



Validation of the revised Mystical Experience Questionnaire in experimental sessions with psilocybin

Frederick S Barrett¹, Matthew W Johnson¹ and Roland R Griffiths^{1,2}

Abstract

The 30-item revised Mystical Experience Questionnaire (MEQ30) was previously developed within an online survey of mystical-type experiences occasioned by psilocybin-containing mushrooms. The rated experiences occurred on average eight years before completion of the questionnaire. The current paper validates the MEQ30 using data from experimental studies with controlled doses of psilocybin. Data were pooled and analyzed from five laboratory experiments in which participants ($n=184$) received a moderate to high oral dose of psilocybin (at least 20 mg/70 kg). Results of confirmatory factor analysis demonstrate the reliability and internal validity of the MEQ30. Structural equation models demonstrate the external and convergent validity of the MEQ30 by showing that latent variable scores on the MEQ30 positively predict persisting change in attitudes, behavior, and well-being attributed to experiences with psilocybin while controlling for the contribution of the participant-rated intensity of drug effects. These findings support the use of the MEQ30 as an efficient measure of individual mystical experiences. A method to score a “complete mystical experience” that was used in previous versions of the mystical experience questionnaire is validated in the MEQ30, and a stand-alone version of the MEQ30 is provided for use in future research.

Keywords

Psilocybin, hallucinogens, entheogen, psychedelic, spiritual, mystical experience, factor analysis, structural equation modeling, psychometrics

Introduction

Reports of mystical-type experiences have been provided by religious and non-religious figures around the world and throughout recorded history (James, 1902; Stace, 1960). Common dimensions in reports of mystical experiences include the experience of profound unity with all that exists, a felt sense of sacredness, a sense of the experience of truth and reality at a fundamental level (noetic quality), deeply felt positive mood, transcendence of time and space, and difficulty explaining the experience in words (ineffability; Stace, 1960). This operational definition was provided by Stace (1960), and formed the basis of different versions of the Mystical Experience Questionnaire (MEQ), which was developed to evaluate the occurrence and character of individual, discrete mystical experiences occasioned by classic hallucinogens (Griffiths et al., 2006; MacLean et al., 2012; Pahnke, 1963).

The MEQ has been administered in various forms in a number of studies over the past 50 or more years (Bogenschutz et al., 2015; Garcia-Romeu et al., 2015; Griffiths et al., 2006, 2008, 2011; Johnson et al., 2014; MacLean et al., 2012; Pahnke, 1963, 1967; Richards, 1975). The most frequently used version of the MEQ is the 43-item Mystical Experience Questionnaire (MEQ43), also called the Pahnke–Richards Mystical Experience Questionnaire. The MEQ43 contains 43 items that were theoretically derived and qualitatively organized into seven subscales (internal unity, external unity, sacredness, noetic quality, positive mood, transcendence of time and space, and ineffability). The MEQ43 has been used in prospective studies of the subjective effects (Griffiths et al., 2006, 2008, 2011) and therapeutic outcomes (Bogenschutz et al., 2015; Garcia-Romeu et al., 2015; Johnson et al., 2014) of ingestion of psilocybin. MEQ43 scores

are shown to be dose dependent (Griffiths et al., 2011) and predict therapeutic outcomes of psilocybin sessions (Bogenschutz et al., 2015; Garcia-Romeu et al., 2015). However, the factor structure of the MEQ43 has only recently undergone psychometric investigation (MacLean et al., 2012).

The most recently developed version of the MEQ (the 30-item revised Mystical Experience Questionnaire, or MEQ30) was developed and validated through factor analysis of retrospective accounts of profound experiences with psilocybin-containing mushrooms (MacLean et al., 2012). That analysis yielded a four-factor structure of the MEQ30, containing 30 items from the previous MEQ43, which was typically administered within the 100-item States of Consciousness Questionnaire (Griffiths et al., 2006, 2011). The four factors of the MEQ30 are: mystical (including items from the internal unity, external unity, noetic quality, and sacredness scales of the MEQ43), positive mood, transcendence of time and space, and ineffability (all three of which include items from their respective MEQ43 scales). Thus, the MEQ30 retains items from each qualitative subscale in the original MEQ43, but in a reduced number of dimensions. The

¹Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA

²Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Corresponding author:

Frederick S. Barrett, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, 5510 Nathan Shock Drive, Baltimore, MD 21224, USA.
Email: fbarret2@jhu.edu

MEQ30 is not an alternate form of the MEQ43, but rather a psychometrically validated instrument that is derived from the MEQ43.

Factor structures underlying questionnaires assessing mystical experiences have received a fair amount of attention in the psychology of religion (Chen et al., 2011a, 2011b, 2012; Hood et al., 2001). In particular, the Hood Mysticism Scale (HMS; Hood 1975, 2009) is a well-validated instrument that was derived from the same theoretical basis as the MEQ, and it consists of similar underlying content domains. However, the HMS and the MEQ differ in specific items, underlying factor structure, and the time frame over which dimensions of mystical experiences are assessed. The HMS has been applied as a measure of phenomena occurring over a lifetime, whereas the MEQ assesses phenomena occurring during a single, discrete experience. Furthermore, the items of the HMS consistently organize into a three-factor structure (Chen et al., 2011a, 2011b, 2012; Hood et al., 2001). The items of the MEQ30 fit well in the four-factor structure described above, and this four-factor structure of the MEQ30 yielded superior model fit when compared with two alternative three-factor structures that were derived from the HMS (MacLean et al., 2012).

The MEQ30 was developed and validated using online surveys of retrospective accounts of mystical and spiritual experiences encountered after ingesting psilocybin-containing mushrooms (MacLean et al., 2012). The conditions under which these experiences occurred were uncontrolled, and reports were based on retrospective accounts of experiences that occurred on average eight years before the given ratings. While psychometric properties of the MEQ30 were found to be stable and reliable across different online survey samples assessing the effects of psilocybin-containing mushrooms, the factor structure of the MEQ30 has not yet been replicated in data from studies of controlled-dose psilocybin administration in a laboratory setting. Also, while earlier versions of the MEQ43 included a method to identify a “complete mystical experience,” which was increased by psilocybin (Griffiths et al., 2006, 2011; Pahnke, 1967), an analogous method for scoring a complete mystical experience has not been demonstrated in the MEQ30. This “complete mystical experience” may serve as a valuable predictor of positive outcomes. Cancer patients who met the criteria for a “complete mystical experience” in a study of psychedelic-assisted psychotherapy showed greater clinical improvements than did those who did not meet these criteria (Richards et al., 1977). Healthy adults who met the criteria for a “complete mystical experience” showed increases in openness after psilocybin administration, while those who did not meet the criteria did not show these changes (MacLean et al., 2011).

The aim of the current study is to replicate and demonstrate the reliability of the factor structure of the MEQ30 in data collected in a controlled laboratory setting. In this study, the MEQ30 is validated with data from five experimental psilocybin studies in which participants received a moderate to high dose of psilocybin (≥ 20 mg/70 kg; Griffiths et al., 2006, 2011; Johnson et al., 2014; cancer-related anxiety/depression study, in preparation; spiritual practice study, in preparation). Confirmatory factor analysis of MEQ30 items demonstrates the internal validity of the MEQ30. The external validity of the MEQ30 is demonstrated using structural equation modeling, regressing ratings of persisting effects of psilocybin on both latent variable scores from the

MEQ30, and participant ratings of the overall strength of drug effects. The specificity of the MEQ30 is established by showing that latent scores, but not participant ratings of overall drug strength, are related to persisting effects attributed to psilocybin. Finally, the MEQ43 method for dichotomously scoring the presence of a complete mystical experience is adapted to the MEQ30.

Methods

Participants

Data were pooled across five laboratory experiments in which participants received a moderate to high oral dose of psilocybin ranging between 20 and 30 mg/70 kg. One observation was included per participant. For each observation, the maximum dose received for that participant was included in the analysis because the probability of fulfilling criteria for having a complete mystical experience on the MEQ43 is an increasing function of psilocybin dose (Griffiths et al., 2011). Each study was conducted in compliance with US laws and approved by the Johns Hopkins University School of Medicine Institutional Review Board. All participants provided informed consent.

Study 1: Griffiths et al. (2006). Thirty-six participants completed a double-blind human pharmacology study comparing the subjective effects of psilocybin and methylphenidate. Thirty participants received either psilocybin (30 mg/70 kg) or methylphenidate (40 mg/70 kg) in counterbalanced order across two experimental sessions. Six participants did not complete all study questionnaires, and were excluded from the analyses. A further six participants completed all questionnaire material and received psilocybin (30 mg/70 kg) after having received methylphenidate on two previous sessions. Included in the analyses were data from the psilocybin session for the 30 participants who completed all the relevant questionnaires.

Study 2: Griffiths et al. (2011). Eighteen participants in this double-blind study received either an ascending or descending sequence of five psilocybin doses (0, 5, 10, 20, and 30 mg/70 kg), each in a separate experimental session, with the placebo inserted in the sequence in pseudo-random order. Data from each participant's high-dose session (30 mg/70 kg) were included in the analysis.

Study 3: Johnson et al. (2014). Fifteen participants completed an open label pilot study investigating the utility of two or three psilocybin sessions (20 or 30 mg/70 kg) as an aid to smoking cessation. Data from each participant's highest-dose session were included (30 mg/70 kg, $n=14$; 20 mg/70 kg, $n=1$). If the participant received the highest dose twice, the first exposure was used as that participant's observation.

Study 4: Spiritual practices study (in preparation). Seventy-five participants completed a double-blind study of the interaction between psilocybin and training in meditation and other spiritual practices. Fifty participants received 20 and 30 mg/70 kg psilocybin on the first and second experimental sessions, respectively. The remaining participants ($n=25$) received a very low dose of psilocybin (1 mg/70 kg) on two experimental

sessions, and 21 of these participants opted to receive a high dose of psilocybin (30 mg/70 kg) on a third experimental session. Data from all 30 mg/70 kg sessions ($n=71$) were included in the analysis.

Study 5: Cancer-related anxiety/depression study (in preparation). Fifty-one participants were enrolled in a double-blind study investigating the effects of psilocybin in people with anxiety and/or depressed mood comorbid with a life-threatening cancer diagnosis. Participants received either a very low dose (1 or 3 mg/70 kg) or a moderately high dose (22 mg/70 kg) of psilocybin over two counterbalanced experimental sessions. One participant received a high dose of psilocybin (30 mg/70 kg). Data from the 50 participants who received the moderately high dose session were included in the analysis.

Measures

The MEQ43. At the end of each experimental session of each study, participants completed a battery of questionnaires that included the States of Consciousness Questionnaire (SOCQ). The SOCQ is a 100-item questionnaire that contains the 43 items of the original MEQ43 (Griffiths et al., 2006). Each item of the MEQ43 was rated on a 6-point scale, where 0="none; not at all," 1="so slight cannot decide," 2="slight," 3="moderate," 4="strong (equivalent in degree to any previous strong experience or expectation of this description)," and 5="extreme (more than ever before in my life and stronger than 4)." The MEQ43 is comprised of seven subscales: internal unity, external unity, sacredness, noetic quality, deeply felt positive mood, transcendence of time and space, and ineffability/paradoxicality. Scale scores for each participant are generated from the sum of responses to all items of a given scale.

The MEQ30. The MEQ30 is constructed from a subset of 30 items from the MEQ43 (MacLean et al., 2012), and is comprised of four factors: mystical (including items from the internal unity, external unity, noetic quality, and sacredness scales of the MEQ43), positive mood, transcendence of time and space, and ineffability (all three of which include a subset of items from the parallel scales within the MEQ43). Responses to the items comprising the MEQ30 in the analysis were taken from participant responses to the MEQ43, which was embedded within the 100-item SOCQ. Responses to the items of the MEQ30 were submitted to confirmatory factor analysis and structural equation modeling for the purposes of validating the MEQ30 within laboratory experimental data.

The Persisting Effects Questionnaire. At a period of approximately three to eight weeks (which varied across studies) after each experimental session, participants in each study completed a questionnaire assessing persisting behavioral and psychological changes that they attributed to their most recent experimental session. In this questionnaire, participants indicated the meaningfulness of their experience ("How personally meaningful was the experience?") using an 8-point ordinal scale (1="no more than routine, everyday experiences," 2="similar to meaningful experiences that occur on average once a week," 3="similar to meaningful experiences that occur on average once a month," 4="similar

to meaningful experiences that occur on average once a year," 5="similar to meaningful experiences that occur on average once every 5 years," 6="among the top 10 most meaningful experiences of my life," 7="among the top 5 most meaningful experiences of my life," and 8="the single most meaningful experience of my life"). Participants also indicated the degree to which their experience was spiritually significant ("Indicate the degree to which the experience was spiritually significant to you:") using a 6-point rating scale (1="not at all," 2="slightly," 3="moderately," 4="very much," 5="among the top 5 most spiritually significant experiences of my life," and 6="the single most spiritually significant experience of my life"). Participants responded to a question inquiring about the effect of the experience on their well-being or life satisfaction ("Do you believe that the experience and your contemplation of that experience have led to a change in your current sense of personal well-being or life satisfaction?") using a 7-point rating scale (+3, "increased very much"; +2, "increased moderately"; +1, "increased slightly"; 0, "no change"; -1, "decreased slightly"; -2, "decreased moderately"; and -3, "decreased very much"). Finally, participants indicated the degree to which their experience had a positive impact on their behavior following the session ("Your behavior has changed in ways you consider positive since the experience."), using a 6-point rating scale (0="none, not at all"; 1="so slight, cannot decide"; 2="slight"; 3="moderate"; 4="strong"; 5="extreme").

The Hallucinogen Rating Scale. The Hallucinogen Rating Scale (HRS; Strassman et al., 1994) is a 100-item questionnaire designed to assess subjective effects of hallucinogenic substances. The HRS includes a 4-item intensity scale. Three items from the intensity scale (intensity, a rush, and high) were rated on a 5-point rating scale (1="not at all," 2="slightly," 3="moderately," 4="very much," 5="extremely") and the fourth item (amount of time between when the drug was administered and feeling an effect) was rated on a 6-point rating scale (1="not applicable, no effect"; 2="0–15 minutes"; 3="15–30 minutes"; 4="30–60 minutes"; 5="60–90 minutes"; and 6=">90 minutes"). Intensity scale scores from the HRS were added to structural equation models of the effects of the MEQ30 on persisting effects attributed to psilocybin. Separate structural equation models were estimated using just the single rating of the intensity item from the HRS (item 99), as this item does not confound subjective strength of drug effects with any other dimensions (such as rush, high, and speed of onset). These variables were included in the structural equation models to rule out the plausible alternative hypothesis that persisting effects attributed to psilocybin are due to the strength of the overall drug effect rather than any attribution or dimension related to mystical experience.

Analysis

Internal and external validity of the MEQ30. MEQ30 items were entered into a series of confirmatory factor analyses to validate the factor structure and internal validity of the MEQ30 against previously proposed and demonstrated alternate models (MacLean et al., 2012). Data were fit to three competing models: the four-factor structure of the MEQ30, a three-factor "Hood" model, and a three-factor "Stace" model.

The four-factor structure of the MEQ30 was the product of exploratory and confirmatory factor analyses of the initial item pool of the MEQ43 (MacLean et al., 2012). In this four-factor structure, items 9, 12, 14, 22, 35, 36, 41, 47, 54, 55, 69, 73, 74, 77, and 83 from the MEQ43 (a subset of items from the internal unity, external unity, sacredness, and noetic quality factors of the MEQ43) loaded onto factor 1, mystical; items 5, 18, 30, 43, 80, and 87 (a subset of items from the positive mood factor of the MEQ43) loaded onto factor 2, positive mood; items 2, 15, 29, 34, 48, and 65 (a subset of items from the transcendence of time and space factor of the MEQ43) loaded onto factor 3, transcendence of time and space; and items 6, 23, and 86 (a subset of items from the ineffability factor of the MEQ43) loaded onto factor 4, ineffability.

The three-factor “Hood” model was based on empirical factors derived from the HMS (Hood, 1975) and applied as a plausible alternative model for the items of the MEQ30 (MacLean et al., 2012). In the three-factor “Hood” model, items 2, 6, 12, 15, 23, 29, 34, 35, 41, 48, 54, 65, 77, 83, and 86 from the MEQ43 (a subset of items from the transcendence of time and space, ineffability, and internal unity factors of the MEQ43) loaded onto an introvertive factor; items 14, 47, and 74 from the MEQ43 (a subset of items from the external unity factor of the MEQ43) loaded onto an extrovertive factor; and items 5, 9, 18, 22, 30, 43, 69, 36, 55, 73, 80, and 87 from the MEQ43 (a subset of items from the positive mood, sacredness, and noetic quality factors of the MEQ43) loaded onto an interpretation factor. The final three-factor “Stace,” which is an alternate interpretation of the three-factor “Hood” model, is identical to the “Hood” model, except that items 6, 23, and 86 from the ineffability factor of the MEQ43 were moved from the introvertive factor to the interpretation factor (MacLean et al., 2012). Model fit indexes were compared to determine the best-fitting model out of these three alternatives.

An additional confirmatory factor model was fit for the MEQ30, extending the best-fitting factor model by adding an additional second-order latent variable (MEQ30-total) that can account for the correlations between the first-level latent variables from the best-fitting model. For example, if the MEQ30 model was shown to be the best-fitting model (over the “Stace” and “Hood” three-factor models), the second-order latent variable MEQ30-total model would be defined as having all four latent variables (factors) of the MEQ30 model (mystical, positive mood, transcendence of time and space, and ineffability) loading onto MEQ30-total.

Structural equation modeling was used to establish the external and convergent validity of latent variable (i.e., factor) scores of the MEQ30. The best-fitting confirmatory factor model was extended by adding a structural regression. Responses to four items in the persisting effects questionnaire (personal meaningfulness, spiritual significance, change in well-being, and positive behavior change) were regressed on the latent variable scores for the MEQ30 and the ratings of intensity of drug effect. Separate models were fit using either the HRS intensity factor or the HRS “intensity” item (item 99) as measures of the intensity of drug effect. These structural equation models were repeated by regressing responses to items in the persisting effects questionnaire on the second-order latent variable (MEQ30-total) and the intensity of drug effect.

Confirmatory factor analyses and structural equation models were fit with maximum likelihood estimation using the lavaan toolbox (Rosseel, 2012) in the R statistical package (R Core Team, 2013). Model identification and latent variable scaling for

the second-order factor model was achieved by setting the loading of the first item for each latent variable to 1. Model fit was evaluated using the comparative fit index (CFI) and the standardized root mean square residual (SRMR; Bentler, 1995; Hu and Bentler, 1999). Change in the chi-square goodness-of-fit index between CFA models is also reported.

CFI may be considered an indicator of acceptable model fit (Brown, 2006), and consideration of a combination of fit indexes, with “acceptable fit” values of SRMR <0.09 and CFI >0.90, has been shown to minimize both type I and type II error, even in models of small samples ($n \leq 250$; Hu and Bentler, 1999). A commonly used model fit index, root mean square error of approximation (RMSEA; Browne and Cudeck, 1993), has been shown to both over-reject true-population models and under-reject misspecified models when the sample size is small ($n \leq 250$), and therefore it was not considered here (Hu and Bentler, 1999). Given the small sample size, chi-square (and, by extension, CFI) was calculated using the used the Satorra–Bentler scaled test statistic (Satorra and Bentler, 2001).

Adaptation of a scoring method for complete mystical experience. The MEQ43 can be scored to yield a dichotomous classification of “complete mystical experience.” Observations with a score $\geq 60\%$ of the maximum possible score on each of the subscales of the MEQ43, discarding the lower score of either internal unity or external unity (yielding a classification based on six of the seven scales of the MEQ43), are considered a complete mystical experience (Griffiths et al., 2006). This classification approach was adapted to the MEQ30. Observations on the MEQ30 that had a score $\geq 60\%$ of the maximum possible score on each of the four subscales of the MEQ30 were considered a “complete mystical experience.” In the case of the MEQ30, differences in scores for the previously identified internal unity or external unity scales were not considered because items from both of these scales load onto the mystical factor in the MEQ30.

For comparison, “complete mystical experience” classifications were calculated separately using six factors of the MEQ43 and four factors of the MEQ30, and these two methods of classification were compared. Classification statistics of the complete mystical scorings derived from the MEQ30 were assessed, using the classifications of complete mystical experience derived from the MEQ43 as the ground truth. Classification statistics consist of a contingency table (true positive, TP; false positive, FP; true negative, TN; and false negative, FN, observations), hit rate (TP/[TP+FN]), false alarm rate (FP/[FP+TN]), proportion correct ((TP+TN)/[TP+TN+FP+FN]), and the index of sensitivity (d') of the MEQ30-derived scores (Macmillan and Creelman, 2005).

Results

Internal validity of the MEQ30

Comparison of model fit indexes between the four-factor MEQ30 model, the three-factor “Stace” model, and the three-factor “Hood” model for MEQ30 data replicated previous findings of best fit for the four-factor MEQ30 model (MacLean et al., 2012). Specifically, the four-factor MEQ30 model demonstrated acceptable model fit (CFI >0.90, SRMR <0.09), while both the Stace and Hood models demonstrated poor model fit (CFI <0.90; Table 1). Standardized confirmatory factor loadings and factor correlations for the four-factor MEQ30 model

are reported in Figure 1. The four factors of the MEQ30 displayed excellent reliability, calculated using Cronbach's alpha ($\alpha_{\text{mystical}}=0.97$, $\alpha_{\text{positive mood}}=0.92$, $\alpha_{\text{trans.time/space}}=0.86$,

$\alpha_{\text{ineffability}}=0.90$). Factor loadings for the four-factor MEQ30 model show high loading of each item onto its intended factor, and support the internal validity of the instrument. Factor loadings and correlations are similar to those reported in the development of the MEQ30 (MacLean et al., 2012), with the exception of a lower factor loading in the current sample (0.46) compared with the previously reported loading (0.70) of MEQ30 item #11. This factor loading would have met inclusion criteria in the original publication (MacLean et al., 2012). Therefore, no modifications of the questionnaire or factor model were introduced.

An additional model was fit (the MEQ30-total model) with all four latent variables (mystical, positive mood, transcendence of time and space, and ineffability) loading onto a second-order latent variable (MEQ30-total). This model is consistent with the theoretical and philosophical literature on mystical experience that proposes an underlying generative experience that is manifest

Table 1. Model fit for confirmatory factor analyses.

Model	CFI	SRMR	χ^2 (df)	$\Delta\chi^2$ (Δ df)
Stace (3 factors)	0.829	0.074	1068.43 (402)	
Hood (3 factors)	0.839	0.071	1029.99 (402)	38.44 (0)***
MEQ30 (4 factors)	0.906	0.062	767.40 (399)	262.59 (3)***
MEQ30 total	0.900	0.068	790.86 (401)	23.46 (2)***

χ^2 and CFI are corrected using the Satorra-Bentler scaling (Satorra and Bentler, 2001).

*** $p<0.001$.

CFI, comparative fit index; SRMR, standardized root mean square residual.

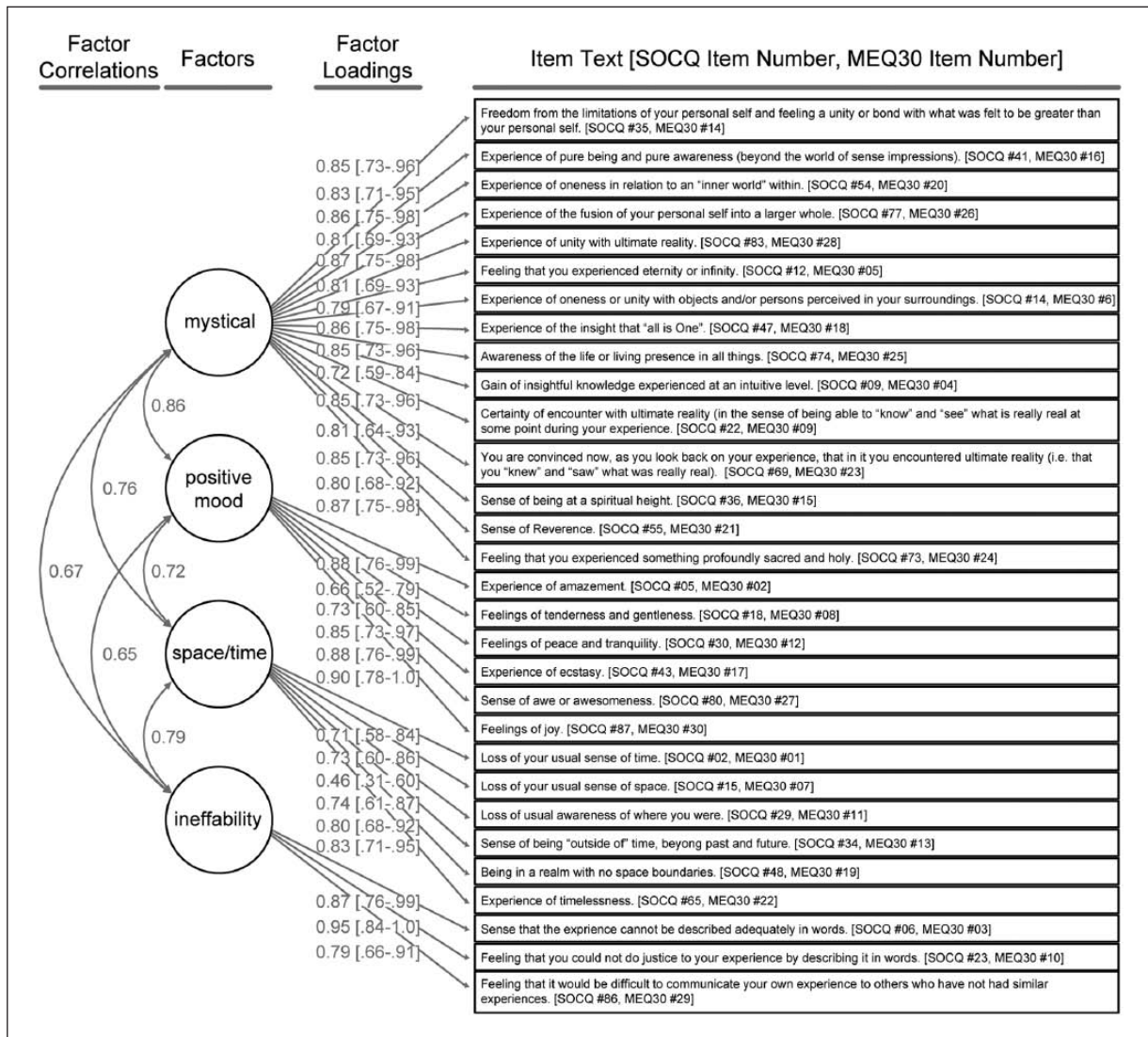


Figure 1. Confirmatory factor structure and factor correlations for the four-factor, 30-item Mystical Experience Questionnaire (MEQ30). For each factor loading, 95% confidence intervals were estimated using bootstrapping in the lavaan toolbox in the R statistical package, and are presented in square brackets next to each factor loading. The States of Consciousness Questionnaire (SOCQ) item number and the MEQ30 item number for each item is presented at the end of the text for that item. Factor loadings are standardized.

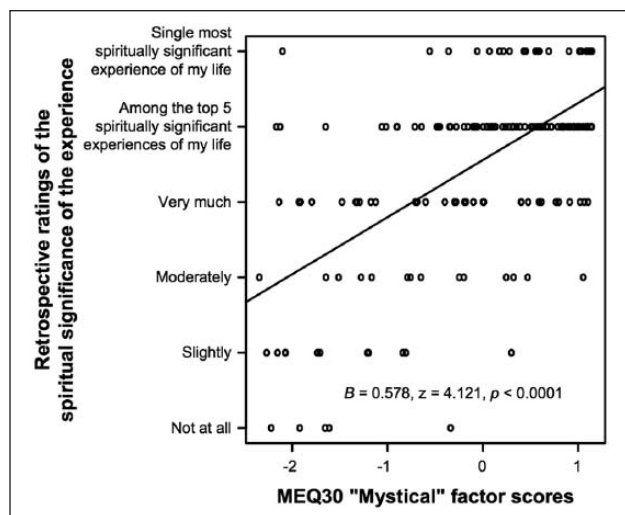


Figure 2. Retrospective ratings of the spiritual significance of psilocybin sessions are predicted by MEQ30 "mystical" factor scores. Retrospective ratings were provided three to eight weeks after each psilocybin session. MEQ30 ratings were provided seven hours after capsule administration. The line of best fit is plotted through the scatterplot. The standardized regression coefficient (B) from the structural equation model and the associated z - and p -values for this relationship are superimposed on the figure.

in multiple facets (e.g., unity, positive mood, transcendence of time and space, and ineffability). This model demonstrated acceptable fit (MEQ30-total; Table 1), supporting the existence of a second-order MEQ30 total score variable.

External validity of the MEQ

The two MEQ30 models above (the four-factor MEQ30 model and the MEQ30-total model) were extended by adding a structural regression to each model. The structural MEQ30 model consisted of multivariate regression of each of four items from the persisting effects questionnaire (personal meaningfulness, spiritual significance, change in well-being or life satisfaction, and positive behavior change attributed by the volunteer to the psilocybin session experience) on the mystical, positive mood, transcendence of time and space, and ineffability latent variables, and the rated intensity of drug effects (using either the intensity factor or intensity item, item 99, from the HRS). The structural MEQ30-total model consisted of a regression of each of the persisting effects items on the MEQ30-total latent variable and the rated intensity of drug effects.

Both the structural MEQ30 model (CFI=0.905, SRMR=0.058) and the structural MEQ30-total model (CFI=0.900, SRMR=0.060) demonstrated acceptable model fit. The HRS intensity item responses did not predict outcome measures in either model. Latent variable scores for the mystical factor of the MEQ30 positively predicted the retrospectively rated spiritual significance of the session experience (Figure 2). Scores on the mystical factor also predicted positive changes in behavior attributed to the session ($B=0.465, z=2.576, p=0.010$). Latent variable scores for the positive mood factor of the MEQ30 predicted ratings of increased sense of personal well-being attributed to the session ($B=0.315,$

$z=2.197, p=0.028$). No first-level latent variable scores (mystical, positive mood, transcendence of time and space, and ineffability) were predictive of the rated meaningfulness of the experience. The MEQ30-total latent variable scores significantly and positively predicted each outcome variable (Figure 3).

Scoring complete mystical experiences

The contingency tables comparing classification of "complete" mystical experience derived from the MEQ30 to classifications of "complete" mystical experience derived from the MEQ43 are presented in Table 2. The hit rate (0.95), false alarm rate (0.14), proportion correct (90.2%), and sensitivity ($d'=3.464$) for complete mystical experience scores derived from the MEQ30 factors indicate that this scoring is comparable to the scoring of complete mystical experience scores derived from the MEQ43, with the MEQ30 exhibiting a slight bias toward scoring more observations as mystical experiences than the MEQ43.

MEQ43 and MEQ30 factor scores for "false alarm" cases (where the MEQ30 factor scores indicate a "complete mystical experience" but the MEQ43 factor scores do not) are presented in Table 3. Closer inspection reveals that each case of a false alarm for complete mystical experience using the MEQ30 factor scores ($n=14$) can be traced to a sub-threshold score on a single scale of the MEQ43. The positive mood factor of the MEQ43 contributed a sub-threshold score for the greatest number false alarm observations ($n=5$), but all other MEQ43 factors contributed a sub-threshold score in at least one false-alarm case (ineffability=3 cases, noetic=2 cases, unity=2 cases, sacredness=1 case, transcendence of time and space=1 case).

Discussion

For this study, end-of-session ratings of mystical experience and intensity of drug effects, as well as longitudinal ratings of persisting effects of psilocybin, were combined across five laboratory experiments. Factor analyses of the items that comprise the MEQ30 support the internal validity of this instrument. Structural equation models support the external validity of latent variable scores on the MEQ30 by demonstrating a significant relationship between MEQ30 latent variable scores (specifically the mystical, positive mood, and MEQ30-total variables) and persisting effects of psilocybin sessions, when controlling for the possible effects of drug intensity on these outcome measures.

The theoretical basis of a one-factor second-level latent structure to the MEQ30 (supporting an MEQ30-total score) lies in the fact that the dimensions of mystical experience outlined by Stace (1960) and codified in the MEQ are theorized to represent expressions of an underlying (latent) unitary mystical experience. The one-factor solution to the MEQ30 in fact represents a quantitative test of this theory, modeling a second-level latent variable that predicts each of the four first-level latent variables (mystical, positive mood, ineffability, and transcendence of time and space). Only two of dimensions of the four-factor model of the MEQ30 (mystical and positive mood) show predictive validity when regressing the retrospective ratings of psilocybin sessions, while the MEQ30-total score showed predictive validity for all retrospective ratings. The MEQ30-total score having the greatest predictive validity is consistent with the premise of a latent mystical experience that drives

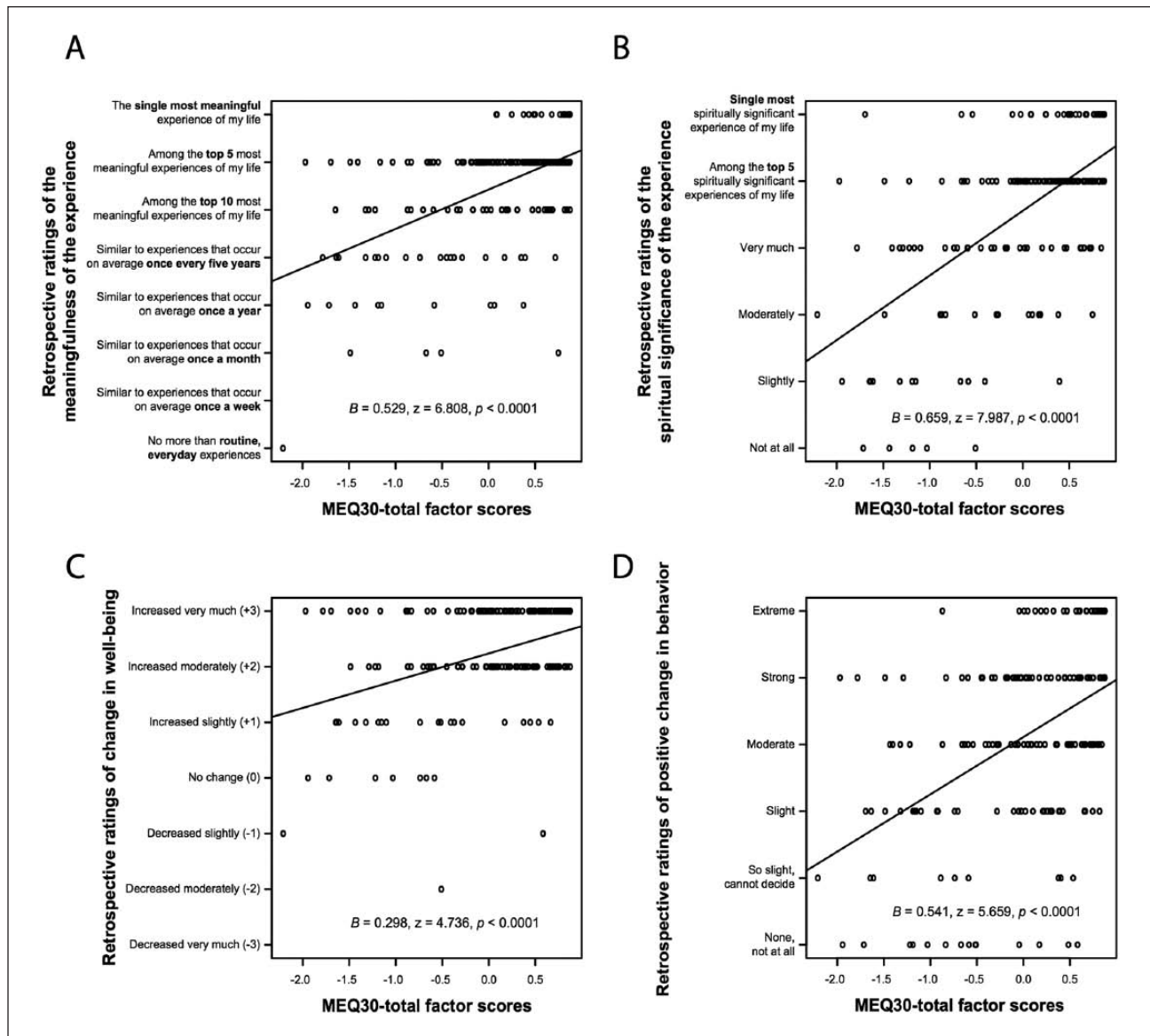


Figure 3. Retrospective ratings of psilocybin sessions are predicted by the MEQ-total factor score. Retrospective ratings were provided three to eight weeks after each psilocybin session. MEQ30 ratings were provided seven hours after capsule administration. Lines of best fit are plotted through each scatterplot. Standardized regression coefficients (B) from the structural equation model and their associated z and p values are superimposed on each axis. (A and B) Retrospective ratings of the (A) meaningfulness and (B) spiritual significance of psilocybin sessions are plotted against standardized MEQ30-total scores. (C and D) Retrospective ratings of (C) change in well-being and (D) positive behavior change attributed to psilocybin sessions are plotted against standardized MEQ30-total scores.

Table 2. Contingency table comparing complete mystical experience classifications between MEQ43-derived classification and MEQ30-derived classification.

		MEQ43 factors		
		Positive	Negative	
MEQ30 factors	Positive	80	14	94
	Negative	4	86	90
		84	100	$N=184$

change in each proposed subscale of mystical experience, and it supports the idea that retrospective ratings of psilocybin experiences and persisting effects that are attributed to those experiences would most strongly be predicted by a measure that incorporates signal from all proposed dimensions of mystical experience.

Mystical experiences are defined as having a number of dimensions (unity, sacredness, noetic quality, ineffability, positive mood, transcendence of time and space), in which the experience of any one of those dimensions alone does not constitute a classic “mystical” experience (Stace, 1960). A scoring system for “complete mystical experiences” derived from the four MEQ30 factor

Table 3. Comparison of factor scores on the MEQ43 and MEQ30 for false positive cases of complete mystical experience derived from MEQ30 factor scores.

False positive case	MEQ43						MEQ30			
	Max. unity	Trans time/space	Ineffability	Sacred	Noetic	Positive mood	Mystical	Positive mood	Trans time/space	Ineffability
1	0.63	0.80	1.00	0.83	0.65	0.43	0.73	0.67	0.83	1.00
2	0.97	0.88	1.00	0.80	1.00	0.49	0.97	0.63	0.83	1.00
3	0.93	0.78	0.84	1.00	0.90	0.54	0.97	0.67	0.73	1.00
4	0.70	0.85	0.88	0.60	0.65	0.54	0.71	0.73	0.87	1.00
5	0.60	0.80	0.64	0.60	0.75	0.54	0.61	0.60	0.87	0.73
6	0.77	0.80	0.80	0.66	0.50	0.74	0.61	0.73	0.83	1.00
7	0.63	0.78	0.60	0.74	0.55	0.69	0.67	0.73	0.80	0.67
8	0.83	0.85	0.80	0.51	0.70	0.63	0.68	0.73	0.83	0.80
9	0.63	0.68	0.48	0.77	0.70	0.80	0.73	0.83	0.63	0.67
10	0.67	0.78	0.48	0.80	0.90	0.74	0.85	0.77	0.70	0.67
11	0.93	0.78	0.56	0.86	1.00	0.97	1.00	0.97	0.70	0.93
12	0.90	0.58	0.76	0.97	1.00	0.86	0.97	0.97	0.63	0.80
13	0.43	0.85	1.00	0.86	0.95	0.80	0.76	0.83	1.00	1.00
14	0.53	0.90	0.92	0.89	0.80	0.80	0.65	0.80	1.00	0.87

Scale scores on the MEQ43 and MEQ30 are expressed as the proportion of the maximum possible score on each scale. The “Max. Unity” column contains the maximum value between the internal unity and external unity scales from the MEQ43, and is considered as the score for the unity factor when determining complete mystical experiences for the MEQ43. Scores on any scale that fall below 60% of the total maximum score for that scale are presented in bold.

scores is demonstrated, and provides a rough classification of a “complete” mystical experience equivalent to the six-factor scoring system for the MEQ43 that has been used in previous research (Griffiths et al., 2006, 2011). Though the false-alarm rate (14%) for classification of complete mystical experiences using the MEQ30 factor scores relative to the MEQ43 factor scores is notable, each false-alarm case may be more likely due to greater sensitivity (ability to detect true positives) of the MEQ30 relative to the MEQ43, rather than a decrease in specificity (ability to identify true negatives) of the MEQ30 relative to the MEQ43. Each case of a false alarm derived from the MEQ30 factor scores is super-threshold ($\geq 60\%$ of the total maximum score) for all but one scale score of the MEQ43, and the one sub-threshold MEQ43 scale score for each false alarm case is not greatly sub-threshold ($M=0.51$, $SD=0.046$, range 0.43–0.58). Given that the MEQ30 was constructed from the most psychometrically valid and reliable items from the entire MEQ43, it is reasonable to assume that items not included in the MEQ30 contribute error variance in the scoring of complete mystical experiences in the MEQ43, and this could explain potential decreased sensitivity of “complete mystical” classification using the MEQ43. From this perspective, removal of weak items in the construction of the MEQ30, and a greater number of observations being judged “complete mystical experience” in the MEQ30, can be considered a positive feature of the revised instrument.

Using a stand-alone version of the MEQ

While the bulk of MEQ data reported to date (Griffiths et al., 2006, 2011; Johnson et al., 2014; MacLean et al., 2012) were collected using the MEQ43 embedded within the full 100-item SOCQ, the second study in the original development of the MEQ30 (MacLean et al., 2012) validated the factor structure of the MEQ30 in a data set in which only the items of the MEQ43 were presented (not including the 57 distractor items of the SOCQ). This demonstrates

the robustness of the MEQ30 factor structure against the presence or absence of the SOCQ distractor items.

In order to replicate the factor structure of the MEQ30 directly in the absence of SOCQ distractor items (as in MacLean et al., 2012: study 2), an additional online survey data set was analyzed (Appendix 1). In this data set, the MEQ43 was presented, rather than the 100-item SOCQ. Confirmatory factor analysis of the MEQ30 items from this data set replicated the factor structure of the MEQ30 as reported in this study and in previous work (MacLean et al., 2012), providing further evidence that the structure of this instrument is resilient to the presence or absence of distractor items from the SOCQ.

A stand-alone version of the MEQ30 is offered in Appendix 2 and is recommended for future experimental use. Though the MEQ30 has not yet been explicitly validated as a stand-alone instrument, it has demonstrated resilience against the presence or absence of distractor items. Considering that the MEQ30 has now been psychometrically validated in both online survey samples and experimental laboratory samples, and in reports of both experiences with psilocybin-containing mushrooms and experiences with controlled-dose psilocybin administrations, use of the MEQ30 is recommended over both the MEQ43 and SOCQ. To facilitate comparison of future studies utilizing the MEQ30 with previously publications that reported on ratings of the MEQ43 (Griffiths et al., 2008, 2011), we provide MEQ30 scores (in Appendix 3) for MEQ43 ratings provided in these previously published reports.

Relationship between mystical and religious experience

The mystical experience is not conceptually limited to religious experience or practice, and the measurement of mystical experience by the MEQ does not require any direct religious or mystical endorsement. The MEQ serves as a psychometrically sound self-report instrument that assesses philosophically and theoretically identified

facets of mystical experiences and, by virtue of scores on these dimensions, can characterize the degree to which a given experience fits the schema of “mystical.”

Conclusion

Although the biological mechanisms underlying mystical experiences have not been identified, mystical experiences have a clear operational definition (Griffiths et al., 2006, 2011; MacLean et al., 2012; Pahnke, 1963; Richards, 1975), and the value of mystical experiences in terms of predicting positive outcomes has been empirically demonstrated (Griffiths et al., 2006, 2008, 2011; Johnson et al., 2014; MacLean et al., 2011). The present study contributes to this literature by demonstrating the psychometric fitness in prospectively obtained experimental data of an instrument that was developed to assess individual mystical experiences (the MEQ30). These findings extend this literature by demonstrating the value of scores on the MEQ30 over a simple measure of intensity of drug effects in predicting positive outcomes related to mystical experiences occasioned by psilocybin. Given the predictive validity of MEQ30-total scores in the current sample, researchers should use the MEQ30-total score over MEQ30 factor scores (identified by the mystical, positive mood, transcendence of time and space, or the ineffability factors) in future investigations of mystical experience, unless they have specific hypotheses regarding the predictive validity of specific dimensions of mystical experience.

Validation of the MEQ30 in experimental data constitutes an important step in developing this questionnaire instrument for use in pharmacological studies of mystical experiences. Further experimental work should validate the MEQ30 in studies of mystical experiences encountered through other means. On the basis of these findings, we recommend use of the MEQ30 for future investigations of individual episodes of mystical experiences.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Dr Griffiths is on the board of directors of the Heffter Research Institute, a funding sponsor for some of the studies included in this report.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the National Institutes of Health [grant numbers R01 DA03889, T32 DA007209], and grants from the Heffter Research Institute and the Council on Spiritual Practices.

References

- Bentler PM (1995) *EQS Structural Equations Program Manual*. Encino, CA: Multivariate Software.
- Bogenschutz MP, Forcehimes AA, Pommy JA, et al. (2015) Psilocybin-assisted treatment for alcohol dependence: a proof-of-concept study. *J Psychopharmacol* 29: 289–299.
- Brown TA (2006) *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Press.
- Browne MW and Cudeck R (1993) Alternative ways of assessing model fit. In: Bollen KA and Long JS (eds) *Testing Structural Equation Models*. Newbury Park, CA: Sage, pp.136–162.
- Chen Z, Hood RW, Yang L, et al. (2011a) Mystical experience among Tibetan Buddhists: the common core thesis revisited. *J Scient Study Religion* 50: 328–338.
- Chen Z, Wen Q, Hood RW, et al. (2011b) Common core thesis and qualitative and quantitative analysis of mysticism in Chinese Buddhist monks and nuns. *J Scient Study Religion* 50: 654–670.
- Chen Z, Zhang Y, Hood RW, et al. (2012) Mysticism in Chinese Christians and non-Christians: measurement invariance of the mysticism scale and implications for the mean differences. *Int J Psychol Religion* 22: 155–168.
- Garcia-Romeu A, Griffiths RR and Johnson MW (2015) Psilocybin-occasioned mystical experiences in the treatment of tobacco addiction. *Current Drug Abuse Rev* 7: 157–164.
- Griffiths R, Richards W, Johnson M, et al. (2008) Mystical-type experiences occasioned by psilocybin mediate the attribution of personal meaning and spiritual significance 14 months later. *J Psychopharmacol* 22: 621–632.
- Griffiths RR, Johnson MW, Richards WA, et al. (2011) Psilocybin occasioned mystical-type experiences: immediate and persisting dose-related effects. *Psychopharmacology (Berl)* 218: 649–665.
- Griffiths RR, Richards WA, McCann U, et al. (2006) Psilocybin can occasion mystical-type experiences having substantial and sustained personal meaning and spiritual significance. *Psychopharmacology (Berl)* 187: 268–283; discussion 284–292.
- Hood RW (1975) The construction and preliminary validation of a measure of reported mystical experience. *J Scient Study Religion* 14: 29–41.
- Hood RW, Ghorbani N, Watson PJ, et al. (2001) Dimensions of the mysticism scale: confirming the three-factor structure in the United States and Iran. *J Scient Study Religion* 40: 691–705.
- Hood RW Jr, Hill PC and Spilka B (2009) *The Psychology of Religion: An Empirical Approach*. 4th ed. New York, NY: Guilford.
- Hu L and Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Eq Model: Multidis J* 6: 1–55.
- James W (1902) *The Varieties of Religious Experience: A Study in Human Nature*. New York: Random House.
- Johnson MW, Garcia-Romeu A, Cosimano MP, et al. (2014) Pilot study of the 5-HT_{2A}R agonist psilocybin in the treatment of tobacco addiction. *J Psychopharmacol* 28: 983–992.
- MacLean KA, Johnson MW and Griffiths RR (2011) Mystical experiences occasioned by the hallucinogen psilocybin lead to increases in the personality domain of openness. *J Psychopharmacol* 25: 1453–1461.
- MacLean KA, Leoutsakos JM, Johnson MW, et al. (2012) Factor analysis of the Mystical Experience Questionnaire: a study of experiences occasioned by the hallucinogen psilocybin. *J Scient Study Religion* 51: 721–737.
- Macmillan NA and Creelman CD (2005) *Detection Theory: A User's Guide*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum.
- Pahnke WN (1963) *Drugs and Mysticism: An Analysis of the Relationship Between Psychedelic Drugs and the Mystical Consciousness*. Cambridge, MA: Harvard University Press.
- Pahnke WN (1967) LSD and religious experience. In: DeBold RC and Leaf RC (eds) *LSD Man & Society*. Middletown, CT: Wesleyan University Press, pp.60–85.
- R Core Team (2013) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Richards WA (1975) *Counseling, Peak Experiences and the Human Encounter with Death: An Empirical Study of the Efficacy of DPT-Assistance Counseling in Enhancing Quality of Life of Persons with Terminal Cancer and Their Closest Family Members*. Los Angeles, CA: UCLA.
- Richards WA, Rhead JC, Dileo FB, et al. (1977) The peak experience variable in DPT-assisted psychotherapy with cancer patients. *J Psyched Drugs* 9: 1–10.
- Rosseel Y (2012) lavaan: an R Package for structural equation modeling. *J Statist Softw* 48(2): 1–36.
- Satorra A and Bentler PM (2001) A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 66: 507–514.
- Stace WT (1960) *Mysticism and Philosophy*. New York: MacMillan Press.
- Strassman RJ, Qualls CR, Uhlenhuth EH, et al. (1994) Dose-response study of N,N-dimethyltryptamine in humans. II. Subjective effects and preliminary results of a new rating scale. *Arch Gen Psych* 51: 98–108.