



NATIONALT  
CENTER FOR  
KUNST  
OG MENTAL  
SUNDHED

# Framework for the assessment of interventions in the field of Art and Mental Health

**An ICAF guide, v. 1.0**

National Centre for Arts and Mental Health,  
Copenhagen, Denmark, 2025





---

## Table of Contents

1.0 Background and Application .....	2
1.1 Bridging practice and research in art and mental health.....	2
1.2 The Framework's Scope: Defining Grey Literature .....	2
1.3 The Tool's Potential Uses .....	3
1.4 What the Tool Does Not Tell You.....	4
1.5 Possible Extensions with Other Frameworks .....	4
2.0 Introduction to the Framework .....	5
2.1 ICAF as a Comprehensive Picture of Potential .....	5
2.2 How to Score: From Indication to Documentation .....	5
2.3 Review of the Individual Criteria.....	6
2.3.1 Design Criteria .....	6
2.3.2 Scalability Criteria .....	8
2.3.3 Evaluation Criteria .....	9
3.0 Acknowledgements .....	10
4.0 Bibliography .....	10
5.0 Scoring Sheet – Intervention Capability Assessment Framework .....	11

---



# 1.0 Background and Application

## 1.1 Bridging practice and research in art and mental health

While the body of academic evidence on the use of art and culture to promote mental health and well-being is growing, the field is still in its early stages. Much of the knowledge about what works is based on experience and has not been translated into academic knowledge. The practical knowledge of how to design and implement effective interventions—the knowledge of what genuinely works in the real world—is also under-documented in traditional scientific literature.

Instead, this valuable knowledge is scattered throughout what is known as "grey literature", which includes reports, project descriptions, informal evaluations, and educational materials from the organisations and practitioners who work with the target group every day. This situation creates a gulf between academic research and practical application. There is a vast amount of innovative projects, each having accumulated unique, local knowledge, but without bridges between them, it's difficult to share experiences, validate approaches, and build a common, robust knowledge base. Traditional evaluation tools designed to assess randomised, controlled trials are unsuitable for capturing the nuances and context-specific value of these practice-oriented interventions.

The motivation for developing the Intervention Capability Assessment Framework (ICAF) is to create a broader analytical framework to systematically assess and compare the practical knowledge generated and described in reports, evaluations, and notes. By adapting standard criteria for critical appraisal, ICAF serves as a tool to measure an intervention's "maturity" and the degree of systematic approach behind it. The tool can thus function as an indicator of the quality of an intervention's design and documentation, allowing for a meaningful comparison across the varying formats of grey literature. It is intentionally designed to use general information often already found in a project description or final evaluation, making it simple to use. It must be emphasised that ICAF is not inherently normative; it doesn't pass judgement on a project's quality but rather creates transparency and comparability to strengthen the entire field.

## 1.2 The Framework's Scope: Defining Grey Literature

To use ICAF correctly, a clear understanding of "grey literature" is essential. The 2010 Prague Definition defines grey literature as:

"[...] diverse document types produced on all levels of government, academics, business and industry in print and electronic formats [...] that are not controlled by commercial publishers" (Schöpfel, 2011).

This covers a wide range of materials not found in traditional academic databases. Examples include reports and evaluations, government policies and guidelines, white papers, annual reports, newsletters, brochures and patient information, conference presentations and posters, websites, and internal memos. Within research and evaluation, it's useful to distinguish between two main types of grey literature:

1. An extension of academic literature: Material that follows a scientific structure but has not been published in a peer-reviewed journal. This can include theses, dissertations, or



unpublished data from clinical trials. This type of literature is important for getting a complete picture and avoiding publication bias. It can also make promising research results available more quickly.

2. Practice-based knowledge: Material that documents experiences, methods, and results from the real world without necessarily following an academic template. This is the knowledge that ICAF is designed to assess. ICAF does not look at scientific validity in the traditional sense, but rather at the intervention's maturity, systematic approach, and documentation level.

ICAF is a specialised tool for unlocking the value of the practice-based knowledge that is crucial for understanding how art interventions for mental health work in practice.

## 1.3 The Tool's Potential Uses

ICAF is more than a checklist. It's a strategic tool that can strengthen work at multiple levels for practitioners, project managers, funding bodies, and researchers. The tool can be used:

- For creating an overview and comparing (Benchmarking):
  - Challenge: You need to choose between three different cultural programmes for vulnerable young people in your council. Their evaluation reports have widely different formats and focuses.
  - ICAF's contribution: The tool provides a "common language" to assess and compare the maturity of interventions across design, scalability, and evaluation. It helps you make a more informed decision.
- For qualifying and selecting (Screening):
  - Challenge: You are a foundation that has received 50 applications. You want to find the most promising and well-thought-out projects.
  - ICAF's contribution: An ICAF assessment can help identify interventions that are built on a systematic approach and have a solid documentation base, and therefore have a stronger potential to succeed and create lasting knowledge.
- For identifying strengths and development potential (Quality Development):
  - Challenge: You run a successful intervention but want to make it even better and perhaps ready to be scaled.
  - ICAF's contribution: By scoring your own intervention, you can see exactly where you are strong (e.g., a solid design) and where there is room for improvement (e.g., a lack of manuals for scaling or a more systematic evaluation).
- For bridging to research and funding (Validation):
  - Challenge: You want your practice-based intervention to be evaluated by researchers or apply for a larger grant, but you are unsure if it is "mature" enough.
  - ICAF's contribution: A high ICAF score signals that your intervention is well-documented and systematic, which makes it an attractive candidate for both research collaborations and larger foundations.



## 1.4 What the Tool Does Not Tell You

To use ICAF correctly, you should be aware of its limitations. The tool is a compass that shows direction and maturity, but it doesn't make a final judgement about an intervention's value. On the contrary, it enables the identification of an intervention's maturity level and where targeted improvements can be made.

ICAF solely measures the extent to which key aspects are addressed in the documentation.

The tool does not tell you:

- Whether the intervention has a positive effect: An intervention can score highly on ICAF because it is well-described and systematically designed, but it may still have a limited or no effect on participants. Conversely, an intervention with a tremendous positive effect can score low because it is poorly documented.
- The quality of the individual elements: The tool registers whether there are references to scientific literature (Q11), but it does not assess whether the references are relevant or of high quality.
- Whether the intervention is the "best": A high score does not mean the intervention is better than one with a low score. It is simply more mature in its structure and documentation.

Remember: A low ICAF score does not necessarily mean the intervention is poor. It most often means that the documentation could possibly be improved and that there are a few steps left towards producing genuine scientific evidence. It is also possible that a project does not have selected criteria as an objective. Conversely, a high-scoring project can have several problems that make it difficult to continue. In any case, an interpretation must be added, and an ICAF score—both high and low—cannot stand alone.

## 1.5 Possible Extensions with Other Frameworks

ICAF can be used as a standalone evaluation tool. Depending on your purpose, it can be supplemented with other frameworks.

Tool	Primary focus	Typical use
ICAF	<b>The intervention's maturity</b> (design, scalability, evaluation)	To assess and compare the systematic approach in practice-based interventions described in grey literature.
AACODS	<b>The source's credibility</b> (Authority, Accuracy, Coverage, Objectivity, Date, Significance)	To assess the quality and reliability of the document itself (the report, website, etc.), especially when the sender is unknown.



Tool	Primary focus	Typical use
<b>CASP</b>	<b>Scientific methodology</b> (especially in qualitative studies)	To conduct an in-depth critical appraisal of research design, data collection, and analysis in documents that have the character of a research study.

## 2.0 Introduction to the Framework

### 2.1 ICAF as a Comprehensive Picture of Potential

The ICAF framework is logically structured around an intervention's natural lifecycle, from the initial idea to a potential rollout. The 11 questions are divided into three phases that together paint a complete picture of the intervention's potential.

1. Design Criteria (Q1-Q5): The Foundation. This phase is about the *idea*. Is the purpose clear? Is it designed for those it is meant to help? Has sustainability been considered, and what is the role of art? Without a solid foundation, it's difficult to build a successful intervention.
2. Scalability Criteria (Q6-Q8): The Growth. This phase is about *dissemination*. Are the experiences so well-described that others can learn from them? Has the model been tested under different conditions, and are there specific tools to ensure a quality rollout?
3. Evaluation Criteria (Q9-Q11): The Evidence. This phase is about *knowledge*. How much experience is the intervention based on? Has knowledge about its effect been systematically collected, and is it anchored in the broader professional and scientific literature?

### 2.2 How to Score: From Indication to Documentation

ICAF scoring is designed to be simple and transparent, making it easy to compare. Each of the 11 criteria is assigned points based on what is explicitly stated in the analysed material. To ensure a consistent assessment (especially when comparing multiple interventions), it is recommended that at least two people score independently of each other and then discuss and align their results.

- 2 points (Yes - Explicitly documented): The criterion is clearly described and met.
  - *Example (Q10)*: The report has a section titled "Evaluation," which states that participants completed a recognised well-being scale (e.g., WEMWBS) before and after the programme.





- 1 point (Probably yes - Indirectly indicated): There are strong indications, but it is not described directly.
  - *Example (Q10):* The report contains a series of positive quotes from participants about increased well-being and mentions that "evaluation interviews" were held, but the method and specific results are not described.
- 0 points (No - Not documented): The criterion is not mentioned, or it appears that it has not been met.
  - *Example (Q10):* The report only describes the project's activities, timeline, and number of participants.

The total score (typically converted to a percentage) provides a quick overview of the intervention's documentation level.

## 2.3 Review of the Individual Criteria

### 2.3.1 Design Criteria

The main question here is whether the intervention is built on a well-thought-out and solid foundation. A high score here indicates a targeted and well-considered intervention with a clear purpose and a strong theoretical or experience-based anchoring. A low score may point to the intervention being more activity-driven than purpose-driven.

Q1: Is the activity designed with a primary ambition to improve mental health or well-being?

This criterion assesses whether mental health is at the core of the intervention's purpose or if it's merely an expected side effect. It's about the intention behind the design.

- In practice: Look for phrases in the project description's purpose and objectives section. Does it explicitly state that the intervention aims to "reduce symptoms of anxiety," "break social isolation," or "strengthen self-esteem in young people with mental health challenges"? Look for whether well-being or health is the focus.
- Why it's important: A clear ambition ensures that all elements of the intervention—from the choice of art form to the facilitators' pedagogy—are aimed at creating mental health effects. It sharpens the focus and makes it easier to evaluate whether the goal is reached.

Q2: Is the initiative designed with the target group in mind (e.g., has the target group been involved)?

This assesses whether the intervention is tailored to the specific needs and reality of the participants. The highest recognition is given for genuine user involvement.

- In practice: Is a process described where young people themselves helped develop the activities (co-design)? Or is the design based on a thorough needs analysis, interviews with the target group, or close collaboration with organisations that know the target group in-depth?



- Why it's important: User involvement increases relevance, ownership, and the likelihood that participants will respond well to the offer and be retained. It ensures that the intervention addresses real needs rather than presumed ones.

Q3: To what extent does the programme incorporate robustness in its design after the programme's completion?

This criterion focuses on the intervention's sustainability for the individual participant. What happens when the programme ends?

- In practice: Are there elements in the design that extend beyond the programme itself? This could be training in specific coping strategies, the establishment of network groups that continue, or a "discontinuation plan" that builds a bridge to permanent services in the council or civil society.
- Why it's important: Without a focus on the time after the intervention, there is a high risk that the positive effects will quickly fade. Robustness in the design is key to creating lasting change for participants.

Q4: Does it build knowledge about systemic prerequisites, e.g., access, dissemination, interest?

This assesses whether the intervention addresses the practical and structural frameworks it operates within. An intervention does not exist in a vacuum.

- In practice: Does the report describe challenges and solutions related to recruitment? Is there a discussion about how to ensure access for young people in rural areas? Are there considerations on how to anchor the project with council partners?
- Why it's important: Knowledge of systemic prerequisites is crucial for an intervention's survival and dissemination. Ignoring these factors is a frequent reason why successful pilot projects never become permanent services.

Q5: To what extent is aesthetics or art a primary influencing factor on the outcome?

This criterion uncovers the specific role that art and creative processes play. Is art the engine for change, or is it merely a vehicle?

- In practice: Does the description argue how the artistic processes (e.g., creating a piece of work, expressing oneself non-verbally, experiencing flow) contribute to the mental health goals? Does it differentiate itself from any other social activity?
- Why it's important: To strengthen the field, it's essential to become sharper on the "active ingredients" in art interventions. This criterion pushes for a deeper reflection on why art specifically works.





### 2.3.2 Scalability Criteria

The main question here is whether the results can be replicated and disseminated to others? A high score shows that there is potential for knowledge and practice to be transferred to other contexts. A low score suggests that the intervention is strongly locally anchored or person-dependent, and that valuable knowledge risks being lost.

Q6: Is the activity documented and well-described?

This is the most fundamental criterion for knowledge to be shared and to survive. If it's not written down, it only exists in the minds of a few people.

- In practice: Is there a clear description of the intervention's purpose, target group, methods, session structure, facilitators' roles, and pedagogical principles? Is it detailed enough for a professional to get a clear picture of what the intervention entails?
- Why it's important: Good documentation is a prerequisite for learning, quality assurance, handover to new employees, and inspiration for others. Without it, replication and scaling are impossible.

Q7: Has the activity been tested in multiple locations and under different circumstances?

This assesses the intervention's robustness and adaptability. A model that works in one place doesn't necessarily work everywhere.

- In practice: Has the intervention been tested in collaboration with different partners (e.g., both a school and a psychiatric department)? Has it been implemented in both a large city and a rural council? Are experiences with adapting the model to new contexts described?
- Why it's important: Testing an intervention in different settings reveals its core elements (what must be preserved) and its flexible frameworks (what can be adapted). This is crucial knowledge for a successful rollout.

Q8: Is scalability incorporated through manuals, certification, training?

This criterion looks at the specific tools that ensure a quality rollout. It is the bridge from a good idea to a standardised practice.

- In practice: Is there a pedagogical manual for facilitators? Is a "train-the-trainer" course offered? Has a certification scheme or a network for professional sparring been established for those who carry out the intervention?
- Why it's important: Without formalised tools for dissemination, there is a high risk of "method drift," where the intervention gradually gets diluted or changes so much that it loses its effect. Manuals and training ensure quality and consistency during scaling.



### 2.3.3 Evaluation Criteria

The main question here is what the basis is for saying that this works, and how do we know it? A high score signals a mature intervention that has invested in collecting knowledge about its own effect and places itself within a broader professional context. A low score indicates a need for more systematic experience collection.

Q9: Has the activity been thoroughly tested? Has the activity had multiple iterations and existed for more than 5 years?

This criterion assesses the intervention's maturity and experience base. It distinguishes between an untested pilot project and a well-established practice.

- In practice: Does it appear that the intervention has been running for several years? Is there a description of how it has been adjusted and improved based on experiences from previous programmes (iterations)?
- Why it's important: A long lifespan often indicates that the intervention has proven its worth to participants and funders and has had time to develop a robust and tested practice.

Q10: Has the activity been evaluated with a focus on results and impact on mental health?

This looks at whether a systematic effort has been made to document the effect. Good intentions are not enough; there must be some form of documentation of the impact.

- In practice: Have recognised methods been used to measure change, such as before-and-after measurements with questionnaires, qualitative interviews, case studies, or systematic collection of feedback? Does the evaluation go beyond pure participant satisfaction?
- Why it's important: A systematic evaluation significantly strengthens the intervention's credibility. It moves the argument from "we think it works" to "we can document that it makes a difference," which is crucial in dialogue with policymakers and foundations.

Q11: Are references to peer-reviewed literature or other scientific sources cited?

This last criterion assesses the intervention's anchoring in and contribution to broader academic knowledge. It's about learning from others and letting others learn from you.

- In practice: Does the intervention's design refer to relevant theory or research in fields such as recovery, social psychology, or art therapy? Are the intervention's own results discussed in light of existing research, or have articles about it been published?
- Why it's important: A link to the research world shows an ambition to be part of a larger field of knowledge. It increases professional weight and ensures that one builds upon the best available knowledge, rather than "reinventing the wheel".



## 3.0 Acknowledgements

ICAF was developed by the National Centre for Arts and Mental Health, Copenhagen, Denmark, which should be credited when used. The work was funded by the Ministry of Health and the Ministry of Culture. There is not considered to be any conflict of interest related to the work.

## 4.0 Bibliography

- *Grey literature rapid review of how art supports mental health and wellbeing among young people in Nordic countries.* (2025). [Submitted].
- Critical Appraisal Skills Programme. (2024). *CASP Checklist: For Qualitative Research*. <https://casp-uk.net/casp-tools-checklists/>
- Schöpfel, J. (2011). Towards a Prague Definition of Grey Literature. *The Grey Journal*, 7(1), 5-18. [https://greynet.org/images/GL12\\_S1P\\_Sch\\_pfel.pdf](https://greynet.org/images/GL12_S1P_Sch_pfel.pdf)
- Tyndall, J. (2008). *How low can you go? Towards a hierarchy of grey literature*. Paper presented at Dreaming08: Australian Library and Information Association Biennial Conference. <https://core.ac.uk/download/pdf/14931482.pdf>
- UWE Bristol & West of England Academic Health Science Network. (2020). *A one-page guidance checklist for your Evaluation Project: What does good look like?* <https://guides.library.ubc.ca/greylitforhealth/eval>



## 5.0 Scoring Sheet – Intervention Capability Assessment Framework

ICAF (Intervention Capability Assessment Framework) was developed to create a broader analytical framework for systematically assessing and comparing the practical knowledge that has been generated and described in reports, evaluations, and notes. ICAF is a specialised tool for unlocking the value of the practice-based knowledge that is crucial for understanding how art interventions for mental health work in practice. By adapting standard criteria for critical appraisal, ICAF serves as a tool to measure an intervention's "maturity," the degree of systematic approach behind it, and where targeted improvements can be made.

ICAF scoring is designed to be simple and transparent, making it easy to compare. Each of the 11 criteria is assigned points based on what is explicitly stated in the analysed material. To ensure a consistent assessment (especially when comparing multiple interventions), it is recommended that at least two people score independently of each other and then discuss and align their results.

- 2 points (Yes - Explicitly documented): The criterion is clearly described and met.
- 1 point (Probably yes - Indirectly indicated): There are strong indications, but it is not described directly.
- 0 points (No - Not documented): The criterion is not mentioned, or it appears that it has not been met.



Intervention Capability Assessment Framework		Score	Comments
<b>Design criteria</b>			
Q1	Is the initiative designed with a primary ambition to improve mental health or well-being?		
Q2	Is the initiative designed with the target group in mind (e.g., has the target group been involved)?		
Q3	To what extent does the programme incorporate robustness in its design after the programme's completion?		
Q4	Does it build knowledge about systemic prerequisites, e.g., access, dissemination, interest?		
Q5	To what extent is aesthetics or art a primary influencing factor on the outcome?		
<b>Scalability criteria</b>			
Q6	Is the activity documented and well-described?		
Q7	Has the activity been tested in multiple locations and under different circumstances?		
Q8	Is scalability incorporated through manuals, certification, training?		
<b>Evaluation criteria</b>			
Q9	Has the activity been thoroughly tested? Has the activity had multiple iterations and existed for more than 5 years?		
Q10	Has the activity been evaluated with a focus on results and impact on mental health?		
Q11	Are references to peer-reviewed literature or other scientific sources cited?		