Brazilian Journal of Psychiatry
bjp
Revista Brasileira de Psiquiatria
Brazilian Psychiatric Association
Fully open access
No APCs

# BJP PRE-PROOF
# (article published as accepted)

Brief Communication

## Addressing the elephant in the screening room: an item response theory analysis of the Prodromal Questionnaire (PQ-16) for at-risk symptoms of psychosis

Christophe Gauld, Pierre Fourneret, Ben Alderson-Day, Emma Palmer-Cooper, Clément Dondé

http://doi.org/10.47626/1516-4446-2024-3614

Submitted: 05-Mar-2024

Accepted: 03-Aug-2024

# Addressing the elephant in the screening room: an item response theory analysis of the Prodromal Questionnaire (PQ-16) for at-risk symptoms of psychosis

Running Title: Discriminating the symptoms of the clinical high risk

Christophe Gauld[1,2], Pierre Fourneret[1,2], Ben Alderson-Day[3], Emma Palmer-Cooper[4], Clément Dondé[5,6,7]

1 Service de Psychopathologie du Développement, Hospices Civils de Lyon, Lyon, France.
2 Institut des Sciences Cognitives Marc Jeannerod & Université Claude Bernard Lyon, Lyon, France.
3 Department of Psychology, Durham University, Durham, United Kingdom.
4 School of Psychology, Centre for Innovation in Mental Health, University of Southampton
5 Univ. Grenoble Alpes, Grenoble, France.
6 INSERM, U1216, Grenoble, France.
7 Psychiatry Department, CHU Grenoble Alpes, Grenoble, France.

Corresponding author:
Clément Dondé
Univ. Grenoble Alpes, F-38000 Grenoble, France
E-mail: clement.donde@univ-grenoble-alpes.fr

## ABSTRACT

**Objective:** Within the context of patients at-risk of psychosis, where a variety of symptoms are present, identifying the most discriminative symptoms is essential for efficient detection and management.

**Methods**: This cross-sectional online study analyzed individuals from the general population in order to better assess their risk of presenting symptoms belonging to the clinical high risk (CHR) for psychosis, called “CHR-related symptoms”. The Prodromal Questionnaire-16 (PQ-16) served as a self-report screening tool. Item response theory (IRT) with a graded response model was used to assess the discrimination and difficulty of its criteria.

**Results**: The analysis included 936 participants (mean age: 21.5 years; 28.1% male, 71.9% female). “Déjà vu” stood out for its high discriminative power, while “Voices or whispers” and “Seen things” demonstrated strong precision relatively to the other CHR-related symptoms. Conversely, “Smell or taste” and “Changing face” were associated with the most severe cases relatively to the other CHR-related symptoms.

**Conclusions**: This study identified the most indicative CHR-related symptoms to emphasize their significance in accurately assessing severity and guiding targeted preventative interventions.


**Keywords:** Clinical high-risk symptoms; discrimination; item response theory; early intervention



## INTRODUCTION

The identification of individuals at risk of developing psychosis has become a central objective in modern clinical practice. The concept of “clinical high risk” for psychosis (CHR), also referred to as “Attenuated Psychosis Syndrome” in the DSM-5, serves as an inclusive clinical construct for identifying potentially prodromal manifestations of psychotic disorders.[1,2] However, community screening approaches are associated with a significant false-positive rate, reaching approximately two-thirds of screened individuals.[3] Recent research indicates that questionnaire items

endorsed by a substantial percentage of the general population who do not seek help may capture more normative experience rather than symptomatology associated with psychosis risk.[4] Notably, some screened participants reported distress but not psychotic-like experiences (PLEs), which are more commonly assessed in the general population. The challenge is further compounded by the prevalence of numerous non-specific clinical manifestations. A recent analysis identified at least 68 distinct symptoms across validated CHR screening questionnaires.[5] One proposed solution to address to this heterogeneity involves identifying a more concise set of discriminating symptoms from existing screening approaches. It therefore seems particularly important to target items from such questionnaires that can best reflect normative experiences and symptoms associated with psychosis risk. To this end, we employed item response theory analysis on the widely used Prodromal Questionnaire in its brief 16-item form (PQ-16), a relevant tool "for screening potential at-risk mental state",[6,7] in large cohorts through population-based web screening.[3] The objective of our study is to identify the most discriminative symptoms belonging the CHR entity, without making any assumption about a formal diagnosis, using the PQ-16.

## METHODS

### *Participants*

We recruited participants aged 18-35 from a non-help-seeking general population through Amazon's Mechanical Turk (MTurk) and university mailing lists in France and the UK, as part of the TONE-P study[8] (see Supplementary Methods for details). To distinguish this sample from individuals definitively meeting CHR criteria and show the risk of presenting symptoms belonging to CHR, we analyze here the CHR-related symptoms. This terminology highlights our interest in symptoms referring to entities such as CHR, schizotypal traits or psychosis, identified as the most relevant on the PQ-16. Indeed, although the PQ-16 only allows the collection of symptoms from individuals at risk of developing psychosis, its factors reflect dimensions such as "perceptual abnormalities/hallucinations" and "general symptoms associated with psychosis-risk".

Ethical approval was obtained from the University Grenoble Alpes, France, Durham University and Southampton University, United Kingdom (UK).

### *Questionnaire*

The PQ-16 is a 16-items self-report screening questionnaire validated in English and French.[7,9] Each item assesses anomalous psychotic experiences and associated distress on a 5-point Likert scale with five options (0 = “none”, 1 = “any distress” [symptom present without distress], 2 = “mild distress”, 3 = “moderate distress”, 4 = “severe distress”). An “option” refers to one of the five possible response options. More specifically, the scale first assesses the presence or absence of a symptom (“0”), and if present, the level of distress associated with it (“1” to “4”)). As a specific screening tool for CHR, the threshold for identifying a sufficient level of distress corresponds to the presence of 6 or more endorsed items.[7]

### *Item response theory*

Item response theory (IRT) analysis, specifically the graded response model (GRM), was employed to evaluate the psychometric properties of each of the 16 ordinal items in the PQ-16. IRT offers clinicians valuable tools for assessing symptom severity with greater precision and differentiating between different levels of severity in patients. It allows to analyze the symptoms directly rather than the construct itself. Within the IRT framework, two key parameters are considered:

- Item “difficulty” reflects the severity level of the CHR-related symptoms set (latent trait) at which respondents have a 50% chance of endorsing a response consistent with the presence of the symptom. It indicates how likely a respondent is to answer in a manner that corresponds to the underlying trait measured by the scale (without a unit of measurement). High difficulty response options (from 0 to 4) are those that are more challenging to select due to factors like complexity, ambiguity, or cognitive demands.
- Item “discrimination” refers to an item’s ability to differentiate between respondents with varying levels of the measured latent trait. It reflects how well an item can distinguish between individuals who are just above or below a specific point on the severity continuum. High discrimination indicates that

the item effectively separates individuals with slightly different levels of the trait.

Within our IRT model, the latent trait, representing the set of CHR-related symptoms, is a continuous variable determined by the relative difficulty and discrimination of each item. Option characteristic curves (analogous to item characteristic curves for ordinal data) are provided with their corresponding coefficients. We also report factor loadings and communalities, which provide insights into the relationship between items and the latent trait (CHR-related symptoms set). Finally, we evaluate model fit indices (infit and outfit) to ensure the IRT model appropriately reflects the data, items, and participant responses. Unidimensionality, a key assumption of the model, was verified through confirmatory factor analysis using established criteria: comparative fit index (CFI) and Tucker-Lewis index (TLI) ≥ 0.95, and root mean square error of approximation (RMSEA) ≤ 0.06.[6,7] All analyses and graphical visualizations were performed with R software (4.3.1).

## RESULTS

### *Sample*

A total of 936 participants were included in the analysis. Socio-demographics and response patterns are reported in Table 1 and Table S1, respectively. 418 subjects (44.42%) exceeded the PQ-16 cut-off (endorsement of at least 6 items).

**Table 1**. Demographic and profiles of study participants (with the Prodromal Questionnaire (PQ)-16 score) (N=936). UK: United Kingdom. SD: Standard Deviation.

| Category | Description | Value |
|---|---|---|
| Age | Mean age | 21.5 years |
| | Standard deviation | 5.1 years |
| Sex | Male (N, %) | 263, 28.1% |
| | Female (N, %) | 673, 71.9% |
| Population Breakdown | France (N) | 367 |
| | UK (N) | 569 |
| Occupation Distribution | Students (N, %) | 764, 81.6% |
| | Employed (N, %) | 119, 12.7% |

| | | |
|---|---|---|
| | Unemployed (N, %) | 53, 5.6% |
| PQ-16 | Mean score (SD) | 5.6 (3.5) |

***Item response theory***

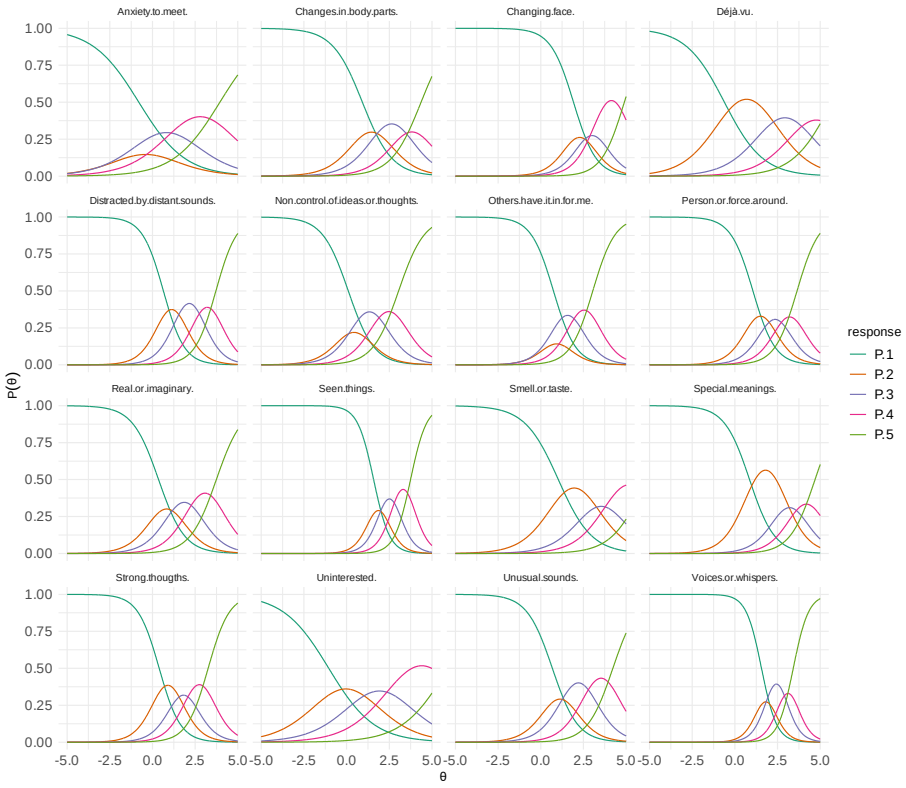
Item discrimination and difficulty parameters estimated using the graded response model (GRM) within the IRT framework are presented in Table S2. “Déjà vu” emerged with the highest discrimination value (3.87), indicating a strong ability to differentiate between participants with varying levels of distress linked to CHR-related symptoms. This suggests a robust relationship between the “Déjà vu” item relatively to all the other CHR-related symptoms (latent trait). Additionally, “Déjà vu” exhibits high uniqueness (0.85) (see Supplementary Materials), signifying that a substantial portion of its variance cannot be explained by the latent trait alone. This implies that while “Déjà vu” effectively discriminates between individuals based on the CHR-related symptom set, it also captures unique experiential aspects not shared by other PQ-16 symptoms. From a clinical standpoint, this suggests that “Déjà vu” might tap into a distinct facet of latent trait, not fully captured by the questionnaire’s main factors, underlining its importance as a unique indicator for clinical assessment. Consistent with this interpretation, steeper slopes in Figure 1 (panel A) depict higher discrimination values. The probability of transitioning between response options on the scale became progressively more difficult (average relative difficulty coefficients, without units: 0.59, 1.70, 2.84, and 4.22). Notably, “Smell or taste” (3.57) and “Changing face” (3.21) were associated with the most challenging response options.

For all items, the response probability curve for the first response options (“None”) indicated that participants without a high level of distress linked to CHR-related symptoms readily endorse this option (Figure 1, panel A.), up to an average level of latent trait severity. The subsequent three options (“Any”, “Mild” and “Moderate”) exhibit good discrimination for most items, but with a shift along the latent trait axis (reflecting difficulty) from left to right. Additionally, endorsement probabilities for these three options are relatively lower compared to “None” and “Severe”. Finally, the “Severe” option shows high endorsement probability for individuals with high level of severity regarding CHR-related symptoms. However, for certain items (“Uninterested”, “Smell or taste”, and “Déjà vu”), endorsement of

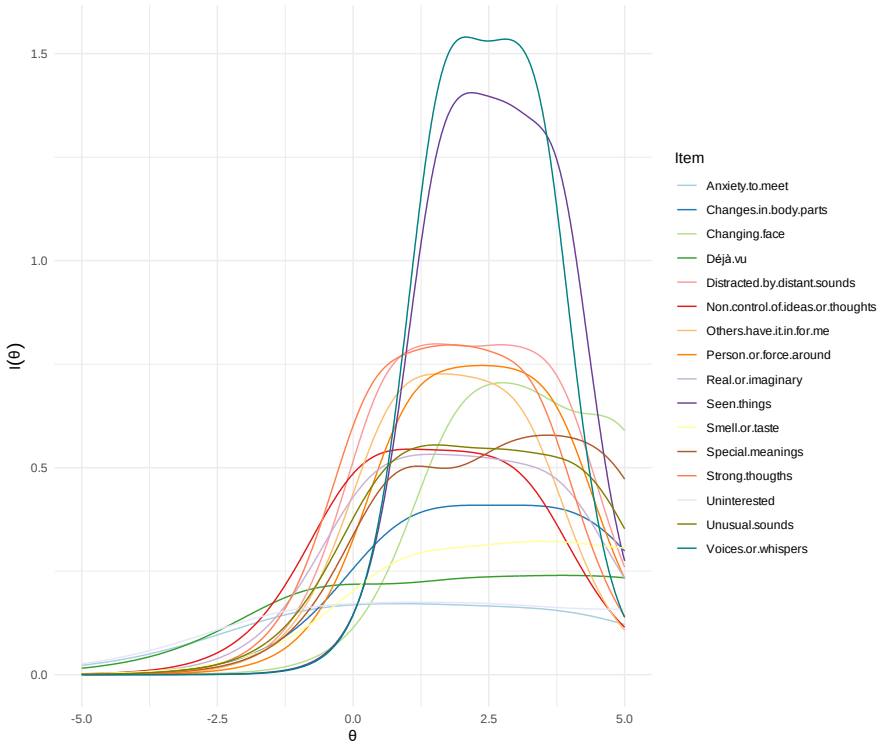
“Severe” becomes progressively more difficult, as evidenced by a sharp rightward shift in the corresponding curve (Figure 1, panel A).

Among the assessed items, “Voices or whispers” and “Seen things” demonstrated the strongest contribution to the precision of measuring the level of distress (Figure 1, panel B & Figure S1). Conversely, “Uninterested” and “Anxiety to meet” exhibited minimal influence on the precision of this measure. Detailed information on factor loadings, communalities (the proportion of variance in an item explained by the latent trait), model fit indices, and applicability conditions can be found in the Supplementary Materials (Tables S3 & S4, Figures S1 & S2, and Box 1).

Notably, all criteria for model interpretability and applicability were met, with Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) exceeding 0.95 (0.96) and Root Mean Square Error of Approximation (RMSEA) below 0.06 (0.046).


Anxiety.to.meet.
Changes.in.body.parts.
Changing.face.
Déjà.vu.
Distracted.by.distant.sounds.
Non.control.of.ideas.or.thoughts.
Others.have.it.in.for.me.
Person.or.force.around.
Real.or.imaginary.
Seen.things.
Smell.or.taste.
Special.meanings.
Strong.thougths.
Uninterested.
Unusual.sounds.
Voices.or.whispers.
P(θ)
θ
response
P.1
P.2
P.3
P.4
P.5

**A.**


I(θ)
θ
Item
Anxiety.to.meet
Changes.in.body.parts
Changing.face
Déjà.vu
Distracted.by.distant.sounds
Non.control.of.ideas.or.thoughts
Others.have.it.in.for.me
Person.or.force.around
Real.or.imaginary
Seen.things
Smell.or.taste
Special.meanings
Strong.thougths
Uninterested
Unusual.sounds
Voices.or.whispers

**B.**

**Figure 1**. **A.** Option characteristic curves to visualize the discrimination and difficulty of each item of the Prodromal Questionnaire-16 (PQ-16) for CHR-related symptoms (N=936). The difficulty parameter is the point on the x-axis where the probability of endorsing a criterion was 0.5, with curve toward the right indicating criteria of greater difficulty relative to the level of severity of the CHR-related symptoms ($\theta$). The discrimination parameter is the slope of the curve at that point, with steeper slopes indicating greater discrimination relative to the level of severity of the CHR-related symptoms ($\theta$). **B.** Contribution of each item to the total precision (or "accuracy") of the Prodromal Questionnaire-16 (PQ-16) for CHR-related symptoms (N=936). As the severity level of the CHR-related symptom set increases (at the right of $\theta$), the probability of endorsing an item increase (except for Likert point 0 "None") – and then decreases as responses move to the next higher Likert point (except for Likert point 3 "Severe"). P.1: None; P.2: Mild; P.3: Moderate; P.4: Severe.

## DISCUSSION

This study employed item response theory (IRT) analysis of the PQ-16 screening tool to identify the most discriminative CHR-related symptoms. We focused on both item difficulty (the level of severity of these symptoms at which endorsement is most probable) and discrimination (the ability to distinguish between individuals with different severity levels). These results individuals demonstrate the model's strong ability to differentiate between individuals across the spectrum of severity. Specifically, as severity related to CHR-related symptoms "increases", the probability of endorsing an item also increases. In other words, this IRT model reveals a gradient in item responses that reflects the underlying latent trait of severity. A key challenge associated with this approach is the selection of the most informative items for clinical use. This information could be valuable for developing and validating shorter, more accurate screening scales.

Among the assessed symptoms, "Déjà vu" emerged with the highest discrimination values, indicating a strong ability to differentiate between individuals with varying level of severity. This distinction highlights the importance of capturing this key symptom for improving identification specificity and, consequently, reducing the risk of overdiagnosis.[10] The high discriminative power of "Déjà vu" suggests potential alterations in source memory/monitoring processes, which aligns with

empirical findings from CHR studies.[11,12] These studies have shown that early auditory processing deficits, such as difficulty discriminating pitches in non-verbal sounds, are linked to broader cognitive dysfunction and poorer functional outcomes in CHR patients.

Another interesting result of our study is that extreme response options ("None" and "Severe") demonstrated strong discrimination in identifying the presence and severity of CHR-related symptoms. Conversely, nuanced response options ("Any", "Mild" or "Moderate" distress) exhibited a finer gradation in discrimination. Interestingly, these three intermediate options also displayed increasing difficulty in terms of endorsement (being "shifted to the right" one the IRT scale). From a clinical perspective, this escalating difficulty in selecting higher response options might reflect the challenge in discerning severity of CHR-related symptom.[13,14] Consequently, participants with an average severity level of severity often endorse the lowest severity option across all items. This could hinder a clinician's ability to detect CHR-related symptoms until the subject reaches a relatively high rating. This finding aligns with prior IRT analyses of the Prodromal Questionnaire.[4,15,16] However, unlike our study, which focused on identifying the most clinically relevant symptoms in the PQ-16, previous research examined the entire PQ,[4] a child-focused version[16] or a specific prenatal cohort.[15]

Our analysis identified "Smell or taste" and "Changing face" as the items with the most difficult response options. This suggests that these perceptual symptoms may be particularly indicative of the most severe cases of patients with CHR-related symptoms. The association of these symptoms with multiple sensory modalities (visual and other) and their appearance only in the most extreme cases aligns with the concept of a higher "disease load" associated with the involvement of multiple sensory dysfunctions. The most informative items for estimating the severity of CHR-related symptoms ("Voices or whispers" and "Seen things") also pertain to perception, specifically auditory and visual alterations. These findings highlight the critical importance for clinicians to assess for such perceptual symptoms. Conversely, symptoms with low discriminatory power, such as "Uninterested" and "Anxiety to meet", appear to be of limited utility in gauging the severity of CHR-related symptoms in clinical practice.

Several limitations should be considered. First, the study lacked clinical investigations to determine if participants scoring above the PQ-16 cut-off had a CHR-P or a psychotic disorder. Moreover, PQ-16 cut-off score itself necessitates cautious interpretation, as it was originally optimized for a help-seeking sample[7] and might not generalize to a broader population. Future studies should incorporate semi-structured CHR diagnosis interviews following the screening of the specific symptoms identified in this study. Second, while multicentric and involving a large number of subjects, our sample may not be entirely representative and generalizable. This is primarily due to the predominantly female sample; however, this gender distribution aligns with existing literature on at-risk symptoms.[1,2] Third, even though the screened sample reflects CHR-related symptoms, potentially representing a clinical population, recruitment bias might be present due to the participants' student status. Consequently, generalizability claims require particular caution. Finally, the innovative application of IRT to identify clinically relevant symptoms necessitates careful interpretation. This is especially true given the model's complexity, requiring a large sample size to estimate the numerous parameters involved. An even larger and more diverse sample would undoubtedly yield more informative results, particularly for a nuanced understanding of the most extreme response options.

By demonstrating that item difficulty and response options vary, these findings reinforce the idea that certain symptoms can better account for the severity of all CHR-related symptoms, advocating for a relevant choice of the most representative symptoms to support the preventive interventions. In a brief executive summary, we present the key insights highlighting the clinical relevance and applicability of these findings:

- This symptom-level study demonstrates to clinicians the ability to differentiate individuals across severity levels; for instance, items like "Smell or taste" and "Changing face" are particularly relevant for identifying severe cases.
- For an accurate evaluation of distress severity, clinicians should prioritize assessing perceptual symptoms such as "Voices or whispers" and "Seen things".

- Clinicians should pay special attention to “Déjà vu” in assessments, as it demonstrated the highest discrimination value among PQ-16 items and captures distinct experiential aspects with high uniqueness.
- Aiding precise clinical assessments, extreme response options (“None” and “Severe”) showed strong discrimination for CHR symptoms, while nuanced options (“Any,” “Mild,” “Moderate”) provided finer gradation in discrimination and increased difficulty.
- From a psychometric perspective, the classification of items based on their degree of discrimination and difficulty provides a better the understanding of the PQ-16 structure and could potentially inform the development of more precise screening tools for clinical use.
- More generally, identifying clinically relevant symptoms to determine individual distress could enhance early identification and intervention strategies.

**Acknowledgements**

The authors thank Rebecca Kraut for her support and valuable advice. CG thanks the Hospices Civils de Lyon for their organizational support.

**Disclosure**

The authors report no conflicts of interest.

**Funding Statement**

This research received a specific grant from Gorilla.sc 2020.

**Ethical standards**

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional guides on the care.

**Conflict of interest**

The authors declare none.

## Author contributions

Author Contributions: CRediT TaxonomyChristophe GauldConceptualization-Equal, Methodology-Equal, Writing – original draft-Equal, Writing – review & editing-EqualPierre FourneretConceptualization-Equal, Validation-Equal, Writing – review & editing-EqualBenjamin Alderson-DayResources-Equal, Supervision-Equal, Validation-Equal, Writing – review & editing-EqualEmma Palmer-CooperMethodology-Equal, Validation-Equal, Visualization-Equal, Writing – review & editing-EqualClément DondéConceptualization-Equal, Methodology-Equal, Resources-Equal, Supervision-Equal, Validation-Equal, Writing – original draft-Equal, Writing – review & editing-Equal

**Edited by**: Dr. Raffael Massuda

## References

1 Fusar-Poli P, McGorry PD, Kane JM. Improving outcomes of first-episode psychosis: an overview. World Psychiatry. 2017;16:251–65.
2 Schultze-Lutter F, Michel C, Schmidt SJ, Schimmelmann BG, Maric NP, Salokangas RKR, et al. EPA guidance on the early detection of clinical high risk states of psychoses. European Psychiatry. 2015;30:405–16.
3 McDonald M, Christoforidou E, Van Rijsbergen N, Gajwani R, Gross J, Gumley AI, et al. Using Online Screening in the General Population to Detect Participants at Clinical High-Risk for Psychosis. Schizophr Bull. 2019;45:600–9.
4 Capizzi R, Pierce KM, Olino TM, Ellman LM. Item-level endorsement on the Prodromal Questionnaire in a large non-clinical sample. Schizophr Res. 2022;248:309–19.
5 Bernardin F, Gauld C, Martin VP, Laprévote V, Dondé C. The 68 symptoms of the clinical high risk for psychosis: Low similarity among fourteen screening questionnaires. Psychiatry Res. 2023;330:115592.
6 Howie C, Hanna D, Shannon C, Davidson G, Mulholland C. The Structure of the Prodromal Questionnaire-16 (PQ-16): Exploratory and confirmatory factor analyses in a general non-help-seeking population sample. Early Interv Psychiatry. 2022;16:239–46.
7 Ising HK, Veling W, Loewy RL, Rietveld MW, Rietdijk J, Dragt S, et al. The validity of the 16-item version of the Prodromal Questionnaire (PQ-16) to screen for ultra high risk of developing psychosis in the general help-seeking population. Schizophr Bull. 2012;38:1288–96.
8 Gauld C, Fourneret P, Alderson-Day B, Palmer-Cooper E, Dondé C. Impacts of Risk Thresholds and Age on Clinical High Risk for Psychosis: A Comparative Network Analysis. Asian Journal of Psychiatry. In Press.
9 Spillebout C, Pelluet A, Bioulac S, Fourneret P, Polosan M, Dondé C. Detection of clinical high risk for psychosis in child and adolescent mental health services: Validation of the first step with the French versions of the Prodromal Questionnaire (fPQ-16) and scale of Perceptual and Cognitive Aberrations (fPCA).

Early Intervention in Psychiatry. 2023;17:708–14.
10 Halpin M. Countervailing medicalization: A relational approach to the medicalization of psychosis. SSM - Qualitative Research in Health. 2022;2:100118.
11 Atkinson RJ, Fulham WR, Michie PT, Ward PB, Todd J, Stain H, et al. Electrophysiological, cognitive and clinical profiles of at-risk mental state: The longitudinal Minds in Transition (MinT) study. PLoS One. 2017;12:e0171657.
12 Hamilton HK, Roach BJ, Bachman PM, Belger A, Carrion RE, Duncan E, et al. Association Between P300 Responses to Auditory Oddball Stimuli and Clinical Outcomes in the Psychosis Risk Syndrome. JAMA Psychiatry. 2019;76:1187–97.
13 Truijens FL, Cornelis S, Desmet M, De Smet MM, Meganck R. Validity Beyond Measurement: Why Psychometric Validity Is Insufficient for Valid Psychotherapy Research. Frontiers in Psychology. 2019;10.
14 Truijens FL, Van Nieuwenhove K, De Smet MM, Desmet M, Meganck R. How questionnaires shape experienced symptoms. A qualitative case comparison study of questionnaire administration in psychotherapy research. Qualitative Research in Psychology. 2022;19:806–30.
15 Levey EJ, Zhong Q-Y, Rondon MB, Sanchez S, Li J, Williams MA, et al. The psychometric properties of the 16-item version of the Prodromal Questionnaire (PQ-16) as a screening instrument for perinatal psychosis in pregnant Peruvian women. Arch Womens Ment Health. 2018;21:563–72.
16 Karcher NR, Perino MT, Barch DM. An item response theory analysis of the Prodromal Questionnaire-Brief Child Version: Developing a screening form that informs understanding of self-reported psychotic-like experiences in childhood. Journal of Abnormal Psychology. 2020;129:293–304.

## Supplementary Materials

**Figure S1**. Contribution of the 16 items of the Prodromal Questionnaire-16 (PQ-16) for clinical high risk (N=936) (equivalent to Figure 1, panel B., but with all the items). The items perform much better in terms of information precision for "high" threshold CHR, and very poor for "low" threshold CHR.

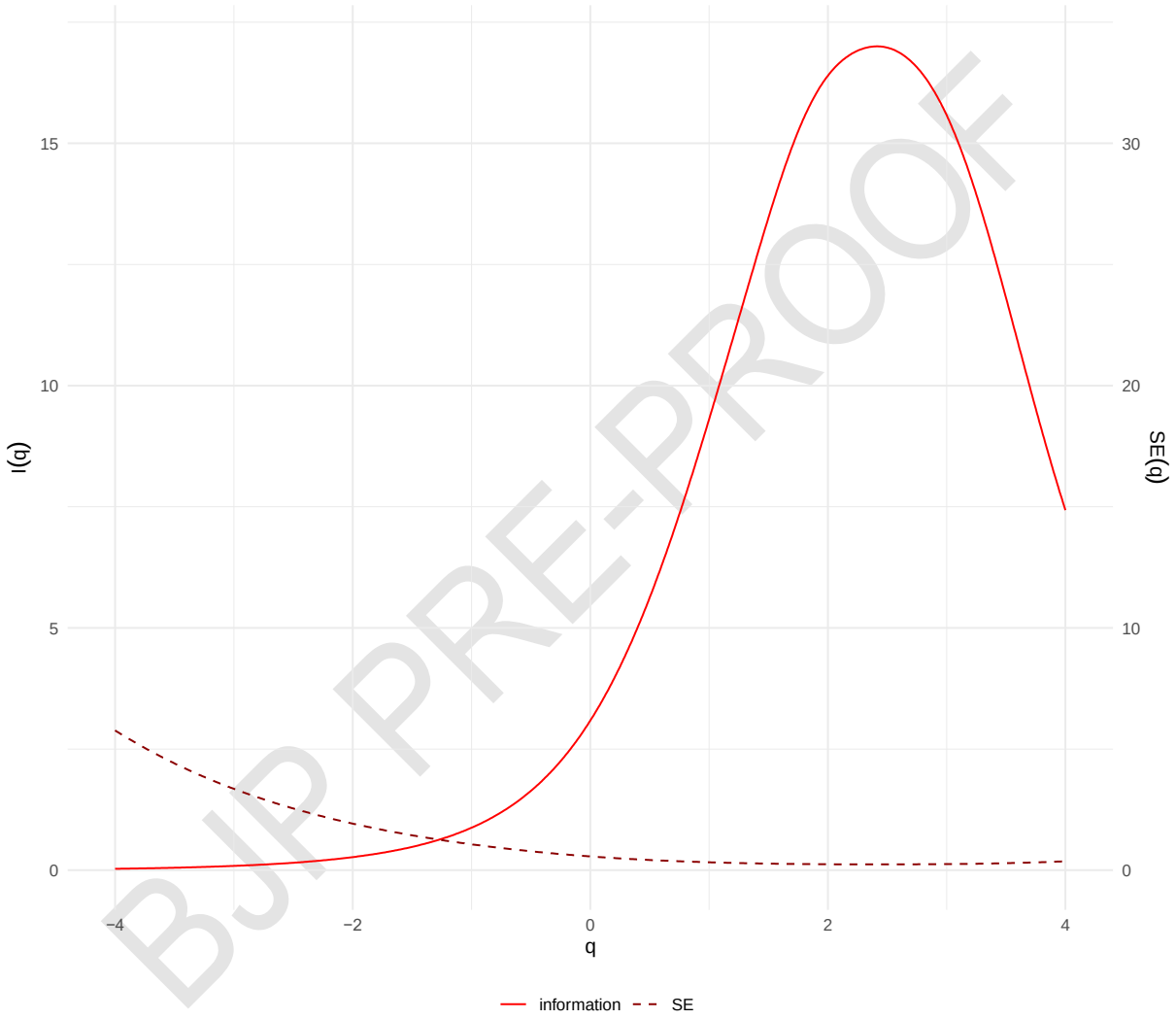


**Figure S2**. Infit and outfit statistics allow to visualize the fit of the items with the model in the Prodromal Questionnaire-16 (PQ-16) (N=936). Non-standardized values should be between .5 and 1.5 to not be "degrading", i.e., "productive" for measurement.

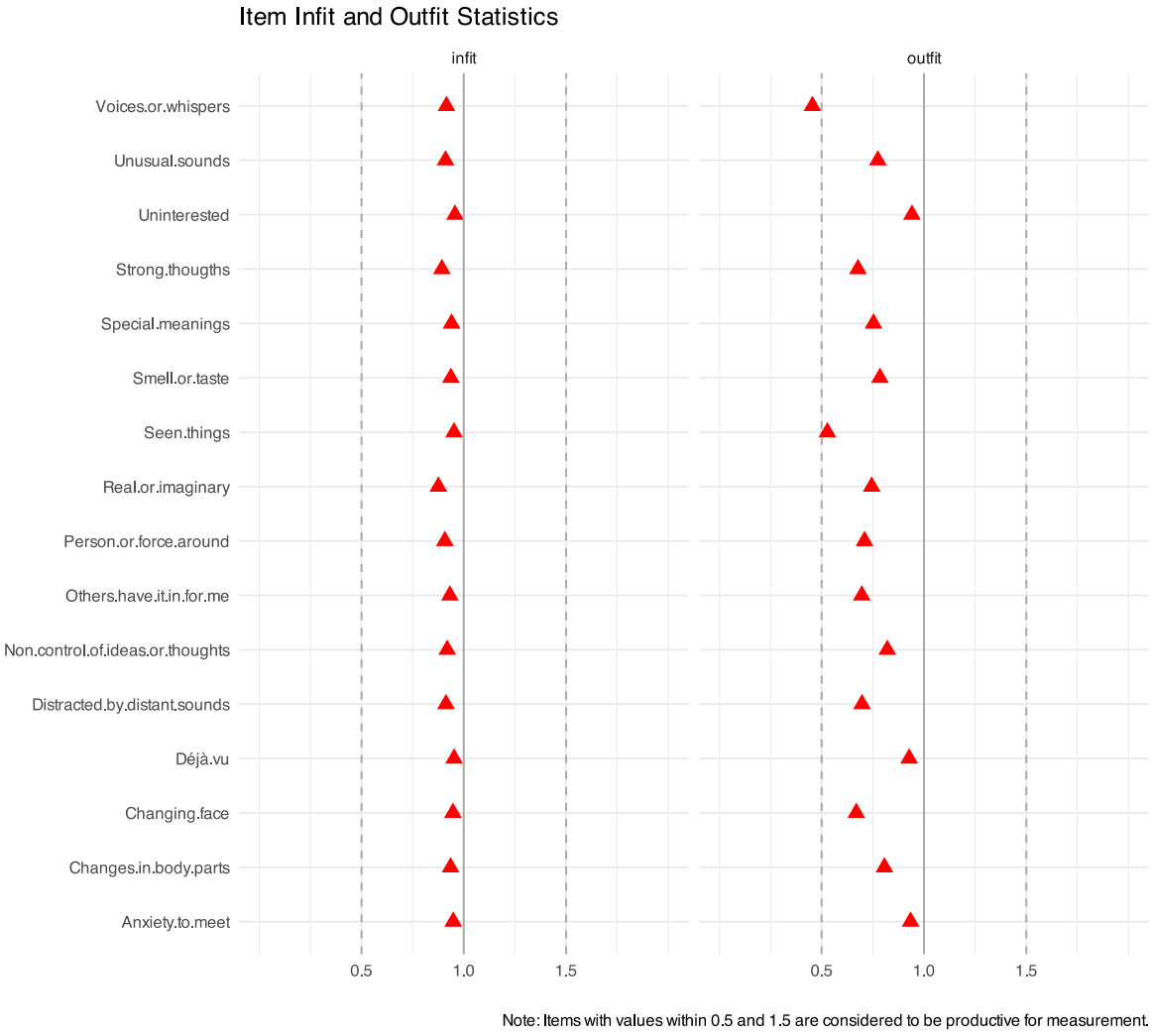


**Table S1**. Frequency and percentage of response of each Likert response modality of the Prodromal Questionnaire-16 (PQ-16) items (N=936).

| Items | None | Any | Mild | Moderate | Severe |
|---|---|---|---|---|---|
| **Uninterested** | 311 (33%) | 305 (32%) | 210 (22%) | 101 (11%) | 14 (1%) |
| **Déjà vu** | 364 (39%) | 411 (44%) | 125 (13%) | 32 (3%) | 9 (1%) |
| **Smell or taste** | 669 (71%) | 205 (22%) | 46 (5%) | 18 (2%) | 3 (0%) |
| **Unusual sounds** | 630 (67%) | 159 (17%) | 110 (12%) | 34 (4%) | 8 (1%) |
| **Real or imaginary** | 564 (60%) | 186 (20%) | 120 (13%) | 55 (6%) | 16 (2%) |
| **Changing face** | 844 (90%) | 54 (6%) | 26 (3%) | 15 (2%) | 2 (0%) |
| **Anxiety to meet** | 353 (38%) | 124 (13%) | 233 (25%) | 172 (18%) | 59 (6%) |
| **Seen things** | 837 (89%) | 56 (6%) | 33 (4%) | 12 (1%) | 3 (0%) |
| **Strong thoughts** | 578 (61%) | 210 (22%) | 89 (9%) | 47 (5%) | 17 (2%) |
| **Special meanings** | 668 (71%) | 227 (24%) | 31 (3%) | 11 (1%) | 4 (0%) |
| **Non-control of ideas or thoughts** | 495 (53%) | 151 (16%) | 180 (19%) | 81 (9%) | 34 (4%) |
| **Distracted by distant sounds** | 636 (68%) | 180 (19%) | 91 (10%) | 26 (3%) | 8 (1%) |
| **Voices or whispers** | 839 (89%) | 51 (5%) | 35 (4%) | 11 (1%) | 5 (1%) |
| **Others have it in for me** | 655 (70%) | 76 (8%) | 124 (13%) | 61 (6%) | 25 (3%) |
| **Person or force around** | 720 (77%) | 129 (14%) | 57 (6%) | 24 (3%) | 11 (1%) |
| **Changes in body parts** | 662 (70%) | 156 (17%) | 86 (9%) | 25 (3%) | 12 (1%) |

**Table S2.** Coefficients for discrimination and of the three difficulty thresholds between the five modalities (Likert point) of each item of the Prodromal Questionnaire-16 (PQ-16) for clinical high risk (N=936). The items are ordered according to the value of discrimination. For example, in the last column, the coefficient indicates the difficulty of moving from the fourth to the fifth modality, but also the ease of moving from the fifth to the fourth. Values in bold are the highest in each column.

| Items | Discrimination (slope) | Difficulty threshold 1 (from "None" to "Any") | Difficulty threshold 2 (from "Any" to "Mild") | Difficulty threshold 3 (from "Mild" to "Moderate") | Difficulty threshold 4 (from "Moderate" to "Severe") |
|---|---|---|---|---|---|
| **Uninterested** | 0.78 | 0.75 | -1.05 | 0.96 | 2.88 |
| **Déjà vu** | 0.85 | 0.89 | -0.61 | 1.99 | 3.87 |
| **Anxiety to meet** | 0.95 | 1.02 | 1.03 | **2.90** | **4.19** |
| **Changes in body parts** | 1.24 | 1.34 | 0.68 | 1.58 | 2.84 |
| **Smell or taste** | 1.04 | 1.31 | 0.36 | 1.31 | 2.41 |
| **Unusual sounds** | 1.46 | 1.50 | **1.92** | 2.64 | 3.39 |
| **Non-control of ideas or thoughts** | 0.59 | 0.74 | -0.80 | 0.00 | 1.64 |
| **Others have it in for me** | **2.46** | 2.13 | 1.58 | 2.14 | 2.87 |
| **Person or force around** | 1.43 | 1.60 | 0.39 | 1.41 | 2.23 |
| **Changing face** | 1.34 | 1.36 | 0.86 | 2.73 | 3.68 |
| **Real or imaginary** | 1.28 | 1.32 | 0.10 | 0.77 | 1.91 |
| **Special meanings** | 1.62 | 1.62 | 0.64 | 1.61 | 2.70 |
| **Distracted by distant sounds** | 2.34 | **2.23** | 1.56 | 2.06 | 2.81 |
| **Strong thoughts** | 1.45 | 1.53 | 0.74 | 1.12 | 2.03 |
| **Seen** | 1.61 | 1.54 | 1.06 | 1.94 | 2.77 |

| | | | | | |
|---|---|---|---|---|---|
| **things** | | | | | |
| **Voices or whispers** | 1.13 | 1.15 | 0.92 | 2.00 | 3.29 |

**Table S3**. Clinical high risk (CHR) factor loadings and communalities on the Prodromal Questionnaire-16 (PQ-16) items (N=936). h2 represents the variance explained in an item by the latent trait.

| Items | CHR Factor loadings (F1) | Community (h2) |
|---|---|---|
| **Uninterested** | 0.404 | 0.163 |
| **Déjà vu** | 0.462 | 0.213 |
| **Smell or taste** | 0.514 | 0.265 |
| **Unusual sounds** | 0.619 | 0.383 |
| **Real or imaginary** | 0.609 | 0.371 |
| **Changing face** | 0.660 | 0.436 |
| **Anxiety to meet** | 0.398 | 0.159 |
| **Seen things** | 0.781 | 0.610 |
| **Strong thoughts** | 0.685 | 0.469 |
| **Special meanings** | 0.625 | 0.390 |
| **Non-control of ideas or thoughts** | 0.613 | 0.376 |
| **Distracted by distant sounds** | 0.689 | 0.475 |
| **Voices or whispers** | 0.795 | 0.632 |
| **Others have it in for me** | 0.668 | 0.446 |
| **Person or force around** | 0.671 | 0.450 |
| **Changes in body parts** | 0.558 | 0.312 |

**Table S4.** Fit of the items with the model in the Prodromal Questionnaire-16 (PQ-16) (N=936). To study the fit of the items, we study the S_X2 of Orlando and Thissen (2000) and the corresponding Root Mean Square Error of Approximation (RMSEA) and p-values. This test have to be insignificant to indicate a good fit of the article. The significant p-values ($p<0.05$, noted "*") have lower fit to the model. Df: degrees of freedom. RMSEA: Root Mean Square Error of Approximation

| Items | S_X2 | Df (for S_X2) | RMSEA (for S_X2) | P-value (for S_X2) |
|---|---|---|---|---|
| **Uninterested** | 94.885 | 71 | 0.019 | **0.031*** |
| **Déjà vu** | 78.823 | 68 | 0.013 | 0.174 |
| **Smell or taste** | 72.934 | 50 | 0.022 | **0.019*** |
| **Unusual sounds** | 76.230 | 69 | 0.011 | 0.257 |
| **Real or imaginary** | 124.851 | 73 | 0.027 | **0.000*** |
| **Changing face** | 55.005 | 46 | 0.014 | 0.170 |
| **Anxiety to meet** | 118.279 | 92 | 0.017 | **0.034*** |
| **Seen things** | 47.404 | 43 | 0.010 | 0.298 |
| **Strong thoughts** | 99.088 | 69 | 0.022 | **0.010*** |
| **Special meanings** | 44.498 | 45 | 0.000 | 0.493 |
| **Non-control of ideas or thoughts** | 84.031 | 84 | 0.001 | 0.479 |
| **Distracted by distant sounds** | 89.327 | 62 | 0.022 | **0.013*** |
| **Voices or whispers** | 50.223 | 43 | 0.013 | 0.209 |
| **Others have it in for me** | 103.945 | 75 | 0.020 | **0.015*** |
| **Person or force around** | 86.685 | 66 | 0.018 | **0.045*** |
| **Changes in body parts** | 46.916 | 35 | 0.019 | 0.086 |

**Box 1 – Supplementary Materials: Details of the methods, model and results**

**Methods – Ethical statements**. The Tone-P Study is a cross-sectional online study to investigate early auditory processing in non-help seekers screened for CHR. The Tone-P study is funded by Gorilla.sc. Informed consent for the study was provided online, followed by a socio-demographic assessment and by two screening questionnaires: the 16-item PQ-16 and the 9-item PCA scale.

**Methods – PQ-16.** The items are summarized as follows in our study: (1) "Uninterested", (2) "Déjà vu", (3) "Smell or taste", (4) "Unusual sounds", (5) "Real or imaginary", (6) "Changing face", (7) "Anxiety to meet", (8) "Seen things", (9) "Strong thoughts", (10) "Special meanings", (11) "Non-control of ideas or thoughts", (12) "Distracted by distant sounds", (13) "Voices or whispers", (14) "Others have it in for me", (15) "Person or force around", (16) "Changes in body parts". The total score was the result of the sum of the scores obtained for each of the 16 items. It was developed on the basis of the Prodromal Questionnaire (PQ), a 92-item self-report measure (Loewy et al., 2005), itself based on the Schizotypal Personality Questionnaire (Raine, 1991). Tested in a general non-help-seeking population, the initial validation study revealed a three-factor structure (perceptual abnormalities/hallucinations, unusual thoughts and negative symptoms). It was translated and validated in French in both adult and adolescent populations (Lejuste et al., 2021; Spillebout et al., 2023).

**Methods – Participants**. Nine hundred forty-eight (948) participants were included in the study and, after processing missing data, 936 participants were analyzed in the study. The mean age was 21.5 years, with a median of 20.0 and a standard deviation of 5.1. Three hundred sixty-seven (367) were included in France (39.2%) and 569 in the UK (60.8%) [Chi2 = 43.59, $p < 0.001$]. Two-hundred sixty-three (263) were male (28.1%) and 673 female (71.9%) [Chi2 = 179.59, $p < 0.001$]. Seven-hundred sixty-four (764) were students (81.6%), 119 were employed (12.7%) and 53 unemployed (5.6%) [Chi2 = 989.21, $p < 0.001$]. Between the French and UK groups, there was a significant difference for age ($t = 15.82$, $p < 0.001$ [4.74 – 6.09]), but not for sex (Chi2 ~ 0, p ~ 1) or occupation (Chi2 = 6, $p = 0.20$).

**Methods – IRT**. We conducted an Item Response Theory (IRT) analysis using the Graded Response Model (GRM). The Graded Response Model is suitable for analyzing ordinal data. We are particularly interested in the factor loadings (F1) and the communalities (h2). In the

context of IRT, `F1` should be greater than 0.5, indicating adequate factor loading. “F1” represents the saturation of the factor (how well the item represents the underlying factor), while “h2” (which is “F1” squared) represents the communality, indicating the variance explained in an item by the factor.

**Results**. The analysis was conducted using full-information item factor analysis with a single factor. It successfully converged after 52 iterations of the Expectation-Maximization (EM) algorithm, with a tolerance level of 1e-04. This analysis was performed using version 1.41 of the mirt software. For optimization in the M-step, the BFGS method was utilized, and Ramsay's method was employed for EM acceleration. The number of rectangular quadrature points used in the analysis was 61, and the latent density was assumed to be Gaussian.
The information matrix was estimated using Oakes' method. The second-order test suggests that the model could be a possible local maximum, and the condition number of the information matrix was found to be 574, indicating potential numerical issues in the inversion or calculation of the matrix.
The log-likelihood of the model was -7339.27. In total, 64 parameters were estimated. Based on these estimates, the Akaike Information Criterion (AIC) was calculated to be 14806.54, while the Bayesian Information Criterion (BIC) and Sample-Size Adjusted BIC (SABIC) were 15116 and 14913, respectively.

**Validity condition**. To observe how much the model fits the data, rather than using a $\chi 2$, we use a specific index, M2, which is specifically designed to assess the fit of item response models. M2 is at 414.9, suggesting a moderate fit between the model and the observed data. The Confirmatory Factor Analysis (CFA) (lavaan 0.6.15 – optimization method with NLMINB) ended after 44 iterations. The Tucker-Lewis Index (TLI) is 0.96 and the Comparative Fit Index (CFI) is also at 0.96, both of which are above the common threshold of 0.95, further confirming the good fit of the model to the data. The log-likelihood values for the user model and the unrestricted model gave an Akaike Information Criterion (AIC) of 20952.556 and a Bayesian Information Criterion (BIC) of 21107.488. The Root Mean Square Error of Approximation (RMSEA) is 0.0464, with a 90% confidence interval ranging from [0.0406 – 0.0523], indicating a good model fit, and the probability of RMSEA being less than or equal to 0.05 was zero. The model had a relatively low Standardized Root Mean Square

Residual (SRMR), at 0.060, indicating a good fit in terms of the standardized difference between observed and predicted correlations.

The CFA also shows that two factors fit the data best, consistent with a number of previous studies on the PQ-16 (Howie et al., 2020). The chi-square statistic of 292.21 with a p-value < 0.001 (1.82e-23) indicates that the two-factor model is a good fit for the data. The factors account for approximately 30.5% of the total variance, with the first factor explaining 16.8% and the second factor explaining 13.7%. Such an analysis reveals the uniqueness of each item in the dataset, which indicates how much variance in the item is not explained by the factors. The items with the highest uniqueness values are “Déjà vu” (0.849), “Special meanings” (0.854), and “Uninterested” (0.761), implying that these items are less well explained by the two factors compared to others. On the other hand, the items with the lowest uniqueness, and hence most influenced by the factors, are “Voices or whispers” (0.470), “Non-control of ideas or thoughts” (0.510), and “Seen things” (0.602). These items are primarily loaded on the first factor, except “Non-control of ideas or thoughts”, which is more associated with the second factor.

**Correspondence Table**

| Items of the PQ-16 | Summary of items |
|---|---|
| I feel uninterested in the things I used to enjoy. | Uninterested |
| I often seem to live through events exactly as they happened before (déjà vu). | Déjà vu |
| I sometimes smell or taste things that other people can't smell or taste. | Smell or taste |
| I often hear unusual sounds like banging, clicking, hissing, clapping or ringing in my ears. | Unusual sounds |
| I have been confused at times whether something I experienced was real or imaginary. | Real or imaginary |
| When I look at a person, or look at myself in a mirror, I have seen the face change right before my eyes. | Changing face |
| I get extremely anxious when meeting people for the first time. | Anxiety to meet |
| I have seen things that other people apparently can't see. | Seen things |
| My thoughts are sometimes so strong that I can almost hear them. | Strong thoughts |
| I sometimes see special meanings in advertisements, shop windows, or in the way things are arranged around me. | Special meanings |
| Sometimes I have felt that I'm not in control of my own ideas or thoughts. | Non-control of ideas or thoughts |
| Sometimes I feel suddenly distracted by distant sounds that I am not normally aware of. | Distracted by distant sounds |
| I have heard things other people can't hear like voices of people whispering or talking. | Voices or whispers |
| I often feel that others have it in for me. | Others have it in for me |
| I have had the sense that some person or force is around me, even though I could not see anyone. | Person or force around |
| I feel that parts of my body have changed in some way, or that parts of my body are working differently than before. | Changes in body parts |