

Factorial validity, measurement equivalence and cognitive performance of the Cambridge Neuropsychological Test Automated Battery (CANTAB) between patients with first-episode psychosis and healthy volunteers

L. Haring^{1*}, R. Möttus^{2,3}, K. Koch¹, M. Trei^{3,4} and E. Maron^{1,5}

¹Psychiatry Clinic of Tartu University Hospital, Tartu, Estonia

²Department of Psychology, University of Edinburgh, Edinburgh, UK

³Department of Psychology, University of Tartu, Tartu, Estonia

⁴Psychiatry Clinic of North Estonia Medical Centre, Tallinn, Estonia

⁵Centre for Mental Health, Imperial College London, London, UK

Background. The purpose of this study was to use selected Cambridge Neuropsychological Test Automated Battery (CANTAB) tests to examine the dimensional structure of cognitive dysfunction in first episode of psychosis (FEP) patients compared with cognition in healthy subjects.

Method. A total of 109 FEP patients and 96 healthy volunteers were administered eight CANTAB tests of cognitive function. Principal components analysis (PCA) was used to estimate dimensionality within the test results. The dimensions identified by the PCA were assumed to reflect underlying cognitive traits. The plausibility of latent factor models was estimated using confirmatory factor analysis (CFA). Multi-group CFA (MGCFA) was used to test for measurement invariance of factors between groups. The nature and severity of cognitive deficits amongst patients as opposed to controls were evaluated using a general linear model.

Results. Amongst subjects PCA identified two underlying cognitive traits: (i) a broad cognitive domain; (ii) attention/memory and executive function domains. Corresponding CFA models were built that fitted data well for both FEP patients and healthy volunteers. As in MGCFA latent variables appeared differently defined in patient and control groups, differences had to be ascribed using subtest scores rather than their aggregates. At subtest score level the patients performed significantly worse than healthy subjects in all comparisons ($p < 0.001$).

Conclusions. Results of this study demonstrate that the structure of underlying cognitive abilities as measured by a selection of CANTAB tests is not the same for healthy individuals and FEP patients, with patients displaying widespread cognitive impairment.

Received 3 April 2014; Revised 13 November 2014; Accepted 21 November 2014; First published online 29 December 2014

Key words: CANTAB, cognitive function, first-episode psychosis, measurement invariance.

Introduction

Abnormalities in cognitive function have long been recognized as one of the key features of schizophrenia spectrum disorders (Saykin *et al.* 1991; Gold & Harvey, 1993), but the specific nature of this dysfunction is still not fully understood. Although there is substantial cognitive heterogeneity among schizophrenia patients (Joyce *et al.* 2005), the typical cognitive deficit associated with chronic psychotic disorders tends to be at least

moderate (Bilder *et al.* 2000; Reichenberg *et al.* 2002; Addington *et al.* 2003). Cognitive deficits in schizophrenia patients may be related to underlying neuronal dysfunction (Kéri & Janka, 2004). Importantly, cognitive deficits may be already present before the onset of the schizophrenia (Bilder *et al.* 2000) and tend to remain relatively stable from the first episode of psychosis (FEP; typically in young adulthood) through to late middle age (Heaton *et al.* 2001).

It is unclear whether schizophrenia-related cognitive deficit could be characterized as a generalized process cutting across all cognitive domains (Dickinson *et al.* 2004; Leeson *et al.* 2009a) or as a set of relatively independent deficits in different cognitive domains (Saykin

* Address for correspondence: L. Haring, Psychiatry Clinic of Tartu University Hospital, Tartu, Estonia.
(Email: Liina.Haring@kliinikum.ee)

et al. 1991; Hutton *et al.* 1998), but this question might be answered by investigating the profile of cognitive deficits across multiple domains. The most frequently used method for neuropsychological data reduction involves grouping cognitive tests into conventional domains (e.g. attention, memory, executive functioning) and averaging the scores of individual standardized tests within each domain (Jaeger *et al.* 2003). An analytical study of MATRICS (Measurement and Treatment Research to Improve Cognition in Schizophrenia) data identified speed of processing, attention/vigilance, working memory, verbal learning and memory, visual learning and memory, and reasoning and problem solving as the cognitive factors that best mark the fundamental dimensions of cognitive deficit in schizophrenia sufferers (Green *et al.* 2004).

Despite evidence that schizophrenia-related cognitive impairments can be mapped into conventional domains (Genderson *et al.* 2007; Dickinson *et al.* 2011), other studies have revealed that domains overlap and are not clearly distinguishable (Gold *et al.* 1997). Schizophrenia-related cognitive impairment may have a hierarchical structure akin to how cognitive abilities are usually conceptualized in healthy people. Specific cognitive functions can be subsumed under a general cognitive (impairment) factor (Deary *et al.* 2010) and may need to be understood within a background of a more general cognitive decline. One aim of the present study was to assess the factorial structure of cognitive functioning among patients suffering from FEP and compare it with that in healthy controls.

Results of factorial analyses depend on the variables being analysed. This study is based on the computer-interfaced Cambridge Neuropsychological Test Automated Battery (CANTAB) tests (Robbins & Sahakian, 1994), which have been extensively validated for assessing brain-behaviour relationships in adult populations (Robbins *et al.* 1994, 1998) and shown to be sensitive to brain dysfunctions of psychiatric disorders including schizophrenia (Elliott *et al.* 1995; Pantelis *et al.* 1997; Stip *et al.* 2008) and FEP (Hutton *et al.* 1998; Barnett *et al.* 2005). Eight CANTAB tests considered likely to reflect a wide spectrum of cognitive dysfunctions among FEP patients early in their illness (before long-term antipsychotic treatment impact) were selected: pattern recognition memory (PRM); spatial recognition memory (SRM); paired associates learning (PAL); spatial span (SSP); spatial working memory (SWM); Stockings of Cambridge (SOC); intra/extradimensional set shift (IED); and rapid visual information processing (RVP). Our aim was to investigate how these CANTAB tests grouped into principal components and thereby purported latent factors.

When scores of ostensibly latent cognitive variables are compared between groups (e.g. patients and

controls), researchers assume that the observed variables define latent traits in exactly the same way for all groups (Meredith, 1993), an assumption called measurement invariance (MI). Unless MI is established, one may be comparing 'apples with oranges'. It may, for example, be that selected tests form a unitary trait factor in one group but not another. To the best of our knowledge there are no available studies examining MI in cognitive-factor comparisons between healthy subjects and FEP patients. Testing for MI in patient-control comparisons was another aim of this study.

After establishing MI we planned to evaluate differences in cognitive factors between groups, which if MI was lacking was to be achieved by comparing patients and controls based on individual test scores. Based on previous studies (Mohamed *et al.* 1999; Bilder *et al.* 2000) we expected patients' performance to be lower compared with control subjects across all cognitive domains.

One aim of this study was to replicate previous research on psychotic disorders related to cognitive impairment. Given the alleged replicability crisis in psychology (Pashler & Wagenmakers, 2012), such studies are badly needed. Additionally we extended the previous research of Leeson *et al.* (2009a) by considering structural differences in cognition between FEP patients and healthy people in order to empirically and explicitly select the most appropriate level of description for assessing psychotic disorder-related cognitive impairment. The question of structural similarity across FEP patients and healthy controls, or lack of it, is also of substantive interest. For example, lack of structural similarity suggests that patients' cognition may differ qualitatively from that of healthy controls.

Method

Participants

The patient sample consisted of 109 in-patients or out-patients (54.1% males) with FEP from two psychiatry clinics in Estonia. Mean patient age was 26.9 years (s.d. = 7.0, range 18–43 years) and 91.7% were right-handed. The patients fulfilled the following inclusion criteria: aged between 18 and 45 years; experience of a first psychotic episode; duration of untreated psychosis less than 3 years; no antipsychotic treatment received before the first contact with medical services for psychosis. When recruited, patients were in the stabilization phase of the first psychotic episode. Diagnoses were based on clinical interview according to International Classification of Diseases (ICD)-10 (World Health Organization, 1992) criteria, medical chart review, information from collateral informants,

and were consented within two clinical psychiatrists. In the psychosis group the diagnoses were F23.0 ($n=20$), F23.1 ($n=22$), F23.2 ($n=31$), F23.3 ($n=7$), F23.8 ($n=5$), F23.9 ($n=3$), F20.09 ($n=19$), F20.29 ($n=1$) and F20.39 ($n=1$). Patients were taking antipsychotics during the neuropsychological testing. The duration of medication use did not exceed 3 months. Patients were excluded from the study if they had psychotic disorders due to a general medical condition or substance-induced psychosis.

A sample of 96 healthy volunteers (controls), of which 40.6% comprised males, was recruited by advertisement from hospital staff and the general public. The mean age of the control group was 25.7 years (s.d. = 6.4, range 18–44 years) and 97.9% were right-handed. The controls were questioned regarding the state of their health and medical history to exclude those with conditions that might interfere with cognitive performance. Conditions that resulted in rejection of control subjects included neurological disorders, mental retardation or significant learning disorder, and major sight and hearing impairment. Exclusion criteria for the control group also included psychotic disorder among close relatives. Both FEP and control subjects were required to have knowledge of the Estonian language. There were no statistically significant differences between controls and patients in terms of age, gender or handedness. The average formal educational experience was 12.9 (s.d. = 2.4) years for patients and 14.0 (s.d. = 2.0) years for healthy subjects, a difference that was significantly significant ($t = -3.51$, $p < 0.01$).

All data were collected cross-sectionally. Patients and healthy subjects were enrolled between January 2009 and March 2013. All participants gave written informed consent to take part in the study and did not receive compensation. Ethical approval was granted by the Ethic Review Committee on Human Research, University of Tartu, Estonia.

Measures and procedures

Computerized neuropsychological assessment

Our clinical study started before the latest version of the CANTAB Schizophrenia Battery was available. The strategy we chose was to design a CANTAB-based battery of tasks that would specifically assess cognitive deficit characteristics of chronic psychotic disorder. Eight computerized tasks (see above) from the CANTABeclipse version 3.0.0 were run on a personal computer with a high-resolution touchscreen. All task stimuli were visual in nature, consisting of geometric designs or simple shapes, and required non-verbal responses. Instructions were given in Estonian from a literal translation of the CANTAB test manual produced by three clinical psychologists

fluent in both English and Estonian. To ensure semantic equivalence of the translated and original test instructions a consensus meeting of translators was held. The battery of tasks took approximately 1 h to administer. During test sessions participants were offered a short break. One of the tests (RVP) required a response key. Participants were tested in two different research centres. The neuropsychological tasks that were employed are briefly described in the online Supplementary material. For more detailed descriptions of these tests, see the CANTAB® website (<http://www.cambridgecognition.com>).

Statistical analysis

First, we compared patients and controls in terms of their demographic characteristics using Pearson's χ^2 test for categorical variables and an independent-samples t test for continuous variables. Second, an analysis of the covariance structure of the measured neuropsychological tests was performed using a series of principal components analyses (PCAs). For each PCA, a parallel analysis (Horn, 1965) helped determine the most appropriate number of components to be retained. As the components were expected to be correlated, each PCA was followed by an oblique (oblimin) rotation. Third, to determine if PCA results were plausible reflections of latent cognitive constructs, a confirmatory factor analysis (CFA) was conducted. PCAs and CFAs were done separately in healthy controls and FEP patients. Fourth, multi-group CFA (MGCFA) (Joreskog, 1971; Byrne *et al.* 1989; Widaman & Reise, 1997) was used to assess whether: (a) the structure of latent cognitive traits was similar in patients and controls; (b) the mean scores of each group could be meaningfully compared (MI).

As is common in MI testing (Wicherts & Dolan, 2010) a series of MGCFA models was fitted with systematically increasing parameter equality constraints across groups (Horn & McArdle, 1992; Vandenberg & Lance, 2000). During MI testing configural invariance criteria were met if the same variables were associated with the same latent factors in each group. No parameter equality constraints across groups were imposed at this point other than that the same tests defined the same latent constructs. The configural invariance model served as a baseline for further comparisons. Weak invariance was achieved when the factor loadings of the CANTAB tests on the latent variables could be held constant across groups without a significant deterioration of model fit. Weak invariance provides evidence that latent factors have the same meaning across groups. To establish a stronger form of invariance (scalar invariance), both factor loadings and intercepts of tests were constrained to be the

same across groups. In the case of no significant deterioration in model fit, scores of latent factors could be considered comparable across the groups. Strict invariance (residual variance invariance), which assumed the residual variances of observed variables to be the same across groups, was also explored. No deterioration of model fit with strict invariance indicated that neuropsychological variables were measured with the same precision in both groups. Finally, variances and covariances of the latent traits were constrained to be equal across groups to test whether the variability and intercorrelations of the latent variables were similar. Models were fitted using the robust maximum-likelihood (MLR) estimator in the 'lavaan' package (Rosseel, 2012). Model fit was estimated using the χ^2 goodness-of-fit statistic (Hu & Bentler, 1999), the comparative fit index (CFI) (Bentler, 1990) and the root mean square error of approximation (RMSEA) (Browne & Cudeck, 1993; Hu & Bentler, 1999; Steiger, 2000). Any given type of MI was supported when the fit of the more parsimonious model (i.e. the model with intercept equality constraints) was not significantly poorer than that of the less constrained model (i.e. the one without intercept equality constraints). Differences in model fit were tested using the χ^2 difference test (Horn & McArdle, 1992), where a statistically significant ($p < 0.05$) $\Delta\chi^2$ indicated a difference in fit. Group differences in latent factors could be estimated by fixing the mean in one group at zero and freely estimating the mean of the other group.

General linear models (GLMs) were used to investigate group differences in subtest scores between FEP patients and controls. Subtest scores were standardized using the mean and standard deviation of the control group. Age, gender and years in education were used as covariates in comparisons of cognitive functioning. The average number of missing subtest scores per participant was very low (0.05%). All available test scores were used for all analyses. Statistical analyses were conducted using the R Statistical software package (R Development Core Team, 2013).

Ethical standards

All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Results

PCA

As the scree plot test and parallel analysis suggested two higher-order components in controls and one

component in FEP patients, both one- and two-component solutions in both groups were tested. In healthy controls five CANTAB subtest variables (PRM, PAL, SRM, RVP and SSP) primarily defined a component representing attention/memory (factor loadings 0.46–0.73) and four variables (SWM errors, SWM strategy, SOC and IED) primarily defined a component called executive function (factor loadings 0.40–0.84). The two-component solution accounted for 45% of the total variance among the nine scores in the control group and 46% in the patient group (with primary loadings ranging from 0.42 to 0.76). The factor-loading pattern of the patient group was different compared with the control group: two variables (PRM and PAL) primarily defined the memory component, whereas seven variables (SRM, RVP, SSP, SWM errors, SWM strategy, SOC and IED) defined the attention/executive function. Component intercorrelations also differed across groups, with 0.15 for controls and 0.30 for patients. Intercorrelations suggested a higher-order factor (a single common cause for all tests, underlying the memory/attention- and executive function-related sources of variance), especially among patients. Therefore, we also examined one-component solutions in both groups.

The one-component PCA also indicated a single high-order factor, with all variables loaded on a broad cognitive variable, with 0.19–0.66 in the control group and 0.30–0.85 in the patient group. The solution accounted for 28% and 35% of the total variance among the indicators, respectively. Thus, although the selected CANTAB test tended to cluster into two (somewhat different) groups in controls and patients, the tests could also be grouped into a single, overarching cognitive-functioning domain, especially among patients.

CFA

One- and two-component models derived from the PCA results (see above) were subsequently converted to latent trait models for single-group CFAs. In the two-factor model latent factors (attention/memory and executive function) were defined by the same variables for both the patient and control groups (see Fig. 1a and b for factor loadings and covariance estimates). All factor loadings (excepted IED) were significant at $p < 0.01$ (z -values ranged from 1.17 to 5.38 for controls and from 2.33 to 13.14 for patients). Attention/memory and executive function had an extremely high intercorrelation ($r = 0.83$) in the patient group, making it difficult to determine whether these factors measure meaningfully different constructs; in the control group the intercorrelation of attention/memory and executive function was much lower ($r = 0.31$). In other words, the two cognitive domains were effectively more coherent among patients,

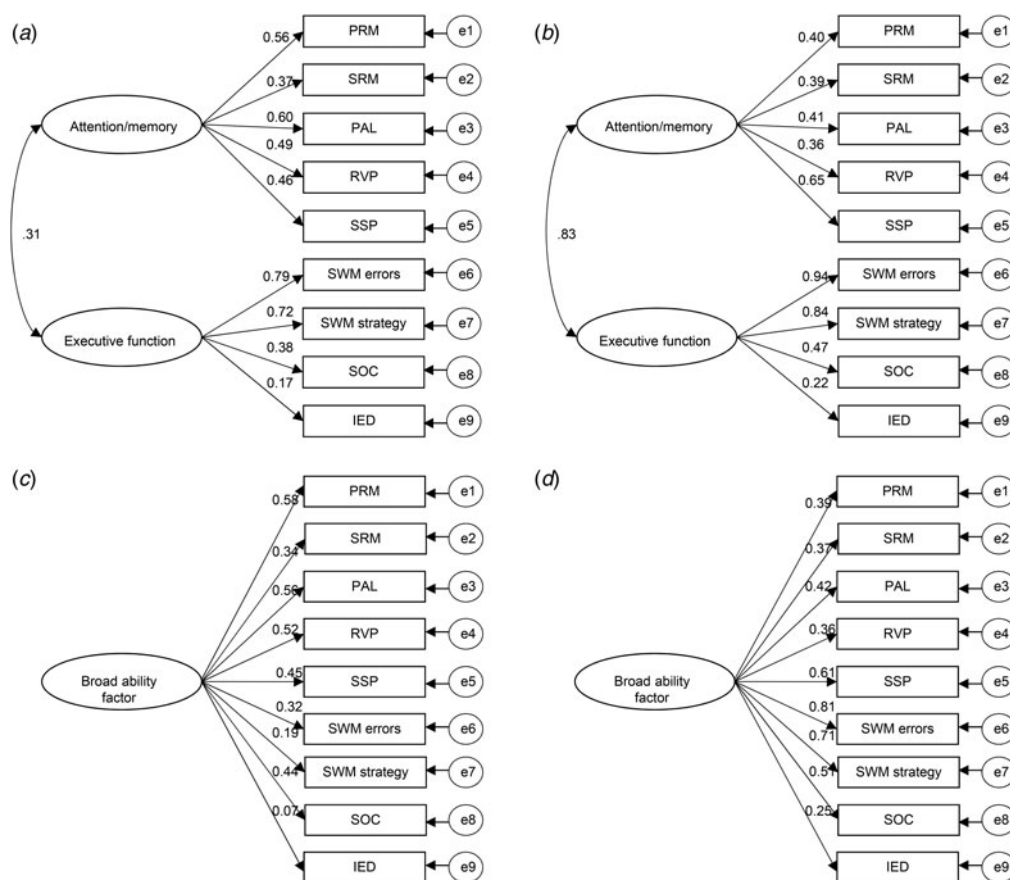


Fig. 1. Representation of the two- (*a, b*) and one- (*c, d*) latent factor structural models derived from the exploratory factor analysis for the control (*a, c*) and first-episode psychosis (*b, d*) samples, respectively. Variables in boxes represent observed measures and variables in ovals represent latent variables. The paths from the latent constructs to the observed variables demonstrate the parameter estimates onto its representative constructs. Two-headed arrows connecting latent variables represent correlations between the constructs. The 'e' represents the unique variance and error associated with each observed variable. PRM, Pattern recognition memory; SRM, spatial recognition memory; PAL, paired associates learning; RVP, rapid visual information processing; SSP, spatial span; SWM, spatial working memory; SOC, Stockings of Cambridge; IED, intra/extra-dimensional shift.

suggesting stronger evidence for a higher-order trait among them.

In the one-factor model (see Fig. 1c and d) all loadings were significantly different from zero in the patient group (*z*-values ranged from 2.14 to 9.00), whereas SWM errors, SWM strategy and IED had non-significant factor loadings ($p=0.09$, $p=0.23$ and $p=0.66$, respectively) in the control group, indicating that this model may be less appropriate than the two-factor model for the latter group. As, however, both one- and two-factor models fitted well in both samples (Table 1), we decided to input both models into MGCFA.

MI

For the two-factor solution, the fit of the configural MI model was good [$\chi^2=42.610$, degrees of freedom (*df*) = 50; RMSEA = 0.000; CFI = 1.000], suggesting it could be

considered a feasible representation of the data in both groups (Table 2) and justifying the evaluation of more restrictive invariance models. Weak MI was marginally supported ($\Delta\chi^2=15.676$, *df* = 9, $p=0.07$), indicating that factor loadings were more or less similar across groups. Strong MI was clearly not supported in the data ($\Delta\chi^2=200.730$, *df* = 7, $p<0.00$; CFI = 0.539; RMSEA = 0.143), indicating that the same observed CANTAB test scores corresponded to different latent trait levels in the two groups, making comparisons of their mean latent scores effectively meaningless. As strong MI was not met, testing for stricter forms of MI was not justified.

In MGCFA specifying just one latent factor, the configural MI model fitted data well (Table 2). The weak MI model was accompanied by a clear drop in model fit, suggesting that stricter forms of MI would not be met and latent factor means would not be comparable across groups.

Table 1. Goodness-of-fit statistics for the structural models adapted from alternative exploratory factor analysis

Single-group model	χ^2 (df)	CFI	RMSEA estimate (90% CI)
Controls ($n = 96$)			
Two-factor solution	29.071 (25)	0.961	0.041 (0.000–0.097)
One-factor solution	28.947 (25)	0.962	0.041 (0.000–0.097)
Patients ($n = 108$)			
Two-factor solution	14.853 (25)	1.000	0.000 (0.000–0.000)
One-factor solution	13.539 (25)	1.000	0.000 (0.000–0.000)

df, Degrees of freedom; CFI, comparative fit index; RMSEA, root mean square error of approximation; CI, confidence interval.

Comparison of cognitive performance

As controls and patients could not be compared based on latent traits, group differences in cognitive performance were tested using GLM based on observed test scores. For each cognitive measure, age, gender and years in education were included as covariates in group comparisons (for the results, see Table 3). In general, patients exhibited widespread cognitive impairments when compared with healthy control subjects.

Attention and memory component

There was a significant main group effect for visual memory (PRM) and SRM ($F_{4,200} = 5.65$, $p < 0.001$ and $F_{4,200} = 7.73$, $p < 0.001$, respectively), indicating that healthy controls gave a higher number of correct responses than FEP patients. Healthy controls also gave a higher number of correct responses than patients for the episodic memory and learning task (PAL) ($F_{4,200} = 11.98$, $p < 0.001$), had greater sensitivity for detecting important sequences in the sustained attention task (RVP) ($F_{4,199} = 18.02$, $p < 0.001$) and had a longer spatial span length in the working memory capacity task (SSP) ($F_{4,200} = 10.4$, $p < 0.001$).

Executive function

In the cognitive planning task (SOC), healthy subjects completed more stages in the least number of moves than patients ($F_{4,200} = 18.28$, $p < 0.001$). In the cognitive shifting and flexibility task (IED), there was a main group effect for the total reverse errors measure ($F_{4,200} = 11.38$, $p < 0.001$). Healthy controls gave more correct

responses in the SWM task (SWM errors) that measured a subject's ability to retain spatial information and manipulate remembered items in working memory ($F_{4,200} = 9.92$, $p < 0.001$), and also used heuristic strategies (SWM strategy) more efficiently than patients ($F_{4,200} = 8.92$, $p < 0.001$).

Overall, the profile of neuropsychological impairment in FEP patients (Fig. 2) was characterized by diminished processing speed (RVP) and impaired executive functioning (SWM errors, SOC and IED).

Discussion

The purpose of this study was to investigate the structure and possible impairment of cognitive abilities of FEP patients compared with healthy similar-aged peers. Our investigation was based on nine CANTAB tests scores (PRM, SRM, PAL, IED, SOC, SSP, SWM errors, SWM strategy and RVP) performed by tapping a computer touchscreen to measure a wide range of cognitive skills considered potentially sensitive to psychotic disorders. This study emphasizes the importance of establishing MI, which can cover nuanced group differences that might otherwise remain undetected. In this study patients and controls could not be compared in terms of their mean latent cognitive factors because the structural relationships among the cognitive tests were different between the groups.

The results of the exploratory PCA and subsequent single-group CFA suggested the selected CANTAB tests considered may group into two different cognitive factors in both groups. Whereas two relatively distinct factors (attention/memory and executive factor) appeared to be a tenable solution among controls, a single broad ability factor, however, was clearly evident among patients.

Consistent with some previous studies (Gladsjo et al. 2004; Dickinson et al. 2006; Burton et al. 2013) our research shows that intercorrelations of cognitive domains are higher for patients with psychotic disorder than for healthy controls. In other words, patients appeared to rely more heavily on general cognitive ability than on individual cognitive processes. The more homogeneous cognitive profile of patients that we found may reflect a similar impairment of cognitive skills resulting from disease-related or disease-preceding processes.

Our study replicated the findings of CFA for FEP patients (Leeson et al. 2009a) that revealed that cognitive functioning in control and patient groups could not be explained by similar theoretical models. Studies that replicate previous research are arguably extremely valuable in their own right (Pashler & Wagenmakers, 2012). We also extended the previous

Table 2. Summary of tests of factorial invariance by groups according to the two- and one-factor solutions

Invariance	χ^2 (df)	$\Delta\chi^2$ (Δ df)	p^a	CFI	RMSEA
Two-factor solution					
Configural	42.610 (50)			1.000	0.000
Weak	59.561 (59)	15.676 (9)	0.07	0.998	0.010
Strong	202.903 (66)	200.730 (7)	0.00	0.539	0.143
One-factor solution					
Configural	41.968 (50)			1.000	0.000
Weak	66.029 (59)	21.885 (9)	0.01	0.976	0.034

df, Degrees of freedom; CFI, comparative fit index; RMSEA, root mean square error of approximation.

^a p value corresponds to $\Delta\chi^2$.

analysis by importantly formally testing for the presence or lack of MI across patient and control groups. Our results indicate that the cognitive differences between patients and healthy individuals may not be limited to general levels of cognition; there may also be structural differences.

Although the MGCFA results indicated that the cognitive domains could be constructed in the same way in controls and patients (configural invariance held for both one- and two-trait models), the nature of the relationships between observed test scores and their purported underlying construct tended to be dissimilar. This suggested that the latent factor scores were not comparable because observed test scores were probably influenced by characteristics other than the latent ability. That patients' cognitive profiles were less diverse than those of healthy individuals may be a result of psychosis-related processes making an impact on cognitive domains in similar ways.

To our knowledge this is the first study to evaluate MI when CANTAB is used to compare FEP patients with healthy individuals.

Our results reinforce the view that there are broad cognitive deficits associated with FEP, although the deficits could not be similarly ascribed to underlying broad cognitive domains in both FEP patients and healthy individuals. At the group level, patients exhibited worse performance than healthy controls on all measured CANTAB subtest scores, indicating substantial cognitive impairment. Performance differences remained significant even after adjusting for years of education, age and gender, which is consistent with a number of other studies (Heinrichs & Zakzanis, 1998; Townsend & Norman, 2004; Dickinson *et al.* 2007).

Our study found IED test scores the most discrepant variable in the CFA models, suggesting the ability to inhibit improper response and shift attention diverge from the other variables of executive function. Previous research (Murray *et al.* 2008; Leeson *et al.* 2009b) has demonstrated that impaired performance in attentional

set-shifting tasks (IED) is already present at the beginning of chronic psychotic disorder and remains stable over time. One explanation for early impairment of IED is that set-shifting tasks require the contribution and co-working of numerous complex cognitive processes (e.g. attention, working memory, learning, problem solving, reasoning and inhibition). Studies have suggested that impaired set shifting correlates with working memory deficits, explaining the reversal learning difficulties especially at the early stage of the illness (Leeson *et al.* 2009b; Pantelis *et al.* 2009).

This study attests that the CANTAB can be used in a variety of cultural contexts. Furthermore, consistency in evidence such as structural similarity across samples and widespread cognitive deficit in patients suggests that the applications of the tests in different setting may be more or less comparable.

This study does have limitations that require consideration when interpreting the results. First, the recruitment of subjects was based on opportunity rather than random sampling. Subjects in the healthy control group came from a subpopulation and results may not be extrapolatable to the general Estonian populace. The clinical sample was restricted to a group of patients that were clinically stable and willing to participate in the testing. Our findings may thus not reflect the overall cognitive characteristics of patients with FEP in Estonia. The recruited patients had a large degree of heterogeneity in terms of diagnosis and medication. We did not exclude participants with co-morbid conditions, for example cannabis use in the previous anamnesis, nor adjust for specific demographic characteristics as our focus was on the general factor structure of the CANTAB test battery and the comparison of differences in the selected neuropsychological test scores between patients and healthy subjects.

Second, the limited sample size may have reduced statistical power for the factor analyses. Although the number of latent dimensions may have been

Table 3. Neuropsychological profile comparisons between FEP patients and control subjects^a

	PRM number correct	SRM number correct	PAL memory score	RVP sensitivity for detecting sequences	SSP span length	SWM total errors	SWM strategy score	SOC problems solved in minimum moves	IED total reverse errors
Parameter	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)	Estimate (95% CI)
Group ^b	0.69 (0.32–1.05)***	0.78 (0.41–1.14)***	0.73 (0.43–1.03)***	1.17 (0.84–1.50)***	0.61 (0.32–0.90)***	–0.94 (–1.30 to 0.58)***	–0.67 (–0.96 to 0.38)***	1.36 (0.98–1.69)***	–1.25 (–1.71 to 0.79)***
Education in years	0.07 (–0.02 to 0.15)	0.08 (–0.01 to 0.16)	0.06 (–0.01 to 0.13)	0.10 (0.03–0.18)**	0.06 (–0.01 to 0.13)	–0.04 (–0.12 to 0.04)	0.01 (–0.05 to 0.08)	0.07 (0.00–0.16)*	–0.13 (–0.24 to 0.03)*
Age	–0.01 (–0.04 to 0.02)	–0.02 (–0.05 to 0.01)	–0.04 (–0.06 to 0.02)***	–0.01 (–0.04 to 0.01)	–0.36 (–0.06 to 0.01)**	0.01 (–0.02 to 0.04)	0.02 (–0.01 to 0.04)	0.01 (–0.02 to 0.04)	0.01 (–0.02 to 0.05)
Gender	0.10 (–0.26 to 0.45)	0.07 (–0.28 to 0.42)	0.07 (–0.22 to 0.37)	–0.14 (–0.46 to 0.18)	–0.40 (–0.68 to 0.12)**	0.57 (0.22–0.92)**	0.50 (0.22–0.77)***	–0.20 (–0.55 to 0.13)	0.37 (–0.07 to 0.82)
Better performance	Controls	Controls	Controls, younger	Controls, more educated	Controls, younger, men	Controls, women	Controls, women	Controls	Controls, more educated

FEP, First-episode psychosis; PRM, pattern recognition memory; SRM, spatial recognition memory; PAL, paired associates learning; RVP, rapid visual information processing; SSP, spatial span; SWM, spatial working memory; SOC, Stockings of Cambridge; IED, intra/extra-dimensional shift; CI, confidence interval.

^a Negative group parameter estimates for SWM and IED indices demonstrate lower scores but better performance in the control group.

^b All group comparisons were made controlling for the effects of education, age and gender.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

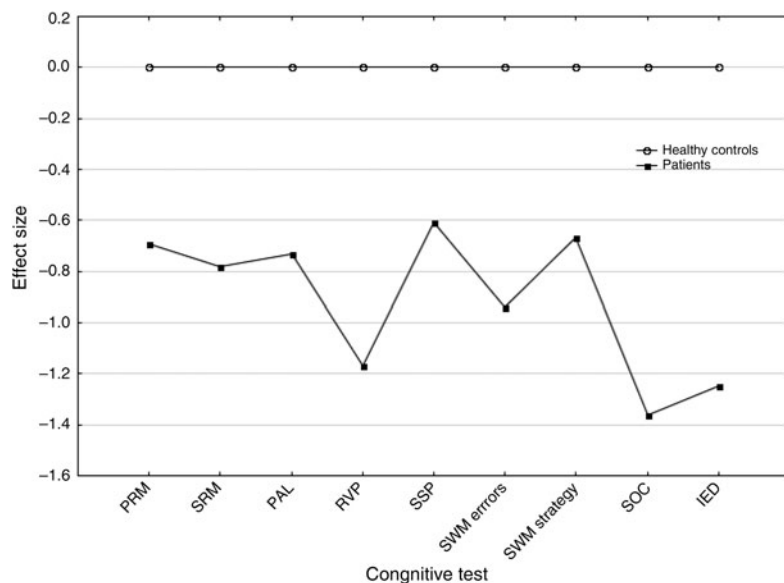


Fig. 2. Cognitive impairment profile: performance of first-episode psychosis patients expressed as effect sizes. PRM, Pattern recognition memory; SRM, spatial recognition memory; PAL, paired associates learning; RVP, rapid visual information processing; SSP, spatial span; SWM, spatial working memory; SOC, Stockings of Cambridge; IED, intra/extra-dimensional shift. The sign of the effect size values was changed for IED and SWM domains in order to have the dysfunctional poles as negative values.

underestimated or factor loadings biased, most of the loadings across factors were at least moderate, indicating that factors were reasonably stable. The factor solutions identified in this investigation, however, accounted for only 35% (one-factor solution for patients) and 45% (two-factor solution for healthy subjects) of the total variance among the nine CANTAB tests, indicating that a substantial amount of variance was not accounted for by the identified factors.

Third, the current study did not assess the pre-morbid cognitive functioning of the patients as we lacked properly adapted existing instruments in Estonian.

Despite potential limitations, however, we believe our present study offers interesting results that are useful in everyday psychiatric practice. Our findings have practical significance for the broader use of CANTAB neuropsychological tests for assessing cognition at the early stage of psychotic disorder. We also recommend that studies which use CANTAB batteries do not combine latent domains when comparing FEP patients and controls, but restrict analyses to differences in the subtests.

Conclusion

To conclude, our study addresses the often-ignored but critical consideration in research employing neuropsychological test batteries of a lack of MI in comparisons of psychometric analyses between non-clinical (healthy) and clinical (patient) samples. We found that there are probably qualitatively and quantitatively different

cognitive patterns in the FEP patients compared with healthy subjects, and that patients exhibit widespread cognitive impairments.

Our findings support continued efforts to elucidate cognitive dysfunction as a biomarker of early-stage schizophrenia.

Supplementary material

For supplementary material accompanying this paper visit <http://dx.doi.org/10.1017/S0033291714003018>

Acknowledgements

This research received no specific grant from any funding agency, commercial or not-for-profit sectors. The authors thank the participants who volunteered their time and effort for this study, as well as colleagues for their contributions in assessing the patients. The authors thank Alexander Zharkovsky for helpful comments on an earlier draft of this article.

Declaration of Interest

None.

References

- Addington J, Brooks BL, Addington D (2003). Cognitive functioning in first episode psychosis: initial presentation. *Schizophrenia Research* **62**, 59–64.

- Barnett JH, Sahakian BJ, Werners U, Hill KE, Brazil R, Gallagher O, Bullmore ET, Jones PB** (2005). Visuospatial learning and executive function are independently impaired in first-episode psychosis. *Psychological Medicine* **35**, 1031–1041.
- Bentler PM** (1990). Comparative fit indexes in structural models. *Psychological Bulletin* **107**, 238–246.
- Bilder RM, Goldman RS, Robinson D, Reiter G, Bell L, Bates JA, Pappadopulos E, Willson DF, Alvir MJ, Woerner MG, Geisler S, Kane JM, Lieberman JA** (2000). Neuropsychology of first-episode schizophrenia: initial characterization and clinical correlates. *American Journal of Psychiatry* **157**, 549–559.
- Browne MW, Cudeck R** (1993). Testing structural equation models. In *Alternative Ways of Assessing Model Fit* (ed. K.A. Bollen and J.S. Long), pp. 136–162. Sage: Beverly Hills, CA.
- Burton CZ, Vella L, Harvey PD, Patterson TL, Heaton RK, Twamley EW** (2013). Factor structure of the MATRICS Consensus Cognitive Battery (MCCB) in schizophrenia. *Schizophrenia Research* **146**, 244–248.
- Byrne BM, Shavelson RJ, Muthén B** (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin* **105**, 456–466.
- Deary IJ, Penke L, Johnson W** (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience* **11**, 201–211.
- Dickinson D, Goldberg TE, Gold JM, Elvevåg B, Weinberger DR** (2011). Cognitive factor structure and invariance in people with schizophrenia, their unaffected siblings, and controls. *Schizophrenia Bulletin* **37**, 1157–1167.
- Dickinson D, Iannone VN, Wilk CM, Gold JM** (2004). General and specific cognitive deficits in schizophrenia. *Biological Psychiatry* **55**, 826–833.
- Dickinson D, Ragland JD, Calkins ME, Gold JM, Gur RC** (2006). A comparison of cognitive structure in schizophrenia patients and healthy controls using confirmatory factor analysis. *Schizophrenia Research* **85**, 20–29.
- Dickinson D, Ramsey ME, Gold JM** (2007). Overlooking the obvious: a meta-analytic comparison of digit symbol coding tasks and other cognitive measures in schizophrenia. *Archives of General Psychiatry* **64**, 532–542.
- Elliott R, McKenna PJ, Robbins TW, Sahakian BJ** (1995). Neuropsychological evidence for frontostriatal dysfunction in schizophrenia. *Psychological Medicine* **25**, 619–630.
- Genderson MR, Dickinson D, Diaz-Asper C, Egan MF, Weinberger DR, Goldberg TE** (2007). Factor analysis of neurocognitive tests in a large sample of schizophrenic probands, their siblings, and healthy controls. *Schizophrenia Research* **94**, 231–239.
- Gladsjo JA, McAdams LA, Palmer BW, Moore DJ, Jeste DV, Heaton RK** (2004). A six-factor model of cognition in schizophrenia and related psychotic disorders: relationships with clinical symptoms and functional capacity. *Schizophrenia Bulletin* **30**, 739–754.
- Gold JM, Carpenter C, Randolph C, Goldberg TE, Weinberger DR** (1997). Auditory working memory and Wisconsin Card Sorting Test performance in schizophrenia. *Archives of General Psychiatry* **54**, 159–165.
- Gold JM, Harvey PD** (1993). Cognitive deficits in schizophrenia. *Psychiatric Clinics of North America* **16**, 295–312.
- Green MF, Nuechterlein KH, Gold JM, Barch DM, Cohen J, Essock S, Fenton WS, Frese F, Goldberg TE, Heaton RK, Keefe RSE, Kern RS, Kraemer H, Stover E, Weinberger DR, Zalcman S, Marder SR** (2004). Approaching a consensus cognitive battery for clinical trials in schizophrenia: the NIMH-MATRICES conference to select cognitive domains and test criteria. *Biological Psychiatry* **56**, 301–307.
- Heaton RK, Gladsjo JA, Palmer BW, Kuck J, Marcotte TD, Jeste DV** (2001). Stability and course of neuropsychological deficits in schizophrenia. *Archives of General Psychiatry* **58**, 24–32.
- Heinrichs RW, Zakzanis KK** (1998). Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology* **12**, 426–445.
- Horn JL** (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika* **30**, 179–185.
- Horn JL, McArdle JJ** (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research* **18**, 117–144.
- Hu L, Bentler PM** (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling* **6**, 1–55.
- Hutton SB, Puri BK, Duncan LJ, Robbins TW, Barnes TRE, Joyce EM** (1998). Executive function in first-episode schizophrenia. *Psychological Medicine* **28**, 463–473.
- Jaeger J, Czobor Pá, Berns SM** (2003). Basic neuropsychological dimensions in schizophrenia. *Schizophrenia Research* **65**, 105–116.
- Joreskog KG** (1971). Simultaneous factor analysis in several populations. *Psychometrika* **36**, 409–426.
- Joyce EM, Hutton SB, Mutsatsa SH, Barnes TRE** (2005). Cognitive heterogeneity in first-episode schizophrenia. *British Journal of Psychiatry* **187**, 516–522.
- Kéri S, Janka Z** (2004). Critical evaluation of cognitive dysfunctions as endophenotypes of schizophrenia. *Acta Psychiatrica Scandinavica* **110**, 83–91.
- Leeson VC, Robbins TW, Franklin C, Harrison M, Harrison I, Ron MA, Barnes TRE, Joyce EM** (2009a). Dissociation of long-term verbal memory and fronto-executive impairment in first-episode psychosis. *Psychological Medicine* **39**, 1799–1808.
- Leeson VC, Robbins TW, Matheson E, Hutton SB, Ron MA, Barnes TRE, Joyce EM** (2009b). Discrimination learning, reversal, and set-shifting in first-episode schizophrenia: stability over six years and specific associations with medication type and disorganization syndrome. *Biological Psychiatry* **66**, 586–593.
- Meredith W** (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* **58**, 525–543.
- Mohamed S, Paulsen JS, O’Leary D, Arndt S, Andreasen N** (1999). Generalized cognitive deficits in schizophrenia: a study of first-episode patients. *Archives of General Psychiatry* **56**, 749–754.
- Murray GK, Cheng F, Clark L, Barnett JH, Blackwell AD, Fletcher PC, Robbins TW, Bullmore ET, Jones PB** (2008). Reinforcement and reversal learning in first-episode psychosis. *Schizophrenia Bulletin* **34**, 848–855.

- Pantelis C, Barnes TR, Nelson HE, Tanner S, Weatherley L, Owen AM, Robbins TW** (1997). Frontal–striatal cognitive deficits in patients with chronic schizophrenia. *Brain: A Journal of Neurology* **120**, 1823–1843.
- Pantelis C, Wood SJ, Proffitt TM, Testa R, Mahony K, Brewer WJ, Buchanan J, Velakoulis D, McGorry PD** (2009). Attentional set-shifting ability in first-episode and established schizophrenia: relationship to working memory. *Schizophrenia Research* **112**, 104–113.
- Pashler H, Wagenmakers E** (2012). Editors' introduction to the special section on replicability in psychological science: a crisis of confidence? *Perspectives on Psychological Science* **7**, 528–530.
- R Development Core Team** (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria (<http://www.R-project.org/>).
- Reichenberg A, Weiser M, Rabinowitz J, Caspi A, Schmeidler J, Mark M, Kaplan Z, Davidson M** (2002). A population-based cohort study of premorbid intellectual, language, and behavioral functioning in patients with schizophrenia, schizoaffective disorder, and nonpsychotic bipolar disorder. *American Journal of Psychiatry* **159**, 2027–2035.
- Robbins TW, James M, Owen AM, Sahakian BJ, Lawrence AD, McInnes L, Rabbitt PM** (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: implications for theories of executive functioning and cognitive aging. Cambridge Neuropsychological Test Automated Battery. *Journal of the International Neuropsychological Society: JINS* **4**, 474–490.
- Robbins TW, James M, Owen AM, Sahakian BJ, McInnes L, Rabbitt P** (1994). Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. *Dementia (Basel, Switzerland)* **5**, 266–281.
- Robbins TW, Sahakian BJ** (1994). Computer methods of assessment of cognitive function. In *Principles and Practice of Geriatric Psychiatry* (ed. J. R. M. Copeland, M. T. Abou-Saleh and D. G. Blazer), pp. 205–209. John Wiley & Sons, Ltd: Chichester, UK.
- Rosseel Y** (2012). Iavaan: an R package for structural equation modeling. *Journal of Statistical Software* **48**, 1–36.
- Saykin AJ, Gur RC, Gur RE, Mozley PD, Mozley LH, Resnick SM, Kester DB, Stafiniak P** (1991). Neuropsychological function in schizophrenia. Selective impairment in memory and learning. *Archives of General Psychiatry* **48**, 618–624.
- Steiger JH** (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: some comments and a reply to Hayduck and Glaser. *Structural Equation Modeling* **7**, 149–162.
- Stip E, Lecardeur L, Sepehry AA** (2008). Computerised assessment of visuo-spatial cognition in schizophrenia – an exploratory meta-analysis of CANTAB findings. *European Psychiatric Review* **1**, 48–54.
- Townsend LA, Norman RMG** (2004). Course of cognitive functioning in first episode schizophrenia spectrum disorders. *Expert Review of Neurotherapeutics* **4**, 61–68.
- Vandenberg RJ, Lance CE** (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods* **3**, 4–69.
- Wicherts JM, Dolan CV** (2010). Measurement invariance in confirmatory factor analysis: an illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice* **29**, 39–47.
- Widaman KF, Reise SP** (1997). Exploring the measurement invariance of psychological instruments: applications in the substance use domain. In *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research* (ed. K. J. Bryant, M. Windle and S. G. West), pp. 281–324. American Psychological Association: Washington, DC.
- World Health Organization** (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. World Health Organization: Geneva.