1021–1031.

Cain, W. S. (1979, February 2). To know with the nose: Keys to odor identification. *Science, 203*(4379), 467–470.

Clark, C., Kopala, L., Hurwitz, T., & Li, D. (1991). Regional metabolism in microsmic patients with schizophrenia. *Canadian Journal of Psychiatry, 36,* 645–650.

Connolly, A. J., Nachtman, W., & Pritchett, E. M. (1971). *Keymath: Diagnostic Arithmetic Test*. Circle Pines, MN: American Guidance Service.

Cummings, J. L. (1993). Frontal-subcortical circuits and human behaviour. *Archives of Neurology, 50,* 873–880.

Doty, R. L. (1995). *The Smell Identification Test administration manual.* Haddon Heights, NJ: Sensonics.

Doty, R. L., McKeown, D. A., Lee, W. W., & Shaman, P. (1995). A study of the test–retest reliability of ten olfactory tests. *Chemical Senses, 20,* 645–656.

Doty, R. L., Newhouse, M. G., & Azzalina, J. D. (1985). Internal consistency and short-term test–retest reliability of the University of Pennsylvania Smell Identification Test. *Chemical Senses, 10,* 297–300.

Doty, R. L., Shaman, P., & Dann, M. (1984). Development of the UPSIT: A microencapsulated test of olfactory function. *Physiological Behavior, 32,* 489–502.

Dunn, T., & Weller, M. (1989). Olfaction in schizophrenia. *Perceptual and Motor Skills, 69,* 833–834.

Eslinger, P., Damasio, A., & van Hoesen, G. (1982). Olfactory dysfunction in men: Anatomical and behavioural aspects. *Brain Cognition, 1,* 259–285.

Frye, R. E., Schwartz, B. S., & Doty, R. L. (1990). Dose-related effects of cigarette smoking on olfactory function. *JAMA, 263,* 1233–1236.

Geddes, J., Huws, R., & Pratt, P. (1991). Olfactory acuity in the positive and negative syndromes of schizophrenia. *Biological Psychiatry, 29,* 774–778.

Good, K. P., Martzke, J. S., Frankland, B. W., & Kopala, L. C. (1998). Deterioration of olfactory identification abilities in patients with schizophrenia. *American Journal of Psychiatry, 155*(10), 1463.

Good, K. P., Martzke, J. S., Honer, W. G., & Kopala, L. C. (1998). Left nostril olfactory identification impairment in a subgroup of male patients with schizophrenia. *Schizophrenia Research, 33*(1–2), 35–43.

Hawkes, C. H., Shephard, B. C., & Daniel, S. E. (1997). Olfactory dysfunction in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry, 62,* 436–446.

Houlihan, D. J., Flaum, M., Arnold, S. E., Keshavan, M., & Alliger, R. (1994). Further evidence for olfactory identification deficits in schizophrenia. *Schizophrenia Research, 12*(2), 179–182.

Hurwitz, T. A., Kopala, L., Clark, C., & Jones, B. (1988). Olfactory deficits in schizophrenia. *Biological Psychiatry, 23,* 123–128.

Isseroff, R. G., Stoler, M., Ophir, D., Lancet, D., & Sirota, P. (1987). Olfactory sensitivity to androstenone in schizophrenic patients. *Biological Psychiatry, 22,* 922–925.

Kopala, L., Clark, C., & Hurwitz, T. A. (1989). Sex differences in olfactory function in schizophrenia. *American Journal of Psychiatry, 146*(10), 1320–1322.

Kopala, L. C., Clark, C. M., & Hurwitz, T. (1992). Olfactory deficits in neuroleptic naive patients with schizophrenia. *Schizophrenia Research, 8*(3), 245–250.

Kopala, L. C., Good, K. P., & Honer, W. G. (1994). Olfactory hallucinations and olfactory identification ability in patients with schizophrenia and other psychiatric disorders. *Schizophrenia Research, 12*(3), 205–211.

Kopala, L. C., Good, K., & Honer, W. G. (1995). Olfactory identification ability in pre- and post-menopausal women with schizophrenia. *Biological Psychiatry, 38,* 57–63.

Kopala, L., Good, K., Martzke, J., & Hurwitz, T. (1995). Olfactory deficits in schizophrenia are not a function of task complexity. *Schizophrenia Research, 17*(2), 195–199.

Linacre, J. M., & Wright, B. D. (1998). *A user's guide to Winsteps, Rasch Model Computer Program* (Version 2.81). Chicago: MESA Press.

Malaspina, D., Lignelli, A., Pererra, G., Marshall, R., Gorman, J., & van Heertum, R. (1996). *SPECT brain activation by odor discrimination in schizophrenia and controls.* Paper presented at the Academy of Neuropsychopharmacology, San Juan, Puerto Rico.

Malaspina, D., Wray, A. D., Freidman, J. H., Amador, X., Yale, S., Hasan, A., et al. (1994). Odor discrimination deficits in schizophrenia: Association with eye movement dysfunction. *The Journal of Neuropsychiatry and Clinical Neurosciences, 6,* 273–278.

Mead, R. L. (1976). *Assessing the fit of data to the Rasch model through analysis of residuals.* Unpublished doctoral dissertation, University of Chicago.

Moberg, P. J., Agrin, R., Gur, R. E., Gur, R. C., Turetsky, B. I., & Doty, R. L. (1999). Olfactory dysfunction in schizophrenia: A qualitative and quantitative review. *Neuropsychopharmacology, 21*(3), 325–340.

Moberg, P. J., Doty, R. L., Mahr, R. N., Mesholam, R. I., Arnold, S. E., Turetsky, B. I., & Gur, R. E. (1997). Olfactory identification in elderly schizophrenia and Alzheimer's disease. *Neurobiology of Aging, 18*(2), 163–167.

Moberg, P. J., Doty, R. L., Turetsky, B. I., & Arnold, S. E. (1997). Olfactory identification deficits in schizophrenia: Correlation with duration of illness. *American Journal of Psychiatry, 154*(7), 1016–1018.

Pansky, B., & Allen, D. (1980). *Review of neuroscience.* New York: MacMillan Press.

Pantelis, C., & Brewer, W. (1995). Neuropsychological and olfactory dysfunction in schizophrenia: Relationship of frontal syndromes to syndromes of schizophrenia. *Schizophrenia Research, 17*(1), 35–45.

Park, S., & Schoppe, S. (1997). Olfactory identification deficit in relation to schizotypy. *Schizophrenia Research, 26*(2–3), 191–197.

Potter, H., & Butters, N. (1980). An assessment of olfactory deficits in patients with damage to prefrontal cortex. *Neuropsychologia, 18,* 621–628.

Price, J. (1990). Olfactory system. In G. Paxinos (Ed.), *The human nervous system* (pp. 390–403). San Diego, CA: Academic Press.

Purdon, S. E. (1998). Olfactory identification and Stroop interference converge in schizophrenia. *Journal of Psychiatry & Neuroscience, 23*(3), 163–171.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Danmarks Paedogogiske Institut.

Rasch, G. (1966a). An individualistic approach to item analysis. In P. F. Lazarsfeld & N. W. Henry (Eds.), *Readings in mathematical social science* (pp. 89–108). Chicago: Science Research Associates.

Rasch, G. (1966b). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology, 19*(1), 49–57.

Rasch, G. (1968). *A mathematical theory of objectivity and its consequences for model construction.* Report from European Meeting on Statistics, Econometrics and Management Sciences, Amsterdam.

Rausch, R., & Serafetinides, E. A. (1975, June). Specific alterations of olfactory function in humans with temporal lobe lesions. *Nature, 255,* 557–558.

Rupp, C. I. (2003). Dysfunctions in olfactory processing in schizophrenia. *Current Opinion in Psychiatry, 16,* 181–186.

Saoud, M., Hueber, T., Mandran, H., Dalery, J., & d'Amato, T. (1998). Olfactory identification deficiency and WCST performance in men with schizophrenia. *Psychiatry Research, 81*(2), 251–257.

Seidman, L. J., Goldstein, J. M., Goodman, J. M., Koren, D., Turner, W. M., Faraone, S. V., & Tsuang, M. T. (1997). Sex differences in olfactory identification and Wisconsin Card Sorting performance in schizophrenia: Relationship to attention and verbal ability. *Biological Psychiatry, 42,* 104–115.

Seidman, L. J., Kalinowski, A. G., Kremen, W. S., Faraone, S. V., Oscar-Berman, M., Ajilore, O., & Tsuang, M. T. (1995). Experimental and clinical neuropsychological measures of prefrontal dysfunction in schizophrenia. *Neuropsychology, 9*(4), 481–490.

Seidman, L. J., Talbot, N. L., Kalinowski, A. G., McCarley, R. W., Faraon, S. V., Kremen, W. S., et al. (1992). Neuropsychological probes of fronto-limbic system dysfunction in schizophrenia: Olfactory identification and Wisconsin Card Sorting performance. *Schizophrenia Research, 6*(1), 55–65.

Serby, M., Larson, P., & Kalkstein, D. (1990). Olfactory sense in the psychoses. *Biological Psychiatry, 28,* 830.

Spitzer, R. L., & Endicott, J. (1978). Schedule for Affective Disorders and Schizophrenia—Lifetime Version. *Archives of General Psychiatry, 35,* 837–844.

Sreenivasan, K., Abraham, A., & Verghese, A. (1987). Right temporal lobe functions in psychiatric disorders. *Indian Journal of Clinical Psychology, 14,* 40–42.

Wilmott, A., & Fowles, D. (1974). *The objective interpretation of test performance: The Rasch model applied.* Atlantic Highlands, NJ: NFER.

Woodcock, R. W. (1974). *Woodcock Reading Mastery Tests*. Circle Pines, MN: American Guidance Service.

Wright, B. D. (1968). Sample-free test calibration and person measurement. In B. Bloom (Ed.), *Proceedings of the ETS 1967 Invitational Conference on Testing Problems* (pp. 85–101). Princeton, NJ: Educational Testing Service.

Wright, B. D., & Douglas, G. A. (1975). *Better procedures for sample-free item analysis* (Research Memorandum No. 20). Chicago: University of Chicago, Department of Education, Statistical Laboratory.

Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis: Rasch Measurement*. Chicago: MESA Press.

Wright, B. D., & Mead, R. I. (1975). *CALFIT: Sample-free calibration with a Rasch measurement model.* (Research Memorandum No. 18) Chicago: University of Chicago, Department of Education, Statistical Laboratory.

Wright, B. D., & Mead, R. I. (1977). *BICAL: Calibrating items and scales with the Rasch model* (Research Memorandum No. 23). Chicago: University of Chicago, Department of Education, Statistical Laboratory.

Wright, B. D., & Panchapakesan, N. A. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement, 29,* 23–48

Wright, B. D., & Stone, M. H. (1979). *Best Test Design.* Chicago: Mesa Press.

Wu, J., Buchsbaum, M., Moy, K., Denlea, N., Kesslak, P., Tseng, H., et al. (1993). Olfactory memory in unmedicated schizophrenics. *Schizophrenia Research, 9*(1), 41–47.

Received April 12, 2002
Revision received September 16, 2003
Accepted September 26, 2003 ■