Validation of the Manuals and Inventories ��� Practical Advice for Future Researchers

Validation of the Manuals and Inventories — Practical Advice for Future Researchers

From The Mars Colonist: Training for the Symbiocene *— Book Three*

Why This Section Exists

Everything in the manuals and the complete instruments reproduced in Book Three is content-complete but not yet formally validated. That distinction matters, and this section exists to keep it visible rather than let a polished manuscript quietly imply more certainty than the evidence currently supports.

What follows is not a claim of validation. It is the working paper trail: a protocol for validating the training manuals themselves, a proposed convergent-validity battery for the psychometric instruments, a completed copyright audit of that battery's anchor scales, a live research proposal to a genuine analogue-isolation research programme, and an honest statement of what statistical groundwork has and has not yet been done. It is offered here in full — questions, open flags, and unresolved decisions included — because a future researcher picking this work up needs the actual state of things, not a tidied summary of it.

Protocol A: Validating the Training Manuals

Content and Criterion Validation of the Mars/Moon Readiness Training Manuals (The Aspirant's Manual and The Trainer's Manual). Version 1.0 — draft for statistical review. Prepared by Dinu-Stefan Teodorescu, Ph.D., in collaboration with Claude. Prepared for the review of Tore Wentzel-Larsen, Statistician. Draft date: 6 August 2026.

Purpose of This Document

This protocol proposes a two-phase validation study for the two Mars/Moon Readiness training manuals developed alongside *The Human Imperative* and the Mars Colonist Symbiocene Inventory (MCSI): The Aspirant's Manual (for individual trainees preparing for long-duration extreme-environment missions) and The Trainer's Manual (for facilitators delivering the programme). The manuals are content-complete but have not yet been formally validated. This document is a draft for Tore's review — it proposes a design and a statistical approach, and asks him three specific questions (Section 9) where his judgement should shape the final protocol before any data collection begins.

A secondary purpose of validation, beyond scientific rigour, is credibility: should discussions with a commercial or institutional partner (e.g. a space-industry training programme) become concrete, a manual with documented content and outcome validity is a fundamentally different proposition than an unvalidated one.

1. Background and Rationale

The two manuals translate the book's Nine Archetypes framework and the constructs measured by the MCSI into a structured training curriculum: psychological, contemplative, and relational preparation for the isolation, confinement, and extreme conditions of long-duration Moon or Mars missions. The MCSI itself is undergoing formal psychometric validation (WLSMV-primary, MLR-sensitivity CFA, dual omega reliability) in parallel work with Tore. The training manuals are a distinct object of validation: the question is not whether the

MCSI measures the intended construct, but whether completing the training programme the manuals describe actually moves trainees toward the readiness the programme claims to build.

Validation evidence for a training programme conventionally rests on two separate pillars, and this protocol treats them as two sequential phases:

- **Content validity** — do subject-matter experts agree that the manuals' modules and competencies are the right ones, and are they described clearly and completely enough to be delivered as intended?
- **Criterion (outcome) validity** — do people who complete the training show a measurable, attributable change on relevant readiness indicators, ideally corroborated by behaviour in an analogue extreme environment?

This mirrors the standard applied framework for training evaluation (Kirkpatrick's four levels: Reaction, Learning, Behavior, Results). Phase 1 below addresses the first two levels; Phase 2 addresses the latter two.

2. Objectives

- **Primary:** Establish the content validity of the competencies and modules in both manuals through structured expert panel review.
- **Secondary:** Establish preliminary criterion validity through a pre/post pilot study, using the MCSI as the primary outcome measure.
- **Exploratory:** Assess the feasibility of a field validation in a genuine space-analogue environment — Concordia Station (Antarctica) is the aspirational target, given its established use by ESA for exactly this category of psychological and behavioural research; realistic contingency analogues are proposed in Section 5.4 in case a Concordia partnership cannot be secured on the relevant timeline.

3. Study Design Overview

Phase	Design	Question answered	Output
1	Cross-sectional expert panel review	Is the manual content the right content, clearly described?	Content Validity Index (I-CVI, SCVI), modified Kappa, qualitative revisions
2a	Single-arm pre/post pilot	Does completing training move MCSI scores in the intended direction?	Effect size estimate, feasibility data for Phase 2b
2b	Quasi-experimental, comparison cohort where feasible; analogue-environment field pilot if access is secured	Is the change attributable to the training, and does it correspond to real-world behaviour under isolation/confinement?	Preliminary criterion validity evidence

4. Phase 1 — Content Validity

4.1 Expert Panel Composition. A target panel of 6–10 experts, deliberately spanning the domains the manuals draw on, so that no single disciplinary blind spot goes unchecked: clinical or contemplative psychology (construct fidelity to the Nine Archetypes framework); aerospace human factors / astronaut selection and training design; polar or extreme-environment behavioural research (ideally someone with direct Concordia, Antarctic, or comparable analogue experience); adult education / instructional design (is the material deliverable as written?); and a statistician for the scoring methodology itself (Tore, in an advisory rather than rated-item capacity).

4.2 Materials to Be Rated. Each discrete competency, module, and practice exercise in both manuals is listed as a separate item for rating (the current Trainer's Manual competency list and Aspirant's Manual module list will populate Appendix A once finalised — this protocol proposes the rating method rather than the item list itself, since the manuals' own structure should determine the item count).

4.3 Rating Procedure. Each expert rates each item independently, before any group discussion, on a 4-point relevance scale — the standard approach for content validity indexing:

Rating	Meaning
1	Not relevant to Mars/Moon readiness as currently written
2	Item needs major revision to be relevant
3	Relevant, but needs minor revision for clarity or fit
4	Highly relevant and clearly written as-is

An item's I-CVI (Item-level Content Validity Index) is the proportion of experts rating it 3 or 4. The scale-level S-CVI/Ave is the mean of all item I-CVIs. Each expert also provides open-ended comments per item, which matter as much as the numeric score for a manual this dependent on tone and delivery.

4.4 Decision Rules.

I-CVI	Decision
≥ .78	Retain as written, incorporate minor qualitative feedback where offered
.70–.77	Borderline — revise based on panel comments, re-circulate to panel for a second-round rating
< .70	Substantially revise or remove; do not retain without a second validation round

(The .78 threshold for panels of 6–10 experts follows Lynn's widely used criterion; Tore's view on whether this threshold is appropriate for this panel size and content type is one of the three questions in Section 9.)

4.5 Inter-Rater Agreement. Because I-CVI alone does not correct for the possibility of high agreement by chance on a small 4-point scale, a modified kappa statistic (correcting for chance agreement, as used in comparable training-programme validation studies) will be calculated alongside the I-CVI for each item, not as a replacement for it.

5. Phase 2 — Criterion (Outcome) Validity Pilot

5.1 Design Logic. Phase 2 is proposed as two steps rather than one, because access to a genuine analogue environment (Concordia or equivalent) cannot be guaranteed on the timeline of this protocol, and a feasibility step should not wait on a partnership that may take a year or more to secure.

- **Phase 2a** (feasibility, no analogue access required): a single-arm pre/post pilot with a small volunteer cohort — not necessarily astronaut candidates, but individuals with a genuine, motivated interest in extreme-environment or long-duration mission preparation.
- **Phase 2b** (field validation, contingent on partnership): if a genuine analogue placement is secured — Concordia Station is the aspirational target — trainees who completed the manual-based programme are followed into the analogue environment itself, with MCSI administered pre-training, post-training, and at intervals during the isolation period.

5.2 Primary and Secondary Outcome Measures. Primary: MCSI total and subscale scores, pre- and post-training. Secondary (Phase 2b only, where available): crew cohesion ratings, supervisor or mission-control performance assessments, sleep and mood tracking already standard at analogue stations, and — if ethically and logistically feasible — the ESI as a complementary measure of ecological/relational readiness.

5.3 Target Population and Sampling. Aspirational: Concordia Station winter-over crew or ESA-affiliated analogue-mission candidates, contingent on a formal research partnership with IPEV/PNRA/ESA. Realistic contingency analogues: other established Mars-analogue facilities with more accessible research-partnership pathways (e.g. the Mars Desert Research Station in Utah, or comparable Austrian Space Forum analogue missions). Minimum viable Phase 2a sample: a non-analogue volunteer cohort, purely to establish whether the training produces any measurable pre/post effect before investing in analogue access negotiations.

5.4 Analogue Partnership Contingency. Because Concordia access is not guaranteed, this protocol treats Phase 2b as contingent and sequenced after Phase 2a, not as a precondition for beginning validation work. Outreach to IPEV/PNRA/ESA for a formal research collaboration can proceed in parallel with Phase 1 and Phase 2a, on its own institutional timeline.

5.5 Sample Size. Analogue-environment research is characteristically small-N (Concordia winter-over crews are typically 12–15 people total). This protocol does not propose a formal power analysis at this draft stage — that is the second question posed to Tore in Section 9 — but flags explicitly that Phase 2 should be understood and reported as a feasibility/pilot study, not a definitively powered trial, unless and until sample size can be meaningfully increased across multiple analogue rotations or cohorts.

6. Statistical Analysis Plan

- **Phase 1:** I-CVI and S-CVI/Ave calculated per item and overall; modified kappa for chance-corrected agreement; qualitative comments coded thematically per module.
- **Phase 2a/2b:** paired pre/post comparison of MCSI scores (exact test — paired t-test vs. Wilcoxon signed-rank — pending distributional assessment and Tore's recommendation); effect size (Cohen's d or a matched non-parametric equivalent); reliability of the pre/post administration itself checked via the MCSI's existing dual-omega reliability framework.
- **Where a comparison cohort is feasible in Phase 2b:** between-group comparison of change scores, with the explicit caveat that group assignment is not randomised and any between-group claim is correlational, not causal.

7. Ethical Considerations

Standard informed consent for all participants at every phase; voluntary participation with no penalty for withdrawal. Formal ethics committee review required before Phase 2 data collection begins — REK (Norway) for any Norway-based cohort, and the relevant partner institution's own ethics board (e.g. ESA/IPEV/PNRA) for any Concordia-affiliated cohort. Cross-border data handling: a Concordia-affiliated cohort would involve multiple nationalities and institutions; data protection agreements will need to be settled as part of any formal partnership discussion. Transparency: any resulting publication or presentation of this validation work should disclose the AI-collaborative origin of the manuals and the book they accompany.

8. Limitations

Small sample sizes are structurally inherent to analogue-environment research and cannot be fully resolved within this protocol's scope. Self-selection: volunteers for analogue missions are very unlikely to be representative of the general population — any effect found may not generalise beyond similarly self-selected groups. Phase 2b is contingent on a partnership that may not materialise on any predictable timeline; this protocol is designed so that Phase 1 and Phase 2a produce genuinely useful validation evidence on their own, independent of whether Phase 2b ever proceeds.

9. Questions for Tore

Question 1 — Content validity threshold and method: Is the I-CVI ≥ .78 / S-CVI ≥ .90 combination (Lynn's criterion) appropriate for a 6–10 member expert panel rating this kind of experiential/practice-based content, or would an alternative approach (e.g. Aiken's V) be more defensible here?

Question 2 — Phase 2 statistical test and power: Given the near-certainty of a small sample (single digits to low tens) in any real pre/post pilot, what would you recommend as the primary test — paired t-test, Wilcoxon signed-rank, or a different approach entirely — and is a meaningful minimum sample size specifiable in advance, or is this genuinely an exploratory-only design until sample size grows across cohorts?

Question 3 — Comparison group strategy: Given that a randomised or even a well-matched comparison group is unlikely to be feasible in Phase 2b, what is the most defensible fallback — a historical/normative benchmark from existing MCSI data, a matched cohort from

a different analogue site, or reporting Phase 2 findings as purely within-subject with the causal limitation stated plainly?

10. Proposed Timeline

Stage	Dependent on	Approximate timing
Tore's review of this protocol	This document	Immediate
Phase 1: expert panel assembled and item list finalised	Tore's Q1 answer; manual competency list finalised	Following protocol approval
Phase 1: rating and CVI analysis	Panel assembled	4–8 weeks after panel starts
Phase 2a: feasibility pilot cohort recruited	Revised manual from Phase 1; Tore's Q2 answer	Parallel with or following Phase 1
IPEV/PNRA/ESA outreach for Concordia (or contingency analogue outreach)	None — can begin immediately, in parallel	Ongoing, own institutional timeline
Phase 2b: field validation	Successful partnership; Tore's Q3 answer	Contingent, not scheduled

---

Protocol B: A Proposed Test Battery for Validating the MCSI/ESI Subscales

Developed in collaboration with DeepSeek. A self-report inventory that only correlates with itself proves nothing about construct validity — external anchors are needed: established, peer-reviewed instruments that should correlate with specific MCSI/ESI subscales and with the Arcadia Index's stage placement.

The Principle

Each MCSI/ESI subscale makes an implicit claim: that it measures something real, and that something should overlap with what established science already measures under a different name. A convergent-validity pilot tests that overlap — convergent validity (correlates with related constructs) and discriminant validity (doesn't just collapse into personality traits already known, like Big Five). Twelve subscales plus the Arcadia Index is too many constructs to anchor individually without making the battery unbearably long for a real pilot with limited crew time — so the design question is which constructs are load-bearing enough to justify the respondent burden, and which can wait for a later, larger validation study.

Proposed Core Battery (Lean, Defensible Version)

MCSI/ESI Construct	Anchor Instrument	Items	Why This One
The Witnessing Self + Contemplative Depth	Experiences Questionnaire (Fresco et al.) — decentering subscale	11	Decentering is the closest peer-reviewed construct to "witnessing awareness"
Mortality Integration	Death Attitude Profile–Revised (Wong et al.) — Neutral Acceptance subscale	5 (of 32)	Distinguishes genuine acceptance from avoidance or fear-driven approach
Ecological Belonging	Nature Relatedness Scale, short form (NR-6)	6	Most widely cited, brief, strong psychometrics
Relational Depth	Experiences in Close Relationships–Short Form (ECR-S)	12	Attachment security is the best-established proxy for honest, repairable relationship under pressure
Free Will/Destiny Paradox	FAD-Plus (Paulhus & Carey) — Free Will + Fatalistic Determinism subscales	14 (of 27)	Only well-validated instrument separating agency belief from fatalism

AI Partnership Capacity	GAAIS (General Attitudes to AI Scale)	12	Nearest validated construct, and the newest/least-precedented subscale
Arcadia Index (stage placement)	Washington University Sentence Completion Test (WUSCT), short form, or Cook-Greuter Leadership Maturity Profile	~18 sentence stems	The actual gold-standard adult-development instrument
Discriminant control	Big Five Inventory–2 Short Form (BFI-2-S)	30	Shows MCSI/ESI isn't just repackaging Openness/Conscientiousness
Cross-check on the built-in lie scale	Marlowe-Crowne Social Desirability Scale, short form	13	External validation of the instrument's own Social Desirability subscale accuracy

That's roughly 120 additional items, comparable in length to the MCSI itself. Realistic total pilot session: MCSI/ESI (132) + Arcadia (14) + this battery (~120) ≈ 265 items — long, but doable as a single sitting for a paid analog-mission pilot population, especially split across two sessions.

What to Leave Out, Deliberately

Grief Tolerance, Embodied Presence, Isolation Resilience, and Integral Altitude all have plausible anchors (Distress Tolerance Scale, MAIA-2, UCLA Loneliness Scale, Defining Issues Test), but adding all of them pushes the battery past what a pilot can justify. These are held for a full validation study post-pilot, not an 8–12 week logistics test.

A Note on Sequencing

Right now the pilot as initially scoped tests logistics, not validity. Adding this convergent-validity battery is what turns a feasibility/operations trial into a genuine validation pilot.

Copyright and Licensing Audit

Before committing to any external anchor instrument, each one was checked individually for licensing status rather than assumed free to use — "everyone cites it in papers" and "free to use" are not the same thing.

Instrument	Status	What This Means
NR-6 (Nature Relatedness, short)	Free — CC BY licensed	Published open-access in Frontiers. Use freely, cite the paper.
BFI-2-S (Big Five, short)	Free for research	Soto & John explicitly grant permission for personal and research use.
Marlowe-Crowne SDS, short form	Free — explicitly public domain	The scale's own distributor states in writing that no permission is required.
GAAIS (General Attitudes to AI)	Free — CC licensed, open access	Found during the audit; also a better construct fit than the originally proposed alternative.
FAD-Plus (Free Will/Determinism)	Technically copyrighted, low friction	Standard journal copyright (Taylor & Francis), but freely reused worldwide. A courtesy email before use is still advisable.
Experiences Questionnaire (decentering)	Permission requested as standard practice	The instrument's official listing asks users to contact the author for a copy.
DAP-R (Death Attitude Profile–Revised)	Requires permission	Explicitly copyrighted (© Wong & Reker, 1994).

ECR-S (attachment, short form)	Requires permission	Explicitly stated as copyrighted; permission must be obtained through Taylor & Francis.
Trust in Automation Scale (Jian et al.)	Not publicly available	Items are not publicly posted and require direct author contact. Not usable as originally proposed.
WUSCT / Cook-Greuter Leadership Maturity Profile	Commercial, paid, licensed scoring	Loevinger's original requires purchasing the manual and trained raters; Cook-Greuter's version is an outright commercial product.

Resulting change to the battery: the originally proposed Trust in Automation Scale is swapped for GAAIS, which is free, open, and arguably the better construct match. The WUSCT/Cook-Greuter developmental anchor is treated as a separate line-item cost-and-vendor decision. Everything else in the core battery — NR-6, BFI-2-S, Marlowe-Crowne — is ready to use today, no permission required. FAD-Plus and the Experiences Questionnaire each need one short, low-friction email to their respective authors before use. DAP-R and ECR-S require formal written permission from the copyright holders before any data collection using them can proceed.

The Concordia Station Field Proposal

Research proposal submitted in response to the ESA / IPEV / PNRA Announcement of Opportunity for Human Research at Concordia Station. Reproduced here as submitted.

Dinu-Stefan Teodorescu, Ph.D. — Clinical Psychologist & Transpersonal Therapist (Ret.), Skreia, Norway · din.nu

Title: Validating the Symbiocene Inventories — A Self-Report Instrument Battery for Psychological Readiness in Long-Duration Isolation and Confinement

Principal Investigator: Dinu-Stefan Teodorescu, Ph.D., Skreia, Norway

1. Background and Rationale. Analog environments such as Concordia Station have been central to ESA's isolation and confinement research for two decades, generating 36 biomedical and psychological research projects to date. Existing work at Concordia has largely measured outcomes — mood, cognitive performance, team conflict, physiological stress markers. This proposal offers a complementary measure: self-report instruments assessing the underlying individual psychological capacities that plausibly predict those outcomes, administered before and across a winter-over deployment.

2. Objectives. (a) To administer the Mars Colonist Symbiocene Inventory (MCSI), a 142-item, thirteen-subscale self-report instrument, to Concordia crew members at pre-departure and at intervals across the winter-over period. (b) To administer the Arcadia Index, a fourteen-vignette forced-choice narrative instrument assessing developmental stage along a distinct axis from the MCSI. (c) To evaluate whether pre-departure subscale scores — particularly Isolation Resilience & Solitude, Relational Depth, and Ecological Belonging — predict trajectory of loneliness, conflict, and cohesion across the mission. (d) To contribute genuine reliability and validity data toward a psychometric instrument intended, longer-term, for use in psychological preparation ahead of crewed lunar and Martian missions.

3. Instruments and Methodology. Both instruments are deployed online, self-administering, and self-scoring, requiring no specialized equipment. The MCSI takes approximately 20–25 minutes to complete; the Arcadia Index approximately 15–20 minutes. A single timepoint requires roughly 35–45 minutes of crew time. The proposed protocol would follow a four-timepoint design (pre-departure, and three points across the winter-over). All response data would be anonymized; no personally identifying information is collected or required.

4. Relevance to ESA's Exploration Goals. ESA has stated that psychological and physiological challenges of long-duration exploration missions will surpass anything previously undertaken. This proposal directly serves that goal: a validated pre-departure psychological readiness instrument would offer crew selection and preparation programs a practical tool, grounded in Concordia-derived data.

5. Resource Requirements. Minimal. No equipment, laboratory space, travel, or on-site researcher presence is required. Crew members would complete each instrument via a web-connected device during already-scheduled research or personal time.

6. Timeline. The Principal Investigator is available to begin coordination immediately and can accommodate whatever crew rotation and review timeline ESA, IPEV, and PNRA judge appropriate.

7. Investigator Background. The Principal Investigator holds a Ph.D. in clinical psychology and has prior peer-reviewed publications in *Cognitive Science*, *BMC Psychology*, the *Scandinavian Journal of Psychology*, *Health and Quality of Life Outcomes*, and the *International Journal of Transpersonal Studies*, including a 2002 study on eye movements during visualization meditation and a 2011 psychometric scale development study. The instruments described in this proposal form part of a larger book-length work on psychological preparation for interplanetary settlement.

Statistical Analysis: Status So Far

No empirical validation data has been collected for any of the three instruments or either manual as of this writing. What exists is design and planning work:

Piece of work	Status
MCSI/ESI validation study (external, general-population)	Analysis plan finalized with statistician Tore Wentzel-Larsen. WLSMV primary (polychoric correlations, ordinal five-point Likert data), MLR sensitivity model, dual-omega reliability. A constrained pseudorandomized item-presentation order has been built with a fixed seed.
Training manual validation (Protocol A)	Design-complete, awaiting statistician review of three open questions (content-validity threshold, Phase 2 statistical test and power, comparison-group strategy). No expert panel yet assembled.
Convergent-validity battery (Protocol B)	Design-complete and copyright-audited. Permission requests to DAP-R and ECR-S authors not yet sent. No pilot data collected.
Concordia proposal	Submitted; response pending. Contingent on a multi-institutional partnership (ESA/IPEV/PNRA).

Practical Advice for Future Researchers

A few things worth stating plainly for anyone — including a future version of this research team — picking this work up:

Separate the content from the claim. Both manuals and all three instruments are, at this stage, well-constructed hypotheses about what psychological readiness for isolation and confinement looks like, not confirmed measures of it. Keep that distinction visible in any future presentation of this work.

Check copyright before you build the battery, not after. The audit in Protocol B caught one instrument that would not have been usable at all, and two that require formal written permission before any data collection. This is cheap insurance early and expensive to discover late.

Small-N is the honest starting condition for analogue research, not a flaw to engineer away. Concordia winter-over crews run 12–15 people total. Protocol A treats its Phase 2 pilot explicitly as feasibility/pilot work, and that framing should carry forward into any future write-up.

Sequence the decisions that are expensive to reverse. Whether the MCSI/ESI move from grouped-and-labelled to interleaved-and-unlabelled item presentation is a choice that must be made once, before validation data collection begins, and held constant throughout.

Let the AI-collaborative origin of this work be stated, not discovered. Both manuals, the instruments, and this validation section itself were developed through sustained collaboration between the Principal Investigator and multiple AI systems (Claude and DeepSeek). That commitment should hold for whatever comes next.

Treat disagreement between the Arcadia Index and the MCSI/ESI as data, not noise. The two instruments measure genuinely different things — capacity pattern versus inferred developmental stage — and a real person can legitimately score differently on each.