

Psychometric Properties of “Community Assessment of Psychic Experiences”: Review and Meta-analyses

Winifred Mark¹ and Timothea Touloupoulou^{*,1-3}

¹Department of Psychology and ²The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, SAR, China; ³Department of Basic and Clinical Neuroscience, the Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

*To whom correspondence should be addressed; Department of Psychology, 6/F, Centennial Campus, The University of Hong Kong, Pokfulam Road, Hong Kong; tel: (852) 39178927, fax: (852) 28583518, e-mail: timothea@hku.hk

The Community Assessment of Psychic Experiences (CAPE) has been used extensively as a measurement for psychosis proneness in clinical and research settings. However, no prior review and meta-analysis have comprehensively examined psychometric properties (reliability and validity) of CAPE scores across different studies. To study CAPE's internal reliability—ie, how well scale items correlate with one another—111 studies were reviewed. Of these, 18 reported unique internal reliability coefficients using data at hand, which were aggregated in a meta-analysis. Furthermore, to confirm the number and nature of factors tapped by CAPE, 17 factor analytic studies were reviewed and subjected to meta-analysis in cases of discrepancy. Results suggested that CAPE scores were psychometrically reliable—ie, scores obtained could be attributed to true score variance. Our review of factor analytic studies supported a 3-factor model for CAPE consisting of “Positive”, “Negative”, and “Depressive” subscales; and a tripartite structure for the Negative dimension consisting of “Social withdrawal”, “Affective flattening”, and “Avolition” subdimensions. Meta-analysis of factor analytic studies of the Positive dimension revealed a tridimensional structure consisting of “Bizarre experiences”, “Delusional ideations”, and “Perceptual anomalies”. Information on reliability and validity of CAPE scores is important for ensuring accurate measurement of the psychosis proneness phenotype, which in turn facilitates early detection and intervention for psychotic disorders. Apart from enhancing the understanding of psychometric properties of CAPE scores, our review revealed questionable reporting practices possibly reflecting insufficient understanding regarding the significance of psychometric properties. We recommend increased focus on psychometrics in psychology programmes and clinical journals.

Key words: psychosis proneness/reliability/validity/
reliability generalization

Psychosis proneness refers to unusual experiences and psychotic-like symptoms that do not meet clinical threshold for psychotic disorders. Despite being relatively common in adolescence and young adulthood,¹ psychosis proneness is not necessarily innocuous.² Isolated psychotic symptoms, even those insufficient in severity or impairment to warrant a clinical diagnosis of psychotic disorders, are associated with increased risk for suicidal behavior,³ nonpsychotic psychiatric disorders,⁴⁻⁶ and functional disability.⁷

Shared etiological risk factors,^{1,8-10} cognitive correlates¹¹, and demographic characteristics^{1,11} between sub-clinical psychotic experiences and psychotic disorders suggest that psychotic experiences exist along a continuum, whereby psychotic disorders differ only *quantitatively* from psychosis proneness.^{1,11} As such, individuals with psychosis proneness in the general populace could be a valid population for studying the etiology of psychosis.^{12,13} This has sparked the need for valid measurements to track the presence and trajectory of the psychosis proneness phenotype in community samples.

Psychosis proneness is typically measured with self-report questionnaires or structured interviews. Interview-based measures are typically comprehensive, but lengthy and require training in administration and scoring. On the other hand, self-report instruments are brief and more user-friendly for research and low-intensity clinical settings. A popular psychosis proneness self-report tool that has been used extensively in schizophrenia research is the Community Assessment of Psychic Experiences (CAPE),¹⁴ which measures lifetime psychotic experiences. CAPE taps frequency and distress of psychotic-like feelings, thoughts, or mental experiences (eg, “Do you ever feel as if people seem to drop hints about you or say things with a double meaning?”) relating to dimensions of psychosis symptomatology, namely, positive symptoms

(excessive behaviors not seen in normal individuals), negative symptoms (disruptions to normal behavior), and depressive symptoms.

More recently, CAPE has been utilized in supplementing clinical diagnosis for improving detection of first episode psychosis in mental health care services.¹⁵ However, a comprehensive examination of psychometric properties of CAPE scores has yet to be conducted. Psychometric properties are concerned with whether a scale measures what it was designed to measure. Two such important properties are “internal reliability”, which is concerned with how consistently scale items measure the same construct; and “factorial validity”, which is concerned with whether a scale contains dimensions measuring what it purports to assess. Knowledge of psychometric properties of scores of CAPE helps researchers and clinicians make informed decisions about whether the instrument could fulfill the examiner’s assessment needs; such as whether scores obtained could be trusted, and which psychotic-like-symptom dimensions could be measured with the instrument.

Despite the importance of psychometric properties, problems in reporting internal reliability and factorial validity in published research are common.¹⁶⁻²² Here, we hope to provide new insights concerning the utility of CAPE through a review and meta-analysis of reliability coefficients and factor structures. We also hope to alert readers to the importance of psychometric properties in clinical research and practice. The following research questions are addressed:

1. What is the typical reliability of CAPE scores and subscale scores?
2. What is the factor structure of CAPE and its subscales?

Psychometric Properties

Internal Reliability. To date, there are no comprehensive reports on the internal reliability of CAPE scores. This needs to be rectified because score reliability could implicate on the extent to which we can accurately interpret the data collected.^{21,22} One important type of reliability is internal reliability, which addresses the question: are items in the scale consistently measuring the same construct? The principal is that, if items on a test assessed the same construct, their scores should be highly correlated.

Internal reliability could be estimated in the form of “reliability coefficients”, which is a statistic calculated from correlations between item scores of a scale. The higher the reliability coefficient of scores, the lower the possibility that results obtained might reflect variance attributable to random error; so the better we can trust our results. Scales that produce scores with low reliability coefficients are of limited utility, because this implies that participants’ responses to related items appeared contradictory (eg, endorsing the item “I am often suspicious of

those around me” but not the item “I find it difficult to trust others”) in the same administration. In such cases, researchers or clinicians could not interpret the test scores accurately.²³

The most common reliability coefficient reported is Cronbach’s alpha,²⁴ which gives an overall picture of how every possible pairs of item scores are related (eg, correlation between the first- and second-half of item scores or between odd- and even-numbered item scores). Hence, the Cronbach’s alpha value is an estimation of how reliable scores are, with a minimally acceptable value of 0.7²⁵ for research. Previously, it was thought that higher Cronbach’s alpha values were always more desirable. However, some researchers have recently cautioned that alpha values that are “too high” (over 0.9) might indicate item redundancy rather than desirable internal consistency.²⁶ Due to the useful information that could be reflected by internal reliability coefficients, it is important to investigate the typical reliability coefficient of CAPE scores.

Despite its significance in research, reliability is often misunderstood.²⁶⁻³¹ Contrary to popular belief, only *scores*, but not *tests*, can be reliable.^{18,26,32,33} This is because psychometric reliability is a function of sample characteristics (eg, age group) and test forms (eg, language) employed in a study. An implication of this is that reliability estimates should be generated for each administration, rather than cited from previous studies. However, researchers have repeatedly found that reliability reporting for data at hand was the exception rather than the norm in journals.¹⁶⁻²⁰

Given that accurate interpretation of test scores is contingent upon reliability data, we attempted to fill the gap in the CAPE literature by conducting the first study that examines reliability estimates of a large sample of CAPE scores. A review was conducted to identify published studies that have utilized CAPE. A meta-analysis of reliability coefficients, known as a “reliability generalization” (RG) study, was then carried out using Cronbach’s alpha generated from original data to arrive at a meta-analytic “mean reliability estimate” that informs the *typical* score reliability for CAPE. RG could also inform the source of variability in reliability coefficients across studies, such as age group of samples.³⁴

Factorial Validity

In psychological research, latent, unobservable constructs (known as “factors”) such as psychosis proneness are often studied using questionnaires. It is hence important to ask: does the questionnaire measure what it purports to measure? This is known as factorial validity. To possess factorial validity, the scale must demonstrate that the composition of subscales (known as the “factor structure”) fit the theoretical understanding of the construct it supposedly measure.³⁵ Copious validation studies have

attempted to demonstrate factorial validity of CAPE scores in independent samples.^{14,36–51} However, no study has yet to examine whether different studies produced similar factor structures across previous studies.

Issues of factorial validity are commonly addressed statistically by factor analysis.^{52–54} By conducting a factor analysis, one can determine the underlying constructs of the test, thus informing whether the test indeed measures the constructs it purported to measure. Factor analysis can be either “exploratory” or “confirmatory”. Exploratory Factor Analysis (EFA) determines the smallest number of factors that could account for the observed correlations in the test scores.⁵⁵ Such factors are then matched with researchers’ theoretical understanding of the construct in question. For example, if a scale is hypothesized to have 3 subscales, and EFA of the collected data indeed generated 3 factors containing relevant items, the scale’s factorial validity is partially supported. EFA is often used as a preliminary exploration of the data structure, which is later “confirmed” using confirmatory factor analysis (CFA).

However, EFA is heavily biased by researchers’ subjectivity, because the method of deciding the smallest number of factors, known as “factor retention” rules, could result in different numbers of factors. Hence, it is important for researchers to report the exact way in which procedures were carried out in order to facilitate replication and independent evaluation of evidences of factorial validity. However, previous studies have found that a majority of EFA studies failed to adequately report details of how the factor analysis was performed.^{16,18}

CFA conforms better to rigorous hypothesis testing by allowing researchers to statistically compare how their data fit with an a priori hypothesis of the structure of a test. In CFA, researchers first justify theoretically or empirically (eg, preliminary EFA results) how many factors there should be in the instrument and which items should belong together in specific subscales. If the data collected are found to map onto hypothesized factors, as indicated by acceptable model fit statistics (known as “goodness-of-fit indices”) in CFA, the scale is said to possess factorial validity.^{56,57}

In CFA, multiple fit indices should be reported regardless of whether or not they are favorable to the hypothesized model. Such practices reduce self-fulfilling prophecies and allow researchers to make a more balanced decision as to how many factors provide the best fit for the data at hand.^{58–61} A variety of goodness-of-fit indices can easily be calculated in statistical packages such as Mplus⁶² and LISREL,⁶³ though the interpretation of model fit and suggested cut-offs are debated, and could vary in different sources.⁶⁴ The most popular indices are Comparative Fit Index (CFI; acceptable if ≥ 0.95), Tucker-Lewis Index (TLI; acceptable if ≥ 0.95), Root Mean Square Error of Approximation (RMSEA; acceptable if ≤ 0.06), and Standardized root mean square residual (SRMR; acceptable if < 0.08). Researchers are suggested to report a bare minimum of 2 indices consisting of SRMR paired

with TLI, RMSEA or CFI,⁶⁵ in order to reflect different aspects of model fit.⁶⁴

In the original validation study of CAPE, Stefanis and colleagues¹⁴ recruited a sample of 932 Greek conscripts and conducted a CFA. Results demonstrated that a 3-factor model provided better fit to the data than did 1- and 2-factor solutions, as indicated by multiple goodness-of-fit indices. This was in line with the researchers’ theory-driven hypothesis of a 3-dimensional model of positive, negative, and depressive symptom domains based on clinical symptom clusters of patients with psychotic disorders. Subsequently, different research groups have tried to replicate and further explore the factor structures of CAPE.^{36–39} More recently, factor structures of the Positive (CAPE-*pos*) and Negative (CAPE-*neg*) dimensions have also been proposed.^{40–46,50,66}

Despite individual reports on factor structures of CAPE, no previous studies have examined comprehensively the factorial validity of CAPE. In order to review the degree to which published factor analytic studies converged on the same number and nature of factors, a review of factor analytic studies was conducted. In the case of variability in the number of factors proposed, a meta-analysis of factor structures would be used to statistically resolve the discrepancy and to arrive at a basic underlying structure. The results would inform researchers and clinicians which symptom dimensions could be assessed by CAPE and its subscales.

Method

Search of Studies and Selection Criteria

Electronic searches on the PsycINFO and PubMed databases using the search term “Community Assessment of Psychic Experiences” yielded 169 results from peer-reviewed journals, from which studies were selected for review if they meet the following criteria: (1) available in English; (2) reported original research; and (3) used CAPE as an independent or dependent variable. Reference lists of included articles were also reviewed to identify studies missed by the literature search. All in all, 111 studies covering samples from 15 countries published between 2002 and 2014 were included.

Studies were classified on the basis of 4 categories: articles reporting original reliability for the data at hand; articles citing estimates from previous studies; articles mentioning reliability with no reference to specific estimates; and articles making no mention of reliability. For our review and meta-analysis of factor analytic studies, factor analytical studies were catalogued according to their factor analysis methods: EFA or CFA. All EFA articles ($n = 12$) were tabulated according to author’s name, publication year, sample characteristics, test adaptation, and reported factor structure of scores (table 1). The CFA articles ($n = 8$) were tabulated based on author’s name, publication year, sample characteristics, models

Table 1. Author(s), Number of Items Analyzed, Number of Factors Retained and Variance Explained, Form and Response Format, Sample Size, and Description for Exploratory Factor-Analytic Studies ($n = 12$)

Study	Items Analyzed	Factors (Variance Explained)	Form	<i>N</i>	Sample Description
<i>Full-scale</i>					
Brenner et al ³⁸	42	3 (31.50%)	English & French	Random half of 2275	Canadian general population (49.5% male; Mean age = 26)
Fonseca-Pedrero et al ³⁹	42	3 ^b (34.57%)	Spanish	660	Spanish university students ($N = 660$; 29.5% male; Mean age = 20.3)
Verdoux et al ³⁶	42	3 (31.30%)	French	571	French undergraduate female students (Mean age = 19.8)
<i>Positive dimension</i>					
Armando et al ⁴¹	18 ^a	4 (51.50%)	English & Italian	1777	Australian high school students ($N=848$; 47% male; Mean age = 15) and Italian university students ($N=929$; 23.4% male; Mean age = 21)
Armando et al ⁴²	20	4 (51.00%)	Italian	997	Italian university students (23.8% male; Mean age = 21)
Barragan et al ⁴³	20	4 (48.00%)	Spanish	777	Spanish high school students (49.1% male; Mean age = 14.4)
Stefanis et al ⁴⁴	20	4 (N.A.)	Greek	3500	Greek adolescents (45% male; Mean age = 19)
Therman et al ⁴⁰	20	5 (62.80%)	Swedish P&P; O	31822	Swedish women (Mean age = 51.4)
Wigman et al ⁴⁷	20	5 (60.30%)	Language not specified	5422	Adolescents in Europe & North America (50% male; Mean age = 14.0)
Yung et al ⁴⁶	20	3 (52.44%)	English	140	Australian help-seeking youth with a nonpsychotic psychiatric problem (42% male; Mean age = 17.67)
Yung et al ⁴⁵	20	4 (N.A.)	English	875	Australian high school students (46.9% male)
<i>Negative dimension</i>					
Barragan et al ⁴³	14	3 (43.00%)	Spanish	777	Spanish high school students (49.1% male; Mean age = 14.4)
Ziermans ⁴⁸	14	3 (54.00%)	Swedish	1012	Swedish adolescents and young adults (Mean age = 24.4)

Note: P&P, Pencil & Paper; O, Online.

^aThe authors excluded items 15 and 20 from factor analysis because these items reportedly “related more closely to cultural background and age than to psychopathology.”

^bAuthors reported that a tetradimensional solution was implicated for extraction, but only 2 factors loaded on the fourth factor. Hence, a tridimensional solution was preferred.

tested, and fit indices (table 2). The 3 studies that analyzed data using both techniques were cross-tabulated on both tables.

Results

Sample Characteristics

Overall, sample sizes of all studies ($n = 111$) ranged from 22 to 47859 with a median of 533. Participants had an age range of 11.09 and 51.4 years, with a mean age of 25.27 years ($SD = 8.33$). The mean gender distribution in the studies was 43.74%.

In the sample of 12 EFA studies, sample sizes ranged from 777 to 31822. The median sample sizes for CAPE-42,

CAPE-*pos*, and CAPE-*neg* studies were 875, 1012, and 997 respectively. The mean age of participants were 18.52 years ($SD = 4.08$) for CAPE-42 studies, 22.52 years ($SD = 10.94$) for CAPE-*pos* studies, and 19.23 years ($SD = 4.70$) for CAPE-*neg* studies. The average male distribution was 39.50% for CAPE-42 studies, 40.9% for CAPE-*pos* studies, and 47.55% for CAPE-*neg* studies.

In the sample of 8 CFA studies, sample sizes ranged from 566 to 22386. The median sample sizes in CAPE-42 and CAPE-*pos* studies were 1311 and 1012 respectively. Participant age was reported for 3 CAPE-42 studies, with a mean of 21.43 years ($SD = 6.15$). Mean age for CAPE-*pos* samples was 22.47 years ($SD = 5.35$). Gender distribution was reported for 6 CFA studies. The average male

Table 2. Author(s), Best Fitting Model Reported, Sample Size, and Description for Confirmatory Factor-Analytic Studies ($n = 8$)

Study	Form	Best Fitting Model	Goodness-of-fit indices				<i>N</i>	Sample Description
			SRMR ^a	TLI ^b	CFI ^c	RMSEA ^d		
<i>Full scale</i>								
Original validation: Stefanis et al ¹⁴	Greek CAPE-40	3 correlated factors	—	—	—	0.045	932	Greek young men undergoing compulsory military service aged 18–24
Brenner et al ³⁸	English & French	3 correlated factors, with 19 items deleted	—	—	0.739	—	Random half of 2275	Canadian general population (49.5% male; Mean age = 26)
Fonseca-Pedrero et al ³⁹	Spanish	3 correlated factors	1.803	0.885	0.825	0.079	660	Spanish university students (29.5% male; Mean age = 20.3)
Vleeschouwer et al ³⁷	Dutch P&P; O	3 correlated factors with 2 residual correlations	—	P&P: 0.95 O: 0.94	P&P: 0.86 O: 0.80	P&P/O: 0.05	P&P: 796; O: 21590	Dutch general population (38.5% male; Mean age = 41.0)
<i>Positive dimension</i>								
Capra et al ⁴⁹	English	3 correlated factors, with 5 items deleted	—	0.938	0.976	0.027	1610	Australian university students (24% male; Mean age = 22.1)
Wigman et al ⁶⁷	English	5 correlated factors	—	—	0.92	0.038	2230	Dutch adolescents in the general population (49% male; Mean age = 11.1)
Wigman et al ⁵⁰	Language not specified	5 correlated factors (across T1, T2, T3)	—	—	T1: 0.959 T2: 0.987 T3: 0.990	T1: 0.030 T2: 0.019 T3: 0.016	566	Caucasian Belgian women (Mean age = 27.3)
Ziermans ⁴⁸	Swedish	4 correlated factors, allowing error terms of 3 item pairs to correlate	—	—	0.93	0.04	1012	Swedish adolescents and young adults (Mean age = 24.4)

Note: Figures in bold denote acceptable values for goodness of fit indices.

^aSRMR, Standardized root mean square residual; it is generally established that SRMR should be inferior to 0.08 for a good fit of the model data.

^bTLI, Tucker-Lewis Index; it is generally established that TLI should be superior to 0.95 for a good fit of the model data.

^cCFI, Comparative fit index; it is generally established that CFI should be superior to 0.95 for a good fit of the model data.

^dRMSEA, Root Mean Square Error of Approximation; it is generally established that RMSEA should be inferior to 0.06 for a good fit of the model data.

distributions in CAPE-42 and CAPE-*pos* studies were 54.38% and 49.13% respectively.

Language

The Dutch version of CAPE was the most commonly used ($n = 44$), followed by the English ($n = 23$) version, German ($n = 15$), Spanish ($n = 8$), Italian ($n = 5$), Greek ($n = 4$), Swedish ($n = 3$), French ($n = 2$), while 7 did not specify the languages used. Validation studies on translations of CAPE included Spanish,^{39,43} French,^{36,38} Italian,⁴² Greek,⁴⁴ and Swedish.^{40,48}

Reliability Generalization

A meta-analysis of original Cronbach’s alpha values reported in published studies was conducted. Of the

111 studies reviewed, 22 (19.82%) reported Cronbach’s alpha values using the data at hand. However, there was some potential overlap in samples, especially for large-scale local or multinational datasets. After elimination of potential overlaps, 18 unique samples resulted. Five studies (4.5%) induced reliability estimates from previous samples, 39 (35.14%) mentioned reliability but made no reference to specific estimates, and 45 (40.54%) did not mention reliability at all.

A meta-analysis of original reliability coefficients using a varying-coefficient model of aggregation computation was performed in Excel 2007 using the meta-analytic tool Synthesizer 1.0.^{68–70} The level of original alpha values reported for CAPE-42 ($n = 5$) had a meta-analytic mean of 0.91 ($SD = 0.05$), while that of CAPE-*pos* ($n = 9$) had a meta-analytic mean of 0.84 ($SD = 0.1$). Alpha values of

CAPE-*neg* ($n = 5$) had a mean of 0.81 ($SD = 0.10$), and that of the depressive subscale (CAPE-*dep*; $n = 4$) had a mean of 0.76 ($SD = 0.10$). Figure 1 depicts the distribution of original alpha coefficients reported for CAPE-42 and for each dimension.

In order to examine whether reliability estimates of CAPE scores depended on sample characteristics, mean alpha estimates were computed separately for studies that utilized a younger sample (mean age ≤ 25 years old) and those that utilized an older sample (mean age > 25 years old) for each subscale. To estimate the magnitude of the differences, CI values for the younger sample were compared with that of the older sample. Significant results are implicated if the 95% CI of the differences did not include zero.⁶⁹ Results showed that CAPE-*pos* (95% CI = $-.018, -.054$) and CAPE-*neg* (95% CI = $.054, .096$) scores appeared to be more internally reliable in younger samples. Age group of samples did not make a difference to internal reliability for full scale CAPE-42 and CAPE-*dep* scores.

Factor Structure of CAPE-42

The factor structure of CAPE-42 was examined by a review of EFA and CFA studies. In CFA studies, 3 studies in addition to the original Stefanis and colleagues' validation study of the Greek CAPE reported a tridimensional structure as best fitting the data. The Dutch version of CAPE-42 demonstrated acceptable model fit with RMSEA and TLI, though with a below cut-off CFI.³⁷ However, the Spanish, French, and English versions did not demonstrate acceptable model fit.

All 3 EFA studies conducted resulted in 3-factor structures. Two EFA studies were adjuncts to the CFA studies mentioned above: *post hoc* EFA were conducted to find a better-fitting model due to some unacceptable goodness-of-fit indices in the CFA.^{38,39} Both EFAs generated

3-factor models with different item loadings from the original CAPE-42.^{38,39} Fonseca-Pedrero and colleagues³⁹ initially extracted a tetra-dimensional solution; but the 4-factor model was abandoned by the authors because only 2 items loaded onto the fourth factor. Brenner and colleagues³⁸ found that the follow-up CFA on the 3-factor structure resulting from the EFA did not significantly improve model fit in their French or English versions of CAPE.

Factor Structure of the Positive Domain of CAPE

Of the 7 EFA-only studies on the positive dimension, one study reported a tripartite factor structure for CAPE-*pos*,⁴⁶ one reported a 5-factor structure,⁴⁰ while a majority ($n = 5$) reported a tetradimensional solution.^{41,42} In determining the number of factors to retain, 4 studies reported the use of eigenvalues and scree test, Monte Carlo simulations or a combination.^{40-42,45} Three studies failed to report their determination criteria. None of the EFA studies were followed up by CFA to confirm their results.

Four CFA studies reported the factor structure of CAPE-*pos*, one of which tested the resultant factor structure from preliminary EFA.⁴⁷ Three of the studies tested models of 1, 4, and 5 factors^{49,50} while Wigman and colleagues⁷¹ tested a sequence of nested models from 1 to 6 factors. All CFA studies reported multiple goodness-of-fit indices.

The number of CAPE-*pos* factors was not consistently replicated, although there were some similarities in nature of factors, with *perceptual anomalies/hallucinatory experiences*, *persecutory ideation*, and *bizarre experiences* frequently emerging as factors. Capra and colleagues proposed a "CAPE-p15", deleting questions on magical thinking, grandiosity and paranormal beliefs, resulting in 3 subscales (*Persecutory ideation*, *Perceptual abnormalities*, and *Bizarre experiences*) with inter-correlations between error terms for 4 item pairs.⁴⁹ Ziermans tested 3 4-factor models,⁴⁸ while Wigman and colleagues⁴⁷ provided evidence for 5-factor models (*Hallucination*, *Delusions*, *Paranoia*, *Grandiosity*, and *Paranormal beliefs*) in 2 large adolescent general population samples, and over 3 time points.⁵⁰

Given the inconsistencies, a meta-analysis of the 8 EFA studies on CAPE-*pos* was conducted to statistically resolve the discrepancy in factor structures, based on the Shafer method.⁷² An index of similarity between factor structures proposed in the individual EFA studies was computed based on simple proportion of co-occurrence for each pair of items—ie, (Number of times a pair of scale items loaded highest on the same factor) \div (Total number of times that pair of test items was measured). A Principal Component Analysis (PCA) was then conducted on the standardized similarity matrix.

A 3-factor structure emerged in the Varimax-rotated PCA of co-occurrence matrix according to the eigenvalues

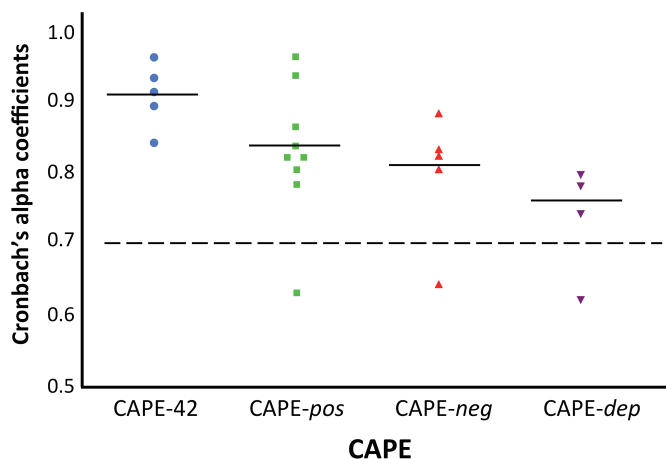


Fig. 1. Distribution of original Cronbach's alpha values of CAPE-42, CAPE-*pos*, CAPE-*neg*, and CAPE-*dep*. The solid and dashed lines represent the group mean and the acceptable level of reliability respectively.

and scree plot. Because factor loading co-occurrences are not expected to correlate, orthogonal rotation was employed. Altogether, the 3 factors explained 93.1% of the variance: Component 1 (39.53%) was defined by 7 items and corresponded to *Bizarre experiences*. Component 2 (32.15%) was defined by 9 items and corresponded to *Delusional ideations*, including grandiose and persecutory content. Component 3 (21.42%) was defined by four items and corresponded to *Perceptual anomalies*. Table 3 displayed the rotated 3-component solution.

Factor Structure of the Negative Domain of CAPE

Both EFA studies on CAPE-*neg* reported a model consisting of 3 factors, namely *Avolition*, *Affective flattening*, and *Social withdrawal* with PCA. Eigenvalues, scree test, and Monte Carlo simulations were used to confirm the 3-factor structure as best representing the data in the Ziermans study; while Barragan and colleagues failed to report their determination criteria. To date, no CFA studies of CAPE-*neg* have been published.

Summary

In sum, based on our review and meta-analyses, CAPE was found to consist of 3 subscales that could be further divided into 3 subdimensions each (figure 2).

Discussion

CAPE is one of the most frequently used instruments in psychosis proneness research. Despite increasing citations and usage in the literature, there is a paucity of documentation about psychometric properties of the scores generated by CAPE. There have been various attempts to examine the reliability and validity of CAPE scores in individual data sets; however, no prior reviews and meta-analytic studies have provided an overview of the psychometric properties of CAPE reported across different studies in the literature. Without knowing the typical reliability of CAPE scores, what factors could affect reliability of the scores, and whether CAPE’s factor structures are generalizable across samples, researchers and clinicians cannot know whether CAPE can accurately measure the psychosis phenotype.

As such, we conducted a review of the literature and a reliability generalization study to establish whether CAPE scores are typically reliable, and whether its scores might be more valid in younger vs older samples. Overall, our results indicate that CAPE scores were generally reliable in previous administrations, with the full scale CAPE-42 being reliable across age groups while positive and negative subscales tended to be more reliable in younger samples. The factorial validity of CAPE-42 scores was also satisfactory, with all EFA and CFA studies resulting in a 3-factor structure. Our review and meta-analysis generated new insights particularly in relation to the age group for which CAPE scores are

Table 3. CAPE Positive Dimension (CAPE-*Pos*) Exploratory Factor Analytic Studies Varimax Rotated Factors

CAPE- <i>Pos</i> Item	Bizarre Experiences	Delusional Ideations	Perceptual Anomalies
People seem to drop hints about you or say things with a double meaning	-.293	-. 857	-.399
Things in magazines or on TV were written especially for you	-. 633	-.597	
Some people are not what they seem to be	-.308	-. 905	-.262
Being persecuted in some way	-.289	-. 859	-.389
Conspiracy against you	-.310	-. 904	-.260
Destined to be someone very important	-.451	-. 683	-.470
You are a very special or unusual person	-.435	-. 706	-.453
Communicate telepathically		-. 837	
Electrical devices such as computers can influence the way you think	-. 931		
Believe in the power of witchcraft, voodoo or the occult	-.432	-. 656	
People look at you oddly because of your appearance	-.302	-. 856	-.396
Thoughts in your head are being taken away from you	-. 991		
Thoughts in your head are not your own	-. 987		
Thoughts so vivid that you were worried other people would hear them	-. 991		
Hear your own thoughts being echoed back to you	-. 992		
Under the control of some force or power other than yourself	-. 980		
Hear voices when you are alone			. 990
Hear voices talking to each other when you are alone			. 990
A double has taken the place of a family member, friend or acquaintance			. 966
See objects, people or animals that other people cannot see			. 939

Note: Only factor loadings >.25 are shown. The values in bold denote items loaded onto the respective factors.

more reliable, as well as the factor structure of CAPE-*pos*. Given the comprehensive nature of CAPE in tapping both the frequency *and* distress resulting from psychotic-like experiences, as well as its robust psychometric properties, we highly recommend CAPE for clinical and research use.

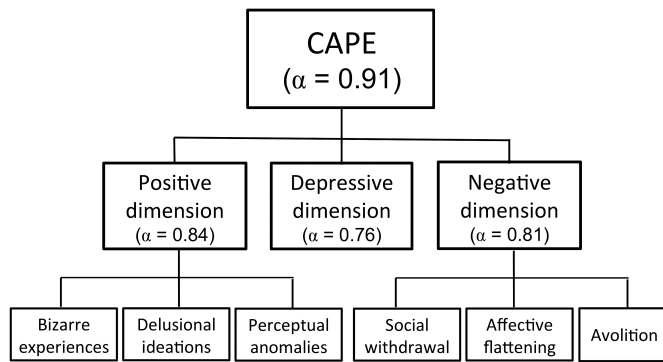


Fig. 2. Structure of CAPE-42 consisting of CAPE-*pos*, CAPE-*neg*, and CAPE-*dep* with their respective subdomains.

Internal Reliability of CAPE Scores

To shed light on the typical psychometric reliability of CAPE-42 items defined in terms of Cronbach's alpha, a meta-analysis of original Cronbach's alpha values reported in published studies was conducted. In studies that reported reliability coefficients for their own data, the majority of studies obtained Cronbach's alpha levels above the recommended 0.70.¹⁰ In addition, though Cronbach's alpha values were high, they were not over 0.9, indicating items were not overly redundant. This was confirmed by our meta-analytic work on the positive and negative dimensions. The depressive dimension received less support though the possibility that this relates to the small number of items cannot be excluded.

We also found that CAPE scores were more reliable for adolescents and young adults for the positive and the negative dimensions, but no significant difference was found for the full scale CAPE-42 or the depressive dimension. This was understandable because younger individuals have a higher tendency of currently experiencing psychotic-like symptoms,¹ which might be reflected in their response through increased identification with items. On the other hand, adults might not remember that they had such experiences, because psychotic-like experiences are usually transient and dismissed promptly (eg, attributed to tiredness) in normal populations. The nonsignificant difference in reliability of CAPE-*dep* scores was consistent with the higher prevalence of depressive symptoms across the lifespan, while the nonsignificant difference in reliability of CAPE-42 scores could be related to the inclusion of CAPE-*dep* scores, or the higher total number of items, which could pull up score reliability and mask differences reflected in subscale score reliability.

Taken together, CAPE scores appeared to be psychometrically reliable for general population samples, especially in younger populations. Hence, CAPE scores can typically be trusted and interpreted to reflect positive and negative psychotic-like experiences. Our results also showed that CAPE might be particularly useful for tapping positive and negative psychosis proneness features in younger populations. This is useful information

for researchers and clinicians, who are ethically bound to have a certain degree of confidence in an instrument before administering it to their participants or patients.

Despite the importance of psychometric reliability, reports of internal reliability coefficient were the exception rather than the norm. This is consistent with other psychological research. Less than 20% reported original reliability estimates, and more than 40% of studies made no mention of reliability at all. The lack of original reliability estimates reported in the literature casts doubts on the accuracy of measurements in findings, rendering results obtained questionable.⁷³ This phenomenon could stem from lack of understanding regarding the significance of psychometric properties, both on the part of authors as well as journal editors.

Factorial Validity of CAPE

In order to examine the factorial validity of CAPE, a review on factor analytic studies of CAPE-42 and its subscales were conducted. The latent factor structures emerging from EFA studies and the well-fit models in CFA collectively inform what structures underlie the CAPE, hence informing researchers and clinicians what psychotic-like experiences could be tapped by CAPE and its subscales.

Studies on CAPE-42

Our review on factor analytic studies of CAPE-42 revealed that all EFA studies included in this sample replicated a tridimensional structure reported by Stefanis and colleagues with Positive, Negative, and Depressive dimensions. However, it should be noted that none of the EFA conducted on translated CAPE-42 closely replicated the original proposed structure of the Greek CAPE-42, nor did any CFA study of the translated CAPE-42 achieve optimal model fit for the original structure.

Taken together, the results suggested that the factor structure of CAPE-42 contained 3 dimensions, but the exact composition of items in each subscale could vary across samples. The fact that the 4 CFA studies each used a different translation of the questionnaire could also contribute to the variability found. Hence, it could not be ruled out that the original 42-item scale might not be sensitive to psychosis proneness across cultures, possibly because of differential understanding of scale items. Hence, translation and validation of CAPE across cultures is warranted.

Regarding extent of replication of the original 20*pos*-14*neg*-8*dep* structure as originally proposed by Stefanis and colleagues, CAPE-*dep* had the highest replication fidelity to the original structure. Items proposed to be in CAPE-*dep* were mostly found to load together. However, item 21 (*lacking in energy*), originally proposed to belong to CAPE-*neg*, was found to load more consistently with Depressive items. This was not surprising because general fatigue is 1 feature of depressed individuals. That said,

such cross-loading highlighted the difficulty of clearly differentiating positive, negative, and depressive symptoms based on self-report questionnaires. Hence, CAPE should be used as a preliminary screening for secondary clinical attention rather than for diagnostic purposes.

Studies on CAPE-Pos

CAPE-*pos* structure was not consistently replicated. Our review revealed that a 4-factor structure was the most frequently-reported model of CAPE-*pos* while a 5-factor structure was also posited^{40,47,50,71} and replicated across 3 time points.⁵⁰ The 3 factors of *perceptual anomalies/hallucinatory experiences*, *persecutory ideation*, and *bizarre experiences* were consistently reported across studies, while the nature of the fourth or fifth factor varied. However, both EFA studies reporting 5-factors did not report their determination criteria for retaining factors. One could not exclude therefore the possibility that the variability in factor structure could be related to differential methods.

Given the high replicability of the 4-factor structure but greater psychometric robustness for the 5-factor structure, we resolved the discrepancy statistically by conducting a meta-analysis of EFA studies. Our meta-analysis indicated that 3 factors corresponding to *Bizarre experiences*, *Delusional ideations*, and *Perceptual anomalies* underlie the EFA studies on CAPE-*pos*. The emergence of the *Bizarre experiences* and *Perceptual anomalies* subscales corroborated previous factor analyses, suggesting that these 2 factors reflected stable constructs. Our delusional ideation subscale contained items referring to both grandiose and persecutory content, which was at odds with previous individual studies that usually found them to load separately on 2 subscales. Our meta-analysis results suggested that items on delusional ideations might have a common loading pattern that is sometimes masked in individual samples.

One important point to note is that our meta-analysis should be taken as reference for the aggregate underlying structure of positive symptoms of psychosis proneness as measured by CAPE, rather than as a gold-standard for the factorial structure of CAPE-*pos*. No meta-analytic results could replace CFA or EFA in establishing factorial validity for different translations or cultural groups and specific samples.

Studies on CAPE-Neg

There is a relative paucity of research on the factor structure of the negative dimension of CAPE, with only 2 published studies to date. The 2 studies showed high consistency in terms of CAPE-*neg* factor structures, both reporting factors of *Social withdrawal*, *Affective flattening*, and *Avolition*.

Conclusion

Our study is the first to comprehensively examine the psychometric properties of CAPE scores, a widely used

self-report assessment tool for subclinical psychotic-like experiences. Our RG study and meta-analysis of factor analysis led us to conclude that CAPE-42 scores were typically reliable, and its 3-factor structure had factorial validity. Moreover, we confirmed CAPE's subscale structures to facilitate the use of CAPE for diverse assessment objectives in clinical and research use.

Our results also highlighted questionable reporting practices of internal reliability in the current literature. Despite the availability of comprehensive guidelines, the field still showed an under-appreciation of the importance of psychometric properties.⁷⁴⁻⁷⁸ We believe that rectification of the state-of-art would require a 2-pronged approach. On one hand, budding researchers need to be better educated in psychometrics to produce more psychometrically knowledgeable researchers in the long run. To this end, we advocate for increased focus on psychometric properties and appropriate reporting strategies in undergraduate- and graduate-level programmes. On the other hand, we encourage experimental and clinical psychology journals to welcome more psychometric reviews and meta-analytic studies to increase readers' awareness of the importance of psychometrics.

Funding

General Research Fund grant from the Research Grant Council of Hong Kong (to T.T.). W.M. was supported by a rewarding internationalization scheme applied by T.T.

Acknowledgment

The authors have declared that there are no conflicts of interest in relation to the subject of this study.

References

1. Linscott RJ, van Os J. An updated and conservative systematic review and meta-analysis of epidemiological evidence on psychotic experiences in children and adults: on the pathway from proneness to persistence to dimensional expression across mental disorders. *Psychol Med*. 2013;43:1133–1149.
2. Murray GK, Jones PB. Psychotic symptoms in young people without psychotic illness: mechanisms and meaning. *Br J Psychiatry*. 2012;201:4–6.
3. Kelleher I, Lynch F, Harley M, et al. Psychotic symptoms in adolescence index risk for suicidal behavior: findings from 2 population-based case-control clinical interview studies. *Arch Gen Psychiatry*. 2012;69:1277–1283.
4. Yung AR, Buckby JA, Cosgrave EM, et al. Association between psychotic experiences and depression in a clinical sample over 6 months. *Schizophr Res*. 2007;91:246–253.
5. Werbeloff N, Drukker M, Dohrenwend BP, et al. Self-reported attenuated psychotic symptoms as forerunners of severe mental disorders later in life. *Arch Gen Psychiatry*. 2012;69:467–475.
6. Kelleher I, Keeley H, Corcoran P, et al. Clinicopathological significance of psychotic experiences in non-psychotic young people: evidence from four population-based studies. *Br J Psychiatry*. 2012;201:26–32.

7. Rössler W, Riecher-Rössler A, Angst J, et al. Psychotic experiences in the general population: a twenty-year prospective community study. *Schizophr Res*. 2007;92:1–14.
8. Fanous A, Gardner C, Walsh D, Kendler KS. Relationship between positive and negative symptoms of schizophrenia and schizotypal symptoms in nonpsychotic relatives. *Arch Gen Psychiatry*. 2001;58:669–673.
9. Cantor-Graae E, Selten JP. Schizophrenia and migration: a meta-analysis and review. *Am J Psychiatry*. 2005;162:12–24.
10. Read J, van Os J, Morrison AP, Ross CA. Childhood trauma, psychosis and schizophrenia: a literature review with theoretical and clinical implications. *Acta Psychiatr Scand*. 2005;112:330–350.
11. van Os J, Linscott RJ, Myin-Germeys I, Delespaul P, Krabbendam L. A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. *Psychol Med*. 2009;39:179–195.
12. Linscott RJ, van Os J. Systematic reviews of categorical versus continuum models in psychosis: evidence for discontinuous subpopulations underlying a psychometric continuum. Implications for DSM-V, DSM-VI, and DSM-VII. *Annu Rev Clin Psychol*. 2010;6:391–419.
13. Polanczyk G, Moffitt TE, Arseneault L, et al. Etiological and clinical features of childhood psychotic symptoms: results from a birth cohort. *Arch Gen Psychiatry*. 2010;67:328–338.
14. Stefanis NC, Hanssen M, Smirnis NK, et al. Evidence that three dimensions of psychosis have a distribution in the general population. *Psychological Medicine*. 2002;32:347–358.
15. Boonstra N, Wunderink L, Sytema S, Wiersma D. Improving detection of first-episode psychosis by mental health-care services using a self-report questionnaire. *Early Interv Psychiatry*. 2009;3:289–295.
16. Meier ST, Davis SR. Trends in reporting psychometric properties of scales used in counseling psychology research. *J Couns Psychol*. 1990;37:113–115.
17. Thompson B, Snyder PA. Statistical significance and reliability analyses in recent Journal of Counseling & Development research articles. *J Couns Dev*. 1998;76:436.
18. Vacha-haase T, Ness C, Nilsson J, Reetz D. Practices regarding reporting of reliability coefficients: a review of three journals. *J Exp Educ*. 1999;67:335–341.
19. Willson VL. Research techniques in AERJ articles: 1969 to 1978. *Educ Res*. 1980;9:5–10.
20. Vacha-Haase T, Thompson B. Score reliability: a retrospective look back at 12 years of reliability generalization studies. *Meas Eval Couns Dev*. 2011;44:159–168.
21. John OP, Soto CJ. The importance of being valid: reliability and the process of construct validation. In: Robins RW, Fraley RC, Krueger RF, eds. *Handbook of Research Methods in Personality Psychology*. New York, NY: Guilford; 2007:461–494.
22. Peterson RA. A meta-analysis of Cronbach's coefficient alpha. *J Cons Res*. 1994;21:381–391.
23. Henson RK. Understanding internal consistency reliability estimates: a conceptual primer on coefficient alpha. *Meas Eval Couns Dev*. 2001;34:177–189.
24. Hogan TP, Benjamin A, Brezinski KL. Reliability methods: a note on the frequency of use of various types. *Educ Psychol Meas*. 2000;60:523–531.
25. Nunnally JC. *Psychometric Theory*. 2nd ed. New York, NY: McGraw-Hill; 1978.
26. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess*. 2003;80:99–103.
27. Cortina JM. What Is coefficient alpha? an examination of theory and applications. *J Appl Psychol*. 1993;78:98–104.
28. Schmitt N. Uses and abuses of coefficient alpha. *Psychol Assess*. 1996;8:350–353.
29. Russell DW. In search of underlying dimensions: the use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Pers Soc Psychol Bull*. 2002;28:1629–1646.
30. Henson RK, Roberts JK. Use of exploratory factor analysis in published research: common errors and some comment on improved practice. *Educ Psychol Meas*. 2006;66:393–416.
31. Jackson DL, Gillaspay JA, Purc-Stephenson R. Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychol Methods*. 2009;14:6–23.
32. Caruso JC. Reliability generalization of the NEO personality scales. *Educ Psychol Meas*. 2000;60:236–254.
33. Yin P, Fan X. Assessing the reliability of Beck Depression Inventory Scores: reliability generalization across studies. *Educ Psychol Meas*. 2000;60:201–223.
34. Vacha-Haase T. Reliability generalization: exploring variance in measurement error affecting score reliability across studies. *Educ Psychol Meas*. 1998;58:6–20.
35. Howitt D, Cramer D. *Introduction to Research Methods in Psychology*. Essex: Pearson Education; 2005.
36. Verdoux H, Sorbara F, Gindre C, Swendsen JD, van Os J. Cannabis use and dimensions of psychosis in a nonclinical population of female subjects. *Schizophr Res*. 2002;59:77–84.
37. Vleeschouwer M, Schubart CD, Henquet C, et al. Does assessment type matter? A measurement invariance analysis of online and paper and pencil assessment of the Community Assessment of Psychic Experiences (CAPE). *PLoS One*. 2014;9:1–9.
38. Brenner K, Schmitz N, Pawliuk N, et al. Validation of the English and French versions of the Community Assessment of Psychic Experiences (CAPE) with a Montreal community sample. *Schizophr Res*. 2007;95:86–95.
39. Fonseca-Pedrero E, Paino M, Lemos-Giráldez S, Muñiz J. Validation of the Community Assessment Psychic Experiences -42 (CAPE-42) in Spanish college students and patients with psychosis. *Actas Esp Psiquiatr*. 2012;40:169–176.
40. Therman S, Suvisaari J, Hultman CM. Dimensions of psychotic experiences among women in the general population. *Int J Methods Psychiatr Res*. 2014;23:62–68.
41. Armando M, Nelson B, Yung AR, et al. Psychotic-like experiences and correlation with distress and depressive symptoms in a community sample of adolescents and young adults. *Schizophr Res*. 2010;119:258–265.
42. Armando M, Nelson B, Yung AR, et al. Psychotic experience subtypes, poor mental health status and help-seeking behaviour in a community sample of young adults. *Early Interv Psychiatry*. 2012;6:300–308.
43. Barragan M, Laurens KR, Navarro JB, Obiols JE. Psychotic-like experiences and depressive symptoms in a community sample of adolescents. *Eur Psychiatry*. 2011;26:396–401.
44. Stefanis NC, Delespaul P, Henquet C, et al. Early adolescent cannabis exposure and positive and negative dimensions of psychosis. *Addiction*. 2004;99:1333–1341.
45. Yung AR, Nelson B, Baker K, et al. Psychotic-like experiences in a community sample of adolescents: implications for the continuum model of psychosis and prediction of schizophrenia. *Aust N Z J Psychiatry*. 2009;43:118–128.

46. Yung AR, Buckby JA, Cotton SM, et al. Psychotic-like experiences in nonpsychotic help-seekers: associations with distress, depression, and disability. *Schizophr Bull.* 2006;32:352–359.
47. Wigman JTW, Vollebergh WAM, Raaijmakers QAW, et al. The structure of the extended psychosis phenotype in early adolescence—a cross-sample replication. *Schizophr Bull.* 2011. 37:850–860.
48. Ziermans TB. Working memory capacity and psychotic-like experiences in a general population sample of adolescents and young adults. *Front Psychiatry.* 2013;4:1–10.
49. Capra C, Kavanagh DJ, Hides L, Scott J. Brief screening for psychosis-like experiences. *Schizophr Res.* 2013;149:104–107.
50. Wigman JTW, Vollebergh WAM, Jacobs N, et al. Replication of the five-dimensional structure of positive psychotic experiences in young adulthood. *Psychiatry Res.* 2012;197:353–355.
51. Daneluzzo E, Stratta P, Di Tommaso S, et al. Dimensional, non-taxonic latent structure of psychotic symptoms in a student sample. *Soc Psychiatry Psychiatr Epidemiol.* 2009;44:911–916.
52. Goodwin LD, Goodwin WL. Focus on psychometrics. Estimating construct validity. *Res Nurs Health.* 1991;14:235–243.
53. Dixon J. Grouping techniques. In: Munro B, ed. *Statistical Methods for Health Care Research.* 3rd ed. Philadelphia, PA: Lippincott; 1997:310–340.
54. Bryman A, Cramer D. *Quantitative Data Analysis for Social Scientists.* New York, NY: Routledge; 1990.
55. Fabrigar LR, MacCallum RC, Wegener DT, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods.* 1999;4:272–299.
56. Brown TA. *Confirmatory Factor Analysis for Applied Research.* New York, NY: Guilford; 2006.
57. Kline RB. *Principles and Practice of Structural Equation Modeling.* 2nd ed. New York, NY: Guilford; 2005.
58. MacCallum RC, Austin JT. Applications of structural equation modeling in psychological research. *Annu Rev Psychol.* 2000;51:201–226.
59. DiStefano C, Hess B. Using confirmatory factor analysis for construct validation: an empirical review. *J Psychoeduc Assess.* 2005;23:225–241.
60. Schreiber JB, Nora A, Stage FK, Barlow EA, King J. Reporting structural equation modeling and confirmatory factor analysis results: a review. *J Educ Res.* 2006;99:323–337.
61. Worthington RL, Whittaker TA. Scale development research: a content analysis and recommendations for best practices. *Couns Psychol.* 2006;34:806–838.
62. Muthén LK, Muthén BO. *MPlus User's Guide.* 6th ed. Los Angeles, CA: Muthén & Muthén; 1998–2011.
63. Jöreskog KG, Sörbom D. *LISREL 8.8 for Windows [Computer Software].* Skokie, IL: Scientific Software International; 2006.
64. Hooper D, Coughlan J, Mullen MR. Structural equation modelling: guidelines for determining model fit. *Eur J Bus Res Methods.* 2008;6:53–60.
65. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling.* 1999;6:1–55.
66. Armando M, Lin A, Girardi P, et al. Prevalence of psychotic-like experiences in young adults with social anxiety disorder and correlation with affective dysregulation. *J Nerv Ment Dis.* 2013;201:1053–1059.
67. Wigman JTW, van Winkel R, Ormel J, Verhulst FC, van Os K, Vollebergh WAM. Early trauma and familial risk in the development of the extended psychosis phenotype in adolescence. *Acta Psychiatr Scand.* 2012;126:266–273.
68. Bonett DG. Varying coefficient meta-analytic methods for alpha reliability. *Psychol Methods.* 2010;15:368–385.
69. Krizan Z. Synthesizer 1.0: a varying-coefficient meta-analytic tool. *Behav Res Methods.* 2010;42:863–870.
70. Bonett DG. Meta-analytic interval estimation for standardized and unstandardized mean differences. *Psychol Methods.* 2009;14:225–238.
71. Wigman JT, van Winkel R, Raaijmakers QA, et al. Evidence for a persistent, environment-dependent and deteriorating subtype of subclinical psychotic experiences: a 6-year longitudinal general population study. *Psychol Med.* 2011;41:2317–2329.
72. Shafer AB. Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *J Clin Psychol.* 2006;62:123–146.
73. Vassar M, Hale W. Reliability reporting across studies using the Buss Durkee Hostility Inventory. *J Interpers Violence.* 2009;24:20–37.
74. Schmitt TA. Current methodological considerations in exploratory and confirmatory factor analysis. *J Psychoeduc Assess.* 2011;29:304–321.
75. Kaiser HF. The application of electronic computers to factor analysis. *Educ Psychol Meas.* 1960;20:141–151.
76. Cattell RB. The scree test for the number of factors. *Multivariate Behav Res.* 1966;1:245–276.
77. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika.* 1965;30:179–185.
78. Thompson B, Daniel LG. Factor analytic evidence for the construct validity of scores: a historical overview and some guidelines. *Educ Psychol Meas.* 1996;56:197–208.