

Proposed Battery of Screening, Outcome, Experience, and Safety Measures

New Mexico Medical Psilocybin Program | Discussion Draft

1. Selection Principles

Parsimony and burden: Collect the minimum information necessary. Avoid unnecessary duplication and favor shorter measures when they provide adequate information.

Psychometric quality and longitudinal utility: Favor reliable and valid measures and, for outcomes, measures capable of detecting meaningful change over time. Where possible, use the same measure at baseline and follow-up.

Feasibility and accessibility: Consider training requirements, cost, licensing, and availability.

Stepped assessment: Use a core set of brief measures, with more detailed extended measures when initial responses indicate they may be needed.

2. Psychological Screening and Clinical Assessment

Screening data could be particularly valuable to the research program. Clinical trials have generally studied highly selected populations, often excluding people with comorbid conditions or other potential risk factors. New Mexico's program will provide an opportunity to learn from a much broader real-world population, including people who have not been well represented in psychedelic research. Sharing screening data would allow outcomes and adverse experiences to be understood in the context of clients' characteristics and comorbidities, building an evidence base that can ultimately be shared back with practitioners to inform their work. If practitioners also share relevant information from their intake forms, demographic information already collected at intake would not need to be requested from clients again.

The aim is a learning system in which practice informs research and research returns useful information to practice.

For each domain, measures raised or discussed by the workgroup were compared using the selection criteria above. Where multiple measures were under consideration, they are presented in a brief comparison table followed by a recommendation and rationale. As a due diligence step, a limited review also identified established alternatives that could materially affect the choice of measure. These additional candidates are included in the relevant tables and marked with a dagger (†).

Depression

Measure	Format	Items / Time	Access / Training	Main advantage	Main limitation
PHQ-9	Self-report	9 / 2 to 3 min	Widely accessible; no specialized training	Brief, validated, familiar, repeatable	Less direct comparability with some psilocybin trials
MADRS	Clinician-rated	10 / 15 to 20 min	Trained rater	Strong research and trial comparability	Greater burden
QIDS-SR16	Self-report	16 / about 5 min	Verify use conditions	Strong depression measure	Longer without clear routine advantage over PHQ-9
GRID-HAMD	Clinician-rated	17 to 21 / 20 to 30 min	Training required	Detailed clinician-rated assessment	High burden for routine program use

Favored measure: PHQ-9. It provides the strongest overall combination of psychometric quality, brevity, familiarity, ease of administration, and longitudinal utility. MADRS remains the principal alternative when direct comparability with psilocybin depression trials is a high priority. MADRS and GRID-HAMD provide more detailed clinician-rated assessments at substantially greater burden.

Anxiety

Measure	Format	Items / Time	Access / Training	Main advantage	Main limitation
OASIS	Self-report	5 / about 1 min	Brief; verify use conditions	Broad measure of anxiety severity, avoidance, and functional interference across anxiety presentations	Less familiar than GAD-7
GAD-7	Self-report	7 / 2 to 3 min	Widely accessible; no specialized training	Highly familiar and comparable across settings	More focused on generalized anxiety symptoms
STAI	Self-report	40 / about 20 min	Licensing conditions apply	Separates state and trait anxiety	Too burdensome unless state versus trait anxiety is a specific research question

Favored measure: OASIS, with GAD-7 as a strong alternative. OASIS is especially attractive for a heterogeneous statewide population because it assesses anxiety across diagnoses and presentations rather than focusing on one specific disorder. GAD-7 offers greater familiarity and comparability across settings.

PTSD

Measure	Format	Items / Time	Access / Training	Main advantage	Main limitation
PCL-5	Self-report	20 / 5 to 7 min	Widely available	Established, repeatable PTSD symptom measure	Not a full diagnostic interview
CAPS-5	Clinician interview	45 to 60 min	Training required	Comprehensive diagnostic assessment	High burden for routine use

Favored measure: PCL-5 for routine baseline and outcome measurement. CAPS-5 provides a more comprehensive clinician-administered diagnostic assessment, but at substantially greater administration and training burden.

Wellbeing

Measure	Format	Items / Time	Access / Training	Main advantage	Main limitation
WHO-5	Self-report	5 / 1 to 2 min	Widely accessible	Brief positive wellbeing measure; useful across conditions including end-of-life care	Does not provide a detailed multidimensional wellbeing profile
SWLS	Self-report	5 / about 1 min	Verify use conditions	Measures cognitive evaluation of life satisfaction	Life satisfaction may not track near-term wellbeing, particularly in end-of-life contexts

Favored measure: WHO-5. It captures recent positive wellbeing rather than asking clients to evaluate their life as a whole. This distinction is especially important in end-of-life care, where wellbeing may improve even

when a person's overall life circumstances remain profoundly difficult. SWLS measures the related but distinct construct of overall life satisfaction rather than providing a more detailed version of WHO-5.

Substance Use Disorder

Measure	Primary role	Items / time	Main advantage	Main limitation
AUDIT	Alcohol screening	10 / 3 to 5 min	Established alcohol screen already in the workgroup battery	Alcohol only; longer than AUDIT-C and does not cover other substances
DAST-10	Drug problem screening	10 / 2 to 3 min	Brief screen for non-alcohol drug problems already in the battery	Separates alcohol from other drugs and provides limited substance-specific detail
TLFB	Detailed consumption assessment	Interview / 15 to 30 min	Detailed frequency, quantity, and pattern of use over a defined period	High administration burden for routine statewide use
TAPS†	Broad substance screening	Brief, staged self-report/interview	Covers tobacco, alcohol, prescription medication, and other substances in one screen	Designed primarily for screening rather than longitudinal treatment outcomes
AUDIT-C†	Alcohol screening	3 / about 1 min	Adds strong alcohol-specific screening with very little burden	Alcohol only and not a comprehensive SUD assessment
BAM†	Treatment baseline and outcome monitoring	17 / about 5 min	Designed for repeated SUD progress monitoring; covers use, risk, and protective recovery factors	Not intended for case-finding or diagnosis at program entry

Favored measures by stage: TAPS plus AUDIT-C at diagnostic entry, and the Brief Addiction Monitor (BAM) at treatment baseline and follow-up once a client is accepted into the SUD pathway.

TAPS plus AUDIT-C are screening measures that identify potentially problematic substance use relevant to the qualifying diagnosis. BAM is designed for longitudinal SUD progress monitoring and captures substance use together with risk and protective recovery factors.

TLFB provides more detailed frequency and quantity data, but at substantially greater administration burden.

Suicidality

Measure	Format / burden	Main advantage	Main limitation
C-SSRS	Structured assessment; length varies by version and responses	Assesses ideation and behavior and is designed for safety assessment and triage	Requires training; administration burden varies by version and responses
SIDAS†	5-item self-report	Very brief measure of suicidal ideation severity with strong psychometric support	Measures ideation, not the full history of suicidal behavior or a complete safety assessment
Beck Scale for Suicide Ideation (BSS)†	19-item self-report or clinician-rated versions	Detailed and well-established assessment of suicidal ideation severity	Longer than SIDAS; access conditions should be verified

Favored measure: C-SSRS. It assesses both suicidal ideation and suicidal behavior within a structured safety framework. SIDAS is a very brief measure of ideation severity, while the Beck Scale for Suicide Ideation provides a more detailed assessment of suicidal ideation. Neither provides the same combination of ideation and behavior assessment as the C-SSRS.

Dissociation

Measure	Format / burden	Primary role	Main advantage	Main limitation
CDS-2	2 items / less than 1 min	Initial screening	Extremely brief screen for clinically significant depersonalization and derealization	Narrower than broad dissociation measures

Measure	Format / burden	Primary role	Main advantage	Main limitation
CDS-29	29 items / about 12 to 15 min	Detailed follow-up	Provides detailed assessment of depersonalization and derealization when CDS-2 flags concern	Greater burden than CDS-2
DSS-B	8 items / about 2 to 3 min	Post-treatment assessment when indicated	Brief, validated measure covering depersonalization, derealization, gaps in awareness/memory, and dissociative re-experiencing	Measures a broader dissociation construct than CDS-2/CDS-29
DES-II†	28-item self-report / about 5 to 10 min	Broad dissociation screening	Very well-established broad measure of dissociative experiences	Longer and broader in scope than the more targeted measures under consideration

Favored stepped measurement approach: CDS-2 as the initial brief screen for depersonalization and derealization, with CDS-29 providing a more detailed assessment of the same symptom domain if the CDS-2 indicated a propensity. DSS-B offers a brief broader measure of clinically significant dissociative symptoms and is particularly relevant to post-treatment assessment. DES-II is an established broad dissociation measure added during due diligence, but its 28-item burden and broader trait-like scope offer less advantage for this specific measurement framework. Dissociation is distinct from psychosis because depersonalization and derealization can occur with preserved reality testing.

Bipolar Spectrum

Measure	Format / burden	Main advantage	Main limitation
MDQ	13 symptom items plus supplementary questions / about 5 min	Brief, familiar, and generally more specific; already in workgroup battery	May be less sensitive to bipolar II and hypomanic presentations
HCL-32†	32-item self-report / about 5 to 10 min	High sensitivity to hypomanic symptoms and evidence of stronger detection of bipolar II in some settings	Longer and generally less specific, creating more false positives
BSDS†	19-item narrative checklist plus global rating / about 5 to 10 min	Designed to capture a broad bipolar spectrum, including softer presentations	Less brief than MDQ; broader spectrum approach may increase positive screens

The HCL-32 and MDQ present a meaningful tradeoff. In practical terms, the HCL-32 favors not letting a potential bipolar-spectrum condition slip through screening, while the MDQ favors not stopping or delaying people unnecessarily because of a false alarm.

Psychosis Risk

Measure	Format / burden	Main advantage	Main limitation
PQ-16	16-item self-report / about 3 min	Very brief first-stage psychosis-risk screen already in workgroup battery	Like other psychosis-risk screens, can generate substantial false positives
PQ-B†	21-item self-report with distress ratings / about 5 min	Well-studied brief prodromal screener; distress ratings add clinically useful context	Longer than PQ-16
PRIME Screen-Revised†	12-item self-report / about 5 min	Very brief established screen developed for attenuated psychotic symptoms	Evidence does not clearly establish a decisive advantage over PQ measures; false positives remain a concern

Favored measure: PQ-16. The due diligence review confirms that PQ-B and PRIME Screen-Revised are established alternatives, but none eliminates the central limitation of brief psychosis-risk screening: false positives can be common, particularly when the underlying prevalence of true high-risk states is low. PQ-16 retains an advantage in burden and is already in the proposed battery.

Borderline Personality Features

Measure	Format / burden	Main advantage	Main limitation
MSI-BPD	10-item self-report / about 2 min	Very brief, well-studied BPD-specific screener and already in workgroup battery	Screening accuracy varies by population
BPQ†	80-item true/false self-report / about 15 to 20 min	Detailed coverage of all nine borderline personality criterion domains and strong psychometric performance in some studies	Substantially greater burden than MSI-BPD

Favored measure: MSI-BPD. The due diligence review identified the Borderline Personality Questionnaire (BPQ) as an established alternative with broader criterion-level coverage, but at 80 items it imposes substantially greater burden. MSI-BPD therefore remains the more parsimonious measure of borderline personality features.

3. Clinical Outcome Measures

Outcome domain	Favored measure / approach	Rationale
Depression	PHQ-9	Brief, validated, repeatable, and practical for routine longitudinal use.
Anxiety	OASIS	Very brief and applicable across different anxiety presentations; GAD-7 remains a strong alternative for familiarity and comparability.
PTSD	PCL-5	Established self-report measure suitable for repeated symptom assessment.
Wellbeing	WHO-5	Adds a positive wellbeing outcome so success is not defined only as reduction in pathology.
Substance use	BAM	Designed for repeated SUD progress monitoring; measures substance use together with risk and protective recovery factors.

4. Psychedelic Experience, Integration, and Safety Outcomes

Psychedelic Experience and Integration

Measure	Format / burden	Primary role	Main advantage	Main limitation
MEQ-4 + CEQ-7	11 items	Acute psychedelic experience	Very low burden while capturing meaningful/mystical and challenging experience	Less detail than the full forms
MEQ-30 + CEQ-26	56 items	Acute psychedelic experience	Fuller characterization of both meaningful/mystical and challenging experience	Higher burden
PIS-6	6 items, ~2-3 min	Psychedelic insight	Brief measure of insight arising from the experience	Adds a separate construct and measure

IES	12 items, ~4-6 min	Integration behavior	Captures activities used to integrate the experience	Adds burden beyond experience and clinical outcome measures
EIS	12 items, ~4 min	Experienced integration	Captures the person's sense that the experience has been integrated	Some overlap with perceived benefit and broader outcomes

The long-form MEQ-30 and CEQ-26 are favored for fuller characterization of the psychedelic experience. The additional value of PIS-6, IES, and EIS should be considered against the added burden.

Adverse Effects and Post-Treatment Challenges

The single challenge/harm item and CDS-2 provide brief first-stage screening, with PATAE and DSS-B available as extended measures when indicated.

Component	Proposed role
SPSI	32 items
Single challenge/harm item	1 item
PATAE	Extended assessment
CDS-2	2 items
DSS-B	8 items

The final adverse-event reporting architecture may ultimately be developed in collaboration with Palitsky and colleagues. This section therefore reports the proposed measurement components and their respective roles rather than specifying the final reporting mechanism.

5. Medical and Medication Safety Screening

Source Frameworks

The medical screening domains summarized here are drawn from three established psychedelic-treatment frameworks: MAPS protocols for MDMA-assisted therapy, COMPASS Pathways clinical protocols for COMP360 psilocybin therapy, and the Australian psychedelic-assisted therapy framework and training materials used by Mind Medicine Australia (MMA).

These sources provide complementary perspectives. COMPASS offers the most directly applicable psilocybin-specific clinical-trial precedent. MAPS provides a mature and extensively developed model of medical and medication safety screening in psychedelic-assisted therapy. MMA provides an important example of translating psychedelic-assisted therapy from research settings toward regulated clinical practice.

The comparison highlights areas of convergence across these frameworks while recognizing that New Mexico is developing a real-world clinical program rather than a clinical trial. Research exclusion criteria identify medical risk domains but are not equivalent to statewide clinical criteria. MAPS is an MDMA source, while COMPASS provides psilocybin-specific precedent, an important distinction where substance-specific medical risks differ.

Medical screening domain	Source precedent	How represented in source frameworks
Medical history and current conditions	MAPS, COMPASS, MMA	Included in routine medical history.
Current medications and supplements	MAPS, COMPASS, MMA	Included in medication and interaction review.

Medical screening domain	Source precedent	How represented in source frameworks
Cardiovascular and cerebrovascular history	MAPS, COMPASS, MMA	Included in medical history, with additional evaluation in some protocols.
Blood pressure and pulse	MAPS, COMPASS, MMA	Vital-sign assessment is included before dosing.
Neurological and seizure history	MAPS, COMPASS	Neurological and seizure history is included, with additional assessment in some protocols.
Hepatic and renal disease	MAPS, COMPASS	Included through history and, in some protocols, laboratory testing.
Pregnancy and reproductive considerations	MAPS, COMPASS, MMA	Included in screening frameworks where applicable.
ECG	MAPS, COMPASS	ECG requirements vary by protocol and risk profile.
Laboratory testing	MAPS, COMPASS	Laboratory testing requirements vary by protocol and risk profile.
Substance use and intoxication	MAPS, COMPASS	Substance use and intoxication are represented as safety domains.
Medication changes or tapering	MAPS, COMPASS	Medication restrictions and tapering requirements vary across protocols.
Further medical evaluation	MAPS, COMPASS, MMA	Source frameworks include additional medical evaluation for selected findings.

This section is intended to establish the provenance and major domains represented in established medical screening frameworks. Exact New Mexico screening requirements, thresholds, tests, and reporting pathways remain outside the scope of this measure-comparison document.

6. Evaluation and Accountability Items

The Participant-Centered Reporting Framework includes information collected across preparation, administration, integration, and follow-up. Information is intended to be collected by the person best positioned to provide it and, wherever possible, collected once and reused rather than requested repeatedly.

Practitioner-Reported Items

Preparation

- Goals and intentions
- Preferences and boundaries
- Cultural and spiritual considerations
- Support planning
- Readiness for treatment
- Hopes, concerns, and expectations

Administration Session

- Accommodations provided
- Support interventions
- Safety concerns
- Notable events
- Cultural or spiritual elements incorporated into treatment, such as music, language, ceremonial elements, spiritual supports, or environmental modifications

Integration

- Integration themes
- Meaning-making
- Challenging experiences
- Significant life changes
- Emerging concerns
- Referrals and ongoing support needs

Participant Feedback

The adopted framework identifies four brief client-reported questions:

1. Did you feel respected, understood, and supported throughout the process?
2. Were your cultural, spiritual, accessibility, or personal needs appropriately considered?
3. Were there any aspects of your experience that you found particularly meaningful, challenging, or concerning?
4. Is there anything else you would like us to know about your experience?

These items provide a brief source of participant feedback alongside practitioner documentation and standardized outcome and safety measures. The framework deliberately limits client reporting to information that can most meaningfully be provided by the client.

These are proposed reporting domains rather than finalized questions. Specific wording, response options, timing, and burden should be trialed with participants and practitioners before the program commits to them.