



A generalisability theory approach to quantifying changes in psychopathology among ultra-high-risk individuals for psychosis

Zohreh Dobarjeh^{1,2}, Oleg N. Medvedev³, Maryam Dobarjeh¹, Balkaran Singh¹, Alexander Sumich^{4,5}, Sugam Budhraj¹, Wilson Wen Bin Goh^{6,7,8,9,10}, Jimmy Lee^{6,11}, Margaret Williams¹², Edmund M-K Lai^{1,13} and Nikola Kasabov^{1,13,14}

Distinguishing stable and fluctuating psychopathological features in young individuals at Ultra High Risk (UHR) for psychosis is challenging, but critical for building robust, accurate, early clinical detection and prevention capabilities. Over a 24-month period, 159 UHR individuals were assessed using the Positive and Negative Symptom Scale (PANSS). Generalisability Theory was used to validate the PANSS with this population and to investigate stable and fluctuating features, by estimating the reliability and generalisability of three factor (Positive, Negative, and General) and five factor (Positive, Negative, Cognitive, Depression, and Hostility) symptom models. Acceptable reliability and generalisability of scores across occasions and sample population were demonstrated by the total PANSS scale ($G = 0.85$). Fluctuating symptoms (delusions, hallucinatory behaviour, lack of spontaneity, flow in conversation, emotional withdrawal, and somatic concern) showed high variability over time, with 50–68% of the variance explained by individual transient states. In contrast, more stable symptoms included excitement, poor rapport, anxiety, guilt feeling, uncooperativeness, and poor impulse control. The 3-factor model of PANSS and its subscales showed robust reliability and generalisability of their assessment scores across the UHR population and evaluation periods ($G = 0.77$ – 0.93), offering a suitable means to assess psychosis risk. Certain subscales within the 5-factor PANSS model showed comparatively lower reliability and generalisability ($G = 0.33$ – 0.66). The identified and investigated fluctuating symptoms in UHR individuals are more amendable by means of intervention, which could have significant implications for preventing and addressing psychosis. Prioritising the treatment of fluctuating symptoms could enhance intervention efficacy, offering a sharper focus in clinical trials. At the same time, using more reliable total scale and 3 subscales can contribute to more accurate assessment of enduring psychosis patterns in clinical and experimental settings.

Schizophrenia (2024)10:87; <https://doi.org/10.1038/s41537-024-00503-y>

INTRODUCTION

Psychosis is a complex condition characterised by a disconnection from reality, with symptoms that can differ in their stability and how they respond to environmental factors^{1–3}. Differentiating stable (trait-like) and fluctuating (state-related) illness features can be challenging due to the complex interplay of clinical, environmental, and biopsychosocial mechanisms⁴. A well-established literature indicates that positive symptoms, such as delusions and hallucinations, vary more compared to negative symptoms, such as affective blunting and social withdrawal, which tend to be enduring^{5,6}. This underpins a prevalent clinical perception that negative symptoms are intractable. Stable symptoms, such as emotional withdrawal, poverty of speech, and blunted affect, have been reported to show less improvement with standard therapeutic approaches^{4,7} compared to more fluctuating symptoms such as hallucinations or delusions^{8,9}. Accordingly, fluctuating symptoms may be more sensitive to situational stressors or contextual changes¹⁰. Recent meta-analyses support the potential

for improvement in negative symptoms across diverse treatments⁵, and have re-opened the debate as to which symptoms are amenable to change.

Stable and fluctuating psychopathological symptoms across UHR

Understanding symptom progression in youth individuals who are Ultra High Risk (UHR) of developing psychosis¹¹ is essential for devising prompt, precise, and effective diagnostic and management strategies that might avert the onset of illness. UHR individuals are characterised by subthreshold symptoms and/or genetic risk factors that place them at a heightened likelihood of transitioning to full psychosis¹¹. Early identification of UHR individuals allows for timely interventions aimed at altering the disease trajectory and improving outcomes¹². Therefore, the distinction between stable and fluctuating symptoms in this population is important. Fluctuating symptoms, characterised by

¹Knowledge Engineering and Discovery Research Institute, School of Engineering, Computer, and Mathematical Sciences, Auckland University of Technology, Auckland, New Zealand. ²School of Population Health, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand. ³School of Psychological and Social Sciences, The University of Waikato, Hamilton, New Zealand. ⁴NTU Psychology, School of Social Sciences, Nottingham Trent University, Nottingham, United Kingdom. ⁵Department of Psychology and Neuroscience, Auckland University of Technology, Auckland, New Zealand. ⁶Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. ⁷School of Biological Sciences, Nanyang Technological University, Singapore, Singapore. ⁸Center for Biomedical Informatics, Nanyang Technological University, Singapore, Singapore. ⁹Center of AI in Medicine, Nanyang Technological University, Singapore, Singapore. ¹⁰Division of Neurology, Department of Brain Sciences, Faculty of Medicine, Imperial College London, London, United Kingdom. ¹¹Institute of Mental Health, Singapore, Singapore. ¹²School of Public Health and Interdisciplinary Studies, Auckland University of Technology, Auckland, New Zealand. ¹³School of Software Engineering, Dalian University, Dalian, China. ¹⁴Institute for Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria. ✉email: zohreh.dobarjeh@aut.ac.nz; wilsongoh@ntu.edu.sg

their variability and sensitivity to situational factors over time, can provide early indicators of impending psychosis onset, enabling proactive intervention strategies. In contrast, stable symptoms, despite being less responsive to standard treatments, reflect psychosis risk associated with the enduring clinical challenges in managing UHR individuals¹³.

Although prior studies have indicated that negative symptoms may have greater stability and resistance to therapy compared to positive symptoms¹⁴, recent analysis using Generalisability Theory (G-theory) provides a more nuanced picture. Medvedev et al. (2020) found that certain symptoms such as lack of attention, blunted affect, poor rapport, and delusions were among the most stable. However, they also identified fluctuating symptoms across positive, negative, and general domains. This underscores the importance of examining stability at the individual symptom level rather than broad symptom categories.

So far, stable versus fluctuating clinical and cognitive features have been identified using G-theory in relation to schizophrenia¹⁵, mood disorders¹⁶, therapy response (e.g. mindfulness)¹⁷ and ageing¹⁸. G-theory is a specialised method for determining the evaluation scores' generalisability and overall reliability¹⁵. This approach allows for more precise measurement of symptom stability and change compared to traditional methods of Classical Test Theory (CTT).

G-Theory for reliability and generalisability of the assessment in UHR

While previous studies have examined symptom stability in UHR populations using traditional methods, the application of G-theory to this context represents a novel and potentially more informative approach^{15,19,20}. G-theory offers several advantages over classical test theory in analysing symptom fluctuations. Firstly, it allows for the simultaneous evaluation of multiple sources of variance, including person-level traits, time-dependent fluctuations, and measurement error¹⁵. This is particularly relevant for UHR populations, where symptoms may be more labile and influenced by various factors. Secondly, G-theory can distinguish between group-level trends and individual-level variability, providing a better understanding of symptom dynamics. Thirdly, it provides a framework for calculating change score reliability, which is necessary to accurately monitor the progression or improvement of symptoms over time. The application and use of G-theory to UHR samples addresses a significant limitation in the current psychiatric research.

Therefore, the current study applied G-theory to understand symptom fluctuation in individuals at UHR for psychosis.

The 3 and 5 factor models of PANSS for UHR Individuals

The Positive and Negative Syndrome Scale (PANSS) is a well-studied measure in measuring schizophrenia symptoms studies²¹. PANSS contains of 30 items that are classified into three groups of psychopathologies including positive, negative, and general domain. Other factor structures were developed, including 5, 6, and 7 factor models, following later factor analytic research.

Among these, the 5-factor model has gained particular prominence. The 5-factor model includes the symptom domains of Positive; Negative; Cognitive; Depression; and Hostility. It is frequently utilised in studies exploring treatment response²², psychotic disorders²³, and social cognition²⁴. Several studies propose that PANSS is better represented with a 5–7 (rather than 3 subscales) factor model^{25,26}. For example, Lim et al. proposed a 5-factor model comprising i) Positive (5 items), ii) Negative (7 items), iii) Cognitive and disorganisation (7 items), iv) Depression and anxiety (5 items) and v) Excitement and hostility (6 items) factors that showed good symptoms indication in participants with schizophrenia²⁷.

Many research on schizophrenia have developed models with a 5-factor structure²⁸; however, as the items loaded in each domain vary throughout sample populations, there is still no general agreement on what should be included in each domain²⁶.

While generally this model is considered as reliable and valid, its test-retest reliability and internal consistency present challenges, especially in clinical studies when sensitivity to real symptom alterations is required^{21,29}. Reporting practices in the context of the PANSS sometimes fail to capture the unique and subtle variations in symptoms experienced by individuals^{30–32}. Therefore, to establish the true reliability of the PANSS, it is important to differentiate between stable and fluctuating symptomatology, achievable through repeated measurements¹⁹.

Current reporting practices using the PANSS often rely on subscale or total scores, which may not fully capture the unique and subtle variations in symptoms experienced by individuals³³. For instance, Medvedev et al. (2020) demonstrated that while some symptoms (e.g., poor attention, blunted affect) remain relatively stable, while others (e.g., grandiosity, somatic concerns) can be quite fluctuating. However, these subtleties are often obscured when only reporting composite scores¹⁵.

Additionally, the PANSS was originally developed for use with individuals diagnosed with schizophrenia, and its applicability to UHR populations has been questioned³⁴. Some symptoms that are particularly relevant to the UHR state, such as basic symptoms or subtle cognitive changes, may not be adequately captured by the PANSS³⁵. By applying G-Theory to PANSS data, we aimed to validate the scale with this population, and provide a more detailed analysis of symptom stability and variability at the individual item level, potentially revealing patterns that are not apparent when using traditional scoring and reporting methods. Addressing these limitations is essential for developing precise and effective intervention strategies for those at the highest risk of psychosis. In addition, the current research compared the reliability and generalisability of the 3-factor and 5-factor PANSS models in assessing these features in UHR individuals.

Our objective was to enhance understanding and aid in the development of fluctuating subscales that may augment sensitivity in detecting group variations in cutting-edge treatment studies aimed at delaying or preventing the onset of illness. We hypothesised that fluctuating symptoms would demonstrate more variability over a two-year period than stable symptoms. Also, the 3-factor model of the PANSS would demonstrate higher reliability and generalisability compared to the 5-factor model in assessing symptomatology among UHR individuals.

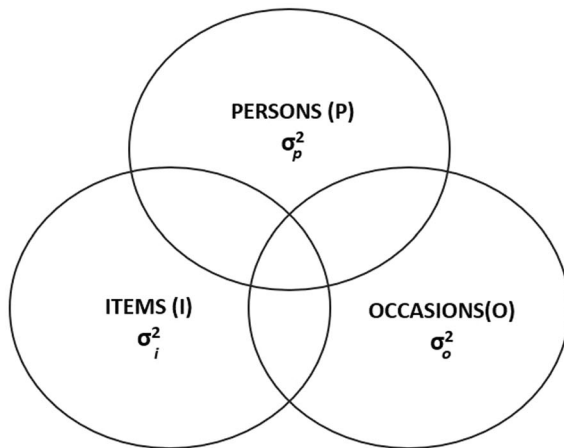
METHOD AND PROCEDURE

Participants and demographics

The Longitudinal Youth-at-Risk Study (LYRIKS) comprised 173 young individuals at UHR with the Mean age of 21.3 years and Standard Deviation of 3.5; and $n = 384$ healthy control individuals with the Mean age of 21.7 years and Standard Deviation of 3.4. The research was a prospective observational investigation of youths at risk for psychosis³⁶. The participants' ages ranged from 14 to 29 years old.

Current inclusion criteria. The study only included participants whose data were gathered for all variables over a period of two years. For psychosis, a cohort of 159 individuals was chosen from the UHR population.

This cohort comprised 105 male and 54 female individuals, ethnically diverse with 113 Chinese, 24 Malay, 16 Indian, and 5 participants of other ethnic backgrounds. In terms of educational background, the sample included individuals from various levels: 46 had completed secondary education, 37 had attended a polytechnic institution, 26 had primary education, 17 had been enrolled in a vocational institute, 13 had undertaken pre-university



Persons (P)	Averaged deviation from grand mean over items and occasions
Items (I)	Averaged deviation from the grand mean over persons and occasions
Occasions(O)	Averaged deviation from grand mean over persons and items
P x I	Effect of interaction between person p and item i averaged over occasions
P x O	Effect of interaction between person p and occasion o averaged over items
I x O	Effect of interaction between item j and occasion o
P x I x O, e	Effect of interaction between person p , item j and occasion o , containing a random error e

Fig. 1 Venn diagram of two-facet design person x item x occasion ($P \times I \times O$) and the components definitions for two-facet generalisability analysis.

studies, 10 were university-educated, 8 had diverse educational backgrounds, and 1 individual had no formal education.

The Comprehensive Assessment of At-Risk Mental States (CAARMS) was used to assess for risk at five distinct intervals during a two-year period. The data from the PANSS were measured every 6 months and denoted as Baseline (T0), 6-months (T1), 12-months (T2), 18-months (T3), and 24-months (T4) to provide a comprehensive understanding of their mental state and progression over time.

Measures and assessments selection

The PANSS comprises of 30 questions that are categorised into three symptom domains: positive (7 items, e.g. delusions, conceptual disorganisation, hallucinations, excitement, grandiosity, and hostility), negative (7 items, e.g. blunted affect, poor rapport, emotional withdrawal, lack of spontaneity and flow in conversation), and general (16 items, e.g. depression, anxiety, and lack of attention). Each item is rated based on a seven-point scale ranging from 1 as absent to 7 as severe. In the current study, we also utilised the proposed 5-factor model from Lim et al.²⁷.

While the CAARMS was designed specifically for UHR populations and provides a sensitive measure of subthreshold psychotic symptoms, we chose to use and validate the PANSS in our study for several reasons. First, the PANSS offers a broader range of symptom measures, including general psychopathology items, which allows for a more comprehensive analysis of symptom stability across various domains. This is particularly valuable for our G-theory approach, which benefits from a wider range of items to analyse sources of variance.

Second, the common use of the PANSS in psychosis research and its earlier examination using G-theory facilitates comparisons with other studies and populations, potentially allowing us to track symptom trajectories from the UHR state through to potential transition to full psychosis. This continuity in measurement could provide valuable insights into the evolution of symptoms over time.

Third, while the PANSS was developed for individuals with diagnosed psychotic disorders, recent research has suggested that it may still be sensitive to subthreshold symptoms in UHR populations. For example, Fulford et al. (2014) found that the PANSS demonstrated good reliability and validity in clinical high-risk samples, albeit with some limitations³⁴.

Data analyses

Comparing the reliability and generalisability of the 3 and 5 factor models of the PANSS using G-theory. In the current study, all

generalisability analyses were performed using the EduG 6.1-e software, in accordance with the procedures outlined in Cardinet et al.³⁷.

G-theory includes two studies: D (decision) and G (generalisability). D-Studies design the measurement process optimally, whereas G-Studies estimates the variance components and the sources of measurement error. In both studies, the variables are person (P), item (I), and occasion (O); and the random effects model is described as $P \times O \times I$. The (I) is fixed, and (P) and (O) are treated as infinite facets.

The (P) is considered as the primary object of measurement rather than a potential source of error, whereas the (I) and (O) are considered as instrumentation facets (Fig. 1). The error variance caused by the interaction between (P) and (O) reflects individual fluctuations over time¹⁷. Brennan's equations are used to calculate the variance components for each facet and their interactions³⁸. Whimbeys correction is applied to adjust variance estimates³⁷ using the formula $((N(f)-1)/N(f))$, where $N(f)$ is representing the population size of the f facet.

Then, using generalisability analysis, the relative and absolute error variances as well as G-coefficients for the differentiation facet (persons) are used to evaluate the contribution of each facet to the overall variance scores.

The absolute G-coefficients (G_a) contain all sources of variation that indirectly affect an absolute measure, whereas relative G-coefficients (G_r) explicitly account for variance linked to the object of measurement³⁷.

Identifying the stable and fluctuating psychopathological features

Using formulae developed by Medvedev et al.¹⁷, a state component index (SCI) and trait component index (TCI) were computed to assess the capacity of scales/items to measure both stable and fluctuating symptoms¹⁷. These indices show how much of a measure's variation is attributable to its stable and fluctuating variables. The D-study optimised assessment by adjusting and altering facet designs to investigate the characteristics of individual items reflecting certain illnesses.

Comparing the reliability and generalisability of the 3 and 5 factor models of the PANSS using classical test theory

The internal consistency Cronbach's alpha (α), descriptive statistics, test-retest, and McDonald's omega (ω) coefficients were calculated for the 3 and 5 factor models of PANSS, as well as the overall scale ($n = 159 \times 5$ instances). The results were compared to the G-theory analysis.

Table 1. G-study estimates of variance components^a.

Scale/subscale	P	I	O	PxI	PxO	IxO	PxIxO	G-r	G-a	SCI
3-Factor Model										
PANSS Total (PT)	0.02	<0.001	0.01	<0.00	0.004	<0.00	<0.00	0.85	0.56	0.17
PT Variance %	56.0	0.0	34.0	0.0	10.0	0.0	0.0	–	–	–
Positive Symptoms (PS)	0.06	0.001	0.01	0.004	0.004	0.002	0.009	0.77	0.64	0.06
PS Variance %	64.0	1.4	12.5	4.6	4.5	2.3	10.6	–	–	–
Negative Symptoms (NS)	0.06	0.001	0.01	0.004	0.004	0.002	0.01	0.77	0.68	0.05
NS Variance %	68.0	0.7	8.9	4.5	3.8	2.3	11.7	–	–	–
General Symptoms (GS)	0.03	<0.001	0.01	0.003	0.004	0.001	0.003	0.77	0.56	0.12
GS Variance %	56.0	0.8	25.0	4.7	7.5	1.1	4.8	–	–	–
General (excluding 15) (G15)	0.035	<0.001	0.014	0.003	0.004	0.001	0.003	0.78	0.58	0.10
G15 Variance %	58.0	0.0	33.5	0.3	9.8	0.0	0.3	–	–	–
5-Factor Model										
Positive Symptoms (PS)	0.04	0.007	0.007	0.003	0.004	0.003	0.01	0.66	0.56	0.09
PS Variance %	56.0	2.5	8.4	5.5	5.8	4.3	17.6	–	–	–
Negative Symptoms (NS)	0.03	<0.00	0.01	0.008	0.004	0.002	0.01	0.56	0.44	0.12
NS Variance %	44.0	0.8	16.2	12.3	6.2	4.0	16.5	–	–	–
Cognitive/Disorganisation (CD)	0.01	<0.00	0.01	0.008	0.004	0.002	0.01	0.33	0.24	0.12
CD Variance %	24.0	0.0	19.4	18.1	8.7	3.9	21.9	–	–	–
Depression/Anxiety (DA)	0.07	0.003	0.009	<0.00	0.07	0.004	0.02	0.68	0.58	0.08
DA Variance %	58.0	2.9	7.9	8.5	5.2	3.6	13.9	–	–	–
Hostility (H)	0.03	0.002	0.03	0.02	0.003	0.002	0.01	0.40	0.27	0.09
H Variance %	27.0	1.7	28.6	22.3	3.1	2.3	15.0	–	–	–

^aDifferentiation variance component of person (P); Absolute error variance of item (I), occasion (O), interaction between person and item (PxI), person and occasion (PxO), item and occasion (IxO) and person, item and occasion (PxIxO); Relative G-coefficient (Gr); Absolute G-coefficient (Ga); and State Component Index (SCI).

RESULTS

G-Study

Table 1 shows the actual values of the variance components and the percentage of variation that each contributes to the assessment scores, as well as the findings of the G-Study analysis of the PANSS total scale and the subscales of the 3 and 5 factor models.

With just a modest amount of variation attributable to fluctuations (SCI = 0.17), the total PANSS scale demonstrated acceptable generalisability of scores across several occasions and the sample population (Gr = 0.85 and Ga = 0.56). The highest proportion of error variation in the total assessment score (34.0%) was attributed to (O), which is essentially the overall change of score as a function of time for all participants.

The 3-factor model of PANSS (positive, negative, and general symptoms subscales) showed acceptable relative generalisability of scores across (P) and (O) (Gr: positive = 0.77, negative = 0.77, and general = 0.77) respectively, just under 0.80 benchmark. However, none of the subscales of the 5-factor model reached the reliability standard, with Gr values falling below the 0.70 threshold (Table 1). The true variance explained by individual differences ranges from as low as 24% (cognitive/disorganisation subscale) to 56% (positive symptoms subscale), while at least 70% is required to reliably differentiate between individuals.

D-Study

The D-study sought to identify PANSS items that are susceptible to fluctuating changes. In Table 2, the (P) (stable symptoms) and (O) × (P) interaction variance components (fluctuating symptoms), along with the SCI values that were used to find items that represent these fluctuations are presented. With a mean of 0.38

and a range of 0.27 to 0.50, the SCI values indicate that stable patterns account for the majority of symptom variation. The results showed that the following six items had the highest level of stability (SCI ≤ 0.40): anxiety, guilt feelings, uncooperativeness, excitement, poor rapport, and poor impulse control. On the other hand, twelve items—delusions, hallucinatory behaviour, conceptual disorganisation, suspiciousness, difficulty thinking abstractly, emotional withdrawal, lack of spontaneity, motor retardation, somatic concern, disorientation, active social avoidance, and disturbance of volition—showed the highest variability, with a higher state-related variance (SCI = 0.50–0.68).

Combining items with a SCI above the average of 0.38 did not improve the scale's sensitivity to temporal changes. This resulted in a greater proportion of trait-related (P) variation (e.g., SCI < 0.30). Modifications to positive and negative factors did not improve reliability, except for the general factor, where reliability increased when somatic concern was included. Additionally, the 5-factor model was found to lack acceptable generalisability, and no further modifications were pursued.

Comparison with conventional results

We compared G-theory results with conventional results obtained using CTT.

Tables 3 and 4 present the descriptive statistics and test-retest reliability coefficient that were calculated using the current sample for the 3 and 5 PANSS assessments.

The 3-factor model demonstrated a statistically significant decrease in positive, general, and total symptoms at each evaluation point when compared to the baseline (Table 3). In the 5-factor model, a decrease in all assessment's points across positive, negative, cognitive, and mostly in hostility can be observed (Table 4).

Table 2. Variance components and the SCI for the individual PANSS items^a.

Item/scale	P	PO	SCI
Positive symptoms			
P1: Delusions	0.03	0.06	0.68
P2: Conceptual Disorganisation	0.09	0.10	0.53
P3: Hallucinatory Behaviour	0.08	0.13	0.61
P4: Excitement	0.24	0.12	0.34
P5: Grandiosity	0.05	0.04	0.46
P6: Suspiciousness/Persecution	0.06	0.08	0.56
P7: Hostility	0.11	0.07	0.40
Negative symptoms			
N1: Blunted Affect	0.12	0.10	0.44
N2: Emotional Withdrawal	0.07	0.14	0.67
N3: Poor Rapport	0.24	0.12	0.33
N4: Passive/Apathetic Social Withdrawal	0.05	0.05	0.49
N5: Difficulty in Abstract Thinking	0.07	0.10	0.60
N6: Lack of Spontaneity & Flow of Conversation	0.04	0.07	0.64
N7: Stereotyped Thinking	0.10	0.10	0.51
General symptoms			
G1: Somatic Concern	0.08	0.14	0.63
G2: Anxiety	0.23	0.13	0.35
G3: Guilt Feelings	0.07	0.05	0.39
G4: Tension	0.10	0.09	0.47
G5: Mannerisms and Posturing	0.08	0.06	0.45
G6: Depression	0.12	0.08	0.40
G7: Motor Retardation	0.10	0.13	0.57
G8: Uncooperativeness	0.30	0.13	0.30
G9: Unusual Thought Content	0.07	0.05	0.40
G10: Disorientation	0.08	0.08	0.50
G11: Poor Attention	0.06	0.06	0.48
G12: Lack of Judgment and Insight	0.14	0.10	0.42
G13: Disturbance of Volition	0.09	0.11	0.56
G14: Poor Impulse Control	0.21	0.13	0.38
G15: Preoccupation	0.07	0.05	0.43
G16: Active Social Avoidance	0.08	0.08	0.52

^aDifferentiation variance of person (P); Absolute error variance of person and occasion interaction (P × O); and the State Component Index (SCI).

Figures 2 and 3 show that the symptoms decrease compared to the baseline at the following assessments with very similar patterns. This reflects the variance component of (O) in G-theory results. Test-retest reliability coefficients progressively declined as a function of time between baseline and assessment timepoint for both models, suggesting that the fluctuating nature of symptoms in this populations is consistent with the G-theory results (Tables 3 and 4). Internal reliability of the total PANSS ($\alpha = 0.86$, $\omega = 0.87$) was consistent with Gr, measuring the overall reliability at the relative level. Consistent with G-theory results, for 3-factor model, the internal reliability was lower compared to the total scale (positive $\alpha = 0.62$ and $\omega = 0.64$; negative $\alpha = 0.69$ and $\omega = 0.74$; general $\alpha = 0.79$ and $\omega = 0.81$).

Similar to G-theory results, compared to the 3-factor model, the 5-factor model displayed lower internal reliability (positive $\alpha = 0.69$ and $\omega = 0.71$; negative $\alpha = 0.79$ and $\omega = 0.80$; cognitive $\alpha = 0.62$ and $\omega = 0.72$; depression $\alpha = 0.69$ and $\omega = 0.69$; and hostility $\alpha = 0.52$ and $\omega = 0.58$).

DISCUSSION

This study applied G-theory to validate PANSS in individuals at UHR for psychosis, quantify differences between stable and fluctuating symptom patterns, and compare the reliability and generalisability of the 3- and 5-factor models. Our findings reveal several key insights into the nature of symptomatology in this population.

Overall, the total PANSS score showed acceptable reliability in our UHR sample and outperformed the individual subscales. The overall PANSS score indicated stable symptoms more than fluctuating symptoms, and it was unaffected by other measurement errors.

The overall symptoms operationalised by the PANSS items cannot be classified as fluctuating in this population. The most stable items were predominantly general symptoms (anxiety, guilt feelings, uncooperativeness, poor impulse control), with one item each from positive (excitement) and negative (poor rapport) subscales.

The most fluctuating symptoms were positive (4 items including conceptual disorganisation, delusions, suspiciousness, and hallucinatory behaviour), negative (4 items including difficulty in abstract thinking, emotional withdrawal, stereotyped thinking, and lack of spontaneity), and general (5 items including somatic concern, motor retardation, active social avoidance, disorientation, and disturbance of volition).

Reliability and generalisability of the 3 and 5 factor models of PANSS

The 3-factor model demonstrated acceptable score reliability and generalisation across occasions and sample populations. The 5-factor model demonstrated reduced reliability, falling below the required benchmark 0.70. This implies that measurement error attributable to subscale features, such as interaction between (P), (O), and (I), exceeded 30% for these subscales³⁹. These results indicate that the true differences between individuals was as low as 24% for cognitive/disorganisation subscale and only exceeded 50% for the positive symptoms subscale of the 5-factor model. Our findings suggest that measurement errors affecting 5-factor model are inherent to the combination and number of items rather than the conceptual validity of 5-factor model. Therefore, using a 5-factor solution over a 3-factor model in UHR is not supported by current G-theory analysis. Future studies may develop more reliable 5-factor model by including additional items with better psychometrics properties to ensure reliable and valid assessments. Modifications to enhance reliability in positive and negative factors yielded limited success, except for a minor improvement in the general factor with the exclusion of somatic concern. This finding prompts a consideration of the purpose of fluctuating nature of this symptom. Additionally, 5-factors fell below acceptable generality, and while no modifications were pursued, acknowledging these limitations underscores the complexity of UHR symptomatology and the ongoing need for refinement in assessment tools and conceptual frameworks.

Stable and fluctuating symptom patterns in UHR individuals

Although negative symptoms were largely stable, focusing only on subscale-level reporting would overlook important fluctuating changes. For instance, several negative symptoms showed variability, such as those on the general subscale (disturbance of volition and motor retardation). Individual item scores, as seen in Table 2, may be more helpful for monitoring certain symptoms that fluctuate or remain stable, even if the PANSS and its subscale scores are beneficial for evaluating overall symptomatic changes.

A more thorough assessment of individual symptom patterns instead of subscale and total scores allows for a more detailed

Table 3. Means, standard deviation (SD), and test-retest coefficients for the 5-factors PANSS subscales and the total scale ($n = 159 \times 5$ occasions).

Scale/assessment	Baseline	6 Months	12 Months	18 Months	24 Months
PANSS Positive					
Mean (SD)	8.51(2.6)	7.37**(2.43)	6.90**(2.07)	6.97**(2.40)	7.26**(2.3)
Test-retest (r^a)		0.46	0.38	0.42	0.28
PANSS Negative					
Mean (SD)	12.05(4.82)	10.35**(3.9)	10.23**(3.93)	9.19**(3.0)	9.68**(3.1)
Test-retest (r^a)		0.5	0.33	0.4	0.32
PANSS Cognitive					
Mean (SD)	9.72(2.84)	9.05**(2.15)	8.92**(1.92)	8.71**(1.9)	8.92**(2.3)
Test-retest (r^a)		0.66	0.60	0.58	0.52
PANSS Depression					
Mean (SD)	9.94(3.64)	8.27**(3.1)	8.14**(2.84)	8.24**(2.53)	8.55**(2.76)
Test-retest (r^a)		0.38	0.36	0.27	0.31
PANSS Hostility					
Mean (SD)	8.13(2.46)	7.53**(1.95)	7.57**(1.7)	7.50**(1.8)	7.58**(2.12)
Test-retest (r^a)		0.27	0.27	0.07	0.02

^aTest-retest bivariate correlations between assessments and the baseline.
Note: ** $p < 0.01$. Mean difference is significant compared to the baseline.

Table 4. Means, standard deviation (SD), and test-retest coefficients for the 3-factors PANSS subscales (positive, negative, and general) and the total scale ($n = 159 \times 5$ occasions).

Scale/assessment	Baseline	6 Months	12 Months	18 Months	24 Months
PANSS Total					
Mean (SD)	48.35**(11.50)	42.57**(9.70)	51.75**(8.76)	42.62**(8.16)	42.01**(8.70)
Test-retest (r^a)	--	0.66	0.51	0.46	0.38
PANSS Positive					
Mean (SD)	10.71**(2.75)	9.57**(2.6)	9.01**(2.03)	9.03**(2.38)	9.35**(2.53)
Test-retest (r^a)	--	0.57	0.35	0.38	0.31
PANSS Negative					
Mean (SD)	12.16**(4.17)	11.15**(3.66)	11.14**(3.47)	10.19**(2.90)	10.67**(3.07)
Test-retest (r^a)	--	0.58	0.37	0.38	0.36
PANSS General					
Mean (SD)	25.48**(6.93)	21.86**(5.6)	21.60**(5.14)	21.40**(4.84)	21.99**(5.23)
Test-retest (r^a)	--	0.66	0.57	0.50	0.41

^aTest-retest bivariate correlations between assessments and the baseline.
Note: ** $p < 0.01$. Mean difference is significant compared to the baseline.

evaluation and treatment responses. However, this approach has not been widely reported in psychosis research.

In our study, poor rapport and blunted affect were observed to be stable, however, Fusar-Poli et al.⁴⁰ found that negative symptoms in schizophrenia changed with time. They utilised the CTT method, which, although controlling for person variance, cannot distinguish clearly between variation resulting from changes at the group level and variability of particular items indicating symptoms at the individual level¹⁶. Thus, Fusar-Poli et al.⁴⁰ findings on evolving negative symptoms may indicate how effectively negative items reflect change at the group level but how poorly they reflect individual change when unwanted error variation is taken into consideration.

Contrary to earlier studies based on the CTT approach, our study using the G-theory technique revealed that both stable and fluctuating symptoms were observed on the PANSS positive and

general subscales. This shows that positive symptoms are more changeable in UHR individuals⁵.

This discrepancy is most likely due to CTT constraints, emphasises the need for further G-theory replication of these findings with different sample sizes and various symptom severity. Moreover, it emphasises the significance of assessing changes in symptoms at the individual item level as opposed to scale and subscale scores.

Our study presents interesting findings that delusions had a fluctuating nature in youth at risk of psychosis ($SCI = 0.68$), in contrast to previous research¹⁵ which has identified delusions as relatively stable features in adults diagnosed with schizophrenia. This could be attributed to several factors, including developmental differences, the neuroplasticity present in younger individuals, and potentially differing stages in the progression of psychosis that may or may not be developed.

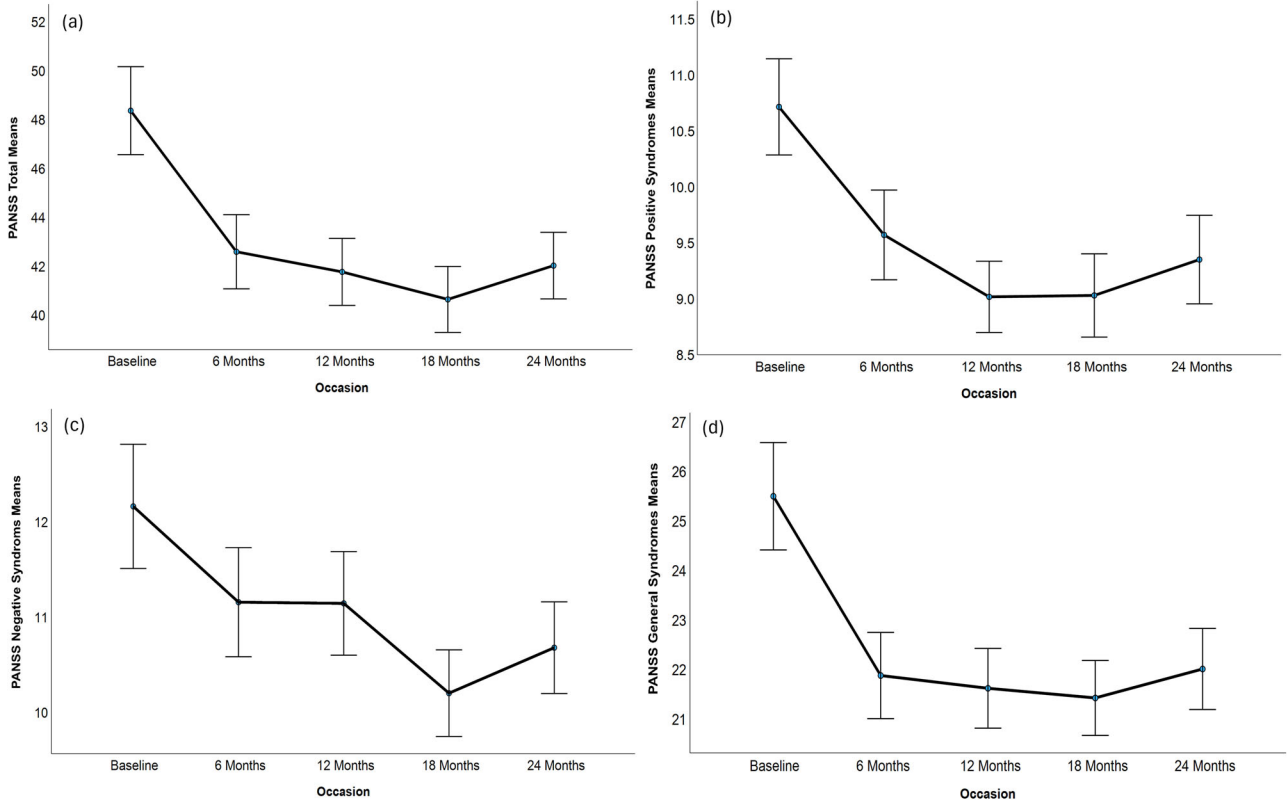


Fig. 2 Psychotic symptoms mean scores computed at 5 time points for the following measures. **a** PANSS total and PANSS 3 factors model: **b** positive, **c** negative and **(d)** general symptoms.

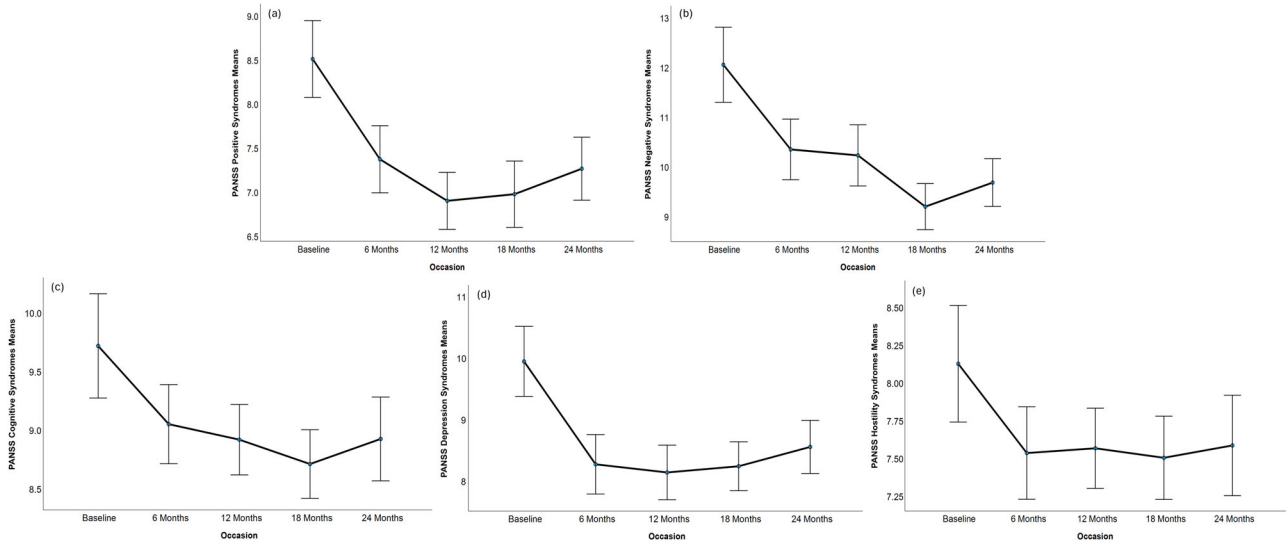


Fig. 3 Psychotic symptoms mean scores computed at 5 time points for PANSS 5-factors model. **a** positive, **b** negative, **c** cognitive, **d** depression, and **(e)** hostility symptoms across two years.

Implications for clinical practice

The observation of considerable variability in delusional experiences among young individuals implies a potential opportunity for early intervention and intervention efficacy. Timely identification of such prodromal signs and prompt intervention strategies could potentially alter the trajectory of psychosis progression. This prospect holds significant promise, as it offers tailored and age-appropriate interventions to address the needs of young individuals more effectively.

Limitations and future research

Firstly, the present study has a limitation related to cultural diversity and interpret, particularly in the context of Singapore. Secondly, G-theory is a powerful tool to assess reliability and identify sources of measurement error, but it is limited in assessing construct validity and the interactive network of symptoms in this population. And finally, we recognise that the PANSS may have limitations when applied to UHR populations. Therefore, future research may benefit from the development of instruments

specifically designed to capture the complete range of symptom intensity and variability in UHR populations while maintaining comparability with measures used in established psychotic disorders.

Future studies should apply network analysis to investigate construct validity and interactive links between symptoms in this population over time⁴¹.

This study's findings will help future researchers design a reliable tool for measuring stable and fluctuating psychotic symptoms. The current study of the PANSS utilising G-theory offers a methodological foundation for developing such a tool.

This methodology holds potential to advance clinical practice by enabling precise longitudinal tracking of patients' symptomatology through dual mechanisms. First, a fluctuating symptoms scale is used to evaluate individual symptom change, followed by an assessment of overall psychopathology using a stable symptoms measure.

In conclusion, the present study utilised G-theory to analyse PANSS data in individuals at UHR for psychosis, assessing fluctuating versus stable symptom patterns and evaluating reliability and generalisability of 3 and 5 factor models. While the PANSS demonstrates reliability in capturing stable symptoms, the 5-factor model's limitations underscore the need for refinement in assessment tools. Moreover, the identification of fluctuating symptom patterns highlights the importance of individual-level evaluation for tailored interventions. These findings not only enhance our theoretical grasp of psychosis progression but also hold immense potential in guiding more targeted and effective interventions.

The used in the paper G-theory is a complementary approach when compared to the previously used spiking neural network (SNN) approach, that first represents changes of variables over time encoded into spikes, before a predictive SNN model is created^{42,43}. Integrating the two approaches, making use of their advantages, would be a viable step forward.

DATA AVAILABILITY

The participant permission statement prevents the dataset from being made publicly available; however, the corresponding author may provide it upon reasonable request and with NTU and AUT's approval as part of a process for data sharing agreements.

Received: 16 May 2024; Accepted: 29 August 2024;

Published online: 04 October 2024

REFERENCES

- Coutts, F., Koutsouleris, N. & McGuire, P. Psychotic disorders as a framework for precision psychiatry. *Nat. Rev. Neurol.* **19**, 221–234 (2023).
- Arciniegas, D. B. Psychosis. *Continuum: Lifelong Learn. Neurol.* **21**, 715 (2015).
- Punsoda-Puche, P., Barajas, A., Mamano-Grande, M., Jiménez-Lafuente, A. & Ochoa, S. Relationship between social cognition and premorbid adjustment in psychosis: a systematic review. *Schizophrenia* **10**, 36 (2024).
- Golay, P. et al. Symptom dimensions stability over time in recent onset psychosis: a prospective study. *Schizophr. Res.* **246**, 126–131 (2022).
- Savill, M., Banks, C., Khanom, H. & Priebe, S. Do negative symptoms of schizophrenia change over time? A meta-analysis of longitudinal data. *Psychol. Med.* **45**, 1613–1627 (2015).
- Arndt, S., Andreasen, N. C., Flaum, M., Miller, D. & Nopoulos, P. A longitudinal study of symptom dimensions in schizophrenia: prediction and patterns of change. *Arch. Gen. Psychiatry* **52**, 352–360 (1995).
- Kirkpatrick, B., Mucci, A. & Galderisi, S. Primary, enduring negative symptoms: an update on research. *Schizophr. Bull.* **43**, 730–736 (2017).
- Correll, C. U. & Schooler, N. R. Negative symptoms in schizophrenia: a review and clinical guide for recognition, assessment, and treatment. *Neuropsychiatr. Dis. Treat.* **16**, 519–534 (2020).
- Kalisova, L. et al. Impact of cognitive performance and negative symptoms on psychosocial functioning in Czech schizophrenia patients. *Schizophrenia* **9**, 43 (2023).
- Mucci, A., Merlotti, E., Üçok, A., Aleman, A. & Galderisi, S. Primary and persistent negative symptoms: concepts, assessments and neurobiological bases. *Schizophr. Res.* **186**, 19–28 (2017).
- McHugh, M. J. et al. The ultra-high-risk for psychosis groups: evidence to maintain the status quo. *Schizophr. Res.* **195**, 543–548 (2018).
- Cravaro, V. R. & Raballo, A. Early detection of schizophrenia: a clinical-psychopathological revision of the ultra-high risk approach. *J. Psychopathol.* **20**, 442–450 (2014).
- van der Tuin, S. et al. Dynamic symptom networks across different at-risk stages for psychosis: an individual and transdiagnostic perspective. *Schizophr. Res.* **239**, 95–102 (2022).
- Stahl, S. M. & Buckley, P. F. Negative symptoms of schizophrenia: a problem that will not go away. *Acta Psychiatr. Scand.* **115**, 4–11 (2007).
- Medvedev, O. N. et al. A novel way to quantify schizophrenia symptoms in clinical trials. *Eur. J. Clin. Invest.* **51**, e13398 (2021).
- Paterson, J. et al. Distinguishing transient versus stable aspects of depression in New Zealand Pacific Island children using Generalizability Theory. *J. Affect. Disord.* **227**, 698–704 (2018).
- Medvedev, O. N., Krägeloh, C. U., Narayanan, A. & Siebert, R. J. Measuring mindfulness: applying generalizability theory to distinguish between state and trait. *Mindfulness* **8**, 1036–1046 (2017).
- Truong, Q. C. et al. Clinical investigation of dynamic and enduring aspects of global cognition in aged population. *Eur. J. Clin. Invest.* **52**, e13681 (2022).
- Vangeneugden, T., Laenen, A., Geys, H., Renard, D. & Molenberghs, G. Applying concepts of generalizability theory on clinical trial data to investigate sources of variation and their impact on reliability. *Biometrics* **61**, 295–304 (2005).
- Khan, A. et al. Assessing the sources of unreliability (rater, subject, time-point) in a failed clinical trial using items of the Positive and Negative Syndrome Scale (PANSS). *J. Clin. Psychopharmacol.* **33**, 109–117 (2013).
- Kay, S. R., Fiszbein, A. & Opler, L. A. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* **13**, 261–276 (1987).
- Jerrell, J. M. & Hrisko, S. Utility of two PANSS 5-factor models for assessing psychosocial outcomes in clinical programs for persons with schizophrenia. *Schizophr. Res. Treat.* **2013** (2013).
- van der Gaag, M. et al. The five-factor model of the Positive and Negative Syndrome Scale I: confirmatory factor analysis fails to confirm 25 published five-factor solutions. *Schizophr. Res.* **85**, 273–279 (2006).
- Rodríguez-Jimenez, R. et al. Cognition and the five-factor model of the positive and negative syndrome scale in schizophrenia. *Schizophr. Res.* **143**, 77–83 (2013).
- Jiang, J., Sim, K. & Lee, J. Validated five-factor model of positive and negative syndrome scale for schizophrenia in Chinese population. *Schizophrenia research* **143**, 38–43 (2013).
- Wallwork, R., Fortgang, R., Hashimoto, R., Weinberger, D. & Dickinson, D. Searching for a consensus five-factor model of the Positive and Negative Syndrome Scale for schizophrenia. *Schizophr. Res.* **137**, 246–250 (2012).
- Lim, K. et al. Large-scale evaluation of the Positive and Negative Syndrome Scale (PANSS) symptom architecture in schizophrenia. *Asian J. Psychiatry* **62**, 102732 (2021).
- Fountoulakis, K. N. et al. Staging of schizophrenia with the use of PANSS: an international multi-center study. *Int. J. Neuropsychopharmacol.* **22**, 681–697 (2019).
- Peralta, V. & Cuesta, M. J. Psychometric properties of the positive and negative syndrome scale (PANSS) in schizophrenia. *Psychiatry Res.* **53**, 31–40 (1994).
- Pignon, B. et al. A latent class analysis of psychotic symptoms in the general population. *Aust. N Z J. Psychiatry* **52**, 573–584 (2018).
- Edgar, C. J. et al. Reliability, validity and ability to detect change of the PANSS negative symptom factor score in outpatients with schizophrenia on select antipsychotics and with prominent negative or disorganized thought symptoms. *Psychiatry Res.* **218**, 219–224 (2014).
- van der Gaag, M. et al. The five-factor model of the Positive and Negative Syndrome Scale II: a ten-fold cross-validation of a revised model. *Schizophr. Res.* **85**, 280–287 (2006).
- Obermeier, M. et al. Is the PANSS used correctly? A systematic review. *BMC Psychiatry* **11**, 1–5 (2011).
- Fulford, D. et al. Symptom assessment in early psychosis: the use of well-established rating scales in clinical high-risk and recent-onset populations. *Psychiatry Res.* **220**, 1077–1083 (2014).
- Levine, S. Z., Rabinowitz, J., Engel, R., Etschel, E. & Leucht, S. Extrapolation between measures of symptom severity and change: an examination of the PANSS and CGI. *Schizophr. Res.* **98**, 318–322 (2008).
- Lee, J. et al. The longitudinal youth at risk study (LYRIKS)—an Asian UHR perspective. *Schizophr. Res.* **151**, 279–283 (2013).
- Cardinet, J., Johnson, S. & Pini, G. *Applying generalizability theory using EduG*. (Routledge, 2011)

38. Brennan, R. L. Generalizability Theory. *Educational Measurement: Issues and Practice* **11**, 27–34 (1992).
39. Bloch, R. & Norman, G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med. Teacher* **34**, 960–992 (2012).
40. Fusar-Poli, P. et al. Treatments of negative symptoms in schizophrenia: meta-analysis of 168 randomized placebo-controlled trials. *Schizophr. Bull.* **41**, 892–899 (2015).
41. Truong, Q. C. et al. Examining the validity of the Mini-Mental State Examination (MMSE) and its domains using network analysis. *Psychogeriatrics* **24**, 259–271 (2024).
42. Kasabov, N. K. *Time-space, Spiking Neural Networks and Brain-inspired Artificial Intelligence* (Springer, 2019).
43. Dobarjeh, Z. et al. Investigation of social and cognitive predictors in non-transition ultra-high-risk individuals for psychosis using spiking neural networks. *Schizophrenia* **9**, 10 (2023).

ACKNOWLEDGEMENTS

This research is supported by the Ministry of Business, Innovation and Employment (MBIE), New Zealand, under the New Zealand-Singapore Data Science Research Programme and the National Research Foundation, Singapore under its Industry Alignment Fund-Pre-positioning (IAF-PP). All findings, interpretations, and conclusions in this study are solely from the author(s) and do not reflect the views of the funders. The National Research Foundation Singapore funded the LYRIKS data as part of the National Medical Research Council Translational and Clinical Research Flagship Programme (NMRC/TCR/003/2008). Dr Wilson Wen Bin Goh acknowledge support from the Ministry of Education (MOE) Tier 1 grant (RT11/21).

AUTHOR CONTRIBUTIONS

Dr. Zohreh Dobarjeh carried out the experiments, analysed data using statistical and G-theory models, evaluated the results, authored the initial draft, and contributed to the project funding procedure. Dr. Oleg Medvedev assessed the proposed the research design, G-theory data modelling, as well as contributed to the manuscript's writing and interpretation of the results and findings. Dr. Maryam Dobarjeh contributed to designing the project, data analysis, writing, and funding initiation. Mr Balkaran Singh contributed to analyse the statistical data and the manuscript. Professor Alexander Sumich contributed to analysis and interpreting the results and manuscript writing. Mr Sugam Budhraj contributed to data preprocessing, cleaning, and statistical analysis. Associated Professor Wilson Wen Bin Goh contributed to the study project's execution, outcome interpretation, and manuscript writing. Dr Jimmy

Lee contributed to initiating the study funding, longitudinal data collection, and manuscript writing. Dr. Margaret Williams contributed to the manuscript's writing and result interpretation. Professor Edmund M-K Lai contributed to the initiating the project, funding and writing the manuscript. Professor Nikola Kasabov, as the project's scientific leader, contributed to the research plan, study design, and manuscript writing. After reading the submitted version of the work, all authors have given their approval.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Zohreh Dobarjeh or Wilson Wen Bin Goh.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



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

© The Author(s) 2024