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

Winifred Mark1 and Timothea Touloupoulou*1,2,3

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

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 characteristics11–13 between subclinical psychotic experiences and psychotic disorders suggest that psychotic experiences exist along a continuum, whereby psychotic disorders differ only quantitatively from psychosis proneness.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. 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. 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. 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.

The most common reliability coefficient reported is Cronbach’s alpha, 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 0.7 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. 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. Contrary to popular belief, only scores, but not tests, can be reliable. 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.

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.

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. Copious validation studies have
attempted to demonstrate factorial validity of CAPE scores in independent samples.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{58-61} A variety of goodness-of-fit indices can easily be calculated in statistical packages such as Mplus\textsuperscript{62} and LISREL,\textsuperscript{63} though the interpretation of model fit and suggested cut-offs are debated, and could vary in different sources.\textsuperscript{64} The most popular indices are Comparative Fit Index (CFI; acceptable if $\geq 0.95$), Tucker-Lewis Index (TLI; acceptable if $\geq 0.95$), Root Mean Square Error of Approximation (RMSEA; acceptable if $\leq 0.06$), 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,\textsuperscript{65} in order to reflect different aspects of model fit.\textsuperscript{64}

In the original validation study of CAPE, Stefanis and colleagues\textsuperscript{14} 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.\textsuperscript{36-39} More recently, factor structures of the Positive (CAPE-pos) and Negative (CAPE-neg) dimensions have also been proposed.\textsuperscript{50-54,56,57}

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
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
W. Mark & T. Toulopoulou

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, French, Italian, Greek, and Swedish.

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. 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

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

<table>
<thead>
<tr>
<th>Study</th>
<th>Form</th>
<th>Best Fitting Model</th>
<th>Goodness-of-fit indices</th>
<th>N</th>
<th>Sample Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>SRMR^a</td>
<td>TLI^b</td>
<td>CFI^c</td>
</tr>
<tr>
<td>Full scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original validation:</td>
<td>Greek CAPE-40</td>
<td>3 correlated factors</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Stefanis et al^14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brenner et al^18</td>
<td>English &amp; French</td>
<td>3 correlated factors, with 19 items deleted</td>
<td>—</td>
<td>0.739</td>
<td>—</td>
</tr>
<tr>
<td>Fonseca-Pedrero et al</td>
<td>Spanish</td>
<td>3 correlated factors</td>
<td>1.803</td>
<td>0.885</td>
<td>0.825</td>
</tr>
<tr>
<td>Vleeschouwer et al^37</td>
<td>Dutch P&amp;P; O</td>
<td>3 correlated factors with 2 residual correlations</td>
<td>—</td>
<td>P&amp;P: 0.95</td>
<td>P&amp;P: 0.86</td>
</tr>
<tr>
<td>Positive dimension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capra et al^49</td>
<td>English</td>
<td>3 correlated factors, with 5 items deleted</td>
<td>—</td>
<td>0.938</td>
<td><strong>0.976</strong></td>
</tr>
<tr>
<td>Wigman et al^67</td>
<td>English</td>
<td>5 correlated factors</td>
<td>—</td>
<td>0.92</td>
<td><strong>0.038</strong></td>
</tr>
<tr>
<td>Wigman et al^50</td>
<td>Language not specified</td>
<td>5 correlated factors (across T1, T2, T3)</td>
<td>—</td>
<td>T1: <strong>0.959</strong></td>
<td>T2: <strong>0.987</strong></td>
</tr>
<tr>
<td>Ziermans^48</td>
<td>Swedish</td>
<td>4 correlated factors, allowing error terms of 3 item pairs to correlate</td>
<td>—</td>
<td>0.93</td>
<td><strong>0.04</strong></td>
</tr>
</tbody>
</table>

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.
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.\(^6\) Results showed that CAPE-pos \((95\% CI = −0.018, −0.054)\) and CAPE-neg \((95\% CI = 0.054, 0.096)\) 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.\(^7\) 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.\(^8,9\) Both EFAs generated 3-factor models with different item loadings from the original CAPE-42.\(^38,39\) Fonseca-Pedrereo and colleagues\(^9\) 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 colleagues\(^10\) 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.

**Figure 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.

**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,\(^48\) 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.\(^40,42,44\) 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\(^9,49,50\) while Wigman and colleagues\(^51\) 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.\(^9\) Ziemans tested 3 4-factor models,\(^49\) while Wigman and colleagues\(^51\) 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) ÷ (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.
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 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.

---

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

<table>
<thead>
<tr>
<th>CAPE-Pos Item</th>
<th>Bizarre Experiences</th>
<th>Delusional Ideations</th>
<th>Perceptual Anomalies</th>
</tr>
</thead>
<tbody>
<tr>
<td>People seem to drop hints about you or say things with a double meaning</td>
<td>−.293</td>
<td>−.857</td>
<td>−.399</td>
</tr>
<tr>
<td>Things in magazines or on TV were written especially for you</td>
<td>−.633</td>
<td>−.597</td>
<td></td>
</tr>
<tr>
<td>Some people are not what they seem to be</td>
<td>−.308</td>
<td>−.905</td>
<td>−.262</td>
</tr>
<tr>
<td>Being persecuted in some way</td>
<td>−.289</td>
<td>−.859</td>
<td>−.389</td>
</tr>
<tr>
<td>Conspiracy against you</td>
<td>−.310</td>
<td>−.904</td>
<td>−.260</td>
</tr>
<tr>
<td>Destined to be someone very important</td>
<td>−.451</td>
<td>−.683</td>
<td>−.470</td>
</tr>
<tr>
<td>You are a very special or unusual person</td>
<td>−.435</td>
<td>−.706</td>
<td>−.453</td>
</tr>
<tr>
<td>Communicate telepathically</td>
<td>−.302</td>
<td>−.856</td>
<td>−.396</td>
</tr>
<tr>
<td>Electrical devices such as computers can influence the way you think</td>
<td>−.931</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Believe in the power of witchcraft, voodoo or the occult</td>
<td>−.432</td>
<td>−.656</td>
<td></td>
</tr>
<tr>
<td>People look at you oddly because of your appearance</td>
<td>−.991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thoughts in your head are being taken away from you</td>
<td>−.987</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thoughts in your head are not your own</td>
<td>−.991</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thoughts so vivid that you were worried other people would hear them</td>
<td>−.992</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hear your own thoughts being echoed back to you</td>
<td>−.980</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Under the control of some force or power other than yourself</td>
<td>−.990</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hear voices when you are alone</td>
<td>−.990</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hear voices talking to each other when you are alone</td>
<td>−.966</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A double has taken the place of a family member, friend or acquaintance</td>
<td>−.393</td>
<td></td>
<td></td>
</tr>
<tr>
<td>See objects, people or animals that other people cannot see</td>
<td>−.393</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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,7 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 20pos-14neg-8dep 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\textsuperscript{40,47,50,71} and replicated across 3 time points.\textsuperscript{39} 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.\textsuperscript{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


